Changes in Mathematics Achievement Over Time in Australia and Ethiopia

Tilahun Mengesha Afrassa

FLINDERS UNIVERSITY INSTITUTE OF INTERNATIONAL EDUCATION
RESEARCH COLLECTION
NUMBER 4

STUDIES IN COMPARATIVE AND INTERNATIONAL EDUCATION

Number 4

Changes in Mathematics Achievement Over Time in Australia and Ethiopia

TILAHUN MENGESHA AFRASSA

FLINDERS UNIVERSITY
INSTITUTE OF INTERNATIONAL EDUCATION

Title: Changes in Mathematics Achievement Over Time in Australia and Ethiopia
Series: Flinders University Institute of International Education Research Collection: Number 4
First Published: April 2002

Produced by the Flinders University Institute of International Education
GPO Box 2100, Adelaide 5001, Australia
Telephone: (+61 8) 8201 2441
Facsimile: (+61 8) 8201 3184
Email: fuiie@flinders.edu.au
WebSite: http://www.flinders.edu.au/education/fuiie

Designed and Edited by Katherine L. Dix
Published by Shannon Research Press, South Australia

ISBN: 0-9580704-4-X

Preface

Three international mathematics studies have been conducted by International Association for the Evaluation of Educational Achievement, at the 13-year-old students level. Australia was one of the countries that participated in the three studies, namely the First, Second and Third International Mathematics Studies. A similar study was conducted in 1996 in Ethiopia. The information which was collected from students in Australia and Ethiopia provided the evidence employed for the analyses of the present study.

Some of the major purposes of the study are listed below.

1. One of the purposes of the study was to investigate the structure underlying the mathematics achievement test items. For this purpose, confirmatory factor analyses of the mathematics tests were undertaken using linear structural relations analysis.
2. An other purpose was to examine possible changes in mathematics achievement of the Australian lower secondary schools over time and between Australia and Ethiopia. The tests that were administered on the three occasions in Australia, and in 1996 in Ethiopia were brought to a common mathematics scale. In addition, the views and attitudes towards mathematics and schooling statements which were administered on two occasions in Australia and on one occasion in Ethiopia were also linked to form common view and attitude scales. The Rasch model was employed for these purposes using the QUEST (Adams and Khoo, 1993) computer program.
3. The other purpose of the present study was to investigate changes in student level factors that influenced mathematics achievement at the lower secondary school level in Australia over the past 30-year period. In addition, similarities and differences between student level factors that influenced mathematics achievement of Year 8 students in Australia and Ethiopia have been examined. Here, the study employed partial least squares path analysis procedures.

Furthermore, the study addressed in detail the questions of gender differences in mathematics achievement, and views and attitudes towards mathematics and schooling, and to determine whether or not the items in the mathematics tests, views and attitudes questionnaires were biased against boys or girls.

The major results reported from the present study are listed below.

1. A nested model was the best fitting model indicating that both a total score and separate subscale scores could be computed.
2. The mathematics items that were administered on the four occasions could be brought to a common mathematics scale. In addition, the views and attitudes statements that were administered on the three occasions could also brought to

seven common view and attitude scales. The common scales were constructed using the Rasch model.

3. The achievement level of lower secondary school Australian students declined over time, between 1964 and 1978 at the 13-year-olds level, and between 1964 and 1994 at the Year 8 level. Furthermore, the Ethiopian Year 8 students achieved at a lower level than their 1964 and 1994 Australian counterparts.
4. The PLSPATH analyses of the Australian data sets indicated that students from higher socioeconomic status background families, students in large class groups, and students who expressed more positive attitudes towards mathematics were likely to achieve at a higher level in mathematics during the past 30-years. Moreover, gender was not a student level factor that influenced mathematics achievement of students in Australia.
5. Home background, attitudes towards mathematics and class size were also student level factors that influenced mathematics achievement, both in Australia and Ethiopia. It is important to point out that in Australia students from larger class groups were likely to achieve at a higher level than students from smaller class groups, however, in Ethiopia students from smaller class groups were likely to achieve at a higher level than students from larger class groups. This difference possibly occurred because of the big differences in the average class size between Australia (average class size 30) and Ethiopia (average class size 74). Thus further investigation would seem necessary in cross-national studies to find out the appropriate class size for effective teaching and learning in mathematics.

Therefore, the present study provides a detailed account of the main student level factors that influenced mathematics achievement level of lower secondary school Australian students over the last three decades, and identifies the similarities and differences in student level factors that influence the achievement level of Australian and Ethiopian students.

Acknowledgments

I would like to express my appreciation to those who helped and assisted in the production of this book.

First of all I would like to thank my supervisors Professor John P Keeves and Professor Jonathan Anderson for their invaluable experience, guidance and assistance. I always appreciate their patience when they help their students. Their assistance and help were not only intellectual guidance, but also in personal matters. I am delighted to have been one of their students.

I would also to thank all academic and administrative staff in the School of Education for their encouragement and stimulating discussion at the ‘morning teas’ and ‘lunch time seminars’ every Thursday. My sincere gratitude also goes to Ramon, with whom I shared a room, for all his cooperation, encouragement, and assistance in everyway. I also would like to thank Robin Ryan for all his friendship and the discussions we had that made my work most stimulating.

My deepfelt thanks go to Dr Ayele Meshesha for making possible the data collection in Ethiopia, without his support the data collection process would have been very difficult. I would also like to thank the Addis Ababa Regional Education Office and its six branch offices, the participating school principals, teachers and students whose cooperation made the study possible. My sincere gratitude also goes to Ato Negussie Yibas, W/ro Alemnesh Yimer and all members of ENEO for their unconditional support during the data collection period. My sincere appreciation goes to ato Kebede

Tiku for his valuable information and comment about curriculum development in Ethiopia and the ICDR Librarians for their supplying reading materials which were valuable for the study. My simcere appreciation also goes to Mrs Anderson for her help in typing all the questionnaires.

I would also like to express my deepfelt thanks to my parents in Ethiopia, especially my elder brother Ato Tesfa Mengesha for introducing me to modern education, and to his support and encouragement to undertake work in higher education. I would also like to express my appreciation to Dr Josie Misko, Michael Perth, John Halsey and Judith Lydeamore for their encouragement.

Last but not least, I would like to express my appreciation to my wife Alem Abebaw, for her emotional support, patience, and sacrifice during my study.

Contents

Figures

Tables

1

Introduction

The Problem and its Setting

Mathematics education is one of the most important areas in the school curriculum. Both in Australia and Ethiopia, mathematics is vital when seeking employment or engaging in further study (Atweh, 1980; Ethiopian School Leaving Certificate Examinations Office, 1985). Costello (1978) reported that three-quarters of the higher education courses in Victoria require a pass in mathematics in the Higher School Certificate Examinations (Year 12) and two thirds of certificate technology courses required a pass in mathematics at the Year 11 level. Mathematics achievement is also important for entry to trade courses, apprenticeships and other employment opportunities (Atweh, 1980; Carss, 1980). In Ethiopia, mathematics is a compulsory subject throughout primary and secondary education. Students who wish to proceed to higher education must sit for the mathematics examination in the Ethiopian Schools Leaving Certificate Examination (ESLCE), which serves purposes of both certification and selection for admission to higher education. Those students who are applying for higher education should obtain a high grade point average including mathematics. It is clearly of interest to undertake a comparative study of mathematics achievement and student level factors influencing achievement in mathematics of students in Ethiopia and Australia as well as students' views and attitudes towards mathematics and schooling.

In order to improve the educational system of a country, it is also important to be aware of what is going on in other countries and why it is occurring. This can be done by comparing that system with others (Mallinson, 1975). Keeves and Adams (1997) contend that comparisons have been made between countries with respect to many different aspects of educational provision. Such comparisons help to bring about an understanding of a particular system and assist with planning for the future. Consequently, comparative studies are important to identify similarities and differences among systems, to know the reasons for the similarities and differences and why different solutions have been undertaken to problems which are common to all (Mallinson, 1975). Jones (1973) argues that the comparative study of education makes scholars clear sighted in the effective analysis of their own educational environment. Such studies can lead, eventually, to the reform of their own educational system.

Comparisons can be made between two or more educational systems. Teese (1988) contended that in Australia it was believed that the method of a series of two-country comparisons could lead to more meaningful observations because national data would be less hindered by the standard instruments used by international agencies and also because greater reference could be made to specific national features of context and structure. By reducing the lag-time in producing comparative data, the two country case study method would allow the origins of measured differences in educational participation and performance to be studied more readily and in depth because of the limited frame of reference. However, such a series of studies would not provide opportunities for the examination of cross-national relationships.

Cross-national studies are a recent phenomenon (Robitaille and Travers, 1992). During the past 30 years, international comparative studies have been conducted in many different subject areas including mathematics. The International Association for the Evaluation of Educational Achievement (IEA) has conducted three international mathematics studies, with Australia participating in each. The First International Mathematics Study (FIMS), was carried out in 1964 (Husén, 1967; Keeves, 1968, Keeves and Radford, 1969; Rosier, 1980). The Second International Mathematics Study (SIMS) was conducted between 1978 and 1982 (Garden, 1987). In 1994/95 the Third International Mathematics and Science Study (TIMSS) was administered in 45 countries (Beaton et al., 1996a, 1996b).

Changes in mathematics achievement over time can be analysed for Australian lower secondary school students, because of their participation in FIMS, SIMS and TIMSS studies. However, this is not the case for Ethiopian students. Ethiopia has not conducted any comparative studies in mathematics education. It has not been a member country of IEA and therefore, did not participate in any of the international studies conducted by IEA. Consequently, it has yet to obtain any of the advantages of comparative international studies. Hence, it is desirable to take the initiative and to introduce the idea of a comparative study to the authorities in Ethiopia. For these reasons the present study was undertaken.

Australia was chosen to be part of this research project for several important reasons. The first and the major reason for the inclusion of Australia in this study was the availability and the accessibility of the 1964, 1978 and 1994 data sets. From those countries which participated in the first three IEA studies, Australia was the only country where there was immediate access to these data. It is impossible to undertake a comparative study without the necessary data for comparison, so it was a sensible idea to include Australia whose data were easily accessible as part of the study.

Another reason for including Australia as part of this study was the awareness by Australian educators of international comparative studies. In Australia, international comparisons have already become part of the vocabulary of educational politics (Teese, 1988). Teese has argued that authorities especially at the school level, had either actively searched for reports on overseas educational policy or had perceived comparisons as an important instrument in formulating educational policy, at least within certain limits. Teese (1988) reported that the then Commonwealth Department of Employment, Education and Training had sought to examine Australia's international position by using a series of cross-country comparisons.

Therefore, because of this understanding of comparative international investigations, it was believed to be desirable to obtain the necessary information for comparison between Australia, a highly developed country, and Ethiopia, a developing country.

Purposes of the Study

This study has five major purposes. The first purpose of the study is to develop a general theoretical model which considers the multivariate structure of the available data. In this general theoretical framework, specific models of the structure of mathematics achievement test items at the lower secondary school level are proposed. The proposed models are examined by employing appropriate procedures using the Australian and Ethiopian data sets. The identification of the structure of mathematics achievement tests at this level is very important for the selection of the most appropriate methods of data analysis and for calculating scores for comparative purposes.

The second major purpose is to examine the changes of the mathematics achievement level of Australian lower secondary school students over time. This investigation would indicate whether the achievement level of students had improved or declined over a 30-year period. Furthermore, the achievement level of Ethiopian students is also compared with that of the 1964 and 1994 Australian lower secondary school students.

The investigation of mathematics achievement over time and across countries requires the development of a common mathematics scale which would be independent of the samples of students tested and the items employed. Hence, the third purpose of the study is to develop a common mathematics scale.

Causal models of student level factors influencing mathematics achievement of students at the lower secondary school level are required to examine the hypothesised interrelationships between variables that are considered important as a result of a theoretical framework and previous research findings. In the present study investigating the type and size of interrelationships between latent variables and their effects on mathematics achievement, whether the relationships are consistent over time and across the two countries, are important. If the same variables indicated relatively stable effects over time in Australia and between the two countries, they could be taken as evidence for the generality of the model and the coherence of the theoretical framework under investigation. Thus the fourth purpose of this study is to develop a theoretical model of student level factors influencing the mathematics achievements of lower secondary students in Australia and Ethiopia and to examine these hypothesised interrelationships between variables.

The final major purpose of the present study is to investigate the views and attitudes of Australian and Ethiopian students towards mathematics and schooling and to develop common scales which are independent of the samples of students tested and the items employed. These scales are used in this study to make comparisons between the attitudes of students across countries and over time.

Significance of the Study

This study is significant for several reasons. It is the first major research project focusing on the development of a common mathematics scale for measuring mathematics achievement of students over time. The 1964 and 1978 Australian data sets were analysed by state, to compare Year 8 students' mathematics achievement on the common items across time by Rosier (1980) and by Moss (1982). However, these studies did not scale the data and a common mathematics scale was not developed to measure the mathematics achievement of students over time. A common mathematics scale has not yet been developed either in Australia or internationally in any of the countries that participated. Therefore, a major focus of this study is to develop a mathematics achievement scale that is independent of both the samples of students tested and the samples of items employed. This work would contribute to the advancement of educational measurement, particularly in the field of mathematics.

The study examines the achievement differences of Australian students over time. In addition, it examines the attitudinal differences of Australian students towards mathematics and schooling over time. Furthermore, the study investigates the student level factors that influence mathematics achievement over time. The outcomes of these findings would help to improve the achievement level of students in mathematics education. The identification of the student level factors that influence mathematics achievement of Australian students over the 30-year period would contribute to the development of a theory of school learning.

Another area of significance is the contribution to the improvement of mathematics education, and educational research in Ethiopia. There are no major studies of student achievement in this country, in general, and in mathematics, in particular (Ademe and Gebre-Meskel, 1989). As this study is the first of its kind in mathematics education in Ethiopia, it would be an important milestone in educational research in that country.

The result of comparisons between Australia and Ethiopia would make a valuable contribution to knowledge by identifying not only the differences in the factors that influence the mathematics achievements of students in developed and developing countries, but also by estimating the comparative magnitudes of the effects of such factors in the two countries.

Limitations of the Study

One of the major limitations of this study is that the analyses undertaken are limited to the available data. Even though some recoding of items could be considered, the development of scales was limited by the items that were administered either in the Australian or Ethiopian studies. It was disappointing to find that the student questionnaire in the Third International Mathematics Study which was administered in 1994 contained a relatively small number of common items with previous testing programs. Even though adequate data on student level factors that influenced mathematics achievement were available to make meaningful analyses, greater consideration in the development stage of a mathematics study for potential longitudinal data collection in the area of mathematics at the three time points would have advanced the scope and strength of these studies. Furthermore, regarding students' views and attitudes towards mathematics and schooling there were insufficient items to make comparisons over time and across countries. Thus, the comparisons of views and attitudes were limited between FIMS and SIMS, and between FIMS and EMS students.

Unfortunately, the sampling procedures employed in the four mathematics studies were not the same. In the 1964 study, students were selected as both age and grade level samples. The age sample included, 13-year-olds in Years 7, 8 and 9 and the grade sample involved, all students in Year 8 including 13-year-olds. By contrast, in 1978, students were chosen as an age sample, that is 13-year-olds in Years 7, 8 and 9. However, in 1994, students were selected as grade level samples, that included students in four states of Australia from Years 7 and 8, and the other four states from Years 8 and 9. Finally, in the 1996 Ethiopian Mathematics Study (EMS), students were chosen as a grade level sample, that involved students in Year 8. Thus, comparisons in this study are limited to between 13-year-old students in 1964 and 1978 on the one hand, and Year 8 students in FIMS, TIMSS and EMS on the other. South Australia, the two territories and all nongovernment schools throughout

Australia did not participate in the 1964 study, therefore, they are excluded from some comparisons.

For calibration and scoring purposes, it was necessary to decide on issues regarding the occurrence and handling of missing data in the mathematics achievement tests. Thus, decisions were made which were based on the results of the Rasch analyses of the available data sets. However, the TIMSS data set was not considered for making the decision, because the data were only made available for analysis about six months prior the completion of this study.

Morrison and Fitzpatrick (1992) and Mohandas (1996) have argued that the concurrent equating procedure is probably the most robust of the current methods of equating several test forms. Even though this method seems strong and robust, it was not possible to employ it in equating EMS, FIMS, SIMS and TIMSS data sets. The main reason was that the TIMS data were not available on time for concurrent equating.

Another major limitation of the present study is that the Ethiopian data collected for the analyses were only obtained from one region. In Australia the data were obtained on three occasions. On the first occasion the data were collected in all states, except South Australia, while on the other two occasions data were drawn from all Australians states and territories. However, due to financial, manpower and time constraints, the Ethiopian data collection was limited to one region, that is from the Addis Ababa region.

The schools and the students in the Addis Ababa region were selected randomly using a random sampling procedure with probability proportional to the size of the school sampling. However, because of lack of cooperation from the principal and staff of one school, one of the selected schools was replaced after four visits. In Australia, in the FIMS and SIMS studies students were sampled randomly from within the schools chosen at the primary sampling stage. However, in TIMSS the procedure of sampling intact classes from within schools was employed at the second stage of sampling. This could be expected to lead to larger errors of sampling.

In addition, it should be noted that school level factors were not reported in this study because of the sensitivity of school authorities in Australia to comparisons that involved important school and teacher level characteristics, such as type of school.

Structure of the Book

The following three chapters review the literature related to this study. Chapter 2 examines the international comparative studies conducted by IEA, and the major findings from FIMS, SIMS and TIMS in Australia are also reviewed. Chapter 3 reviews the development of the mathematics curricula of Australian and Ethiopian schools since the 1960s and the 1940s respectively. Major curricular changes in mathematics occurred between the 1960s and the 1990s and the main reasons for the changes are considered. Chapter 4 examines the research that has been undertaken into student level factors that influence the mathematics achievement of Australian and Ethiopian students. The chapter explores a number of aspects of the previous research regarding the student level factors that influence achievement in mathematics in Australia and Ethiopia.

Chapter 5 describes the data collection procedures employed in the study in both countries, while Chapter 6 presents the theoretical framework undertaken in this project. The analytical procedures employed in the study are described in Chapter 7. The chapter commences with an explanation of the partial least squares path analysis procedure which is used to evaluate the student level factors that influence achievement in mathematics. The chapter then explores the use of the Rasch model which is applied to measure the mathematics achievement of students and their attitudes toward mathematics and the development of common mathematics achievement and attitude scales. The chapter also examines the confirmatory factor analysis procedure, a method which is used in the study to examine the factor structure of the mathematics achievement items. The last part of this chapter presents the test equating, item bias detection procedures and analysing data with complex samples.

The data analysis and the results relating to achievement are addressed in Chapter 8, while, the views and attitudes of students towards mathematics and schooling are discussed in Chapter 9.

The results related to sex differences in mathematics achievement, and views and attitudes towards mathematics and schooling are discussed in Chapter 10. The first part of the chapter presents the sex differences in mathematics achievement, while sex differences in views and attitudes on the different occasions are addressed in the second part of the chapter.

Chapters 11 presents the data analysis and the results of student level factors influencing mathematics achievement on different occasions. The major findings, conclusions and recommendations of the present study are presented in Chapter 12.

2

International Mathematics Studies

The comparison of educational systems between countries is especially important in sharing experiences and obtaining valuable information that could be utilised in developing future strategies for reform in education. This method of comparing different educational systems is called the comparative study of education. The greater importance to some would be the implication that such study could lead, eventually, to the reform of their own educational system. Robitaille (1994) argued that without such comparisons, the country tends not to question customary teaching practices, and may not even be aware of the choices to be made in the process of implementing those practices.

Comparisons may be made between two or more educational systems. This type of comparative study can be called an international comparative survey and is a relatively recent phenomenon (Robitaille and Travers, 1992). The launching of Sputnik, which was the first earth satellite, by the former Soviet Union generated substantial criticisms of the American school system during the late 1950s and early 1960s. In addition, the major industrialised countries were increasingly becoming concerned about the greater costs of providing free, universal public education. These events led to the implementation of international comparative studies (Husén, 1967; Inkeles, 1977).

In addition, Robitaille and Travers (1992) have stressed the importance of international studies in identifying the strong and weak points of the educational systems of other countries; the need to separate fact from fiction about what school systems of other countries are able to accomplish with their students; the value of providing the opportunity for investigators to evaluate the importance of variables that might not be applicable in their own country; and the consequences of establishing a view of what can be achieved in education. Furthermore, Baker (1997) has argued that international studies are scientifically and politically useful when they are employed to shed light on how and why a country produces a particular pattern of achievement.

Although international comparative education studies are useful to improve the understanding of the teaching and learning of mathematics, Theisen et al. (1983) have pointed out several weaknesses. These weaknesses are related to the expense of conducting such studies, the cultural bias in the development of instruments, differences in test taking among students of different cultures and the difficulties in obtaining approvals for research.

More attention would appear to have been given to international comparisons in mathematics than to other areas of the curriculum. There are several likely reasons for this. First, mathematics plays a prominent part in the curriculum of every country, usually second in importance only to that of the mother tongue. Second, there is a great deal of similarity in mathematics curricula internationally. Third, the language of mathematics is, in many ways, truly international. As a consequence, whether a mathematics class is investigated in Addis Ababa or Adelaide, it is possible to grasp the major elements of the course being presented fairly readily because of the universality of mathematical symbolism and notation (Robitaille and Travers, 1992).

Reasons for International Studies in Mathematics

FIMS was the first large project of this kind (Keeves and Radford, 1969) and also included a detailed curriculum analysis (Keeves, 1968). Prior to FIMS there was a lack of comparative international achievement data. For the last 30 years however, the number and nature of the variables included in comparative studies of educational achievement have continued to expand. SIMS included the assessment of the teaching practices, and a longitudinal study of growth in student achievement over the course of the school year, although these aspects were not investigated in the Australian study.

Eckstein (1982) has argued that the importance of international studies are to accommodate investigators with a chance of evaluating the importance of variables which might not be available in their own country; to compare different procedures applied to the same objectives and to evaluate their effects; to provide a view of what can be accomplished, and a context in which decision makers in each participating country can view their own system; to evaluate the outcomes of an educational innovation by examining its implementation and operation in another country; and to discover new methods of teaching by observing the classroom practices of teachers in other countries.

Stigler and Perry (1988) share similar ideas with Eckstein about the importance of comparative international studies on the innovation of teaching methods. They argue that cross cultural comparison also leads investigators to a greater understanding of their own assumptions about how children learn mathematics and to evaluate traditional teaching practices.

Even if the proponents of international comparative studies are able to show the importance of international mathematics studies, some researchers have criticised these international mathematics studies. Among the opponents of FIMS was Freudenthal (1975). His criticisms related to the selection of items used in the study, the inappropriateness of some of the topics tested with respect to the curricula of some of the participant countries; and the failure to consider whether or not students had been taught the content which was tested in the study.

Other opponents of international surveys of achievement argue that such studies require expenditures of time, energy, and resources on the part of participants as well as investigators. They have also questioned the significance and the influence of the findings of such studies (Robitaille and Travers, 1992). Clearly, there are political reasons for applying international comparative study to education. These researchers argue that the political reasons for knowing more about education in other countries are the need to know about the educational systems of other countries in order to find out the strengths and weaknesses of their own system; and the need to identify the fact about what school systems in other countries are able to achieve with their students.

Unfortunately, the results of performance on achievement tests from such international studies are used for the ranking of the educational systems of the participating countries. Such comparisons are inevitable and, to the extent that they raise questions about possible sources of variation in achievement across national systems, their importance should not be down played. On the other hand, such comparisons need to be accompanied by a thorough analysis of the variables that contribute to those differences (Robitaille and Travers, 1992).

While there have been other international studies conducted by different groups of researchers such as the International Assessment of Educational Progress, the Dallas Times-Herald Survey and the Japan/Illinois Study, only FIMS, SIMS and TIMSS are discussed here because their instruments and findings are drawn upon in the present study.

The First International Mathematics Study (FIMS)

The main purpose of FIMS was to investigate differences among different school systems and the interrelations between the achievement, attitudes and interests of 13-year-old students and final-year secondary school students (Husén, 1967; Keeves, 1968: Keeves and Radford, 1969; Rosier, 1980; Moss, 1982). The choice of mathematics for the first international study was largely for convenience. The organisers of the project assumed that it would be easier to make international comparisons in mathematics than in any other area of the curriculum (Husén, 1967). The second reason was the concern of the participating countries for improving their scientific and technical education (Postlethwaite, 1971). However, a third reason was associated with the evaluation of the introduction of “new maths” (Husén, 1967) into the mathematics courses in many countries during the late 1950s and early 1960s.

Countries which participated in the FIMS were Australia, Belgium, England, The Federal Republic of Germany, Finland, France, Israel, Japan, The Netherlands, Scotland, Sweden, and the United States. All these participating countries with the exception of Australia, Israel, Japan and United States were from Europe, and all were highly industrialised countries.

School and students who participated in the study were selected using two stage random sampling procedures, involving age and grade level samples. The age level sample included all 13-year-old students in Years 7, 8 and 9. While, the grade level sample involved Year 8 students including 13-year-old students at that year level. All students in the samples were government school students. In the cluster sample design schools were selected randomly at the first stage and students were selected randomly from within schools at the second stage.

The results of the international analyses of the FIMS data are given in Husén (1967), and Postlethwaite (1967) and summarised in Keeves (1995) as well as in a very large number of journal articles.

The Second International Mathematics Study (SIMS)

SIMS was conducted in the late 1970s and early 1980s in 21 countries. The main purpose of SIMS "was to produce an international portrait of mathematics education with a particular focus on the mathematics classroom" (Garden, 1987, p. 47). The particular feature of SIMS was that it was designed to be a longitudinal study with tests that were administered on two occasions, first as a pretest given at the beginning of the academic year and then as a posttest, given at the end of the academic year.

SIMS was developed to be a study of the mathematics curriculum at three levels: the intended curriculum which was defined at the national or system level, as codified in mathematics curriculum guides, and as operationalized in textbooks approved for teachers' use; the implemented curriculum which was the mathematics curriculum as it was taught by teachers in their own classrooms; and the attained curriculum which was what was learned by students and manifested in their mathematics achievement and attitudes towards mathematics.

Countries that participated in SIMS were Australia, Belgium (Flemish), Belgium (French), British Colombia, England and Wales, Finland, France, Hong Kong, Hungary, Israel, Japan, Luxembourg, The Netherlands, New Zealand, Nigeria, Ontario, Scotland, Swaziland, Sweden, Thailand and United States of America. From these 21 countries, 12 had participated in the FIMS. However, Australia which participated in both the FIMS and SIMS was not included in the analysis of the data by Garden (1987), Hanna (1989) and Robitaille and Travers (1992), since Australia had administered SIMS in 1978, because of financial constraints, which was two and/or three years earlier than in other countries.

The schools and students who participated in the SIMS study were selected using two stage sampling procedure. The students were all 13-year-olds. The students, involved in this study were both from government and nongovernment schools.

The results of the analyses of SIMS data are reported by Garden (1987), Robitaille and Travers (1992), and are summarised in Keeves (1995).

The Third International Mathematics and Science Study (TIMSS)

In 1994/1995 IEA conducted the Third International Mathematics and Science Study (TIMSS) which was the largest of its kind. It is a cross-national study of student achievement in mathematics and science that was administered at three levels of the school systems (Martin, 1996). Forty-five countries participated in this study. Among the countries that took part in the study, only two were from Latin America, namely Argentina and Mexico, and South Africa was the only African country that participated in the study. A majority of the participating countries were from Europe and Asia. The study investigated student achievement in mathematics and science, and differences in curriculum and instruction.

Robitaille and Donn (1992, pp. 204-205) contended that TIMSS would provide the following:

(a) current national and international information which countries could use to compare and contrast their curricula, teaching practices, and student outcomes with those from other countries of interest;

(b) an assessment of the potential impact that alternative curricular offerings, teaching strategies, and administrative arrangements could have on learning;

(c) an identification of what would be possible in the teaching of mathematics and science; and

(d) a greater understanding of how and why student attitudes changed, and what relationship the development of positive attitudes might bear to classroom practices.

Differences between the three studies

The sampling procedure employed in FIMS and SIMS was a two stage simple random sample. In the first stage schools and in the second stage students were selected from the schools chosen in stage 1. However, in TIMSS, there were three stages of sampling (Foy et al., 1996). The first stage of sampling consisted of selecting schools using a probability proportional to size method. The second sampling stage involved selecting classrooms within the sampled schools by employing either equal probabilities or with probabilities proportional to their size. Meanwhile, the third stage involved selecting students from within the sampled classrooms. However, this sampling stage was optional (Foy et al., 1996). The target populations at the lower secondary school level were students in the two adjacent grades containing the largest proportions of 13-year-olds at the time of testing.

A further difference between the last two studies and TIMSS was that performance assessment was part of the TIMSS assessment program. However, due to the cost and the complexity of data collection the performance assessment was an optional part of the TIMSS study in science (Martin, 1996, p. 1-11). Martin further reported that the performance assessment of TIMSS was tested only in 21 of the 45 countries that participated at Population 2, and 11 of the 28 countries that participated at Population 1. All the items in the tests employed were either short-answer or extended-response items. The TIMSS instruments were translated into 30 languages (Martin, 1996; Maxwell, 1996).

The results of the analyses of TIMSS data are presented in Beaton et al. (1996a) for mathematics and Beaton et al. (1996b) for science for Population 2.

The First, Second and Third International Mathematics Studies in Australia

Australia was one of the countries that participated in the FIMS, SIMS and TIMSS projects conducted in 1964, 1978 and 1994. Five Australian states took part in FIMS, six states and the Australian Capital Territory were involved in SIMS, and all states and territories participated in TIMSS. The data analyses for FIMS were reported by Keeves (1968), and Keeves and Radford (1969); while, Rosier (1980) and Moss (1982) compared the first two Australian studies. The findings of TIMSS were reported by Lokan et al. (1996). Some of the research findings presented in these reports are summarised here, because, their instruments and results are drawn upon in the present study.

Major Findings of the Two IEA Mathematics Studies in Australia

There were 70 mathematics items in the 1964 test and 72 items in the 1978 test. The comparisons of mathematics achievement in the two studies were undertaken by Rosier (1980) and by Moss (1982). Some of their findings are summarised as follows.

Differences in Mathematics Achievement

Rosier (1980) reported that at the 13-year-old level there was a slight overall decline from 1964 to 1978 in mathematics achievement throughout Australia. This was true for both sexes. However, it was only in Victoria that the achievement of girls declined more than that of boys (Moss, 1982).

In 1978, the mathematics total scores were markedly lower than in 1964 in Queensland, New South Wales, Tasmania and also in Victoria. These results provided some basis for asserting that there has been a decline in mathematics achievement of students from 1964 to 1978 (Rosier, 1980).

Queensland, New South Wales and Victoria showed a small decrease in the percentage of students obtaining scores above 40. However, Rosier (1980) did not mention why 40 was taken as a cutting score. From 1964 to 1978 an increase in the percentage of students with low scores was reported by Rosier for all states with the exception of Western Australia. He argued that the decline was a result of an increasing number of low achievers and this also suggested that from 1964 to 1978 there was a greater failure to meet the needs of the low achieving students.

In the subtests of basic arithmetic, advanced arithmetic, algebra and geometry on both occasions the Queensland mean scores were in general higher than those of other states. However, there were only small changes over time in Western Australia and Tasmania, and the largest declines occurred in Victoria. There was also an increase in the new mathematics score in Queensland (Rosier, 1980).

In general, the investigation indicated that on the lower mental process items, that is, in algebra and basic arithmetic, girls were slightly superior to boys, and on the higher mental process items which include advanced arithmetic and geometry, boys slightly outperformed girls. However, Moss (1982) reported some inconsistency across states, such as in Victoria where boys outperformed girls, although girls outperformed boys in the Australian Capital Territory. In no part of Australia was there much evidence for changes in sex differences in performance over time.

Time Spent on Mathematics

In 1964, in the lower secondary schools the class time devoted to mathematics varied from a mean of 5.3 hours per week in Queensland, Victoria and Western Australia to 4.8 hours in New South Wales and 4.2 hours in Tasmania (Keeves, 1968; 1976). The time spent on mathematics instruction decreased from 1964 to 1978 (Rosier, 1980).

Keeves (1968) reported that the number of hours spent on mathematics homework per week was not linked with the level of achievement in mathematics for students at the 13-year-old level. However, he reported a positive relationship between achievement in mathematics and time spent each week on all homework. Keeves suggested that the more able students could complete their mathematics homework without spending much time on it. On the other hand, the less able students probably spent more time on their mathematics homework and achieved less. In 1978, it was found that the mean time spent on mathematics homework each week was more than two hours in each of the states.

Keeves (1976) took the time spent each week on mathematics at school and added to the time spent at home on homework a measure of the total number of hours devoted to homework each week is obtained. In his investigation, at the Year 8 level this variable was strongly linked with the level of achievement in mathematics (Keeves, 1968). Rosier (1980) argued that the time spent in mathematics classes had declined from Years 1 to 8 between 1964 and 1978 in all states except Western Australia. Rosier indicated that the main reason for the decline was the large decrease in time spent on learning mathematics at the lower secondary school level.

In 1964, the most important relationships at the 13-year-old level were between performance and the total time spent on mathematics per week at school and on homework at home. Furthermore, those students at the 13-year-old level who wanted to do more mathematics generally performed better than those merely expecting to do less. The relationships between interest and performance were relatively high (Keeves and Radford, 1969).

Students' Attitudes towards Mathematics

The students expressed less favourable attitudes towards the importance of mathematics at Population 1 level in all the five states in 1978 than in 1964. For the scale concerned with the ease of learning mathematics, there were only slight differences between 1964 and 1978. The scale concerned with enjoyment of school showed that students enjoyed school more in 1964 than in 1978. The mean scores on views of mathematics teaching were lower in 1978 than in 1964 in New South Wales and Tasmania but were higher in Victoria and Western Australia (Rosier, 1980).

Students' Liking of Mathematics

The interest of students in learning mathematics was evaluated by direct questions, for example, about their liking of mathematics relative to other subjects. On both occasions in Queensland, New South Wales, Tasmania and Victoria the percentages of students with strong liking for mathematics were about the same. However, there were smaller percentages with a low liking for mathematics in all the five states in 1978 than in 1964 (Rosier, 1980).

Home Background

There was a strong and consistent pattern of relationships between home background and the level of performance of the students in mathematics. Strong relationships were found between the number of years of education of the father and mother of students and the mathematics achievement level of the students, as well as between father's occupation and the achievement level of the students. Generally, the educational level of the father was higher than that of the mother. In 1978, Australian Capital Territory was markedly higher than the other states in the mean level of education of the parents (Rosier, 1980).

Class Size

The mean number of students in a mathematics class in 1978 was less than in 1964, and in 1978, the medium class size was 29 students per class (Rosier, 1980). However, relationships between class size and achievement in mathematics were not reported by Keeves (1968) or Rosier (1980).

The Third International Mathematics Study (TIMS) in Australia

TIMSS involved both mathematics and science achievement, however, the focus for the present study is only on mathematics. Only the findings of the Third International Mathematics Study (TIMS) in Australia are summarised here.

In the Australian Capital Territory (ACT), New South Wales (NSW), Victoria (VIC) and Tasmania (TAS), students who participated in TIMS were in Years 7 and 8, whereas in Queensland (QLD), South Australia(SA), Western Australia (WA) and Northern Territory (NT) students in Years 8 and 9 participated in the study. Lokan et al. (1996) have reported that at the Population 2 level about 14000 Australian students from 587 mathematics classes in 180 schools participated in TIMS. The data collected from these 14000 students were analysed and reported by Lokan et al. (1996) and the summary of their research finding are presented below.

Lokan et al. (1996) have argued that WA, ACT, SA and QLD performed significantly better than the other states. Furthermore, these researchers pointed out that Victoria, Tasmania and the Northern Territory achieved at an equivalent level with each other, but at a lower level than the other states.

Lokan and her colleagues also contended that at the upper level (Years 8 or 9) Australian students were ranked ninth on the ladder of achievement when compared with the 45 participating countries, whereas in the lower level (Years 7 or 8) they were at the eighth place. Furthermore, there was no difference between boys and girls in mathematics achievement in 1994 in Australian schools. Serious doubts exist for these comparisons because of the sample design and response rates in Australia.

Conclusion

In this chapter the importance of international comparative studies, in general, and in particular the international mathematics studies conducted by IEA has been discussed. Australia as a member of IEA, administered the FIMS, SIMS and TIMSS studies in 1964, 1978 and 1994 respectively, at the 13-year old student level. There were sufficient numbers of common items in the three IEA mathematics tests at the 13-year-old level for equating the three tests and examining change in achievement over time. The 1964 and 1978 Australian data had been analysed to compare the 13-year old students' mathematics achievement on these common items across time by Rosier (1980) and Moss (1982). However, these studies did not scale the data and a common mathematics scale was not developed to measure mathematics achievement over time.

Statistical techniques are now available by means of which the mathematics achievement data could be scaled, and as a consequence, performance on the three occasions could be brought to a common scale, thus providing a scale of mathematics achievement that would be independent of both the samples of students tested and the samples of items employed. Hence, the scaling of the mathematics tests across occasions becomes the major task for this study and is discussed in Chapter 8.

One of the main concerns of the TIMS project was for curriculum development. Therefore, the next chapter deals with the curriculum change in Australia and Ethiopia.

3

Changes in Mathematics Education in Australia and Ethiopia

The International Mathematics Studies conducted internationally by IEA is discussed in Chapter 2. These international studies sought to evaluate the mathematics curricula developed and implemented by the participating countries. Australia was one of the countries that participated in all three international studies. A similar study was also conducted in Ethiopia in the first half of 1996 by the investigator. Therefore, it is very important to review the development of the mathematics curricula in both countries, and to examine the relationship between the curricula of the two countries. This review provides information on whether or not it is possible to compare mathematics achievement between the two countries. Hence, the first part of this chapter deals with mathematics curriculum development in Australia since the 1960s, while, the second part reviews the development of the mathematics curricula in Ethiopia since the 1940s.

Changes in Mathematics Education in Australian Schools since the 1960s

There have been fundamental changes in the mathematics curricula in Australian schools over the past 30 years. These reforms were in part influenced by overseas programs (Blakers, 1978; Owen et al., 1983). Different overseas projects, such as the Nuffield and the School Mathematics Projects (SMP) from England and the School Mathematics Study Group (SMSG) and University of Illinois Center for the Study of Mathematics Project (UICSM) from the United States of America (King, 1975; Blakers, 1978; Owen et al., 1983; Clements, 1989) influenced in part the changes that occurred in the mathematics curricula in Australia. However, the developments that occurred in Australia were not a direct copy of changes that had taken place overseas. Moreover, there were substantial differences between different parts of Australia.

In Australia, each state is responsible for the education of its children (Blakers, 1978). Hence, the curriculum changes that have occurred over the last 30 years have not been the same in all states and territories of Australia. The influences of overseas curriculum developments have also been different from state to state. Thus different states were influenced to different extents by different overseas programs. South Australia was partly influenced by SMP and Western Australia by SMSG, UICSM, and the Nuffield Project, Edith Biggs and SMP and other British, American and Canadian projects (Blakers, 1978; Clements, 1989). However, MacDonald (1977) has suggested that Australian educational systems were influenced more by the developments in the United States with respect to the school mathematics curricula than by those from England, but no sound evidence is advanced to support this claim.

In the 1960s the influence of the overseas developments on change in the Australian mathematics curricula was reflected in the "new mathematics" which was the subject of considerable debate at that time (Owen et al., 1983; Clements, 1989). MacDonald (1977) also contends that the outcome of the influence on the transition to new mathematics curriculum in many parts of Australia combined some of the worst features of change in both the United Kingdom and the United States.

The Introduction of New Mathematics in Australia

The origin of the new mathematics was in the United States and in England, and in the early 1960s it was introduced to the Australian States (Clements, 1989). In 1962, the Australian Mathematics Society conducted a seminar on the new mathematics at Sydney University. Over 100 professors, school inspectors and secondary school teachers from all states in Australia and New Zealand participated in the seminar. The seminar participants called for a reappraisal of content and methods of presentation in school mathematics (Cowban, 1987). Changes followed immediately in New South Wales as a consequence of the introduction of a new system of secondary education, the Wyndham Scheme, but changes did not take place in the other states until several years later. In the mid-1960s, as a result of a UNESCO seminar and other external influences, new secondary school mathematics courses were developed across Australia. New text books were also written by Australian authors. In the new mathematics courses students were encouraged to discover mathematical principles by using inductive and deductive methods (Cowban, 1987; Clements, 1989). Mathematics educators, in general, approved the changes that also stressed the necessity for precision of language and symbolism, and the application of integrating concepts such as algebric structure, functions, sets and transformation geometry (Cowban, 1987; Clements, 1989). However, Clements (1989) has suggested that there was a serious shortage of qualified teachers to implement the new mathematics.

Nisbet (1978) has argued that the introduction of the new mathematics into Australian schools was a result of a combination of historical, political, scientific, technological and educational factors. McQualter (1980) has indicated that a growing dissatisfaction with the school mathematics curriculum and the increasing use of mathematics by all sections of society were the major reasons for the introduction of new mathematics into Australian schools. He also pointed out that the essential features of the new mathematics movement in Australian schools followed an overseas trend. Moreover, the six educational systems in Australia had highly centralised systems of educational administration and each had a tendency to look inward and ignore what was going on in the other states. As a consequence, the introduction of the new mathematics began at different times in each state (McQualter, 1980). In 1966, the new mathematics had been introduced into the lower primary grades (Grades 1 and 2) in Queensland as well as in the other states. By 1967, it was introduced into the middle primary schools, and by 1968 into the upper primary schools. The new mathematics program was introduced into secondary schools in all the states by the 1970s.

Even though the introduction of the new mathematics went through successive stages of development in order to implement the changes effectively, there were strong reactions to change in mathematics education in Australian schools. There were six main reasons for these reaction.

1. The courses were tailored only for the top students.
2. An enormous number of poor quality text books overwhelmed the market, and poorly prepared teachers seized on them as compensation for their own lack of understanding of the course changes.
3. The aim of the change was on content with no attention to teaching method and the problems that arose.
4. The applications of mathematics were largely ignored.
5. Rigour in proof, formalism, and symbolism was over emphasised and seen as an end in itself.
6. The reactions of employers were generally unfavourable. (Nisbet, 1978, p. 125)

McQualter (1980) reported that by 1976 the mathematics education in schools in all the states of Australia had changed substantially. The new courses in mathematics education in Australian schools attempted to:

(a) provide variety in mathematics courses;

(b) integrate mathematics courses offered at primary and secondary school levels; and

(c) meet the increasing demand for more persons with mathematical backgrounds to go into a wide range of work and as in Europe and the United States, Australian society had become ‘mathematised’ (McQualter, 1980, p. 58).

Between the mid-1960s and the early 1970s Australia introduced decimal currency and the metric system of measurement in Australian schools. Decimal currency was introduced into schools in 1964 and by 1968 all grades had full decimal currency syllabuses. The time between 1965 and 1967 was planned as a time of transition (Australian Council for Educational Research, 1964). In the early 1970s, Australia started the process of converting to the metric system of measurement (Blakers, 1978). The introduction of decimal currency and the metric system of measurement in Australian schools demanded that changes be made to the mathematics curriculum in schools. The introduction of decimal currency into the mathematics curriculum was expected to save time in the teaching of arithmetic (Australian Council for Educational Research, 1964). It was considered important that this time should be given to teaching new content in mathematics.

Clements (1989) contended that in the late 1960s and early 1970s there were many objections to the “new maths” from university mathematicians and harsh calls were made for change in courses. New textbooks were subsequently published in which algebraic structure, set language and transformation geometry were hardly mentioned. Clements (1989, p. 49) explained the changes in the following way. “The pendulum had swung back, and the New Maths was now identified in the public mind with an over emphasis on abstruse theory and a lack of adequate attention to applicable mathematics.”

The changes that occurred in the Australian mathematics curriculum were not only related to the inclusion of the new mathematics, and are reviewed separately with respect to the primary and secondary school mathematics curricula.

Primary School Mathematics Curriculum

Blakers (1976, 1978) and Owen et al., (1983) reported that in the 1950s the primary school mathematics curriculum in Australian schools consisted fundamentally of arithmetic. The expected outcomes of this subject were speed, accuracy and neatness (Blakers 1976, 1978; Owen et al., 1983; Brinkworth, 1985). When the curriculum of the 1950s was compared with that of the late 1970s, the primary school mathematics curriculum in the late 1970s tended to have more emphasis on structure, some work on set language, basic arithmetic, spatial relations, probability and statistics and less emphasis on computational skills (Blakers, 1978; Owen et al., 1983). The evidence available indicates that the programs reflected substantially formal approaches to number and structure despite different projects like the Nuffield program which emphasised activities and problem solving. Jones (1979) investigated the changes in primary school mathematics education in Queensland since the late 1950s. His findings on the content changes were similar to those reported in the review by Blakers.

The main curriculum change in Australian primary school mathematics commenced in the early 1960s when the Australian Council for Educational Research (ACER) organised a conference which was held in Melbourne in March, 1964 for persons responsible for the mathematics curriculum development for primary schools in Australia (Australian Council for Educational Research, 1964; Blakers, 1976). The agenda for the conference were:

(a) the introduction of decimal currency and its implications for primary school mathematics students;

(b) the place and value of structured aids; and

(c) a consideration of appropriate course of study in mathematics for Australian primary schools students (Blakers, 1976, p. 153).

Blakers (1976) believed that the conference and its published outcome were of significance for the reform of the primary school mathematics curriculum. The report of the conference also indicated the importance of dividing the curriculum into different levels, to break down the grade level structure of the existing curriculum (Australian Council for Educational Research, 1964). In this report the primary school arithmetic course was not separated from the total mathematics course, but it was a part of the whole. On these grounds, the report recommended that the primary school number and arithmetic courses should be changed to a mathematics course (Australian Council for Educational Research, 1964).

Four main reasons were suggested by the report (Australian Council for Educational Research, 1964) for including specific content in the primary school mathematics program. These reasons were:

(a) to contribute to victorious daily living,

(b) cultural importance,

(c) to develop desirable attitudes in the student, and

(d) to give experience with basic patterns of mathematics thinking.

The 1964 conference proposed the content for the new primary school mathematics curriculum and developed a teachers' reference book called *Background in Mathematics*, which was divided into three parts. These were: mathematics, the child, and the curriculum; basic ideas of mathematics; and evaluation, and included an extensive glossary (Australian Council for Educational Research, 1965).

In August 1975, a conference was again organised by ACER to evaluate developments in the teaching-learning processes of mathematics in the primary schools of Australia in the 1960s and early 1970s and to set strategies and a framework for future development of the primary school mathematics curriculum (Jeffery, 1975). The objectives of the conference were to:

(a) review developments over the last decade in the curriculum and outcomes of mathematics in the primary school;

(b) apply this review and the current state of knowledge about mathematics and its applications to assess current approaches to mathematics in the primary school;

(c) propose at least in outline form any recommended changes to present objectives and, if such followed, changes to content, sequence, approaches and assessment (This would in essence be 'Guidelines for the Next Decade');

(d) outline the implications of such changes for the preparation of those teaching mathematics to primary school children, and those advising teachers (These would include both preservice and inservice education); and

(e) arrange for on-going consultations between the curriculum officers concerned, and between them and those preparing teachers, to help to provide adequate resource materials for use in schools. (Jeffery, 1975, p. 2)

Jeffery (1975) cited ten important results of the 1975 conference. The results that are directly related to the primary school mathematics curriculum are summarised here.

1. The expected outcomes (attitude, understanding, skills and knowledge, and intellectual ability) of the primary school mathematics curriculum which were stated in *Background in Mathematics* were accepted and also the conference agreed to give equal emphasis to the four areas of expected outcomes, that is attitude, understanding, skills and knowledge, and cognitive abilities.

2. In the late 1960s and early 1970s emphasis was given to course content and structure and to familiarising teachers with those changes. However, the conference participants agreed to give greater emphasis to the development of more effective learning situations for the child in the primary school mathematics curriculum.

3. Changes or modifications of content were recommended from the 1964 conference content suggestions. The recommended content for change or modification were: area and circumference of circles, irrational numbers, modular arithmetic, operations with fractions, proportion, systems in other bases and work in other bases.

However, the final decision on such kinds of change was the responsibility of the state or authority or individual school.

After the 1975 curriculum conference, education departments in most Australian states started revising their primary school mathematics curriculum. The South Australian Education Department, for example, revised and published the *Primary School Mathematics Interim Revision, 1975,* which was a modification of the previous curriculum, by considering the recommendations of the 1975 primary school mathematics conference review (South Australian Education Department, 1975).

The next important changes started in most states of Australia in the mid 1980s. Lovitt et al. (1985) indicated that in 1985 all states and territories of Australia were involved in the reforming of their mathematics curricula, syllabus development and related support systems. These authors believed that the pressure for change was based on the influences of the different reports and ministerial papers in Australia and from overseas, such as the Curriculum and Evaluation Standards for School Mathematics that was published by the National Council of Teachers of Mathematics (1980) in the United States of America and the Cockroft Report from England. Different and uncoordinated changes occurred in different parts of Australia in response to these calls for revision of the primary school mathematics curriculum.

The 1986 study on *Mathematics Curriculum Materials Evaluation* indicated that primary school mathematics teachers believed that the objectives of the course were achievable regardless of the teaching styles (South Australian Education Department, 1986). Consequently, there was not much change in the primary school mathematics curriculum in South Australia.

Jamieson and Karmelita (1989) indicated that the reform of the primary mathematics syllabus in Western Australia, was expected to be implemented at any time over the three year period from 1989-1991. The revised syllabus was based on the underlying philosophy and aims of the 1978 syllabus, which was developed on the basis of the 1975 primary school mathematics conference recommendations.

It was in 1976 that a major reform of the mathematics curriculum for Grades K-6 took place in New South Wales (Australian Mathematics Teacher, 1989). After ten years, in 1984 the second major revision was started by the Basic Learning in Primary Schools program. The draft document of the primary school mathematics curriculum was completed in 1988 and the final draft was prepared for 1989.

Reeves (1989) pointed out that in 1989, a revision of mathematics education for Grades K to 12 was undertaken in Tasmania. At the time of review *Primary Mathematics Guidelines* were the current syllabuses. These documents had been prepared in the mid 1970s and were published in 1978. In 1989, the review of the primary and secondary school mathematics syllabus started and the main duty of those who were involved in the revision program was to produce a K to 12 Statement of Principle.

From 1960 to the late 1980s, there were thus three primary school mathematics curriculum changes in all states of Australia. The first curriculum change took place from 1965-1968 in most states. The change began after the 1964 primary school mathematics conference organised by ACER. The second primary school mathematics curriculum change started after the second primary school mathematics conference which was held in 1975. In most states by 1978 a new mathematics curriculum was implemented. The third change in most states started in the late 1980s. However, from 1990, onwards there was a national movement for the implementation of the national mathematics framework. A review of this national mathematics framework is discussed in a later section of this chapter.

Secondary School Mathematics Curriculum

Blakers (1977, 1978) reviewed the changes in the content of mathematics education since the 1950s. He indicated that associated changes in attitude were significant. A decrease in Euclidean geometry and an increase in emphasis on probability and statistics were carried out at the middle secondary school level. He also pointed out that the *retention rate,* which is the proportion of the age group that completes 12 years of schooling, in the late 1950s was around 20 per cent in most states. However, by the late 1970s the figure had grown to around 35 per cent. This increase according to Blakers, had resulted in the presence of less able students in mathematics classes in the upper secondary schools. Some states developed new syllabuses which were appropriate to the needs and ability levels of the less able mathematics students. The Australian Academy of Science Mathematics Curriculum Project, the first national mathematics materials development project in Australia, prepared curriculum materials for less able mathematics students in the upper secondary schools across Australia. However, the courses had less acceptance by students, because these were seen to be less relevant when compared to the normal courses (Blakers, 1978; Owen et al., 1983).

Rosier (1980) in his analysis of the First and the Second International Mathematics Study of Australian data which were collected in 1964 and in 1978 respectively, found that at the Year 12 level, the number of topics available in 1978 was more than that in 1964. Firth (1981), and Owen et al. (1983) identified different problems related to the mathematics curriculum. These authors pointed out that there was a shortage of qualified mathematics teachers, hence unqualified teachers were teaching mathematics in the lower secondary schools in Australia. Kennedy (1981) and Owen et al. (1983) also found that about one third of Victorian state secondary schools had at least one unqualified mathematics teacher.

In the 1970s one of the major influences on the school mathematics curriculum was the emphasis on school based mathematics curriculum development, a move that was strongly supported by the then operating Curriculum Development Centre. School based mathematics curricula were developed especially in South Australia and Victoria (Owen et al., 1983). Clarkson (1979) investigated the establishment of the Rusden Activity Mathematics Project (RAMP) that started in 1973. The project was a Victorian attempt to provide a resource network for secondary school teachers who were participating in the development of school based mathematics curricula. Clarkson (1979) found that experienced teachers, who were competent in traditional textbook-oriented styles of teaching, were not effective in implementing the methods contained in the activity-based orientation of RAMP.

The Reality in Mathematics Education (RIME) project in Victoria which focused on problem solving as the main objective of mathematics education for the 1980s was reviewed by Lowe (1979). RIME was involved in building up a bank of problem solving lessons that could be used by schools.

Another program which was proposed in the early 1980s was the Australian Mathematics Education Program(AMEP). AMEP was established in 1980 (Carss, 1984). The major important contribution of this project before it was terminated in 1981 was the production of a statement of 10 basic mathematical skills needed for effective participation in Australian society. The skills were: number skills and computational skills, geometry, measurement, estimation and approximation, alertness to the reasonableness of results, reading, interpreting and constructing tables and graphs, using mathematics to predict, problem solving, applying mathematics to everyday situations and language (Carss, 1984; Clements, 1989).

In New South Wales there was a common course for Year 7 and Year 8 and three equivalent courses in Year 9 and Year 10. The Year 9 and Year 10 courses were introduced in 1984, while the Year 7 and Year 8 courses were introduced in 1988 (Australian Mathematics Teachers, 1989). The syllabuses of Years 7 and 8 included strands in Algebra, Geometry, Measurement, Number and Statistics. Within each strand there were Guidelines and statements of contents, objectives, applications, skills, and discussions of the ideas and processes that formed the topics (Australian Mathematics Teachers, 1989). The Years 11 and 12 courses in mathematics in New South Wales upper secondary schools remained stable, while in 1990 new revised materials were completed.

At all levels in state schools in Victoria there were changes in the mathematics curricula (Horne and Stephens, 1989). The greatest change in the 1980s took place in Years 11 and 12. This change was related to the introduction of the Victorian Certificate of Education (VCE). The new mathematics curriculum for Years 11 and 12 was on trial in Year 11 and Year 12 in some schools in 1989.

From 1983 onwards the Queensland Education Department worked towards developing a single syllabus for the compulsory years of education. Grace (1989) reported that the department subsequently released two documents: *The Years 1 to 10 Mathematics Syllabus* and *The Years 1 to 10 Mathematics Teaching, Curriculum and Assessment Guidelines.* All state secondary schools started using the new syllabus at Year 8 and at Year 9 in 1989.

In 1987, the Mathematics Syllabus Committee of the Education Department of Western Australia took on the responsibility of revising the upper secondary school mathematics syllabuses (Jamieson, and Karmelita. 1989). In 1988, the Committee produced six courses algebra, geometry, and trigonometry; introduction to calculus and statistics; calculus; discrete mathematics; mathematics principles and applications and statistics and modelling (Jamieson, and Karmelita. 1989).

In Tasmania the Year 9 mathematics syllabus was revised in 1989 and implemented in 1990 and the syllabus was the first to be recognised by the Tasmanian Certificate of Education (Reeves, 1989). In 1987 and 1988, the syllabus of Applied Mathematics was developed and introduced to schools in 1989 in Year 11. Reeves reported that the syllabus had the following main topics: algebra models, applied geometry, finance and optimisation.

In general, changes in the secondary school mathematics curriculum took place in the 1980s in most of the Australian states. However, at the end of the 1980s, the Australian Educational Council decided to introduce the Australian Mathematics Framework.

The Australian Mathematics Framework

The Australian Educational Council, which represented the Ministers for Education in all states and territories of Australia at its 57th meeting decided that the directors of curriculum for the various states and territories should undertake a curriculum mapping exercise (Australian Educational Council, 1989). In this project the similarities and the differences in the curriculum between systems across Australia were assessed. This study found that generally the state curriculum goals for primary and secondary schools were similar and mathematics was a subject that was common in all the systems across Australia (Australian Educational Council, 1989). The curriculum mapping study indicated that at the time of the study, primary school mathematics curriculum development was being given priority in South Australia and Tasmania. Secondary school mathematics had the priority in Victoria, while for New South Wales and Western Australia both primary and secondary school mathematics education had been given priority for reform.

Through recognising the different changes undergone by different systems and realising the importance of change in the mathematics curriculum in Australian schools, the Australian Educational Council when it met in April, 1989 proposed the development of a national mathematics curriculum framework (Stephens, 1989a) based on the common areas of agreement between states about the future directions of school mathematics. The council also approved ten common and agreed national goals for Australian schools (Australian Educational Council, 1989). In addition, the Australian Educational Council (1990a) commissioned the development of a National Statement about school mathematics education. The main purpose of the National Statement was to provide a framework so that states and schools might develop their own curricula and identify important components of a mathematics education for the great majority of students.

The importance of mathematics as stated by the Australian Educational Council (1990a) was: to be useful in everyday life, at work, and to be part of the culture of all Australians.

The Australian Educational Council resolved that the school mathematics curriculum needed to be changed because of the changing needs of daily life of the society both at home and at work, the development of mathematics, the impact of computing technologies, and the need for participation in, and success with mathematics. The Council believed that :

> ... mathematics curricula in Australia must respond to the changing nature of mathematics and the mathematical demands on people in Australia cannot be static. It also recognises that schooling cannot prepare people for all the mathematics they are likely to need in their civic and working lives. If people are to continue to use mathematics, they must develop the competence, confidence and interest needed to become lifelong learners of mathematics. (Australian Educational Council, 1990a, p. 10)

In 1990, the Australian Educational Council published *A National Statement on Mathematics for Australian Schools,* which included the goals of school mathematics, the scope of the mathematics curriculum, strands and bands of mathematics. The strands described the contents, mathematical understandings, skills, knowledge and processes which the students should acquire. These strands were divided under eight headings. attitudes and appreciations, mathematical inquiry, choosing and using mathematics, space, number, measurement, chance and probability, and algebra (Australian Educational Council, 1990b). All these eight strands were covered in all the four bands of the curriculum. The only differences were in the level of complexity of the strands on different bands. Table 3.1 shows the complexity of strands at different band levels for one strand (Activities in Space).

The bands are related to years of schooling. Band A includes from Year 1 to 4, Band B includes Years 4, 5 and 6, Band 3 includes Years 6, 7, and 8 and Band 4 extends from Years 9 to 12. The Curriculum Corporation also published *Using the Mathematics Profile* which helped to elaborate the expected performance of students in mathematics at a certain level of development. The materials developed were published by the Australian Educational Council and the Curriculum Corporation and they contributed to the implementation of the national mathematics framework.

Like all the past mathematics curriculum changes, the development of National Statement has been influenced in part by the *Curriculum and Evaluation Standards for School Mathematics* that was published by the NCTM in 1989 in the United States of America (Stephens, 1989b). Stephens suggested that to implement the *National Statement* in schools, there must be resource support from Federal and State governments. Grace and Carss (1989) agreed with Stephen’s suggestion on the funding support from the Federal and State resources. They believed that without the "support the project would be another manifestation of the Australian theory of curriculum development by *wand waving* " (Grace and Carss, 1989, p. 13).

Table 3.1 Examples of design problems in Space Strand at different Band levels

Bands	Activities in Space (Strand 4)
A	Solve simple design problems (eg. use this piece of paper to make a box without a lid).
B	Design and make packages as specified (eg. send an odd-shaped Lego construction through the mail).
C	Build a skeleton models of 3-D objects to specifications (eg. a polyhedron with 20 edges and 12 vertices).
D	Research the origins of different geographic map projections, the mathematics underlying them and the distortions inherent in them.

Source: Australian Education Council and Curriculum Corporation (1991). *A National Statement on Mathematics for Australian Schools.* Carlton: Victoria.

Baxter and Brinkworth (1989) cited questions about the status, specificity, context, validity and implementation which would help establishing minimum standards for the national framework. The authors believed that the proper answer to their questions would help to implement the national framework. Grace and Carss (1989) also indicated that status, specificity and implementation of the *National Statement* should be addressed, otherwise it would be difficult to implement the *National Statement*. Stephens (1989a) pointed out that the national mathematics framework would only be effective if the necessary support were given to mathematics teachers. However, Williams (1991) criticised the *National Statement* as unattainable.

The main criticisms of the national framework have involved the shortage of resources for implementation. If this shortage of resource were resolved either at the Federal or State level, the effective implementation of the framework might lead for the development of a national mathematics curriculum in Australia. The education departments of the states are currently modifying their curricula on the basis of the national statement. It would seem that they are committed to develop curricula that are consistent with the *National Statement*. Reeves (1989, p. 9) reflected the view of the Tasmanian Department of Education concerning the development of mathematics curricula based on the *National Statement* by saying " if there are major discrepancies between the Tasmanian approach and that outlined in the National Statement, 'fine tuning' will be undertaken when and as required." Hence, the Australian primary school mathematics curriculum which was advanced in the 1960s is, in the 1990s moving rapidly towards the development of an Australian national mathematics curriculum.

Summary

The Australian mathematics curriculum has passed through different stages of development over the last 30 years. On the basis of the 1964 ACER conference, the states started to develop their own mathematics curricula, and subsequently revised their curricula in the mid-70s and the late-80s. At this time efforts were made to develop similar mathematics curricula throughout the country. In order to achieve this objective further measures were required by the Australian Educational Council. In the late 1980s the Council issued the Australian Mathematics Framework and at the beginning of the 1990s, it produced the *National Statement* and *Profiles* which was the first major step towards the establishment of a common mathematics curriculum across Australia.

This great effort and experience in Australia of developing a common mathematics curriculum for use throughout the country provides important lessons for countries like Ethiopia which are trying to regionalise their curricula. The next part of this chapter discusses the curriculum changes in Ethiopia since the 1940s.

Mathematics Education in Ethiopia

Before the beginning of the twentieth century, formal education was provided by the Ethiopian Orthodox Church. The church schools were established for the teaching of church music, Ethiopian law, Geez language and literature, poetry and theology (Imperial Ethiopian Government, 1956). The Church apart from its main spiritual role had always acted as the guardian and preserver of Ethiopian traditional culture whose continuity had been assured by maintaining church schools throughout the country (Imperial Ethiopian Government, 1956). Therefore, through the centuries education has been the preserve of the Church. The contribution of the Church was not only to maintain a supply of educated clergy for the various services of the Church, but also to produce a flow of educated men who filled the various ranks of the civil service.

However, traditional Ethiopian education emphasising church subjects as it did, fell short of the requirements of a society that was undergoing rapid modernisation. Furthermore, Ethiopia's increasing contact with the outside world made it more urgent that the Government should supplement the church schools by establishing modern institutions for the teaching of a western type of education (Imperial Ethiopian Government, 1956). The belief that this type of formal education should be given to the Ethiopian children and that in this respect the Ethiopian Government should play a leading role had gradually gained momentum during the latter decades of the nineteenth century and at the beginning of the twentieth century.

Hence, western education was formally introduced into Ethiopia in 1907 when the Menelik II school was opened by the Ethiopian Government. This school was the first Government school built in the capital city, Addis Ababa. The school got its name from Emperor Menelik II who was Emperor of Ethiopia at this time and the first person to introduce modern education to the country. The program to establish government schools was designed to provide selected groups of students with the linguistic and other skills that were necessary to enable Ethiopia to maintain satisfactory relations with other countries. The new schools were largely staffed by foreign personnel and the curriculum presupposed a thorough understanding in the national language and culture. Priority was given to French, while Arabic, English and Italian languages were optional subjects. But their teaching program gradually expanded until it covered the main subjects of mathematics, science, languages, history and geography and the level of schooling was divided into two, namely, elementary and secondary.

From 1907 up to 1935, modern education was introduced to different parts of the Ethiopian Empire. However, this expansion of modern education was disrupted between 1935 - 41, by the war between Ethiopia and the Italian Fascists. During the Italian occupation all the modern schools were closed and a policy of systematic elimination of those Ethiopians who had acquired western education practised. The result was that the network of modern educational institutions was destroyed and large numbers of educated people were killed. The Ministry of Education and Fine Arts (1955, p. ix) reported the situation as follows.

> Programs were developed, new schools were opened for boys and girls, and many students were enabled to pursue specialised courses abroad. The work of these schools was, however, abruptly terminated by the invasion of Ethiopia by Italy. A large proportion of those Ethiopians who had benefited from the educational opportunities provided here and abroad were killed ...

However, the Italians were defeated in 1941, and the Ethiopian Government started reconstructing and developing the infrastructure of the whole country including modern education. Hence, the reintroduction and expansion of modern education started again in 1941 (Ministry of Education and Fine Arts, 1955). Elementary and secondary schools were built in different parts of the country, and the teaching in these schools was undertaken by some Ethiopian and expatriate teaching and administrative staff (Imperial Ethiopian Government, 1956). However, for the first eight years, that is, from 1941 to 1948 there was no central curriculum. Every school had to develop its own curriculum and also adopt textbooks which were needed to implement the school curriculum.

In 1947/48[1] the first central elementary school curriculum was introduced by the then Ministry of Education and Fine Arts (MOEFA). The subjects that were included in the elementary school curriculum were: Amharic, English, Arithmetic, Social Studies and Science (Ministry of Education and Fine Arts, 1947/48).

The Ministry of Education and Fine Arts also developed the first secondary school curriculum in 1951/52 (Ministry of Education and Fine Arts, 1951/52). From the late 1940s up to the present time there were centrally planned curricula to be implemented in all junior and senior secondary schools. However, since 1993 the primary school curriculum has been decentralised. The country was divided into 14 ethnic based regions. As a consequence, the preparation and implementation of elementary school curricula became the responsibility of each ethnic region. It is important to note that unlike Ethiopia, the curriculum in Australia is still school centred and the Federal Government is trying to develop more centralised curriculum.

Among the subjects taught in Ethiopian schools was mathematics. Therefore, the next section discusses the changes in mathematics education in Ethiopian schools from 1941 up to the present time.

Changes in Mathematics Education in Ethiopian Schools0 since 1941

The first Ethiopian modern education curriculum was introduced in 1947/48. Since 1947 there have been many curricular changes in the Ethiopian educational system, including changes in mathematics education.

Elementary School Mathematics Education

In the 1940s, the 12 years of education were divided into eight years of elementary and four years of secondary education. The first mathematics syllabus which was incorporated with the first Ethiopian modern education curriculum was developed in 1947/48 by the MOEFA. At that time the elementary school mathematics curriculum was known as Arithmetic.

The 1947/48 elementary school syllabus stated that the arithmetic course included the various skills, activities and experiences which were to be developed by students under the guidance of teachers. The arithmetic curriculum was framed with the following principles as a guide to include the arithmetic skills:

(a) which the communities, the provinces, or Ethiopia as a whole could put to use in actual living situation; and

---

[1] Ethiopian calendar dates have been converted to dates used in the rest of the World by the addition of seven or eight years

(b) to include the arithmetic skills necessary for higher mathematical attainment in secondary schools. (Ministry of Education and Fine Arts, 1947/48, p. 66)

The syllabus also recognised that instruction in Arithmetic should deal with two major phases of the subject, its mathematical and social phases. In the mathematical phase two outcomes were stated, namely, skill in computation and mathematical understanding.

For mathematical understanding the student had to develop:

(a) meaningful conceptions of quantity and of the nature and structure of the number system;

(b) a vocabulary of useful technical terms which designate quantitative ideas, such as average, and sum;

(c) an understanding of a unit of measure and of the process of measurement;

(d) the ability to make approximations, estimations, and check work; and

(e) a grasp of important arithmetical generalisations and the relationships underlying computational procedures. (Ministry of Education and Fine Arts, 1947/48, pp. 66-67; 1962/63, p. 25; 1967/68a, p. 43; 1970/71a, p. 43; 1972/73a, p. 43)

The outcomes related to the social phase of arithmetic were:

(a) an understanding of the nature and evolution of the institutions and agencies through which number functions in the affairs of daily life, such as the uses of money, methods of telling time, kinds of taxes, tariffs and duties;

(b) a meaningful vocabulary of quantitative terms and knowledge about arithmetic practices utilised in business, economic, social and industrial affairs;

(c) the disposition and opportunity to utilise the precise, definite techniques of mathematics in dealing with problems and activities of daily life, especially those related to production, distribution, and consumption;

(d) the ability to use and interpret graphs, maps, charts, tables, and simple statistical procedures encountered in reading, and to evaluate the information they present; and

(e) the desire to improve and extend the ways in which the individual deals with the quantitative elements of situations encountered in social and intellectual experiences. (MOEFA, 1947/48, p. 67; 1962/63, pp. 25-26; 1967/68a, pp. 43-44; 1970/71a, pp. 43-44; 1972/73a, pp. 43-44).

In accordance with the above principles and specific outcomes, the Ministry of Education and Fine Arts adopted a set of arithmetic textbooks—*Arithmetic We Use* (Brueckner, 1942) published in the United States. It was argued by the Ministry that this set of arithmetic books had a very fine teacher's guide and the textbooks and the teacher's guide provided a very detailed course of study for the elementary school. The main aim of the Imperial Ethiopian Government in the adoption of these books was to make the education system comparable with that of western countries. This was stated by the Ministry of Education and Fine Arts (1947/48, p. 2) in the first elementary school curriculum as follows.

> Ethiopia stands uniquely alone in Africa, and therefore, is not interested in a colonial type of education. The Imperial Ethiopian Government wants its children to compare educationally with those of other lands, and is proud to know that its youth abroad now maintain the same standards as other youth enrolled in the great educational institutions of the world.

These adopted text books were distributed to schools and began to be implemented. However, teachers and students faced many difficulties in using these books. Even if the MOEFA adopted the text books for comparable purposes, in practice the books were not suitable for the Ethiopian situation. The MOEFA reviewed the books and identified the following problems encountered with the *Arithmetic We Use* textbooks. All the examples were Americanised, the expressions were difficult for Ethiopian teachers to understand and the British system of measurement was not explained well (Ministry of Education and Fine Arts, 1950/51).

In order to overcome the problems, the MOEFA started looking for alternatives and decided to change the adopted text from *Arithmetic We Use* to *Highway Arithmetic* (Francis, 1950).

The 1947/48 curriculum served for about ten years and was replaced by a new experimental curriculum in 1958/59. Before 1958/59, the medium of instruction at all levels of education was English, but in 1958/59, the first experimental curriculum was introduced to change the medium of instruction from English to Amharic from Grades 1 to 6 (Ministry of Education and Fine Arts, 1958/59a). Hence, the medium of instruction for mathematics for those grade levels was also changed to Amharic. Therefore, a new mathematics syllabus was introduced in the same year. The new elementary school mathematics program was arranged as a sequence of nine topics ordered according to their level of difficulty. There was also specific division into year levels of work. Meanwhile, it was expected that on average a topic could be completed in one term for the first four years and seven topics could be covered during the last two years. On the assumption that all the topics would be dealt with in a way which would be meaningful to the students, the new mathematics syllabus was given in an outline form (Ministry of Education and Fine Arts, 1958/59a). Those who developed the syllabus thought that mathematics teaching should be integrated with all other subjects, therefore, the practical application of arithmetic could be found in other subjects such as, natural science, geography, arts and crafts, gardening and cooking. The new elementary mathematics program which was given in an outline form included the following topics: the concept of numbers, fundamental operations on integers, decomposition of integers, properties of numbers, measurements, fractions-proper and improper, decimals, percentages, approximations, applications, solving equations, straight lines and angles, the triangle, the circle, polygons, practical uses of geometry, properties of space figures and indirect measurements (Ministry of Education and Fine Arts, 1958/59a, pp. 24-27, 190/61, pp. 2-11).

This syllabus was in use for about two years and was then revised for the third time. Further revisions of the syllabus were carried out in 1962/63, 1967/68a, 1970/71a and 1972/73a. The aims and purposes of the mathematics syllabus which was revised five times stated the following principles of mathematics education in elementary schools. The stated principles were to include the arithmetic skills which the community could put to use in actual living situations and to include the arithmetic skills which are necessary to serve as a basis for higher mathematical attainment in secondary schools. (Ministry of Education and Fine Arts, 1958/59a, p.9; 1962/63, p. 25; 1967/68a, p. 43; 1970/71a, p. 43; 1972/73a, p. 43)

Like the principles, the outcomes and the contents were similar. The contents were: numbers, short multiplication, division with and without remainder, Ethiopian money, length, fractions, time, rate, postage, Roman numbers, scale drawing, capacity, area, decimals, unitary method, proportion, percentage, foreign currency, ready-reckoner and statistical tables, graphs and practical applications (Ministry of Education and Fine Arts, 1962/63; 1967/68a; 1970/71a; 1972/73a). These contents were specified for each year level.

In the early 1960s, introduction of a new mathematics course in schools was a progressive phenomenon in the western world. This world-wide change in the mathematics syllabus in the 1960s led the Ministry of Education and Fine Arts to adopt the new mathematics program in all its schools, since, their main purpose was to make the educational system comparable with that of western countries. Therefore, in the 1967/68 academic year, Ethiopia officially began a new program of Modern Mathematics, based on the African Mathematics Program sponsored by the Education Development Center of United States of America (Ministry of Education and Fine Arts, 1968a, 1968b). The program was based on the Entebbe texts and materials, adopted by the Curriculum Division of the Ethiopian Ministry of Education and Fine Arts. In the 1990s the elementary school mathematics syllabus of Ethiopian elementary schools is still based on this new mathematics program.

Secondary School Mathematics Education

Prior to 1950, there was no centralised curriculum for Ethiopian secondary schools. The first Ethiopian secondary school curriculum was developed in 1951. The purpose of the curriculum was to meet the needs of the Ethiopian students, to give the students the general training that they needed and to lead them to the Ethiopian School Leaving Certificate Examinations (ESLCE) and General Certificate Examination (GCE) of the University of London. Mathematics was one of the main subjects of this curriculum (Ministry of Education and Fine Arts, 1951/52). Therefore, the syllabus of secondary school mathematics education was one of the syllabuses included in the new secondary school curriculum.

In this mathematics syllabus principles, objectives and outcomes were not mentioned and only the contents of text books were listed. All the texts listed in the secondary school mathematics syllabus were from the Durell's series. Some of them were:

1. General Arithmetic for Schools-----Part I (Durell, 1950)
2. General Arithmetic for Schools-----Part II (Durell, 1950)
3. General Arithmetic for Schools-----Part III (Durell, 1950)
4. A New Algebra for Schools----------Part I (Durell, 1930)
5. A New Algebra for Schools----------Part II (Durell, 1930)
6. A New Algebra for Schools----------Part III (Durell, 1936)
7. A New Geometry for Schools--------Stage "B"-Part I (Durell, 1939)
8. A New Geometry for Schools--------Stage "B"-Part II (Durell, 1939)
9. A New Geometry for School---------Stage "B" -Part III (Durell, 1939)

This curriculum was maintained for about eight years. In 1958/59, the second edition of the secondary school curriculum was introduced in Ethiopian secondary schools. Among the subjects treated in the revised curriculum was mathematics education. The aims and purposes of secondary school mathematics education stated in the new curriculum were to:

(a) expose the students to the thinking of mathematics and to acquaint them with the structured fields of mathematical study;

(b) create in the students an awareness of quantitative as well as the spatial relationships in their physical environment;

(c) cultivate in the students attitudes such as orderliness, accuracy, the power of abstraction and logical thinking and to equip them with basic concepts, techniques and skills of the scientific approach to the solution of their everyday problems; and

(d) foster an appreciation of usefulness and the beauty of mathematics and to prepare the students for scientific study at higher levels. (Ministry of Education and Fine Arts, 1958/59b, p. 53)

However, there was no change in textbooks. Durell's books were used as textbooks in all Ethiopian secondary schools from 1951.

There was a regular revision of the secondary school curriculum in 1963/64, 1967/68, 1970/71, and 1972/73. The aims and purposes of mathematics education in all these revision years were the same. The aims and purposes were to:

(a) build logically coherent, unified courses;

(b) introduce at an early level powerful but elementary ideas, first presented by concrete examples, later, abstracted and placed in an axiomatized system;

(c) place greater emphasis on the meaning of mathematics rather than on memorisation and manipulative procedures;

(d) develop in the student the capacity for original mathematical thought; and

(e) prepare the student for scientific study at a higher level. (Ministry of Education and Fine Arts, 1963/64, p. 94; 1967/68b, p. 102; 1970/71b, p. 103; 1972/73b, p. 103)

The General Secondary Schools Mathematics Curriculum of 1967/68 emphasised the development and implementation of Modern Mathematics in Ethiopian schools. This was because, as from 1967 a Modern Mathematics Syllabus was being introduced gradually, beginning in Grade 9 and extending year by year into all senior, junior and elementary grades (Ministry of Education and Fine Arts, 1967/68b).

In 1970, the General Secondary Schools Mathematics Curriculum was revised. At this time modern mathematics was introduced into Junior and Senior Secondary Schools of Ethiopia (Ministry of Education and Fine Arts, 1970/71b). The revised curriculum gave emphasis to this new mathematics curriculum.

The mathematics curriculum was again revised in 1972/73. This revised curriculum did not have any substantial changes compared to that of the 1967 and 1970 curricula.

In 1974, there was a change of Government. The Imperial power was overthrown by a socialist military junta. The new regime wanted to change the curriculum. In 1975, a new curriculum was developed based on socialist ideology (Ministry of Education, 1975). Therefore, the Mathematics Curriculum was revised to fit the new ideology, but there was no basic change (Ministry of Education, 1980). There has been no substantial curriculum revision since 1980 in mathematics education. However, in 1982/83, the Mathematics Panel of the Curriculum Department of the Ministry of Education, responsible for the mathematics syllabus and textbook development cited the following weaknesses of the Ethiopian socialist mathematics curriculum.

1. The preparation and presentation of the courses at different grade levels did not have any relation to objective reality.
2. Most of the topics could not be covered in the allotted time.
3. The books did not meet the Ethiopian situation. They were direct copies of the former Entebbe books.
4. The problems and the exercises were repetitive and some of them were beyond the grade level of the students.

5. The curriculum was abstract and unrelated to the daily life of the students. (Ministry of Education, 1982/83, pp. 10-13)

In order to overcome these and other problems, the Mathematics Panel started revising the syllabus. At the present time the junior and senior secondary schools are using a new revised mathematics syllabus. However, the bases of this revised mathematics syllabus is Modern Mathematics. The only difference is in the arrangements of the content at each grade level.

Modern (Entebbe) Mathematics

Like Australia, Ethiopia also imported the concept of New Mathematics [Modern (Entebbe) Mathematics]. In 1962, Educational Services Incorporated, a non profit-making research body set up the African Education Program (in Science as well as in Mathematics). The Mathematics program was under the chairmanship of Professor W. T. Martin, Head of the Department of Mathematics, Massachusetts Institute of Technology (Ministry of Education and Fine Arts, 1967). A workshop was held in Entebbe, Uganda during the summer of 1962, and in succeeding summers, a series of texts for secondary and primary schools and teacher training institutes were produced. Ethiopia sent four participants to the 1963 workshop - two from the Ministry of Education and Fine Arts and two from the then Haile Selassie I University (HSIU). In January 1964, a two-week seminar was held at Addis Ababa in the University under the guidance of Professor Roy Dubisch, Washington University, Seattle. In the summers of 1964, 1965 and 1966, Ethiopia was represented at the Entebbe workshops. In subsequent years the Bede Mariam and Tafari Mekonnen schools followed a course in Grades 7, 8 and 9 based on the Entebbe Mathematics (Ministry of Education and Fine Arts, and HSIU, Faculty of Education, 1967).

In December 1966, a meeting was called to consider the Entebbe program. The staffs of the University, the Bede Mariam and Tafari Mekonnen schools were present and a sub-committee was set up to consider the possible introduction of the Entebbe course into all Grades 7, 8 and 9 in 1967. The sub-committee made a unanimous and strong recommendation to introduce the four-year course in 1967. The following reasons were advanced by the sub-committee.

1. The text-books used at that time were inadequate, old-fashioned and unsuited to Ethiopia. They contained inappropriate material and were geared to the British system of weights, measures and monetary systems, and hence to the GCE examination. They were not even uniformly used throughout Ethiopia. Some schools were using the Hudson series, others the Durell series, and a few without an adequate supply of text-books of any sort. New text-books could be produced cheaply in Ethiopia to meet quickly the increasing demands; they could include a teacher's guide; and uniformity of curriculum and teaching in the schools would be more easily attained.
2. The past results in the Mathematics A examination were indicative of the bad presentation of material to high-school students. While this could be attributed in some cases to bad teaching, the text-books were also much to blame, being unreadable by the average students. However, the Entebbe texts were in simple English and were highly suitable for private study. Also, it had been shown in the pilot school in Addis Ababa that several concepts presented in the Entebbe books (eg. negative numbers) were assimilated by the students (even the poorer ones) in a fraction of the time that those using the old text-books took.
3. The "new mathematics" was implemented extensively and successfully in other parts of the world, including many countries in Africa. It seemed a pity to leave

Ethiopia in “darkness”. Students who joined the Science Faculty had to study it there in any event, and for those who did not, it would give a far clearer and deeper understanding of everyday mathematics.

4. The machinery for introducing the ‘new mathematics’ into Grade 9 was readily available, with the facility of producing the books locally at little cost. There was a willingness of experts to come to Ethiopia to help the scheme. In addition, university lecturers were available to run seminars in the summer months, and there was an obvious incentive to the secondary school teachers, who did not want to be left behind, to attend.
5. The introduction of the Entebbe books would codify the mathematics curriculum in a way that would not be possible without unique texts for the whole country. The advantage of these books was most important together with the fact that the curriculum obtained would be of a kind which had been internationally adopted. The difficulties which might occur during the first year were minor and could be met with by working out a supplement to the teacher’s guide, stressing what under all circumstances should be taken up and gone through thoroughly, and what, in case of time constraints, could be treated less thoroughly. By providing supplements allowance could also be made for the courses given in the following year. Those teachers present who had experience with the course either from earlier years or in their work in the current year were willing to schedule the time necessary for different sections of the course. It would certainly be advantageous if the participants in the summer course could be given material worked out in these ways (Ministry of Education and Fine Arts, and HSIU, Faculty of Education, 1967, pp. 35-38).

In January 1967, the University- Ministry Seminar on Curriculum Development and the Ethiopian School Leaving Certificate Examinations was held. In this seminar the Mathematics Commission, consisting of Ministry of Education and Fine Arts, and University officials, as well as 15 senior mathematics teachers from all over Ethiopia, recommended the introduction of Modern (Entebbe) Mathematics provided that books could be printed and teachers trained (Ministry of Education and Fine Arts, and HSIU, Faculty of Education, 1967). Meanwhile, the Educational Services Inc. promised to sponsor one or two mathematicians, as well as to give substantial financial assistance to run the necessary course in August. The printing of books was physically possible and relatively inexpensive. The time factor was most important in that the books would need to be in the schools in September.

The Seminar also recommended that Modern (Entebbe) Mathematics should be introduced as soon as possible in teacher training institutes and elementary schools. It was also pointed out that the HSIU Faculty of Education was preparing Ethiopian future mathematics teachers to teach Modern Mathematics (Ministry of Education and Fine Arts, and HSIU, Faculty of Education, 1967).

At the time of the seminar, Modern (Entebbe) Mathematics had been introduced in Ghana, Kenya, Liberia, Malawi, Nigeria, Sierra Leon, Tanzania, Uganda and Zambia. In November 1967, an O level examination was offered to Nigerian students who had been studying the Modern (Entebbe) course for the past four years. The O level syllabus of the Modern (Entebbe) Mathematics was accepted by the West African Examinations Council and the Cambridge Overseas Examination Syndicate (Ministry of Education and Fine Arts, and HSIU, Faculty of Education, 1967).

After a pilot project covering three years, on the recommendation of the University-Ministry seminar and support from the United States, the Ministry of Education and Fine Arts decided to adopt the new mathematics program in all its schools. In the academic year beginning September 1967, Ethiopian schools officially began a new program of Modern Mathematics, based on the African Mathematics Program sponsored by the Education Development Center from the United States. The program was based on the Entebbe texts and materials, adopted by the Curriculum Division of the Ethiopian MOEFA. Teachers were retrained with the assistance of the Education Development Center. Throughout the year, a teacher-assistance project was in operation and proved invaluable in implementing the program and extending inservice training for the future to as many teachers, secondary and elementary, as it was possible to reach. The program was introduced in Grade 9 in September 1967 and preparations were made for its introduction in Grades 7 and 10 at the beginning of 1968-69. During 1969-70, a start was made on implementing the new mathematics program in Grade 1. In 1976/77 modern mathematics had been introduced at all education levels of the country.

The introduction of Modern Mathematics in the school system was well under way and, through the supervisory services and training programs related to this project, information was gathered on the problems of maintaining effective learning in this area. The basis of present day school mathematics in Ethiopia is still Modern Mathematics. There has been no major mathematics syllabus change since the introduction of modern mathematics into Ethiopian schools.

Summary

Western type education was introduced in Ethiopia at the beginning of the 1900s. Until the late 1940s Ethiopian schools developed their own curricula. However, the first centralised curriculum was introduced in 1947. From that time onwards there have been many changes. Mathematics was one of the key subjects in the curriculum. Ethiopian mathematics education has gone through many different curricular changes. The major change was the introduction of Modern (Entebbe) Mathematics in the late 1960s.

As a consequence, the primary and secondary school mathematics curricula in Ethiopia have gone through many changes. Many of the changes have been based on British and American courses and textbooks. For example, when Modern Mathematics was introduced to Ethiopian schools, it followed a tradition in mathematics teaching based on British and American curricula. Furthermore, since no substantial changes have occurred in the mathematics curricula at any level since the reforms of the 1960s, which were consistent with the tests employed in the 1964 First International Mathematics Study, it is argued that these tests are still appropriate in Ethiopia at the present time. These tests were developed at Teachers College, Colombia University in the United States and were carefully field tested in the participating countries before use. Consequently, they were appropriate for use in countries that followed the British and American traditions, even though they were strongly criticised by a mathematician from The Netherlands (Freudenthal, 1975). Such criticisms do not imply that they were inappropriate for use in most of the countries involved in the FIMS study or inappropriate for use in Ethiopia in 1996, which still maintained the British and American mathematics curricula traditions. The criticism merely reflected the idiosyncratic views of a continental mathematician in a country that was considered to have rejected British and American approaches to mathematics teaching.

Conclusions

In the first part of this chapter the mathematics curriculum developments in Australia since the 1960s are described, while, the Ethiopian mathematics curriculum developments are reviewed in the second part of the chapter. In the development of their mathematics curricula both countries had much in common. Some of their similarities are summarised here. Mathematics curriculum development in both countries has followed similar trends. The change of mathematics was influenced by overseas programs. Both were greatly influenced by Britain and the United States. In both countries seminars were employed as a means of introducing the new developments in the mathematics curriculum in each country.

Nevertheless, there were some differences between the two countries. The main difference was the focus of responsibility for developing a curriculum. In Australia states or schools were responsible for developing their own mathematics curriculum, because the curriculum was not centrally prescribed. Therefore, the federal government had no role on curriculum development. However, at the present time the federal government is working for the development of national curriculum. In Ethiopia on the other hand, the curriculum is centrally planned. Until 1992, the curriculum for primary and secondary schools was planned by the central government. However, since 1993, the regional ethnic based governments started developing their own primary school curricula. Nevertheless, junior and senior secondary school curricula are still centrally planned. The interesting point here is that, the Australians are striving to develop a national curriculum and national unity, whereas in Ethiopia the present government is encouraging diversity and decentralisation of the school systems. Countries are choosing their own way according to their political policies. However, it would be of interest to review previous research findings on the achievement of students in mathematics in both countries, and to examine whether or not there exist similarities between the two countries. Therefore, in the next chapter the previous research findings in Australia and in Ethiopia are reviewed.

4

Research into Student Achievement in Mathematics in Australia and Ethiopia

In Chapter 3, the development of the intended mathematics curriculum in Australia and Ethiopia is summarised and similarities and differences between the two countries are identified. In this chapter previous research findings about the achieved curricula in both countries are reviewed. The achieved curriculum is related to the outcomes that the students achieved after studying a particular topic in mathematics. Many studies have been conducted in different parts of the world to investigate the mathematics achievement outcomes of students. Among those countries which conducted research on the achievement level of students were Australia and Ethiopia. Hence, in the first part of this chapter research conducted in Australia on student achievement in mathematics are reviewed, while Ethiopian studies about student level factors that influence mathematics achievement outcomes are reviewed in part two. Part three summarises the findings reported and discussed.

Research into Student Achievement in Mathematics in Australia

Research findings have indicated that there are substantial differences between students in their achievement in school mathematics, and that the factors that influence the achievement level of students need to be investigated. These differences in student achievement in school mathematics in Australia are reviewed below in terms of gender, attitude towards and interest in mathematics, the difference of time and opportunity to learn, language differences and socioeconomic status.

Gender Differences in Mathematics Achievement

Australian researchers such as Carss(1980), Leder (1989), Willis (1989), Ainley et al. (1990), and Leder and Forgasz (1991) reported that there were no significant sex-related differences in mathematics achievement in Australian primary schools. Willis (1989) and Leder (1990) also pointed out that given the same formal preparation in mathematics, until Year 10, the boys and the girls achieve similarly.

When sex-related differences appear, girls generally score better on computation tests and boys score better on applications of mathematics, problem solving and mathematics reasoning (Moss, 1982; Willis, 1989). In their analysis of the 1975 literacy and numeracy data, Keeves and Bourke (1976) reported that at the 10-year-old level, in estimation and time, the number of boys with correct responses was higher than that of the girls. Moreover, on classroom-based tasks which involved multiplication, division and reading of charts and graphs, girls outperformed boys. At the 14-year-old level, boys outperformed girls on items that included fractions, measurement, multiplication, reading a map and a volume calculation, while, girls did better than boys only on item that required the addition of money.

In 1980, the second series of literacy and numeracy tests was administered to the same age groups (10-and 14-year-olds) of students as in 1975, Bourke et al. (1981) reported that on the numeration test at the 10-year-old level the girls outperformed the boys, particularly in calculations with whole numbers in the classroom rather than in a community context. At the 14-year-old level, boys outperformed girls in the measurement subtest that required the reading of graphs, maps and scales.

When comparing the 1975 and 1980 results on the numeration test at the 10-year-old level, on both occasions, girls outperformed the boys in the classroom-based tasks which involved calculation and the boys outperformed the girls on the measurement subtest. At the 14-year-old level, the boys outperformed the girls on the measurement subtest on both occasions. Crawford (1988) also pointed out that more boys outperformed girls on selected mathematics topics such as measurement and problem solving, while girls did better than boys in arithmetic. Furthermore, Vale (1993) reported that girls were more likely to do better than boys on investigative projects, non-routine problem solving and multiple choice tests of basic skills, while boys were more likely to perform better than girls on routine analysis problems. Bishop and Clements (1994) also reported that boys were more likely to do better on complex mathematics tasks than girls, and girls were more likely to do better on standard elementary computational tasks than boys.

Pattison (1992) reported that on a study conducted in two comparable single-sex schools in Melbourne at Year 10 and Year 12 levels the boys consistently outperformed girls in problems dealing with measurement scales, but they performed less well on problems in solid geometry, while the girls did well generally in discrete logical problems and in deductive geometry. The Australian Association of Mathematics Teachers (1990a) pointed out that boys tend to do better in geometry, measurement and ratio, and girls in algebra, logic number and statistics. Moreover, girls are under represented in the top one per cent achievement level.

The results (1979-1982) on each subset of the Australian Scholastic Aptitude Tests (ASAT) indicated sex differences in the ACT, QLD and WA (Adams, 1984). Each year and in each authority, the boys seemed to outperform the girls on the mathematics, science and social science subtests, and the girls outperformed the boys on the humanities subtest. In Australian secondary schools there were at that time more boys than girls who studied mathematics and science(Adams, 1984; Pattison, 1992). Girls preferred to study subjects in the humanities and social science areas .

Carss (1988) reported that when students in Queensland schools taking the same courses were compared the mean score for the group of female students taking Mathematics 1 was consistently higher than the mean score for male students. However, the number of female students taking the subject was less than the number of boys. Carss (1988) also stated that in assessment for the state senior examinations, the average grade for female students was consistently higher than that of the male students. Parker and Offer (1987) investigated the trends in girls' achievement under the Achievement Certificate in WA (1970-1986). The population involved in their analysis was the total group of Year 10 students in each of the years 1972-1986. They found that girls' achievement in English, Mathematics, Science and Social Studies subject areas increased relative to that of the boys.

In the 1992 Victorian Certificate of Education (VCE) mathematics assessment program, four Common Assessment Tasks (CATs) were administered. Cox (1993, 1994) examined the gender differences in the 1992 and 1993 VCE results. The author reported that the girls performed better on CATs 1 and 2 which involved independent mathematical investigation and challenging problems and the boys performed better on CATs 3 and 4 that included facts, skill and analysis tasks. It was argued that the findings for CATS 3 and 4 supported the previous findings that the boys outperformed the girls in secondary school mathematics, and that sex differences at the age of 17 years were consistently in favour of the boys.
Baker and Jones (1993, p. 96) in their article on "Creating Gender Equality: Cross-national Gender Stratification and Mathematical Performance," state that

> ..in countries with large proportions of women in higher education, sex differences in the performance of eighth-grade mathematics are smaller (or move away from a clear male advantage). Also as women are channelled less into nonuniversity training, sex differences decrease, although this correlation is not statistically significant.

Baker and Jones claimed that in any society females were more likely to perform better than males in mathematics when higher percentages of women worked in the formal work force.

The research findings about gender differences in achievement in school mathematics that are reviewed here indicate that in Australia at the primary school level, in general, there was no sex difference, that is the boys and the girls mathematics achievement levels were similar. Differences started to appear at the lower secondary school level.

Participation

There were also differences in Australia between boys and girls in participation in school mathematics. In the 1960s the number of boys taking mathematics in secondary school was much higher than the number of girls. Moreover, from 1964 to 1978 the proportion of girls taking mathematics at the final year of secondary schooling was substantially increased (Keeves and Mason, 1980). However, Morgan (1986) showed that in 1984 in the NSW Higher School Certificate twice as many boys as girls studied higher mathematics courses and three times as many girls as boys did not study mathematics at all.

Leder and Forgasz (1992) argued that in recent times fewer girls had enrolled in advanced and intensive mathematics courses. The analysis conducted by Dekkers et al. (1986, p. 53) about the enrolment pattern in mathematics and science in the upper secondary school in Australia endorsed this claim that "the most significant feature of the mathematically-specialised programs in most states is that they are studied by at least twice as many males as females." The Australian Association for Mathematics

Teachers (1990b) found that girls generally chose lower level mathematics courses and this limited their choices in post-secondary programs.

Retention Rate

Sampson (1982) argued that in 1972 the retention rate and achievement of girls in Australian schools was lower than for boys. However, by 1978 in most Australian schools there were more girls than boys at the final year of secondary schooling (Keeves and Mason, 1980; Leder, 1990; Widdup, 1980; Leder and Forgasz, 1992). In 1996 the retention rate for female students (77%) was higher than the corresponding rate for males (65.9%) (Australian Bureau of Statistics, 1997, p. 6). Consequently, it was possible to say that the girls were no longer disadvantaged as far as retention rate was concerned.

Yield

In making allowance for differences in holding power at the level of student achievement between FIMS and SIMS, Moss (1982) employed the concept of mathematical yield. Mathematical yield measures take into consideration the proportion of the grade cohort studying mathematics and are concerned with the question 'How many of the students of mathematics are brought how far in their learning of mathematics?' (Moss, 1982, p. 58). It was found that between 1964 and 1978 the yield or output in mathematics increased greatly across Australia and particularly in Tasmania, where there was low level of participation at the Year 12 level in 1964.

Subject Choice

From 1970 to 1989, enrolment data of mathematics and science course students in the upper secondary schools in Australia, collected by Dekkers et al. (1991) indicated that more female students chose to study type I mathematics courses at Year 12, which were designed for students who did not intend to proceed to tertiary education. In more recent years almost equal numbers of female and male students were attracted in those mathematics courses which were prerequisites for tertiary education. However, in specialised or advanced mathematics subjects which led to tertiary education in which mathematics was an integral part of the discipline, as in engineering, mathematics or physics, major discrepancies existed between female and male Australian students.

Generally, when boys and girls are compared in their subject choices, more boys choose mathematics courses than do the girls. Clearly, the boys prefer to do more mathematics courses than the girls, presumably because of their different attitudes, interests and choices of future careers.

Students' Attitudes toward Mathematics

Students' attitudes towards mathematics have an influence on their mathematics achievement. If they have a positive attitude, it would seem likely to motivate them to a higher level of achievement. In both primary and secondary education, the importance of expressing positive attitudes to different subjects is being acknowledged. Educators are realising that students with positive attitudes to subjects and who feel good about their learning, will develop more positive feelings about themselves and this would contribute significantly to their personal growth (Dungan and Thurlow, 1989). Therefore, it is important to assess the attitude of the students towards mathematics. As a consequence, previous studies on attitudes of Australian students towards mathematics are reviewed here.

Research studies have in general indicated that attitudes towards schooling decline with increasing age and grade level. Fraser (1979), found a decline in students' attitudes towards mathematics from the lower secondary school level. Attitudes towards schooling are not the same for boys and girls. Research findings have indicated that girls tend to hold more favourable attitudes than boys. Keeves (1972) pointed out that at Years 6 and 7 girls in Australian Capital Territory showed more favourable attitudes to school than boys. Thus sex differences in attitude are different in various subjects. In some subjects boys expressed more favourable attitudes than girls while in other subjects girls expressed more favourable attitudes than boys. In Keeves' (1972) study of a sample of students in the Australian Capital Territory at Year 7, boys showed more positive attitudes towards mathematics than the girls. Fraser (1980) examined the grade level and sex differences in attitudes towards English, mathematics, social studies and art among Melbourne metropolitan area Years 7-10 students. The results of Fraser's investigation indicated that girls showed more favourable attitudes towards English, social studies and art than the boys, while boys showed more favourable attitudes towards mathematics than the girls.

Keeves (1972) developed a path model to test the factors that influence mathematics achievement. The examination of the path model indicated the direct and indirect effects that influenced final achievement in mathematics. One of the factors that influenced final achievement in mathematics was initial attitudes towards mathematics. The results of the analyses of Keeves' path model also demonstrated that sex of the student and initial attitudes were among the direct factors that influenced final attitudes towards mathematics. The boys showed more favourable attitudes towards mathematics than the girls. The investigation also showed that attentiveness influenced attitudes indirectly through achievement. In this path model analysis Keeves (1972) also demonstrated that the classroom behaviour of the student which was influential in changing level of mathematics achievement was not affected by initial attitudes towards mathematics.

Schofield (1981) conducted a study of mathematics attitudes and achievement on a sample of students in Grades 3 to 6. She reported that the relationship between mathematics attitude and achievement for girls was very low, only intermittently significant and at times negative compared with a stronger more consistently significant and positive relationship for boys.

Leder (1979) found that in Australian secondary schools girls were higher on fear of success measures than boys and this attitude was more characteristic of high performing students, especially girls with high educational and vocational intentions. Ballenden et al. (1985) contended that the most important factors in girls' under achievement in mathematics and attitudes to mathematics was their own attitudes to mathematics and others' attitudes towards girls' mathematics performance. As a consequence girls, at the age of 11 years, expressed less confidence than boys in their mathematical ability and under estimated their performance relative to their ability. The authors suggested that strategies to raise girls' confidence in their mathematics and science skill was very important.

Interest

Another variable that has been shown to influence students' achievement in mathematics is students' interest towards the subject. Those students who like mathematics usually think that it is easy, while those who dislike mathematics think that it is their hardest subject.

Kiryluk (1980), and Helfers (1986) reported that students liked mathematics when they understood the subject. New mathematics topics and activities also increased students' liking of mathematics. In his investigation of attitudes towards mathematics, Kiryluk (1980) found that students liked mathematics when they could do some mathematical activities they had not done before. Those students who believed that they were good at mathematics showed that they liked the subject (Helfers, 1986). She also pointed out that primary school aged students who disliked mathematics perceived the subject as hard. Pattison (1992) reported that the differences between boys and girls in interest in mathematics and achievement in mathematics continued until the later years of upper secondary schooling, in spite of girls dropping out from the study of mathematics. Dungan and Thurlow (1989) argued that homework and the daily unchanging class routine of mathematics classes were also aspects of mathematics that students disliked.

Time and Opportunity to Learn

The time allowed and the opportunity given to students to learn are factors that influence achievement in mathematics. Opportunity to learn is sometimes indicated by the time allocated for students to learn a given task. Normally schools give the same amount of time to all groups of students to undertake a particular task. However, all students do not have the same learning rates. The rate of learning is different for different students (Carroll, 1963; Bloom, 1976) and time allotted by a school to learn a specific task may be appropriate for some, in excess, or less than required for others. Those students who are not allowed enough time for learning need to spend more time than the others. Hence, the time allowed for learning also affects the time spent by the student in learning.

Keeves (1976) argued that a distinction must be made between time spent in learning and the opportunity given by the curriculum for students to learn specific material tested. In addition, he argued that, the formal time spent was a measure which was related to the prescribed intended curriculum. However, opportunity to learn the material tested, if assessed at the teacher or student level and aggregated for inclusion in the analysis was a measure related to the implemented curriculum. Thus the implemented curriculum, assessed by opportunity to learn, could be seen as a mediating or intervening variable acting between the intended curriculum which was assessed by time allowed for learning, and educational outcomes which were assessed by the relative levels of performance of the school or system. Keeves (1976) and Rosier (1982) both confirm the existence of these relationships with Australian data. Keeves (1972) on his path model analyses for mathematics achievement demonstrated that one of the factors that directly influenced final achievement in mathematics was attentiveness that is the attention paid by the student to his/her school work. He reported that the evidence showed that attentiveness in the mathematics classroom was largely influenced by the teacher.

Leach and Tunnecliffe (1984) investigated the influence that allocated time and engaged time and selected context variables had on school mathematics achievement in three metropolitan primary schools in Perth. The researchers found evidence that engaged time had the highest correlation of $r=0.76$ with mathematics achievement, accounting for 58 per cent of the variance. In addition, the authors noted that the variables, engaged time and IQ, had higher correlations with mathematics achievement than had other factors. This study also found that the allocated time had a relatively low influence on student achievement.

Bourke (1986) reported that in a study conducted in the Melbourne metropolitan area in 39 government and non-government schools, the actual time spent in mathematics lessons was positively related to achievement in mathematics. Keeves (1968, 1972) reported that the lesson time spent in mathematics as well as time spent in previous years were influential factors in the level of achievement reached by students within each state system. Bourke (1984) also reported that opportunity to learn was related not only to achievement in mathematics but also to positive attitudes towards mathematics.

Language Factors in Mathematics Achievement

Researchers have found that language is one of the factors that influences achievement in school mathematics. It is generally accepted that linguistic and verbal ability affect achievement in mathematics (Turner, 1980). Morris (1978) argued that mastery of language is essential for mathematics learning and that linguistic fluency was essential for success in school mathematics. Morris also claimed that the mother tongue was the best language in which to learn mathematics. Those who learnt mathematics in a second language were educationally handicapped (Morris, 1978; Turner, 1980).

Keeves and Bourke (1976) in their report on a research project into numeracy and literacy in Australia conducted in 1975, reported that some second language children from non-English speaking homes were experiencing severe difficulties in numeracy.

The findings of research in Australia have indicated that all students whose first language was not the medium of instruction were educationally disadvantaged. Students who were competent both in their first language and in English which was the medium of instruction had some advantage in learning mathematics over those students who were competent in only one of these languages. Both groups had a clear advantage over students who had lower levels of competence in both languages (Dawe, 1983; Clarkson, 1991, 1992, 1993). Therefore, the level of competence which students had in each language was a vital factor which needed to be considered (Clarkson, 1992).

Socioeconomic Status

The socioeconomic status of students influences their level of achievement in mathematics. Keeves (1968) reported that the pattern of relationships between parental variables (such as father education, father occupation, mother education, mother occupation) and students' mathematics achievement levels was different at Year 8 and Year 12 levels. At Year 8 there were strong and consistent relationships, while at Year 12 level there was no significant relationship and the correlation coefficients were small. He also indicated that the number of years of education of both the father and the mother and father's occupation showed moderately strong relationships with the mathematics achievement of the students. Ainley et al. (1990) also found that the correlation coefficient between socioeconomic status and mathematics achievement among Year 5 students was 0.28 and 0.22 among Year 6 students.

Keeves (1972) argued that the home environment of the student influenced the change in their achievement in mathematics over time. He stressed that the home environment made a significant contribution to final mathematics achievement. Similarly, Rosier (1980) reported that students from a higher socioeconomic background as measured by their fathers' occupation were higher achievers in mathematics.

Research into Student Achievement in Mathematics in Ethiopia

Unlike Australia, research studies conducted in Ethiopia on student achievement in school learning in general and in mathematics in particular are few in number. These few research findings have indicated that there are substantial differences in students' achievement in school mathematics and the factors that influence the achievement level of students need to be investigated. These differences in student achievement in school mathematics in Ethiopia are reviewed below in terms of gender, attitude towards and interest in mathematics, the influence of time and opportunity to learn, language differences and socioeconomic status.

Gender Differences in Mathematics Achievements

In this section research findings on gender differences in mathematics achievement in Ethiopia are reviewed. Little research has been done in Ethiopia into students' achievement in mathematics, or indeed into students' academic achievement in general. Ademe and Gebre-Meskel (1989, p. 74) put the problem as follows:

> Today, Government and the entire community are pouring a substantial amount of money to finance the schools. Since money and material resources are in short supply in countries like Ethiopia, the fact that they are properly utilized to promote the students' academic achievement is a matter of great concern. This calls for intensive research in the area of academic achievement which will greatly increase the effectiveness of teachers and educators in their teaching and guidance. Unfortunately studies on this subject are scanty in Ethiopia.

Consequently, it is not possible to review many research studies on factors influencing the mathematics achievement of Ethiopian students. The research conducted on mathematics achievement of students in Ethiopian schools is minimal. The findings of these few research studies has indicated that at all educational levels in both rural and urban areas of the country the achievement of girls in mathematics is much lower than that of boys (Seyoum, 1986; Anbessu and Junge, 1988; Atsede and Kebede, 1988; Assefa, 1991; Gennet, 1991a; Tsion, 1990; Behutiye and Wagner, 1993, 1995; Yelfign et al. 1995).

In his study of sex differences in language and mathematics ability, Tesfaye (1987) showed that boys performed better in mathematics achievement than girls. He also reported that girls preferred the objective, true false items rather than the work out part which required mathematical reasoning. The students aged 12 and 13 years were in Grade 7 which was the first year of junior secondary school.

In their study on problems of elementary school participation and performance in Bahir Dar Awraja (an administrative unit in Ethiopia), Anbessu and Junge (1988) also reported a marked discrepancy between achievement of boys and the girls in mathematics. The difference was in favour of boys. In the same year further results were reported by Atsede and Kebede. The findings of Atsede and Kebede (1988) indicated that there was no discrimination between boys and girls regarding the curriculum of the elementary school, but it was found that sex differentiation started at the junior secondary school level when girls joined commercial streams and home-economics. They also reported that the Grades 6 and 8 National Examination results of the past ten years indicated that the performance of girls in mathematics and science education was low when compared to the performance of boys. Similarly, the mathematics and natural science results of the girls in the Ethiopian Schools Leaving Certificate Examinations (ESLCE) for the previous six years were also low when compared to the results of boys.

Derese et al. (1990) also investigated factors affecting achievement of lower primary school pupils, namely Grades 2, 3 and 4 students in Harer-Zuria, Wolaita-Sodo, Kola-Degadamot and Mocha Awrajas, and reported results which were complementary to the findings of Atsede and Kebede. Their findings indicated that in the mathematics test, boys performed better than girls in each of the three grades. In Grade 3 the means of the two sexes differed significantly in favour of the boys. However, the test results for Amharic showed the opposite. In each of the three grades girls outperformed boys, indicating that the higher performance of the boys involved performance in mathematics and not in other subjects such as language instruction.

In his study of female participation and performance in rural primary school in Ethiopia, Assefa (1991) also reported that the performance of girls was significantly lower than that of boys. He reported that girls held poor rank positions as the result of their poor academic performance. Behutiye and Wagner (1993) in their study of some aspects of primary school students' cognitive development and their relationship to scholastic achievement in Ethiopia reported that in all sample grades (Grades 2 to 5) the boys' achievement levels in mathematics were better than that of the girls'. However, the difference was only significant in Grade 5. Meanwhile, in their 1995 study these researchers found significant differences between the boys and the girls in achievement in mathematics in all the Grades. In his study of some factors affecting scholastic achievement of elementary school pupils Sewnet (1995) also reported that gender difference played a significant role in the academic achievement of students.

Similar results were reported by Gennet (1991b). She found that girls performed comparatively poorly in school achievement. This was reflected in the national examination results and in the rates of repetition and dropouts. The result of the Grade 6 National Examination for the years 1978-1987 showed that of the total number of students who passed on average only 35 per cent were girls. Every year a large number of girls fell behind their male classmates in joining the junior secondary school. The performance of girls at the Grade 8 National Examination was no better than at Grade 6. Of the students who passed the Grade 8 National Examination between 1978-87, about 40 per cent of the girls were unable to continue to the next level of education.

Other researchers also reached the same conclusions about secondary school students. In a study conducted by Ademe and Gebre (1990) in Bahir Dar Awraja, which included students in Grades 9 to 12, it was found that on a mathematics achievement test boys performed better than girls at all grade levels.

In her study which was an assessment of academic performance of female students in higher education institutions in Ethiopia, Tsigie (1991), reported that the percentage of girls dismissed from tertiary institutions was always greater than for the boys and the rates of dismissal over enrolment for females were invariably higher than the rates for males. She further indicated that on the average the academic performance of girls in higher education was inferior to that of boys. The comparison of the performance of students with similar grades in Ethiopian Schools Leaving Certificate Examinations indicated that girls were weaker.

Gennet (1991a) reported that the mathematics achievements of girls in government and non-government schools was lower when compared to other subjects, even though the majority of girls indicated that mathematics was their preference. Most girls (81 per cent of the respondents) showed that they liked mathematics, English, chemistry and biology for their future usefulness. However, mathematics and English were compulsory subjects for entrance to tertiary education in the country. She believed that the preference of the girls for these subjects might not be related to their interest or aptitude, they might have been dictated by the perceived usefulness of the subjects.

In another study conducted by Seleshi (1995) which included students in Grades 8 to 11 in the Northern Shewa region, boys performed better than the girls in mathematics achievement scores. Seleshi reported that from Grades 8 to 11 the male-female differences in mathematics achievement were statistically significant. He contended that gender differences in mathematics achievement were observed at Grades 8, 9, 10 and 11, and at each level the difference was in favour of the boys.

In summary, almost all the studies reviewed in this section, indicated gender differences in mathematics achievement starting from the elementary school level. The differences reported were clearly in favour of boys. These findings differed from the findings in Australia. In Australia almost all research findings indicated that sex differences in mathematics starts to emerge at the lower secondary school level. Therefore, further investigation is required to know exactly where sex difference starts to emerge in Ethiopia.

Factors influencing Mathematics Achievement

Research findings about student level factors that influence mathematics achievement of Ethiopian students are reviewed in the sections that follow.

Attitude toward Mathematics

Like Australian research findings, the Ethiopian research outcomes also showed that students who expressed positive attitudes towards mathematics were likely to achieve higher in mathematics. Seleshi (1995) showed that boys had more positive attitudes towards mathematics compared to girls. The gender differences in attitude towards mathematics were observed at Grades 9, 10 and 11, and at each level it was in favour of the boys. However, statistically significance at the Grade 8 level was not found. He argued that one of the factors that influenced mathematics achievement was the students' attitudes towards the subject. Those students who had positive attitudes achieved higher scores. Derese et al. (1990) reported that students with strong positive attitudes towards schooling had a high mean level of achievement in mathematics.

In his study of factors that affected students academic performance in some selected senior secondary schools in Addis Ababa, Endalkachew (1990) found that the students' attitudes towards this learning experiences was low. Similar findings were reported by Tadesse (1993). He argued that in the senior secondary school students of his samples, the majority of the students had a low attitude towards learning. The positive attitude towards mathematics slightly declined among all students with increases in grade level, the decline appeared to be more noticeable in the case of the girls. He argued that this was because the school administration failed to satisfy the needs of the students which led many of them to be disinterested in learning and eventually develop negative attitudes.

Home background

Among other factors that influenced the achievement of Ethiopian students was home background. Endalkachew (1990) investigated the family background of the students and found that it determined their school performance and further the general quality of life and attitude of the students.

Teshome (1993) found out that the low living conditions at home would be likely to hamper the desired level of achievement of students. Behutiye and Wagner (1995) also reported that in Ethiopia students whose parents were alive outperformed those students whose parents were not alive. They also indicated that students whose parents had a higher income performed better than students from lower income family backgrounds.

Berhanu (1986) reported that a positive and favourable relationship was found between the educational background of a family and degree of academic success of the students. Parents' level of education, rather than the number of literate parents with a low level of education, and fathers' level of education, in particular, seemed to have a strong and positive influence on the degree of academic success of students. Berhanu's (1986) findings were similar to those of Damtew's (1972), because Damtew reported that students coming from homes with parents with better education performed better in their scholastic achievements when compared with the results of students coming from homes with parents with less education.

Gennet (1991a) reported that the educational background of parents, especially mothers had a bearing on the academic achievement and participation of female students. Furthermore, Eshetu (1988) found that the majority of high achieving students had good relationships with their parents, while most of the low achievers indicated having conflict with their parents. Most of the respondents in the high achievers group admitted that their interactions with parents took the form of a warm and permissive style which was rare in the response of the low achievers. The great majority of the high achievers reported that they had parents with an educational level of secondary school and above.

Family Size

Another factor that was investigated by researchers was the size of the families of the students. Behutiye and Wagner (1995) showed that better performance was achieved by those students who lived in a smaller family. Teshome (1993) also reported that the relatively high number of siblings at home was likely to hinder the desired academic results of students. Berhanu (1986) in his investigation of the effect of family background on academic success of students, sampled students in Grades 9, 10 and 11. When he compared the high and low achiever groups, the numbers of high achieving students coming from medium and small sized families were found to be much greater than it was for low achieving students. He also reported that almost half the 43 per cent of low achieving students were found in the large family group while this per cent fell to only seven per cent for the high achieving students in large families. Similar results were identified by Eshetu (1988). He noted that the numbers of low achievers coming from a family of more than four children was found to be greater than the numbers of high achievers.

The research outcomes reviewed above show that students from large family groups performed less well than students from small family groups. Therefore, family size was one of the factors that clearly influenced achievement in Ethiopia.

Involvement in home activities

The involvement of students in home activities was seen as a factor that influenced the achievement of Ethiopian students. Damtew (1972) reported that the students who stayed in homes where there was a continued demand for them to work after school hours performed at a relatively low level in their scholastic achievements when compared with those students who had either nothing to do at home or little work to do at home. On the average, the major work that different students accomplished at home after school hours was related to the monthly income and occupation of their parents, but largely to the attitudes of parents.

Derese et al. (1990) found that the overall mean achievement in mathematics for students with high involvement in home activities was low. The authors concluded that engagement in home activities had a strong connection with students' level of performance. These researchers argued that although the boys and the girls had a similar level of participation in home activities, the girls had better scholastic performance in Amharic, and the boys in mathematics.

Gennet (1991a) reported that one of the factors influencing the poor achievement and performance of girls was their participation in household activities. Girls did not have time to do their homework since they had to help their mothers at home. She also reported that in the Ethiopian family girls were considered fully responsible to manage and take care of the house whenever both parents were away. Thus, the girls did not have enough time to study and they gave less attention to study since they got tired as the result of their duties at home.

Daniel (1995) shared similar views with Gennet (1991a). Daniel claimed that the reason for gender differences in achievement was the household activities of students. Daniel argued that the Ethiopian girls were more involved in a large number of household activities compared to the boys. Hence, girls had less time for their studies and play activities during out of school hours, which might have had a negative effect on their academic achievement.

The above research findings have revealed that household activities were one of the factors that influenced achievement in Ethiopian schools. The research findings showed that the girls achievement level was lower than that of the boys, because of the greater involvement of the girls in household activities compared with the boys.

Participation

Studies conducted on the education of boys and girls in Ethiopia have indicated that boys are more favoured than girls in terms of school participation. Seyoum (1986) reported that the traditional Ethiopian church education was purely for boys, and the same was true for Muslims. In this study it was claimed that beginning from the introduction of modern education by Emperor Menelik II, legal provisions were made for the education of girls. However, the participation rate of girls had been low at all levels. Therefore, the proportion of girls to boys was very low at all levels.

Studies have indicated that only one-fourth of elementary school aged girls were actually enrolled in the educational system, while at the junior and senior secondary school levels only 17 per cent and 10 per cent were enrolled respectively (SIDA, 1993). These reports also revealed that enrolment was further declining in most rural schools. At the same time, alarming negative trends were occurring in the female participation rate at tertiary education levels. Between 1984 and 1987, there was a 13 per cent decline in girls' enrolment at the Addis Ababa University and in the 1987/88 academic year only nine per cent of the students were girls.

Gennet (1991b) reported that students' enrolment and participation from elementary grade to the senior secondary schools indicated that the number of girls enrolled remained markedly lower than for the boys. The number of girls who continued to dropout of school or repeat classes was higher than that of the boys.

Abera and Zewdu (1992) indicated that in 1974 the enrolment ratio of female students in Ethiopian junior and senior secondary schools were 4 and 2 per cent respectively. However, the percentage for the boys were 9 and 5 respectively. Meanwhile, the girls' participation increased in 1989 to 17 and 10 respectively, whereas the boys increased to 23 and 14 per cent respectively.

Researchers are looking for reasons for this difference in participation between boys and girls. Almaz (1989) in her nation-wide study of girls' participation and performance in education, argued that absence of female teachers, characteristics of the school, location of the school, economic conditions, early marriage and parents' attitude and beliefs were some of the factors that affected girls' participation and performance in education.

In January 1989 a workshop, sponsored by Ministry of Education and United Nation International Children Education Fund at Bahir Dar on the psychology of girl learners, examined causes of the low achievement of girls in Bahir Dar Awraja. Some of the causes identified during the workshop were similar to those found by Almaz (1989).

Language

In their study of factors affecting scholastic achievement of lower primary school students, Derese et al. (1990) divided their subjects into two groups: those coming from Amharic speaking families and non-Amharic speaking families. In mathematics they did not find significant differences between the two groups of students. These investigators also concluded that there were also no specific differences according to whether the mother or the father spoke Amharic. Students achieved most if both parents could read and write; they achieved least, if neither parent was able to read or write.

Subject Choice

Atsede (1991) argued that in the 1950s the number of girls enrolled in science and technology was reasonably high, but the ratio of girls to boys enrolled in the two faculties in the Addis Ababa University had shown a significant and continuous decline. The enrolment of girls in science and technology compared poorly with that in the social sciences. In 1988/89 the percentage of girls enrolled in science and technology faculties was seven per cent and three per cent respectively, while the comparative figure for the faculty of social science for the year 1989 was 16 per cent. In addition, she contended that apart from the low proportion of enrolments in science and technology, the total number of girls undergraduates who obtained their bachelor degrees in these fields was insignificant. It was also worth noting that in the natural sciences almost half of the girls were involved in biology courses closely followed by chemistry. Physics and mathematics did not seem to appeal to girls.

Hirut (1986) reported that even if girls went to school, they were inclined to choose jobs and professions usually held by society as the area of women's work for example, nursing and secretarial work. Females were rarely seen to occupy jobs in subjects such as engineering, mathematics and physics.

Conclusions

Research findings on Australian students' achievement in mathematics and factors influencing their achievement levels are reviewed in the first part of this chapter, while the second part of the chapter presents the review of the research findings about Ethiopian students' mathematics achievement level and factors influencing their achievement level. In this section a summary of the review is presented. Like the curriculum development discussed in Chapter 3, research findings on students' achievement in mathematics showed similarities and differences between Australia and Ethiopia. In both countries, when there is a difference in achievement between boys and girls, the difference is in favour of boys. The main difference between the two countries is that Australian research findings showed that sex differences started to emerge at the beginning of lower secondary school, while Ethiopian researchers reported that sex difference existed at the primary school level. Therefore, further investigation seems important to identify where sex differences start to emerge and why they develop.

There were differences in participation and subject choice in mathematics courses between Australian and Ethiopian schools. In Ethiopian mathematic is a compulsory subject. However, mathematics is not compulsory in Australian schools.

Attitudes towards mathematics, socioeconomic status, time and opportunity to learn were some of the student level factors that influenced mathematics achievement of students in both countries. In the present study the information collected from the students about themselves and their home background is considered in Chapter 11 in order to examine student level factors that influence the mathematics achievements of Australian and Ethiopian Year 8 students. Thus, the next chapter deals with the data collection procedures employed in this study.

5

Data Collection

The major objectives of this thesis are to examine changes in mathematics achievement over time, to investigate student level factors influencing mathematics achievement, and to assess the attitudes of students towards mathematics and schooling at the 13-year-old and Year 8 level in Australia and in Ethiopia. For these purposes, primary and secondary analyses were undertaken. The primary data were collected in Ethiopia, while the secondary data were derived from the First, Second and Third International Mathematics Studies conducted in Australia in 1964, 1978 and 1994 respectively.

In this chapter, the sampling procedures and the instruments employed in the studies and a brief description of the administrative procedures are presented.

The Sampling Procedures

The sampling procedures for the IEA studies and the Ethiopian study are presented in the sections that follow.

Sampling Procedures Employed for Australia

IEA has conducted three international mathematics studies, and Australia has participated in all the three studies (see Chapter 2). The sampling procedures employed in the three studies in Australia are different although random sampling has always been employed.

The FIMS Sample

The FIMS was conducted in 1964 by the Australian Council for Educational Research (ACER), in the government schools of five Australian states, namely, New South Wales, Queensland, Tasmania, Victoria and Western Australia.

This was the first major international study conducted in Australian lower secondary schools. In order to simplify the administrative procedures, it was decided to include only students in government schools who were 13.00 to 13.11 years old or in Grade 8 at the date of testing (Husén, 1967, vol I, p. 45; Keeves, 1968; Rosier, 1980, p. 45).

The study also excluded handicapped students located in special schools or special classes. In addition, the study was also limited to students at the three year levels containing the majority of the 13-year-old students. Hence, in 1964 in Australia, the target populations were defined as follows.

Population 1A was defined as:

> all students of age 13.0 to 13.11 years on 1 August 1964 in normal classes in Year 7, Year 8 and Year 9 in government schools in New South Wales, Victoria, Queensland, Western Australia and Tasmania. (Keeves, 1968, p. 66; Rosier, 1980, p. 45; Moss, 1982, p. 10)

Population 1B was defined as:

> All pupils at the grade level where the majority of pupils of age 13.0 -13.11 years were to be found (Keeves, 1968, p. 66).

The number of students who participated in the Australian study is given in Table 5.1. A two-stage sampling design was undertaken, involving the selection of a random sample of schools from each state at the first stage and a random samples of students from within these schools at the second stage were selected.

Table 5.1 Number of students participated in FIMS, SIMS and TIMS

		FIMS				SIMS				TIMS			
Age	**Sex**	**7[a]**	**8**	**9**	**Total**	**7**	**8**	**9**	**Total**	**7**	**8[b]**	**9**	**Total**
13-year-olds Government	M	322	874	334	1530	352	1392	351	2095	342	1409	47	1798
	F	237	805	345	1387	259	1279	356	1894	328	1345	60	1733
	Sub total	559	1679	679	2917	611	2671	707	3989	670	2754	107	3531
13-year-olds Non-government	M					94	384	102	580	174	612	36	822
	F					54	408	89	551	189	1071	56	1316
	Sub total					148	792	191	1131	363	1683	92	2138
Not-13-year-olds Government	M		745		745					289	1052	1049	2390
	F		657		657					295	807	1001	2103
	Sub total		1402		1402					584	1859	2050	4493
Not-13-year-olds Non-government	M									120	499	437	1056
	F									188	562	837	1587
	Sub total									308	1061	1274	2643
	Total	559	3081	679	4319	759	3463	898	5120	1925	7357	3523	12805[c]

a = the sex of one student was not identified in FIMS
b = the sex of one student was not identified in TIMS
c = in TIMS 46 students did not indicate their age

The SIMS Sample

The SIMS data were collected in Australia in 1978 by ACER. SIMS was designed to enable comparisons with the results of FIMS, and hence the 1978 sample was defined to match the 1964 definition, relating more specifically to an age sample, rather than a grade sample. Students in both government and non-government schools in the six states and one territory, namely, New South Wales (NSW), Queensland (QLD), South Australia (SA), Tasmania (TAS), Victoria (VIC), Western Australia (WA) and Australian Capital Territory (ACT) participated in the 1978 testing program. In SIMS the target population was defined as:

> all students of age 13.0 to 13.11 years on 1 August 1978 in normal classes in Year 7, Year 8 and Year 9 in all States except the Northern Territory. (Rosier, 1980, 48; Moss, 1982, p. 11)

Table 5.2 shows the number of students in the achieved samples who participated in the study. The study also involved a two-stage sampling design. The first stage was to select schools randomly with a probability proportional to size (pps) from each state, and the second stage was to choose at random students from the schools selected at stage one (see Table 5.2). However, in practice the plan was modified to take fewer schools from the two smaller states, that is ACT and TAS, and to add primary schools to the samples in QLD, TAS and WA. In addition, the samples were stratified by type of school within each state, namely, government, Catholic and independent schools.

Table 5.2 Achieved Samples in FIMS, SIMS and TIMS

Target Population	Label	Size	Sampling Procedure	Primary Unit	Secondary Unit
13-year-old-Government schools	FIMSA	2917	2SRS	School	Student
Year 8-Government schools	FIMSB	3081	2SRS	School	Student
Total-Government schools	FIMS	4320	2SRS	School	Student
13-year-old_Restricted	SIMS	3038	PPS	School	Student
Total-Government and Nongovernment schools	SIMS	5120	PPS	School	Student
Year 8 _Restricted	TIMS	3786	PPS	School	Class
Total-Government and Nongovernment schools	TIMS	12852	PPS	School	Class

2SRS = 2 Stage Random Sampling; PPS = Probability-proportional -to-size

Restricted = Government schools in five states which participated in FIMS, SIMS and TIMS

In both FIMS and SIMS invitations were sent to the selected schools to request their participation in the two studies. In 1964, the schools in the secondary stage were asked to select the sample students, following defined random sampling procedures based on the date of birth of the students. However, in 1978, the participating secondary schools provided a list of all 13-year-old students at the Population 1 level, and the ACER staff, from the lists provided selected students at random using a procedure based on students' birth dates.

The TIMS Sample

The TIMS data were collected in Australia in 1994 by ACER. TIMS was designed "with the aim of documenting the quality of mathematics and science education in the participating countries and further understanding of factors related to student achievement" (Beaton, 1996, p. ix). In order to attain these aims data were collected from students in the participating countries. Students who participated in this study were from the pair of adjacent grades containing the largest proportion of 13-year-olds (Foy et al., 1996; Lokan et al., 1996; Martin, 1996). Lokan and her colleagues (1996, p. 29) reported that in Australia the adjacent year levels that included most 13-year-old students at the time of sampling were Years 7 and 8 [ACT, NSW, VIC, and TAS) and Years 8 and 9 [QLD, SA, WA and Northern Territory (NT)]. Table 5.1 shows the number of students who participated in the TIMS study.

The selection of samples involved a two-stage sampling design. The first stage was to select schools randomly with a probability proportional to size (pps) from each state and the second stage was to choose at random classes from the schools selected (see Table 5.2) at stage one (Lokan et al., 1996). Lokan and her colleagues (1996, p. 29) believed that the pps sampling ensured appropriate representation of government and nongovernment schools in each state and territory's sample.

However, the selected number of classes was not the same throughout the country. Lokan et al., (1996, p. 29) outlined the situation in the following way.

> Two classes at the higher year level were then selected randomly from each school. In those states and territories where Year 8 was the lower level, two classes were similarly selected in each school at that level as well. In those states and territories where the lower level was Year 7, only one class was selected in each school at the lower level.

These research workers also indicated their reasons for the variations:

> for the country as a whole, the single year level containing the greatest number of 13 year olds is Year 8. Selecting two classes per school at the lower year level, where this was Year 8, better represented the year containing most 13-year olds across the country (Lokan et al., 1996, p. 29).

Thus, because of the complexity of the sample and differential response rate from schools and school systems, it was necessary for the data to be carefully and accurately weighted in order to obtain appropriate estimates of student performance.

Test Administration in FIMS, SIMS and TIMS

In both FIMS, SIMS and TIMS the tests were administered in each participating school by the staff members of the school who had taken the responsibility for conducting the testing program.

Sampling Procedure Employed in Ethiopia

The sample required for data collection for the Ethiopian Mathematics Study was 1200 Grade 8 students from 40 government and nongovernment schools in the Addis Ababa Region. In addition, it was planned that 30 students should be selected from each school. The schools and the students were selected randomly using a random sampling procedure with probability proportional to the size of the school (pps sampling). The sampling details involved in selecting the schools and the students are presented in Appendix 1. Table 5.3 shows the number of schools, the number of Grade 8 students, the number of selected schools and students in the Addis Ababa Region.

Table 5.3 Number of schools, Grade 8 students, selected number of schools and students in the Addis Ababa Region

	Number of schools and students in each Zone		**Selected schools and students in each Zone**		**13-year-olds Gov schools**		**13-year-olds Non-Gov schools**		**13-year-olds Total**		**Students Not 13-year-olds in selected schools**		
Zone	Sch	Stu	Sch	Stu	M	F	M	F	M	F	M	F	Total
1	18	8705	6	180	6	11	-	3	6	14	70	88	178
2	36	11898	8	240	4	10	13	17	17	27	78	117	239
3	26	8823	6	180	8	5	7	9	15	14	82	67	178
4	52	16366	12	360	14	13	11	11	25	24	148	159	356
5	30	9244	6	180	3	9	4	21	7	30	60	80	177
6	7	2594	2	60	3	4	-	-	3	4	23	30	60
Total	169	57630	40	1200	90		96		186		1014		1188[a]

Gov= Government; Non-Gov= Non-government; Sch = schools; Stu = students; M= Male; F= Female
a = 12 students did not write the age, therefore are not included in the table

Instruments

The test instruments employed both in the FIMS (Husén, 1967), and SIMS (Rosier, 1980) for the Australian students and for the Ethiopian Students were: Mathematics Test, Opinion Questionnaire, and General Information Questionnaire.

Information was also collected from teachers about the students opportunity to learn the mathematics items included in the test.

Mathematics Test

The 1964 Mathematics test contained 70 items, while the SIMS test and the Ethiopian Mathematics Study test each had 72 items. On all the three occasions most of the items were of multiple choice format with five alternative responses; only 11 items required the students to construct their own responses, which would be scored on a right-wrong basis.

The 1978 Mathematics Test excluded five of the 1964 items dealing with Euclidean geometry and added seven new items. The Ethiopian Mathematics Test excluded both the 1964 five Euclidean geometry items and the 1978 seven new items, it included seven new items from the 1994 Third International Mathematics and Science Study.

The items were classified into content and cognitive processes by IEA and participating countries. This classification of the items is presented in Table 5.4. The content was subdivided into four parts, namely Arithmetic, Algebra, Geometry and New Mathematics. Each had 34, 17, 17 and 13 items respectively. However, some items were classified in more than one content area (see Table 5.6). Items 18 and 23 were classified as parts of Arithmetic and New Mathematics, while Items 12, 32 and 64 were included in both Algebra and New Mathematics, whereas Items 60, 61 and 62 were classified as parts of Geometry and New Mathematics. Therefore, nine of the 13 New Mathematics items were classified either as Arithmetic, Algebra or Geometry.

Table 5.4 Number of Items in the FIMS, SIMS and EMS Mathematics Studies by Content and Cognitive Process

Total 72	**computational process** lower mental processes computation			knowledge		**verbal processes** higher mental processes abilities				
Content	skills in manipulation and computation			information definition, notation		Translation and inter-pretation		Comprehension, follow and construct proofs	analysis	application
Arithmetic 34	1	23	35	22	48	2	25	70	7	57
	3	24	49	31	51	5	29		18	
	4	26	50	39	54	(7)	47		72**	
	6	27	71**	40	55					
	8	30								
	9	33								
Algebra 17	12	37	56	11		15			32	
	14	38	65	68		(43)			64	
	20	43*				46				
	34	45*				63				
Geometry 17	10			42	67	(13)		21	16*	19
	13			52	69				41*	
	35			53					60	
				58					61	
				59*					62	
New Mathe-matics 13	[12]			28					[18]	
	17			44					[32]	
	[23]			66					[60]	
	[65]								[61]	
									[62]	
									[64]	

Items in square brackets [ ] appear in two content areas; Items in round brackets ( ) appear in two process areas; * Different items for different occasions; ** new items used in EMS.

The items were also classified into two major cognitive processes, namely computational or lower order processes and verbal or higher order mental processes.

The lower order processes were divided into two parts, namely computation and knowledge. There were 16, 10, 3 and 4 computation items in Arithmetic, Algebra, Geometry and New Mathematics respectively, whereas there were 8, 2, 7 and 3 knowledge items in Arithmetic, Algebra, Geometry and New Mathematics respectively.

The verbal, or higher order mental process or abilities were also divided into translation, comprehension, analysis and synthesis. There were 6, 4 and 1 translation and interpretation items in arithmetic, algebra and geometry respectively. However, some items were classified in terms of more than one cognitive process, such as items 13 and 43 which were classified as translation and computation items (see Table 5.4). Two items, namely Items 21 and 70 were classified as comprehension items, whereas 10 items were categorised as analysis items and 2 items were included as application items.

The numbers 1 to 70 are the numbers of the items used in FIMS and EMS, whereas 71 and 72 relate to EMS and SIMS. The item numbers 1 to 23 were also used in SIMS, however, due to the introduction of a new item as Item 24, the numbers in the Table are increased by 1 starting from Item 24, that means Item 24 in the Table is Item 25 in SIMS (Adapted from Keeves, 1966a; Rosier, 1980).

Opinion Questionnaire

The 1964 Australian and the 1996 Ethiopian Opinion Questionnaire contained 65 view and attitude statements. These 65 view and attitude statements were subdivided into seven scales (see Table 5.5). The first two scales were descriptive in nature, one measuring students' views about the teaching methods applied by their mathematics teachers, and the other measuring their perception of the climate of the school in terms of its authoritarian or open nature. The other five scales sought information about the attitudes of students towards mathematics and schooling. From the 65 items, only 36 were used in the 1978 Opinion Questionnaire. These 36 items, which were common to the three occasions, were taken into account for comparison purposes. On all the three occasions students were requested to respond to items by indicating whether they agreed with the statement, disagreed with the statement, or could not decide whether they agreed or disagreed with the statement. The view and attitude scales and their items are presented in Chapter 9.

Table 5.5 Attitude scales and number of items employed in different Occasions

	Occasions		
Scale	**FIMS**	**SIMS**	**EMS**
Views about Mathematics Teaching	11	6	11
Views about School and School learning	11	a	11
Attitudes towards Mathematics as a Process	8	a	8
Attitudes towards Facility of Learning Mathematics	7	7	7
Attitudes towards the Place of Mathematics in Society	8	8	8
Attitudes towards School and School Learning	11	9	11
Attitudes towards Man and the Environment	9	6	9
Total	65	36	65

a = scale excluded from SIMS

General Information Questionnaire

On all of the three occasions general information about students' background was collected from those students who participated on the studies using the General Information Questionnaire. Three types of information were collected from the students. First, information was collected about themselves, such as their date of birth, sex, country of birth, and whether or not they spoke English at home with their parents.

Second, information was collected about their schools and mathematics, such as their Year level in school, name of their mathematics teacher, number of students in their mathematics class, periods of mathematics each week, number of hours they had mathematics each week, number of hours they devoted to mathematics homework, number of hours they devoted to all homework, about their liking of mathematics and their mathematics test results.

Third, information was collected about their home background, such as their fathers' occupation and education, mothers' education, number of brothers and sisters, number of books at home, fathers' and mothers' country of birth and languages most often used in their homes.

The purpose of the Questionnaire was to obtain background information about each student in order to develop variables that would help in explaining differences between students in achievement in mathematics. Therefore, the three types of information which were collected from the students are considered in Chapter 11 in order to examine which variables would explain differences between students in mathematics achievement at the 13-year-olds and Year 8 levels.

Teacher Questionnaire

Information was collected about students' opportunity to learn mathematics topics included in the test. The purpose of this section of the Teacher Questionnaire was to collect information about the students' opportunity to learn the specific mathematical content of the items. In order that information would be available concerning the appropriateness of each item in the test for their students, the mathematics teachers were requested to rate each item as to whether or not the topic with which any particular item dealt had been covered by the students to whom they taught mathematics. The mathematics teachers were asked to rate each item in the Mathematics test (the test has 70 items for FIMS and 72 Items for SIMS and EMS) based on the students' opportunity to learn such kinds of concepts and ideas. For each item, they were requested to circle one of the responses **A**, **B** or **C** to indicate that, in their opinion:

A. all or most (at least 75 per cent) of this group of students have had an opportunity to learn this type of problem, or

B. some (25 per cent to 75 per cent) of this group of students have had an opportunity to learn this type of problem, or

C. few or none (fewer than 25 per cent) of this group of students have had an opportunity to learn this type of problem.

Instruments Employed in TIMS

The test instruments employed both in the FIMS and SIMS for the Australian students and for the Ethiopian Students were discussed in the previous sections. In this section the instruments administered in TIMS are addressed. These instruments were: Mathematics Test, Student, Teacher, School and Country Questionnaires.

However, in this section only the first two items are addressed. The remaining questionnaires are not considered, because they do not have major role in the present study.

Mathematics Test

Three types of items were included in the Mathematics test: multiple choice, constructed response items that included short answers and extended responses, and performance tasks (Garden and Orpwood, 1996). A total of 151 unique items were distributed across eight test booklets (see Table 5.6). However, the performance tasks were not included in the 151 items. Hence the discussion in this section and the analyses in the present study do not involve the performance tasks.

The design of the mathematics test was a cluster-based design. In TIMS a test booklet included item clusters and corresponded to the set of items that were be administered to individual students (Adams and Gonzalez, 1996). Adams and Gonzalez (1996, p. 3-4) defined an item cluster as follows: "An item cluster is a small group of items that are collected and then treated as a block for the purpose of test construction." Some clusters appeared in more than one booklet; moreover, when these clusters appeared in more than one booklet, their position in each booklet was different. For example, cluster C appeared third in Booklet 1 and first in Booklet 2.

Table 5.6 Distributions of mathematics items by item type, by booklet and by content areas

Reporting Category	**Booklet**								**Test types**			
	1	2	3	4	5	6	7	8	MC	SA	ER	Total
Fraction and number sense	11	10	11	10	10	11	11	14	41	9	1	51
Geometry	5	6	6	3	6	4	5	6	22	1	-	23
Algebra	8	5	6	8	4	6	6	9	22	3	2	27
Data representation, analysis, and probability	5	4	4	6	7	6	7	5	19	1	1	21
Measurement	5	5	6	4	6	4	4	3	13	3	2	18
Proportionality	3	3	4	3	6	2	4	4	8	2	1	11
Total	37	33	37	34	39	33	37	41	125	19	7	151[a]

a = Total items are 151, while total responses are 158; MC= Multiple Choice, SA= Short Answers, ER = Extended Responses; Source: Modified from Adams and Gonzalez (1996).

All the mathematics test items were grouped into 22 mutually exclusive item clusters, each was identified by a letter of the alphabet (A to V). These 22 mutually exclusive item clusters were divided into four different types of clusters (Adams and Gonzalez, 1996). The four clusters reported by Adams and Gonzalez are summarised below.

1. Core cluster: among the 22 item clusters, there was one core cluster labelled as cluster A which comprised six multiple choice items and appeared in all the eight booklets;
2. Focus clusters: among the 22 item clusters, seven were focus clusters which appeared in three booklets, from booklet 1 through booklet 7 labelled from cluster B to cluster H;
3. Breadth cluster: there were ten breadth clusters that consisted mostly multiple choice items (The clusters each appeared only in one booklet, which were labelled from I to R); and
4. Free response clusters: There were four free response clusters, that contained short answers and extended responses (These four clusters were each assigned to two booklets and were labelled from S to V).

Student Questionnaire

One of TIMS objectives was to investigate the factors at the student level that were likely to influence student learning of mathematics by employing the student questionnaire to collect the necessary information from students. Therefore, Australian students who participated in TIMS responded a questionnaire expected to occupy 30 minutes of the students' time in responding. In the questionnaire, students were requested to respond with respect to their home background, about their attitudes, motivation and interest in mathematics and in other school subjects, their progress in mathematics and other subjects in schools, and how they spent their outside school time (Schmidt and Cogan, 1996).

Conclusion

In this study two types of data were used, primary and secondary data. The secondary data were collected in Australia by ACER in FIMS, SIMS and TIMS. Five Australian states participated in FIMS, whereas six states and one territory took part in SIMS, Meanwhile in TIMS all states and territories in the country participated in the study. For comparability purposes similar data were collected from 40 schools with 30 students from each school, in the Addis Ababa region of Ethiopia at the Year 8 level. The results of the analyses of mathematics test data are presented in Chapter 8, while the students views and attitudes data are considered in Chapter 9. Chapter 11 presents the analyses of students level factors that influence the mathematics achievement level of 13-year-old and Year 8 students in 1964, 1978 and 1994 in Australia and in 1996 in Ethiopia.

The next chapter addresses the theoretical framework developed on the basis of previous research and theory that would be the basis for the analyses of the data collected and discussed in this chapter.

6

Models for Research in Mathematics Education

This chapter has three major purposes. These are to investigate the underlying structure of mathematics achievement test items; to develop a model of individual student level factors influencing mathematics achievement at the 13-year-old and the Year 8 levels; and to develop research questions for examination in this study.

The role of models and the underlying structures are considered in the first part of the chapter. The different models under investigation are discussed in part two in terms of their assumptions and the way in which they fit the theoretical framework. The third part of the chapter addresses the research questions to be examined in the present study.

Model Building in Educational Research

Educational researchers have employed hypothetical models since the early decades of the twentieth century. Researcher workers such as Brown (1917), Rogers (1918) and Mitchell (1938), developed models of mathematics achievement. Their research findings showed that mathematics involved a general factor while verbal ability was found to be a first or lower order factor. Since the early 1970s, consideration has also been given to the use of causal models (Keeves, 1972).

The fundamental idea of a causal model involves the building of a simplified structural equation model of the causal processes operating between the variables under investigation. A model is developed by employing knowledge gained from theory or from previous research (Keeves, 1988a, 1994, 1997a). From the outcomes of the testing of a hypothetical model the model is either confirmed or rejected for reformulation in order to provide an alternative that could be tested with the available data or with another body of data. In addition, choices can be made between alternative models according to the fit of the models to data obtained from the real world.

Models of Mathematics Learning

After the development of the first model of *School Learning* (Carroll, 1963) educational researchers advanced many different models of school learning involving the home and school environment (Bloom, 1976), instruction (Harnischfeger and Wiley, 1978), educational productivity (Walberg, 1981), student aptitude (Reynolds and Walberg, 1991) and in specific school subjects, such as factors influencing mathematics and science achievement (Keeves, 1975). Other uses of models include performance in science (Keeves, 1992a), the effects of student variables on science achievement and participation in science (Kotte, 1992), factors influencing reading achievement (Lietz, 1992, 1996), a structural model of high school mathematics outcomes (Reynolds and Walberg, 1992), and a model of factors influencing mathematics achievement of junior secondary school students (Tilahun and Keeves, 1997a). These models provided answers to questions raised in research into school learning. However, Keeves (1997b, p. 138) has argued that in the study of school learning:

> educational research is concerned with the process of change, and the study of constancy and change, and such study requires that observations are made for at least two points in time. While it is possible to describe the practice of education by means of a cross-sectional study undertaken at a single point in time, it is necessary to conduct investigations that are longitudinal in nature in order both to describe and explain the influence of educative processes on the constancy and change of related outcomes.

Consequently, it is necessary to conduct longitudinal and comparative investigations over time to examine the factors that influence learning outcomes.

Baltes and Nesselroade (1979, p. 4) defined longitudinal research in the following terms: "the entity under investigation is observed repeatedly as it exists and evolves over time." This method of investigation, as Keeves (1997b, p. 138) put it, is an "investigation conducted over time." Therefore, the repeated observation of a given phenomenon and change over time are important components of longitudinal research. However, Rogosa (1979) stressed the importance of causal models in longitudinal research. Furthermore, it is argued that knowledge from substantive theory which is supported by empirical evidence could be profitably incorporated into the analysis of data from longitudinal studies through the formulation of causal models. The use of causal models is a two-stage procedure involving first the development of the structural model, which specifies the relationship between the explanatory factors and the outcome measures. Secondly, there is the testing and refinement of that model and the choosing between alternative models in the examination and estimation of causal effects.

In the following sections, specific models aimed at clarifying issues related to learning outcomes in mathematics are addressed. The first section presents models which represent the underlying structure of the mathematics achievement tests, since the learning outcomes employed in the present study are developed from the assumptions involved in the mathematics test design. The subsequent section discusses the multivariate causal models of student level factors influencing mathematics achievement.

Models of the Underlying Structure of Mathematics Achievement Tests

IEA has conducted three International Mathematics Studies at the 13-year-old student level. The items in the tests employed in these studies were classified into content and cognitive processes by IEA and participating countries (see Chapter 5). Great care was taken to develop the tests on the one hand, measuring the four different content areas, namely Arithmetic, Algebra, Geometry and New Mathematics, and on the other hand, measuring the different cognitive processes. The cognitive processes were classified into several categories, namely computational and verbal processes, lower and higher mental processes and computation, knowledge, translation, comprehension, analysis and application processes (Thorndike, 1967a; Keeves, 1968; Rosier, 1980).

The results of these international studies were reported in different ways. Some reported total mathematics scores, while others reported subscores, such as algebra, geometry, and others reported both the total and the subscores. Thorndike (1967b, p. 32) has reported the distribution of the FIMS total and subscores. The reason for reporting both the total and subscores given by Thorndike was that “in some ways, it is more instructive to compare national and other groups on part scores for a mathematics test than on a single total score.” However, he did not provide further evidence as to why he reported both scores, although Peaker (1969) subsequently studied the interrelations between whole and part scores to confirm the use of the total score.

Rosier (1980, p. 62) has also stated that “the sets of items in the various content and processes categories were regarded as constituting subtests, for which corresponding scores were also calculated.” However, no argument was advanced as to why scores should be calculated for the subtests, other than that the curriculum had been for convenience subdivided into such strands.

Meanwhile researchers who have undertaken data analyses conducted by IEA have presented the scores based on either a total, the content areas, the cognitive processes or all three classifications. Pidgeon (1967), Thorndike (1967b), Rosier (1980) and Moss (1982) reported the total score and the subscores for both the content and cognitive processes, while Beaton et al. (1996), and Lokan et al. (1996), have presented the total score and only the subscores for the content areas. Others have also reported only the subscores for the content areas (Hanna, 1989; Robitaille, 1990; Xin Ma, 1995). However, none of the studies cited considered whether it was valid to have reported either the total score, the subscores or both the total and the subscores.

Thus consideration must be given as to whether it is appropriate to calculate a total score or whether to calculate the students’ scores on different content and cognitive processes separately. This issue has considerable effect on the manner in which certain multivariate analyses must be conducted and on whether or not certain analyses should be undertaken. Critical issues which require consideration in the present study are summarised below.

1. *Is it appropriate to calculate a total score for the IEA Mathematics test based on all items, when the items in the test are selected to measure different content areas and cognitive processes?*

Hypotheses associated with content areas and cognitive processes involved in responding to mathematics tests can be represented by factor-analytic models (Gustafsson, 1997). Figure 6.1 shows four different hypothetical models of the factor structure underlying the mathematics test employed in the International Mathematics Studies and the Ethiopian Mathematics Study. While these models consider only three content areas associated with the learning of mathematics, similar models could be developed which involved cognitive processes. The four models labelled Figures 6.1a to 6.1d were developed from consideration of the framework underlying the design of the tests. The graphical representations of the models show observed and latent variables. The rectangular shapes represent the observed variables, that is the mathematics test items administered to the students, whereas the ellipse shapes represent the latent variables. Arrows in the figure show a direction of influence of the responses of students to items which reflect the influence of the hypothesised factors.

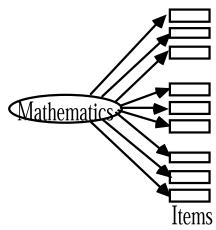


Figure 6.1a: One Factor Model (1FM)

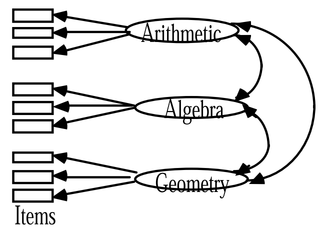


Figure 6.1b: Three Factor Model (3FM)

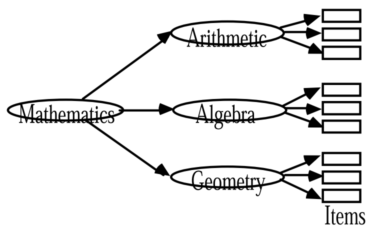


Figure 6.1c: Hierarchical Model (HM)

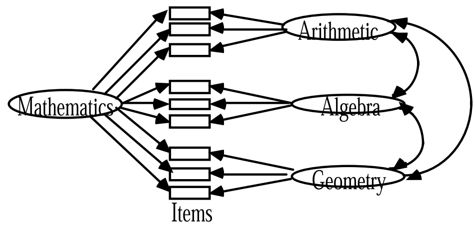


Figure 6.1d: Nested Model Correlated (NMC)

Figure 6.1 Hypothetical factor structures of Mathematics test items

Figure 6.1a presents a One Factor Model (1FM), in which observed items are assigned to only one-single-first order factor labelled as *Mathematics*. Figure 6.1b shows a Three Factors Model (3FM), in which items in the Mathematics test are assigned to three content areas, Arithmetic, Algebra and Geometry. It is important to note that each item is assigned only to one factor and the content area variables are considered to be correlated.

Figure 6.1c presents a Hierarchical Model (HM). HM is an extension of the 3FM in which the covariation between the first-order factors is assumed to be explained by one general higher order factor, *Mathematics*. It should be observed that by moving the correlations in 3FM to the HM, the three correlations are replaced by three factor loadings, and make the HM more meaningful than the correlated factor model. However, with only three first or lower order factors, the two models are highly similar.

Figure 6.1d shows a Nested Model (NMC), where all the items are assigned to a general factor, *Mathematics,* and to one of the three correlated specific factors, namely Arithmetic, Algebra and Geometry which are orthogonal to the *Mathematics* factor. Gustafsson and Balke (1993) have pointed out that a nested model is a preferred to a hierarchical model, because the NMC is specified in terms of relations between latent and observed variables, whereas in HM some relations are indirect. Therefore, a decision was made to compare the HM, and NMC in the present study. Criteria for examining the relative appropriateness of the hypothesised models between countries are addressed in Chapter 7. It should be noted that the three specific factors are assumed to be correlated, in a way that is contrary to models advanced by Gustafsson and Balke (1993), but is consistent with models advanced by Byrne (1996).

It is important to note that the hypothesised models discussed above were based on the total score and only the three content areas. Four content areas were mentioned above. However, nine of the 13 New Mathematics items were classified either as Arithmetic, Algebra or Geometry (see Chapter 5), and there were only four New Mathematics items that were solely assigned to this category. Consequently, it was decided to remove New Mathematics from the analysis.

In addition, the mathematics tests were categorised into several types of cognitive processes. Thorndike (1967b, p. 31) reported on the grouping of the FIMS items in terms of cognitive processes as follows:

> The items were grouped in several different ways. First, they were classified (by the pooled judgement of several reviewers) into items calling for higher mental processes and those calling for lower mental processes. The latter are judged to call for relatively routine application of previously learned techniques, while the former call for a greater amount of ingenuity and inventiveness in the attack upon novel or complex problems. A second subdivision of the items was between those that were verbally formulated and those that involved primarily computation or solution of a problem expressed in numbers or symbols.

Therefore, similar types of hypothetical models were also developed and examined based on the total score and the three major cognitive processes, namely comprehension and verbal ability, lower and higher order mental processes and comprehension, translation, analysis and application.

2. *Is it meaningful to compare student achievement over time (between 1964, 1978 and 1994) and across- countries (between Australia and Ethiopia)?*

If the different content areas and cognitive processes involved the mastering of different skills, it might be argued that it would not be reasonable to compare the total scores over time. However, it would be possible to compare the total scores and measure changes in student achievement level between two or more occasions and between two or more countries, if the mathematics test involved a general performance factor of *Mathematics*.

3. *Is it meaningful to compare the student level factors influencing mathematics achievement between studies and between two or more countries? If so, are there changes over the 30-year period, and between the two countries? What are the differences between the countries and over time?*

Such comparisons could be considered if the measures of student achievement in mathematics were comparable over time and across countries. Furthermore, if the measures were comparable, then it would be possible to investigate if there were changes in the effects of background factors, such as student variables, in their influence on the observed changes in achievement. The manner in which student level factors were assumed to influence mathematics achievement is addressed in the next section.

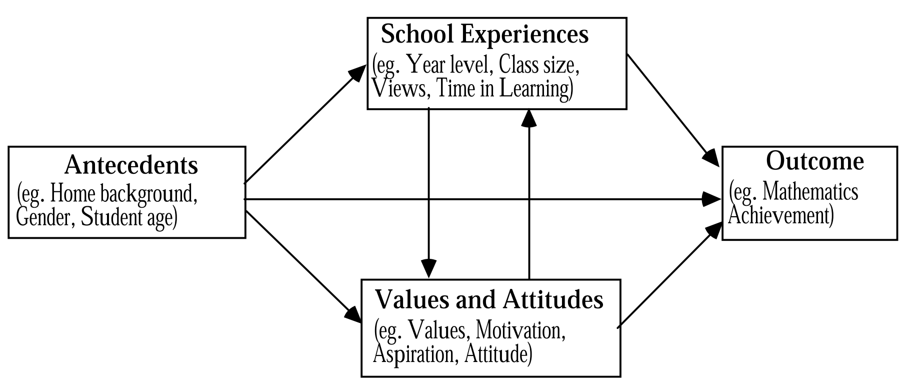


Figure 6.2 Model of School Learning in Mathematics

Models of Student Level Factors influencing Mathematics Achievement

Carroll (1963) developed a model of factors that influence school learning. The factors identified by him are divided into two levels, namely student and school level factors. The student level factors are aptitude (home background), ability and perseverance (motivation and attitude). While, the school level factors are time for learning (including time for homework) and quality of instruction. Carroll (1963, p. 723) in his article *A model of school learning* has argued that the primary task of educational researchers is to develop and apply knowledge concerning the success or failure of students' learning at school, and to assist in the prevention and remediation of learning difficulties. In order to achieve this goal, Carroll suggested that researchers should develop a theoretical model of factors influencing success in school learning and of the way the factors interact. In the first stage of development of a model, an attempt must be made to structure the variables systematically for particular phenomena related to school learning based on theoretical considerations and previous research findings. Thus, in developing a causal model the variables must be arranged in a sequential order.

This process of developing a hypothetical model is the first important step in the study of causality in non-experimental studies in educational research. Once the hypothetical model has been developed it can be shown by graphical representation in a path model. The model is tested by using a path analysis procedure which is employed to examine direct and indirect effects of predictor variables on the criterion. Figure 6.2 shows a general hypothetical path model of school learning in mathematics which was developed based on Carroll's (1963) model of school learning. The left hand side of the figure shows the *Antecedent* latent variables that are not affected by any other variable. They include student's home background, gender and age. These variables are assumed either positively or negatively to influence a student's school experiences, values and attitudes and the outcome variable of mathematics achievement. The variable *School Experiences* includes observed measures such as student's year level, class size, time in learning and views about school and mathematics learning. These factors are hypothesised to influence the outcome variable, that is, *Mathematics Achievement* and *Values* and *Attitudes* directly. Furthermore, *Values* and *Attitudes*, that is, the student's values, motivations, aspirations and attitudes towards school and mathematics learning are assumed to influence the individual student's achievement level in mathematics, which is the outcome variable as well as *School Experiences* possibly in a reciprocal relationship. Meanwhile, the *Antecedent* variables such as *home background*, *gender* and *age* of the student also influence the outcome variable directly and indirectly through the mediating variables *School Experiences, Values* and *Attitudes*. Consequently, the outcome measure is influenced directly by *School Experiences*, *Values* and *Attitudes* and by the *Antecedent* both directly and indirectly through *School Experiences,* and *Values* and *Attitudes*. It is important to point out that this model measures outcomes of school learning at the individual student level, because it is at this level that learning takes place. Therefore, any study of student learning in mathematics must consider information on cognitive and affective outcomes of learning at the student level.

Consequently, one of the purposes of the present study is to examine the changes in the effects of student level factors influencing mathematics achievement between 1964, 1978 and 1994 as well as the differences between school learning in Australia and Ethiopia based on Carroll's (1963) model of school learning. From the general model presented in Figure 6.2, more detailed models were developed and are presented in Figure 6.3. In both the International and Ethiopian Mathematics Studies, information was collected from students by employing the General Information Questionnaire (see Chapter 5). Table 6.1 lists the latent and manifest variables which were employed in the hypothesised path models of student level factors that were hypothesised to influence mathematics achievement for EMS, FIMS, SIMS and TIMS data sets. The first column of the table indicates the latent variables (LVs), while the second column lists the manifest variables (MVs) which were related to the latent variables.

Figure 6.3 show the outer and inner model relationships of the hypothesised student level factors thought to be influencing mathematics achievement of students in Australia and Ethiopia. The outer model indicates the manifest variables that are shown by rectangular boxes, whereas the inner model presents the latent variables that are shown in ellipse shapes. The assignment of manifest variables to latent variables is based on the information in Table 6.1. Figure 6.3 show the outward and inward mode for the exogenous (that act as cause) and endogenous (that act as effect) latent variables. The outward mode is depicted by the arrows pointing from the latent variables to their respective manifest variables, whereas the inward mode is showed by the arrows pointing from the manifest variables (MV) to the corresponding latent variable.

Figure 6.3 show the MVs and the corresponding LVs and the estimation modes specified to estimate the constructs. In order to increase the predictive power of the path model, Sellin (1992) has suggested the use of the inward mode for exogenous variables and the outward mode for endogenous variables. In accordance with his suggestion, the exogenous variables in this study were defined with inward estimation(see Figure 6.3). The endogenous variables were defined with outward estimation including those constructs that consisted of only one MV except *Time in Learning* (Timlearn), because *Hourmath* and *Hourmhmw* are viewed as alternatives and the inward mode was chosen.

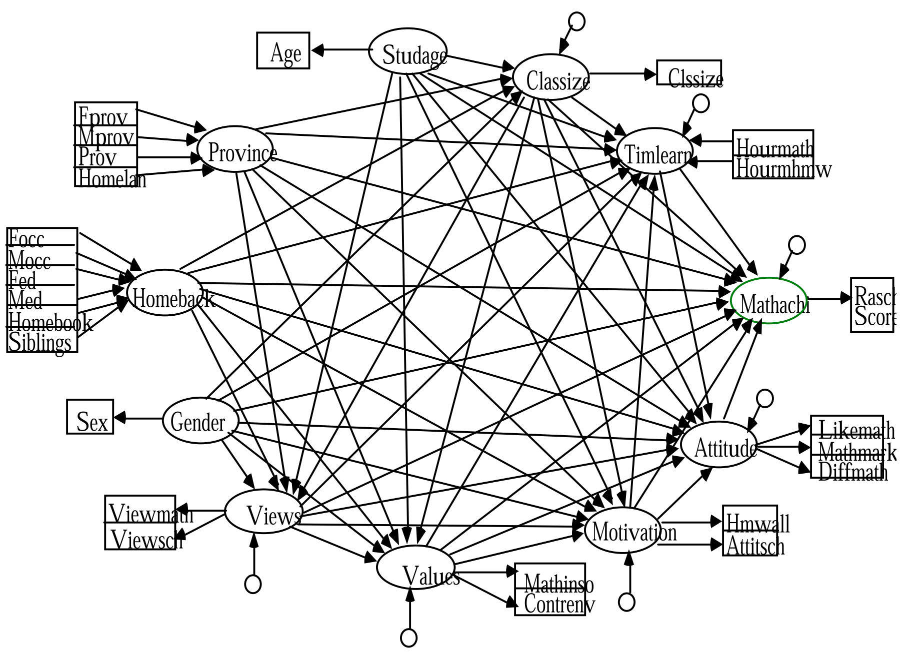


Figure 6.3 Outer and Inner model relationships of hypothesised student factors predicting Mathematics Achievement

The FIMS data set was divided into two subsets, FIMSA and FIMSB. FIMSA includes all 13-year-old students in Years 7, 8 and 9, while FIMSB involves all students in Year 8 in the sample. The only differences between the two groups were that in FIMSA, year level was considered as a latent variable, while student age was not a variable, since all students were 13-year-olds, but differed in year levels. Meanwhile in FIMSB, age was considered as a variable, but not year level, because all the students in this group were Year 8 students. However, the students were of different age levels. Therefore, to examine the effects of differences in year level and age, these two groups of students were treated separately. In EMS and TIMS, since all students were Year 8 students, year level was not considered as a variable. However, age was a variable, because the students were drawn from very different age groups. Moreover, in SIMS all the students were 13-year-old students in Years 7, 8 and 9. Therefore, year level was considered as a variable.

The inner model relationships between latent variables are indicated by the theoretical framework presented in Figure 6.3a through 6.3e. The latent variables *Homeback*, *Gender, Province* and *Studage* for EMS data set, *Homeback* and *Gender* for FIMSA, and *Gender, Homeback* and *Studage* for FIMSB data sets, *Homeback*, *Gender* and *Ethnicity* for SIMS data set, *Homeback, Gender, Studage,* and *Ethnicity* for TIMS data set are exogenous variables or antecedents. They are latent variables that are not influenced by any other latent variable in the hypothesised model. The remaining latent variables in the models are called endogenous variables, that is they are latent variables which influence and are influenced by one or more other latent variables. In all models mathematics achievement (*Mathachi*) is the outcome variable. In the report of this study, both manifest and latent variable names are italicised and indicated by an initial capital letter. Any latent variable in the model may influence achievement either directly, indirectly or both directly and indirectly. The reasoning for the ordering of variables and the way in which latent variables might influence each other and eventually *Mathematics Achievement* is presented as follows.

In all the four data sets, *Homeback* and *Gender* are exogenous variables because they are not influenced by any other variable in the model. The construct *Homeback,* which is represented by *Parental Education* and *Father's Occupation*, is considered to influence *Views* positively, that is students from a higher socioeconomic status background are likely to express more positive *Views* about mathematics and schooling than students from lower socioeconomic family backgrounds. In all the four data sets *Homeback, Gender, Classize, Views, Motivation, Timlearn, and Attitude* are seen as factors that have an effect on *Mathachi.* Among these latent variables only *Gender, Classize, Views, Motivation and Attitude* might show a direct effect on *Mathachi.* However, *Homeback* and *Timlearn* might have an indirect effect on *Mathachi.* For example, if *Homeback* influences *Attitude* positively, this means that those students from higher socioeconomic backgrounds are likely to express more positive attitudes towards mathematics than students from a lower socioeconomic backgrounds. Moreover, if *Attitude* influences *Mathachi* positively, then students who express more positive attitudes towards mathematics are likely to achieve at a higher level in mathematics. Thus, *Homeback* might influence *Mathachi* through *Attitude*. In this way, the model allows endogenous latent variables to mediate the effects of other latent variables earlier in the model on latent variables later in the model and finally on *Mathematics Achievement*, which is the outcome variable.

Attitude is argued to be influenced by *Homeback.* That is, a higher level of parental education and occupation are likely to be related to more favourable students' attitudes towards mathematics and schooling.

Table 6.1 Latent and manifest variables employed in the Path models for EMS, FIMS, SIMS and TIMS data sets

Latent Variables	Manifest Variables
Homeback[i] (Home background of student)	Measuring the socioeconomic status of student's parents • Focc (Father's occupation); • Mocc (Mother's occupation) • Fed (Father's level of education); • Med (Mother's level of education) • Homebook (Number of books in home); • Siblings (Number of siblings)
Gender[o] (Sex of student)	Identifying whether the student is female or male • Sex (Sex of student)
Studage[o] (Age of student)	Identifying student's age) • Age (Age of student
Yearlevel[i] (Year level of students)	identifying the level of students in school • Year 7; • Year 8; • Year 9
Ethnicity[ai] (Ethnic background of the student and his/her parents)	Identifying the country of origin of the student and his/her parents • Fcntry (Father's country of birth); • Mcntry (Mother's country of birth) • Cntry (Student's country of birth); • Yrscntry (Number of years the student lived in Australia); • English at home (English spoken at home)
Province[bi] (Province where the student and his/her parents came from)	Identifying the province of origin of the student and his/her parents • Fprov (Father's province of birth); • Mprov (Mother's province of birth) • Prov (Student's province of birth); • Homelan (Language spoken at home)
Classize[o] (Number of students in the class)	Identifying the number of students in Mathematics class • Clssize (Number of students in the class)
Views[o] (Student's views about mathematics teaching and school and school learning)	Measuring student's views about mathematics teaching and school and school learning using • Viewmath (Views about mathematics teaching scale) • Viewsch (Views about school and school learning scale) • Studpart (Views about students' participation in math teaching scale)[d]
Values[o] (The values of student towards mathematics)	Measuring student's values about mathematics employing • Mathinso (Student's attitude towards the place of math in society scale) • Contrenv (Student's attitude towards control of the environment scale)
Motivation[o] (Motivation of student)	Measuring the student's level of motivation using • Hmwall (Number of hours a week used by student to do all homework) • Attitsch (Student's attitude towards school and school learning scale) • Motiv1 to Motiv4 (Students need to do well in mathematics)[d]
Timlearn[i] (Time in learning)	Assessing the amount of time used by the student to learn mathematics and to do mathematics home work and the frequency of homework given to students by their teachers • Hourmath (Number of hours in a week used by student in learning math) • Hourmhmw (No of hours a week used by student to do math homework) • Homworkf (Frequency of math homework given for students in a week)[d]
Aspirat[co] (Aspiration of the student)	Measuring the student's level of aspiration • Expted (Student's expected education level) • Desired (Student's desired education level) • Exptocc (Student's expected occupation) • Desirocc (Student's desired occupation)
Futmath[co] (Future mathematics)	Measuring the student's level of expectations and wishes to take more mathematics courses • Expmorma (Student's expectations to take more mathematics courses) • Wishmorma (Student's wishes to take more mathematics courses)
Attitude[o] (Student's attitude towards mathematics)	Examining the attitudes of a student towards mathematics • Belima (Mathematics is student's best liked subject)[f] • Besubma (Mathematics is student's best subject)[f] • Likemath (Mathematics is student's best liked subject) • Likmath (Students' liking of mathematics scale)[d] • Mathmark (Best mark of the students is mathematics) • Diffmath (Student's attitude towards facility of mathematics)[e]
Mathachi[o] (Mathematics achievement of student)	Showing the mathematics achievement level of student by employing • Rasch Score (Rasch estimated scores of the mathematics test)

a = variable considered only in SIMS and TIMS, b = variable considered only in EMS, c = variable considered only in FIMS, d = MVs variables considered only in TIMS, e = MVs only one item was considered in TIMS, f = MVs variables considered only in FIMS, o = Outward mode, i = Inward mode

Furthermore, both *Homeback* and *Attitude* are also considered to have an effect on *Mathachi*, because students from a higher socioeconomic background and who express more favourable attitudes towards mathematics and schooling probably recognise the importance of knowledge and skills obtained from the successful learning of mathematics.

It is necessary to observe that Figure 6.3 depict a path model in which all possible paths between latent variables have been shown. Moreover, it is expected that during analysis the model would be refined as a consequence of the interim results to include only those paths that were supported by the data. In this way, a more parsimonious model would be developed, while still maintaining a coherence within the theoretical framework that has been advanced.

Research Questions

In Chapter 4, previous studies in the area of mathematics in Australia and Ethiopia are reviewed. From these research findings, it was apparent that a number of factors were of interest when considering differences and similarities in mathematics achievement at the 13-year-old student level in both countries and over time in Australia. The mathematics achievement level of students was expected to be influenced by some variables directly, while other variables were expected to influence achievement indirectly through mediating factors. Moreover, some variables would influence achievement in one country but not in the other, and some variables would influence achievement on the first occasion, but not on a later occasion, while others could influence achievement on the second occasion but not on a later occasion.

The theoretical framework which has been developed for the investigation of the underlying structure of mathematics achievement tests and the student level factors influencing mathematics achievement are discussed in the previous section. Whereas in the following section the research questions to be addressed in the present study are presented.

General Research Questions

The research questions to be investigated in this study have been developed on the basis of previous research findings and theoretical considerations. For clarity and simplicity the research questions are divided into general and specific research questions. For reasons of convenience, the research questions have also been grouped either under achievement in mathematics, views and attitudes towards mathematics and schooling, or student level factors influencing mathematics achievement.

Achievement in Mathematics

1. Does mathematics achievement consist of separate skills or one underlying dimension?
2. Have changes occurred in the level of achievement in mathematics of the 13-year-old student between 1964 and 1978 in Australia?
3. Have changes occurred in the level of achievement in mathematics of the Year 8 student between 1964 and 1994 in Australia?
4. Are there difference in the level of achievement in mathematics between Year 8 1994 Australian and 1996 Ethiopian students?
5. Are there gender differences in the achievement in mathematics for 1964, 1978 and 1994 Australian students?

6. Are there gender differences in achievement in mathematics for 1996 Ethiopia students?

Views and Attitudes towards Mathematics and Schooling

7. Are there changes in the attitudes towards facility of learning mathematics between 1964, and 1978 Australian students?
8. Are there differences in attitudes towards facility of learning mathematics between 1964 Australian and 1996 Ethiopian students?
9. Are there gender differences in the attitudes towards facility of learning mathematics for 1964 and 1978 of Australian students?
10. Are there gender differences in the attitudes towards facility of learning mathematics for the 1996 Ethiopian students?
11. How do the 1964 Australian students' attitudes towards schooling compare with those of the 1978 students and 1996 Ethiopian students?
12. Are there gender differences in the attitudes towards schooling between Australian and Ethiopian students?
13. How do the 1964 Australian students' views about the teaching of mathematics compare with those of the 1978 and the Ethiopian students?
14. Do boys and girls share common views about the teaching of mathematics in both countries?
15. How do the 1964 Australian students' views about schooling compare with those of the 1996 Ethiopian students?
16. Are there gender differences between the 1964 Australian and the 1996 Ethiopian students' views about schooling?

Factors influencing achievement in mathematics

17. What student level factors influence student learning in mathematics in Australia and Ethiopia?
18. Are there differences between 1964, 1978 and 1994 in the student level factors that influenced the learning of mathematics achievement in Australia and 1996 in Ethiopia?

Conclusion

In this chapter a general theoretical framework of mathematics learning is developed for investigation in the present study. Different types of hypothetical models are developed for examination. The first models developed for testing seek to represent the structure of the mathematics tests as specific outcomes of school learning. Hence different types of models are proposed for examining the dimensionality of the tests. Causal models have also been developed to examine the student level factors that influence the mathematics achievement level of Australian and Ethiopian students. In developing all the hypothetical models, previous research findings have been taken into account. The procedures employed for testing the hypothetical models discussed in this chapter are considered in the next chapter.

7

Methodological Discussions

The purpose of this chapter is to identify the methods of data analysis which are both appropriate for testing the proposed models and appropriate to address the research questions listed in Chapter 6. The use of inappropriate procedures might provide misleading results that fail to elucidate the research questions. However, it is important to remember that no one method is likely to be suitable to answer the diverse issues and the ways in which student level variables have given rise to possible changes in mathematics achievement over time in Australia and between Australia and Ethiopia. Hence, this chapter presents a discussion of different methods that might be used in this study to examine the models proposed in Chapter 6.

The use of the path analysis is discussed in the first part of the chapter. The second part presents the item response theory or the Rasch model. Confirmatory factor analysis is addressed in part three, while in part four the test equating procedures are discussed. In part five the procedures for identifying item bias and analysing data with complex samples are presented. The last part is the conclusion.

The Use of Path Analysis

Educational researchers have frequently sought to identify the factors that influence student achievement in particular subject area. These researchers have developed their hypotheses and tested them using different methods. In educational research studies it is necessary to develop models from theory and to examine systematically the models in order to advance theoretical understanding (Tuijnman and Keeves, 1994, 1997). Sellin and Keeves (1994, 1997) have argued that the use of path analysis with latent variables using partial least squares analysis (PLS) demands the development from theory of a well-specified model for examination and estimation. The model specifies which variables should be included in the analysis and which variables are assumed to be the causes and which are the effects. Since the early 1970s the acceptance of path analysis and causal modelling in educational research and in research in the social and behavioural sciences (Sellin and Keeves, 1997) has increased. This procedure enables the researcher to shift from verbal statements of interrelationships between variables to accurate mathematical ones and to estimate the magnitudes of the causal links involved.

Partial Least Square Path (PLSPATH) Analysis

Partial Least Square Path (PLSPATH) Analysis "is a general technique for estimating path models involving latent constructs indirectly observed by multiple indicators" (Sellin, 1992, p. 398). It is useful in modelling educational and social systems, by its capacity to examine a large amount of data. Thus, PLSPATH analysis is employed as a method of analysing path models which involve latent (indirectly observed) and manifest (directly observed) variables. Pajares and Miller (1994) have pointed out that PLS procedures are also used to examine the direct and indirect effects between variables. The PLS model includes an inner model which specifies the hypothesised relationships among latent variables (LVs) and an outer model that specifies the relationships between LVs and the manifest variables (MVs) (Sellin, 1992).

Like any other statistical procedure, there are assumptions that underlie the use of path analysis. These assumptions were outlined by Pedhazur (1982), and Pizzini and Shepardson (1992, pp. 249-250) have summarised these five assumptions:

(a) the relationships among variables are linear, additive and causal;

(b) all relevant variables are included in the model;

(c) there is a causal flow in the relations between variables in the model;

(d) the variables are measured on an interval scale; and

(e) the variables are measured without error.

PLS considers an optimal linear relationship between variables and gives model parameter estimates quickly and effectively. PLS has common conceptual origins with alternating least squares (Young, 1981), canonical analysis (Levine, 1977), covariance structure analysis (Long, 1983a, 1983b) and principal components analysis (Meredith and Milsap, 1985). The fundamental characteristic of PLS is the precise estimation of latent variable scores by means of least squares methods (Sellin, 1989; Sellin and Keeves, 1994). The term *partial* refers to the mathematical computation of an estimate for each latent variable from its associated manifest variables. This is done either through methods equivalent to factor analysis or regression analysis. Subsequently in the non-iterative estimation of the inner model, the coefficients and loadings are generated by employing least squares estimates. The PLS procedure determines the weights defining the LV estimates. These weights are obtained iteratively by a series of either simple or multiple ordinary least squares (OLS) procedures applied to each block of MVs in turn.

Jacobs (1991), and Bukowski et al. (1993) have argued that PLS is an ideal procedure, because it provides an index of the adequacy of the model, shows the strength of each individual path in the model, and examines the direct and indirect relations among variables. Kotte (1992) also argued that PLS can be employed efficiently to identify strong and weak relationships between LVs and MVs. Furthermore, the indices of overall adequacy show whether or not the model produces an accurate representation of associations among variables in the model. In a causal model certain variables are singled out as causes and others as effects. So, the strengths of the particular paths in the model show how strongly the linked variables are associated with each other. PLS provides a number of advantages which are most appropriate for this study. It is useful for displaying graphically the pattern of causal relationships among sets of observed and unobserved variables that influence the mathematics achievement level of students. PLS is technically simple, and easy and quick to use. Moreover, it does not require strict distributional assumptions. For these and other reasons PLS is referred to as an initial approach to modelling (Falk, 1987; Sellin, 1989, 1990).

It is beyond the scope of this study to detail the mathematical and technical aspects of PLS. However, references such as Noonan and Wold (1988), Sellin (1989), Cheung and Keeves, (1990); Falk and Miller (1992) and Sellin and Keeves (1997) are suggested for further reading.

PLSPATH Model Construction

A PLSPATH model is defined by two sets of linear equations called the inner and outer models (Noonan and Wold, 1988). The formal specification further includes the relations for substitutive prediction of the latent and manifest variables. Inner and outer relations are subject to predictor specification, and all outer model residuals are assumed not to have correlations across the blocks and with other latent variables. The inner model describes the hypothesised relationships among the LVs, while the outer model indicates the links between latent variables and MVs (Sellin and Keeves, 1997). LVs which are not dependent on other LVs are known as exogenous or antecedent latent variables, while the LVs that are dependent on other LVs are called endogenous or mediating latent variables (Falk, 1987; Keeves, 1988b, Sellin and Keeves, 1997). The former variables are thought not to be influenced by other variables in the path model specified (for example, gender), while the latter LVs are variables which influence and are influenced by one or more other LVs (for example, attitude towards mathematics). A dependent LV influenced by other LVs but not influencing other LV, is known as outcome measure (for example, mathematics achievement). Hence, PLSPATH analysis is a technique that investigates causal relationships between independent and dependent variables that would be applicable in the models of mathematics achievement.

In beginning to use PLS it is necessary to draw a diagram of the causal model to be analysed (Falk, 1987). Falk provided four basic elements and four derivative terms that should be included in the path diagram. These are:

(a) the LVs or predictors each of which is represented by an ellipse;

(b) manifest or observed variables which are represented by rectangles;

(c) arrows that indicate the direction of relationships between variables are interpreted as predictive or casual relations and are represented by one headed arrows; and

(d) spans that are arrows which indicate two-way relationships in the sense of correlations or covariances which represent residuals or uniquenesses. They are graphically represented as two headed arrows.

There are several derivative terms that arise from combinations of the four elements. The arrows and the spans between the latent variables are called the inner model or “the latent variable path model”. The arrows and spans between the MVs and LVs are called the “outer model”. The arrows and spans between a single LV and its indicants are called a “block”. Within a block the arrows may have one of two directions, inner or outer. “Inner directed blocks” have arrows pointing from the squares to the circle. “Outer directed blocks” have arrows pointing from the circle to the squares. A block may have only one direction. A “path diagram” is thus a graphic representation of the relationships between variables (Falk, 1987, pp. 11-14). The word partial is used to indicate that each block is estimated at a time, since the overall path diagram is partitioned into its blocks (Falk, 1987, p. 26).

In the testing of models Tuijnman and Keeves (1994) suggested that the principles of coherence and parsimony would seem the most appropriate to apply. The principle of coherence shows the level of agreement between theoretical considerations and the inclusion of a path in a model on the basis of its magnitude estimated in the testing of the model, while the principle of parsimony refers the deletion of a path and a parameter from a model, if there are only tenuous empirical grounds for supporting its inclusion (Tuijnman and Keeves, 1994, p. 4345).

In this study PLS was employed to identify the student level factors that influenced the mathematics achievement of the 13-year-old and Year 8 students in EMS, FIMS, SIMS and TIMS.

Trimming the model

The modification, trimming and deletion of variables and paths in the path model involve the removing of all paths not contributing to the LVs. The deletion or the removal of paths includes both the outer and inner models. As a result of the trimming procedure some MVs and LVs are also removed from further analysis. The results of the analyses after the misfitting variables are deleted are presented in Chapter 11.

Computer Program Employed

PLSPATH version 3.01(Sellin, 1990) computer program was employed for the PLSPATH analyses in this study. The program estimates path models with LVs measured by multiple indicators using the PLS techniques. PLSPATH was chosen for this study because it provides more flexibility in program handling and is more rapid in its estimation of the path coefficients than the alternative programs. The data used with the PLSPATH computer program can be either in the form of a correlation matrix, as a square or lower triangular matrix, or a raw data file, from which the correlation matrix data are computed automatically by the computer program. In the present study, square correlation matrices were generated by employing SPSS, Version 6.10 (Statistical Package for Social Science, 1993) from raw data files for the mathematics tests items. In the preparation of the correlation matrices missing data were treated by using pairwise deletion procedures.

Summary

PLSPATH was employed to perform the model estimation discussed in this section. PLSPATH is a very important exploratory and confirmatory method to reduce the amount of information associated with many MVs by allowing the construction of LVs. It also examines a theoretical model and estimates the parameters of the model by using least squares procedures which are known to be robust and not to require strict distributional assumptions. The results of the PLSPATH analyses are presented in Chapter 11.

Test Theory

Among the roles of educational research, there is the need to examine change in the levels of student achievement over a period of time. These duties of educational research can be carried using test equating techniques. If a set of anchor items is included in different test forms, then the performance of the candidates on the anchor items can be used to adjust the scores so that they are measured on the same scale.

Research into the procedures for equating tests has been an on going process from the 1920s (Flanagan, 1982; Skaggs and Lissitz, 1986) at a time when the field of educational and psychological measurement was beginning to develop. However, interest in equating research has intensified since the 1960s, because of the development of item response theory (IRT) and the availability of appropriate computer programs for their application. The test theories concerning test equating and related issues are categorised into two groups, namely classical test theory and the item response theory. In the first part of this section the classical test theory, its strength and weaknesses are presented while, the strength and weaknesses of the item response theory are discussed in the second part.

Classical Test Theory

Classical test theory (CTT) assumes that the "observed score on a test is the sum of a true score component and a measurement error component." (Feldt and Brennan, 1989, p. 108). This theory considers that a student's score on a test is given by a true score plus an error score associated with the test. The mathematical representation of the model is:

Observed score = true score + error score (Hopkins and Antes, 1990, p. 296)

The consequence of this assumption is that the error scores associated with the test are linearly independent of the true scores associated with the observed score and independent of the error scores arising from other tests (Feldt and Brennan, 1989). Weiss and Yoes (1991, p. 70) have also indicated that the CTT model is a linear one, specifying that there is "a linear relationship between a person's observed number-correct test score on a test and the error-free true score that it estimates."

Weiss and Yoes (1991) cited the four major features of classical test theory. The features are item difficulty, item discrimination, reliability and number-correct test scores.

Item Difficulty

In CTT item difficulty has been defined as "the proportion of examinees who answer a specified item correctly." The item difficulty is also known as the p-value of an item, in some writings it is also referred as item facility (Weiss and Yoes, 1991, p. 70). Weiss and Yoes pointed out that difficult items are associated with low p-values, whereas easy items are associated with high p-values.

Item Discrimination

Wood (1988, p. 377) has defined item discrimination as "the power of an item in separating the more from the less capable on some latent attribute." Furthermore, he indicated that the procedure calculates a correlation, either biserial or point biserial, between the success and failure on the item and a score on a measure that is considered to represent the latent attribute. Thus, a higher correlation between a student's score on an item and the student's total test score is associated with high discriminating power of the item in separating the better students from the poorer students. Weiss and Yoes (1991) have recommended the deletion of items from a test if their discrimination power is low, in order to improve the measurement characteristics of the test.

However, before doing any analysis, it is necessary to decide whether a biserial or point biserial correlation should be employed. Wood (1988) has argued that the point biserial correlation is a special case of the Pearson product-movement correlation, where one of the variables (test scores) is regarded as continuous and the other (item score) can take one or other of two discrete values, 1 for right and 0 for wrong values. Furthermore, Wood pointed out that the biserial correlation assumes that both variables are continuous, but that one has been divided into two groups, those who got the item right and those who did not. Those who got the item right are thought to differ from those who did not get it right only in having more of whatever the item measures. Wood (1988) emphasised that the biserial correlation is not a product-moment correlation, and it should be considered as only a measure of association.

The biserial and point biserial correlations are closely related and it is possible to convert from one to the other. However, researchers such as Thorndike (1982) and Wood (1988) have recommended that the point biserial correlation is more informative in ordinary circumstances when compared with the biserial correlation.

Reliability

Gruijter and Kamp (1991, p. 45) have defined reliability as “the extent to which observed variation reflects true variation in measurements.” Furthermore, Mehrens and Lehmann (1984, p. 269), and Gruijter and Kamp (1991, p. 45) have given a more technical definition of reliability as the “ratio of true to observed score variance.” The standard error of measurement (SEM) and the reliability coefficient are the two indices cited by Feldt and Brennan (1982) that can be employed to quantify the reliability of a measuring instruments. These researchers have argued that the reliability coefficient summarises the “consistency (or inconsistency) among several error prone measurements”, whereas the SEM summarises the “potential within-person inconsistency in score-scale units” (Feldt and Brennan, 1982, p. 105).

Weiss and Yoes (1991, p. 71) have defined reliability (r_{tt}) the standard error of measurement (SEM) in the classical test theory context as

$$reliability = 1 - (error\ variance/total\ score\ variance)$$

and

$$SEM = SD_X\sqrt{1 - reliability}$$

where SD_X is the standard deviation of the total test scores.

Number correct score

In classical test theory a person’s score is calculated based on the number of items correctly answered by the individual. This means that a person’s observed score on a test is equal to the number of items he or she answered correctly. Scores that are calculated in this manner are known as number-correct scores (Weiss and Yoes, 1991).

Weaknesses of Classical Test Theory

Classical test theory suffers from three major weaknesses, sample dependency, scoring problems and misconception of reliability

1. *Sample Dependency.* Osterlind (1983), Hambleton and Swaminathan (1985), Wright (1988), Hambleton (1989), Weiss and Yoes (1991) have argued that the CTT is dependent upon the sample to which the items were administered. Weiss and Yoes (1991, p. 70) have elaborated the problem as follows.

   > ... If a set of test items were administered to a high-ability group of examinees, the item difficulties (and probably the item discriminations) would be very different than if the same items were administered to a group of examinees of moderate or low ability.

2. *Scoring Problems.* The second major problem with CTT is concerned with the scoring of individuals. Test scores are usually total number-correct scores. Because persons are scored based on the number of items to which they responded correctly, test scores are dependent on the difficulty level of the items in the test (Weiss and Yoes, 1991). For example, if Test A contains easy items

and if Test B contains difficult items the test scores of students who sat for both tests would not be the same. Students would score better on Test A than on Test B, because Test A was easier than Test B.

3. *Misconception of reliability.* Weiss and Yoes (1991) have argued that reliability depends upon the sample of students involved in the total score distribution, because the computation of reliability involves the total score variance. Furthermore, Hambleton (1989) has argued that the problem concerning the concept of reliability stems from the fact that the concept is considered in terms of parallel-forms of a test that are difficult to achieve.

As a result of these weaknesses CTT has failed to provide appropriate solutions for the many problems raised in testing (Hambleton and Swaminathan, 1985; Weiss and Yoes, 1991). Consequently, CTT is gradually being replaced by item response theory (IRT). Weiss and Yoes (1991, p. 70) have argued that "it is at least partly in response to these recognised inadequacies of CTT that IRT was developed."

Item Response Theory

The item response theory assumes that student's performance on a test can be predicted (or explained) by latent traits (Hambleton and Swaminathan, 1985). The relationship between student performance and the probability that the student would provide the correct response is explained by a mathematical function. This mathematical function is known as an item-characteristic function or item response function (Hambleton and Swaminathan, 1985: Weiss and Yoes, 1991; Hambleton, 1994; Molenaar, 1995; Stocking, 1997). The item-characteristic function provides the probabilities of examinees at various ability levels answering the item correctly (Hambleton, 1994, p. 148).

Item response theory seeks to model a candidate's performance on a test as a relative function of the characteristics of the item and the person ability or trait on some unobserved, or latent trait. The model shows the relationship between a latent trait and observed performance on the test that is developed to measure that particular ability. This relationship is usually presented as an item characteristic curve (ICC) which is the graph of the item characteristic function (Skaggs and Lissitz, 1986; Hambleton, 1994).

Item response theory requirements

Hambleton (1993), Weiss and Yoes (1991) have described the three fundamental requirements which must be satisfied when using item response theory. These are as follows.

1. The dimensionality of the latent trait or the probability of a correct response by a candidate is attributable to his or her standing on some specific latent trait. For most IRT applications, the latent trait must satisfy the condition of unidimensionality.
2. Local independence or the probability of a correct response of a candidate to an item is not influenced by responses to the other items in the test.
3. The item characteristic curve (ICC) or item response function (IRF) or the number of parameters in the function used to describe the item dictates the form of the ICC, which results in the one, two or three parameter item response theory as well as the family of one parametric type models.

In addition, Weiss and Yoes (1991) included the *know-correct* assumption, namely if a person knows the correct answer to a test item, then that person will probably answer it correctly.

Before employing an item response theory it is necessary to test for the fulfilment of the requirements. These requirements can be checked by employing different techniques. The unidimensionality requirement can be examined by employing confirmatory factor analysis of an item intercorrelation matrix (Lord, 1980; Marsh and Hocevar, 1983; Hattie, 1985; Vijer and Poortinga, 1991; Weiss and Yoes, 1991). While the second requirement of an item response theory, namely local independence can also be tested by employing confirmatory factor analysis procedures (Weiss and Yoes, 1991). The other requirement of item response theory is the item characteristic curve (ICC) or item response function (IRF). Weiss and Yoes (1991) pointed out that this requirement involves the mathematical form of the ICC. Furthermore, these authors argue that the number of parameters in the mathematical function employed to describe the item dictates the form of the ICC. This has resulted in different measurement models within the family of item response theory.

The Item Response Theory (IRT)

Different kinds of IRT models have been developed since the 1960s, moreover, three models have been used for applications with achievement tests. These models are: the one parameter-logistic or Rasch model (Rasch, 1960), the two parameter logistic model and the three parameter logistic model (Birnbaum, 1968). The emphasis in the work that follows is on the one parameter-logistic model.

One parameter item response theory (Rasch model). The one parameter-logistic model, that is popularly known as the Rasch model, was developed by Georg Rasch (1960, 1966, 1980), a mathematician from Denmark. The Rasch model is the simplest and most popular of all IRT models because of its strong measurement properties (Skaggs and Lissitz, 1986). Wright (1977) argued that from all the IRT models proposed for person measurement, the Rasch model has the fewest components, just an ability parameter β_n for each person n and difficulty parameter δ_i for each item i.

The parameters represent the places of the persons and the items on the latent variable scale that they share. They are applied in the model to determine the probability of person n succeeding on item i (Rasch, 1960; 1966; Wright, 1968, 1977; Hambleton, 1994).

$$P\left(X_{ni} = x_{ni} \mid \beta_{ni}\delta_i\right) = \frac{\exp[x_{ni}(\beta_n - \delta_i)]}{1 + \exp(\beta_n - \delta_i)}$$

Hambleton and Swaminathan (1985) viewed the Rasch model as an item response theory in which the ICC is a one-parameter logistic function. Hence, this model requires that all items in a particular test have equal discriminating power and guessing is minimal (Scheuneman, 1979; Lord, 1980; Hambleton and Swaminathan, 1985).

Some researchers have indicated reservations about the effectiveness of these requirements. Opponents of the Rasch model argue that guessing plays a major role in responding to multiple choice items and the achievement test items differ in the degree to which they correlate with the underlying latent trait and thus it is not appropriate to assume uniformity in discrimination power of the items in the test. Hence, they have criticised this model for failing to account for guessing and unequal item discrimination (Hambleton and Swaminathan, 1985; Skaggs and Lissitz, 1986; Kline, 1993). However, proponents of the model contend that guessing is a characteristic of the respondent and not the item, and that, the Rasch model is fairly robust with respect to departures of model requirements normally observed in actual test data (Hambleton and Swaminathan, 1985; Skaggs and Lissitz, 1986).

Researchers have reported many advantages of Rasch model. Three primary advantages of the Rasch model have been identified by Wright, (1977), Wright and Stone (1979), Hambleton (1993).

1. A person’s estimated performance is independent of the particular sample of test items chosen from the calibrated pool of items.
2. The statistical descriptors of a test and of test items are not dependent on the particular sample of persons drawn from the population for whom the test is intended.
3. A statistic indicating the precision of the estimated performance is provided for each person. The statistic is dependent on the person’s performance and the number and the statistical properties of the items in the test to which that person responded.

Since, the Rasch model involves a fewer number of item parameters, it is easier to work with when compared to the other item response theory (Hambleton and Swaminathan, 1985; Hambleton, 1989). However, Lord (1974) argues that if reliable, sample-free scaling is to be developed huge samples have to be tested and this is certainly so for the two and three parameter models, but not for the one parameter logistic model.

A treatment of the Rasch model in a detailed way is beyond the scope of this thesis. However, Rasch (1960, 1961, 1966, 1980), Wright (1968,1977, 1988), Hambleton and Swaminathan (1985), Andrich and Masters (1988), Hambleton (1993) Molenaar (1995) provide a comprehensive account of the model.

Method employed in this study

From the literature reviewed in this section a decision was made to employ an item response theory in this study rather than to use classical test theory. Among the item response theory the one-parameter logistic model or the Rasch model was selected for this study for the following reasons.

1. Unlike the two-and three-parameter item response theory, the Rasch model allows the item parameters to be estimated independently of the students sampled, and the student parameters to be estimated independently of the sample of items used.
2. The Rasch model can be employed without violating the model’s requirement for minimal levels of guessing in responding to a test. Furthermore, it is argued that guessing should be seen as a characteristic of a relatively few students rather than a characteristic of the items.
3. The Rasch model has fewer item parameters and compared to the other two models, it would be easier to work with.
4. The number of samples employed in the present study (EMS, FIMS, SIMS and TIMS) comprised large data sets and problems of precise estimation are unlikely to occur. Lord (1974), and Kline (1993) have argued that with the Rasch model large samples should be used if reliable, population-free scaling is to be established.

5. There is the absence of strong research evidence on the use of item response theory that identifies the model that would be best employed in this type of study, but Sontag (1984) has shown the strength of the use of the Rasch model in the scaling of tests employed in IEA studies.

Before the Rasch model could be used to analyse the data, however, it was necessary to establish the dimensionality of the mathematics tests. Thus, it was decided to test for unidimensionality of the test items using confirmatory factor analysis procedures.

Test Equating Procedures

The last two sections indicated that the mathematics test data would be analysed using the Rasch model and a method was identified to test the unidimensionality of the items which is one of the requirements of the model. The next step is to identify the appropriate method for the equating of the mathematics tests, and view and attitude scales. Hence in this section different kinds of test equating procedures are reviewed and the appropriate technique is proposed.

Test Equating Issues

Before the equating of tests can be undertaken several issues which arise in the equating of student performance on different tests must be considered.

Meaning of test equating

Test equating has been defined by many researchers in the field. Skaggs and Lissitz (1986) defined test equating as the process of determining the relationship between raw and scaled scores on two or more tests. These tests are linked through common or anchor items (Skaggs and Lissitz, 1986; Morrison and Fitzpatrick, 1992). The raw or scaled scores on a new set of test items are equated to scores on a previous set of test items. Within this definition, the same or different groups of students could take the two tests. The important point here is that the two tests should have a limited number of common or anchor items or common persons who took the two tests. These common or anchor items or common persons are necessary to equate the new test scores with the previous test scores.

Another researcher who defined test equating was MacCann (1989). Within this definition, when group ‘a’ takes test ‘a’ and group ‘b’ takes test ‘b’, the scores on ‘a’ have to be compared to the scores on ‘b’. In order to compare both test scores, one group of scores, say ‘a’, is usually transformed statistically to place it on the same scale as the group ‘b’ scores. The procedure commonly demands the use of an anchor test, ‘c’, taken by both groups. The relative performance of the two groups on ‘c’ is used as the basis for adjusting the group scores on test ‘a’ to be equivalent to group ‘b’ scores on test ‘b’. In MacCann’s definition the two groups should take two types of tests, the first type of test should be different for each group and the second type of test should be the same for both groups. The achievement level of the students on the common test is used to adjust the performance level of students on the different tests. This procedure is conceptually equivalent to common item equating.

Dorans (1990) also defined equating test scores as a process of computing a statistical adjustment to the scores on one form of a test that will make them equivalent, in a particular way, to the scores on another form of the test. Dorans’ definition summarises the definitions given by Skaggs and Lissitz (1986) and MacCann (1989). In accordance with Skaggs and Lissitz’s definition the students could take two different tests, but the two tests should have some items in common, while in

MacCann's definition students should take two different tests with a further test in common. The test which is common for both groups is important for statistical adjustment of scores of students on the two different tests. In both cases the main purpose of the common test or common items is for statistical adjustment of scores of students on the different tests. Hence equating is employed to compare the scores of examinees taking different test forms. Equating provides adjustments for difficulty differences, thus allowing different tests to be used interchangeably (Kolen, 1994). After the equating process students are expected to get the same score regardless of the test form administered (Kolen, 1997).

Requirements for equating

The purpose of equating is to develop as much as possible an effective equivalence between scores on two test forms. To do this equating requires a rule for converting scores on one test form to scores on another. Lord (1980) proposed four requirements for successful test equating. He has argued that scores on test X and test Y are equated if the requirements are met. Petersen et al. (1989, p. 242) summarised the four test equating requirements recommended by Lord.

1. *Same ability:* Both test forms must be measures of the same characteristic (latent trait, ability, or skill).
2. *Equity:* For every group of examinees of identical ability, the conditional frequency distribution of scores on test Y, after transformation, is the same as the conditional frequency distribution of scores on test X.
3. *Population invariance*: The transformation is the same regardless of the group from which it is derived.
4. *Symmetry*: The transformation is invertible, that is, the mapping of scores from test X to test Y is the same as the mapping of scores from test Y to test X.

These researchers argued that in practice it is unlikely that these four requirements would be met. Petersen *et al* (1989, p. 243) indicated that:

> there seems to be a general agreement among practitioners that equated scores should satisfy the population-invariance and symmetry conditions. The disagreements revolve primarily around the equity condition and, to a lesser extent, around the same-ability condition.

Because of the difficulty of meeting test equating requirements Petersen et al. (1989, p. 243) concluded that "there is probably no equating method that will produce truly equivalent scores on two forms of the same test". However, these researchers recommended that because test scores can have important consequences for the examinee "an approximate equating of scores on two forms of a test will generally be more equitable to the examinee than no equating at all".

Types of test equating

Researchers have categorised test equating procedures into two types namely horizontal and vertical equating (Hambleton and Swaminathan, 1985; Skaggs and Lissitz, 1986; Baker and Al-Karni, 1991; Weiss and Yoes, 1991).

Vertical test equating procedure

This type of test equating procedure is employed when the tests to be equated have different difficulty levels and the distributions of the ability levels of the examinees are different. This would apply to tests across different year levels. This means that in the vertical equating procedure the tests to be equated would be administered to two or more groups of students at different year levels. The level of difficulty of these tests may not be the same, but there would need to be some items which were common to both tests. These common items would constitute the anchor items which served to equate the tests. For example, in the 1964 First International Mathematics Study which was administered to Population 1, and Population 3 students, there were common items with an intermediate level test. These common items were necessary for equating the achievement scales of Population 1 and Population 3 students on the Mathematics tests. Keeves and Kotte (1996) employed these common items in the mathematics test for the vertical equating of the mathematics achievement scales of Population 1 and Population 3.

Horizontal test equating procedure

The purpose of the horizontal test equating procedure is to make the tests true measures of the same characteristic with the same degree of accuracy and precision regardless of the group of examinees under consideration. This type of test equating procedure is employed when the tests to be equated have approximately the same difficulty level and the distributions of the scores of the examinees taking the test are similar. This would apply to tests within a single year level (Weiss and Yoes, 1991). This means that the same test would be administered to the same year level students at different times and the two tests would be equated. For example, in the 1964 FIMS a test of 70 items was administered to 13-year-old students and in 1978, SIMS was also conducted by administering a mathematics test to the same age group of students. The two tests had 70 and 72 items respectively. Among these items 65 were common to both the 1964 and 1978 tests, and equating was necessary to compare the achievement levels of the students over time.

The major difference between the two types of test equating procedures relates to the samples of students who take the tests. In vertical equating the samples are from different year and performance levels, while in horizontal equating they are from the same year level or performance level. Thus, this study employed the horizontal test equating procedure. This is because the tests being equated were administered to similar groups of students.

Techniques of test equating

There are several different techniques for test equating. These different techniques may be categorised either under classical test theory (CTT) or to item response theory (IRT). Further information on test equating is provided by Holland and Rubin (1982), Thorndike (1982) Skaggs and Lissitz (1986), Petersen et al. (1989), Kolen (1988, 1994, 1997).

IRT test equating techniques

The techniques of test equating which are employed in IRT correspond to the Rasch model (one parameter logistic model), two-parameter and three-parameter logistic models (Skaggs and Lissitz, 1986; Hambleton et al., 1991, Kolen, 1994). Research findings (Petersen et al., 1983; Cook and Eignor, 1991) have indicated that the IRT (one parameter or Rasch model, two-and three-parameters) test equating techniques are the best techniques when compared with other procedures.

Advantages of IRT Techniques. Cook and Eignor (1991) argued that IRT test equating procedures give appropriate solutions to equating problems because of their ability to place several tests and groups of candidates on a common measurement scale. They also argued that in IRT equating a particular candidate would obtain the same ability estimates regardless of the particular test that he or she had taken. In addition, IRT equating techniques can be used to accomplish both vertical and horizontal equating of tests (Baker and Al-Karni, 1991).

Cook and Eignor believe that IRT equating techniques have the following advantages.

1. IRT provides the best procedures to use when different difficulty level of tests are administered to nonrandom samples of candidates who are different in ability.
2. IRT provides conversions that are independent of the group or groups used to obtain them.
3. IRT gives more accurate equating than do classical methods at the upper ends of the score scales where important decisions are often made.
4. IRT provides greater flexibility in choosing the past test forms for equating purposes.

Skaggs and Lissitz (1986) reported that for horizontal equating, IRT equating techniques seems to produce better results than do classical test equating techniques.

Comparing One, Two and Three parameter IRT Equating Techniques. Research findings (Kolen and Whitney, 1981; Sontag, 1984; Doron, 1986) have indicated that among the three IRT test equating techniques, the Rasch model is the most appropriate test equating procedure. Doron (1986) confirmed that the best known method for test equating is the Rasch model which is a sample free calibration technique. She emphasised that the model can be used whenever a small set of common items appears in two or more tests, the scores of which have to be compared.

Mislevy (1987) reported that it is only the Rasch model that assures that the underlying metric is an interval scale when the data fit the model. Thus, only the Rasch model has strong measurement properties and logits are the units of the Rasch scale which is an interval scale. There is, however, the requirement that the items are associated with a unidimensional scale and that the items fit the model. Beard and Pettie (1979) argued from their study comparing linear and Rasch equating results from the Florida Educational Assessment of third grade 1976 and 1977 communications and mathematics test results and concluded that Rasch equating procedures were the more appropriate for the third grade tests.

A study undertaken by Sontag (1984) using American data from the First IEA Science Study in 1970/71 to determine the appropriate IRT model for resolving vertical equating problems, found that the Rasch model was the most stable procedure followed by the two and three parameter models. Ten years later Wright (1995) using The 1992 National Adult Literacy Survey, NALS data compared the three parameter and the Rasch model test equating methods. Wright reported that the three parameter model had no benefits over the Rasch model.

Rentz and Bashaw (1975, pp. 11-12) argued that in the Rasch model, item calibrations were independent of the calibrating sample and person measurements were independent of the specific set of items. This meant that any :

(a) appropriate collection of persons could be used in the calibration process as opposed to resorting to some elaborate sampling plan; and
(b) subset of a pool of items could be used to measure the ability of an individual, and all such subsets would estimate ability on a common scale.

Slinde and Linn (1978) also found that the Rasch model provided person-free item calibration and item-free person measurement. In person-free calibration the item parameters estimated would be equal for all groups of persons. In item-free person measurement, once items were calibrated, the same measure would be applied for a person regardless of which subset of items were used for the measurement. Slinde and Linn indicated three important benefits of the Rasch model.

1. It is relative simple, since items are characterised by a single parameter.
2. It is unique for dichotomous items.
3. It allows for consistent and sufficient estimation of person and item parameters.

Techniques of test equating employing the Rasch model. There are three different techniques of test equating employed with the Rasch model, namely anchor item equating, common item difference equating and concurrent equating. In all the three techniques, the tests to be equated are calibrated first and then equating is considered. The calibration includes the deletion of misfitting items and misfitting examinees.

1. *Anchor item equating*: This technique of test equating involves the anchoring of the threshold values of the common items obtained from one of the tests in the calibration of the other test. This means that the threshold values of the common items in test X would be anchored in test Y. Then the anchored items would be employed to estimate the threshold values of the remaining items in that test.

   Experience indicates that in horizontal equating some anchored items provide infit mean square values outside the expected range especially when the anchored items are more difficult items. The problem here is what to do with these item. When anchored items are few in number the deletion of the items from the analyses would reduce the number of anchor items and affect results. However, leaving the misfitting items in the analysis might attract criticism, because the misfitting items might not measure the same latent trait as the other items.

2. *Common item difference equating*: In this method the threshold values of the second test are first subtracted from threshold values of the first or base test. Then the differences are summed and divided by the number of anchor test items to obtain a mean difference. Subsequently the mean difference is added to each item threshold value on the second test to obtain the adjusted threshold value (Wright and Stone, 1979; Morrison and Fitzpatrick, 1992). Because there is no direct interaction in the data of the tests to be equated, this technique has no means of testing whether or not the items in the tests measure the same latent trait. However, this procedure does permit an estimate of magnitude of the equating error to be calculated as the error of the mean difference.

3. *Concurrent equating*: When two or more test forms share the same anchor items, it is possible to combine them into a single data set and to calibrate them together. This type of equating is referred to as concurrent equating (Wright and Stone, 1979; Petersen et al., 1989; Morrison and Fitzpatrick, 1992;). The calibration of the tests is then done simultaneously.

In concurrent equating items that do not fit the Rasch model are deleted from the analysis. Hence, this technique seems very useful and more accurate in horizontal equating. Recently Mohandas (1996) employed the Rasch model in horizontal equating to equate five English test forms using concurrent and anchor item equating and found that concurrent equating provided more consistent results.

These IRT test equating procedures are concerned mainly with how the anchoring procedure is employed in equating. They could be applied either in vertical equating or in horizontal equating.

Number of anchor items: There is no general agreement about the number of anchor items necessary to position two different tests on a single scale. Wright and Stone (1979) have recommended that 10 to 20 (17 to 34 per cent of the items in each test) items should be employed for equating two different test forms consisting of 60 items each. Meanwhile, Hambleton et al. (1991) suggested approximately between 20 and 25 per cent of the number of the items in the tests should be common. While, Smith and Kramer (1992) have argued that as few as a single item is required.

Conclusion

Research into the methods of test equating has been an ongoing process since the beginning of this century in order to examine change in the levels of student achievement over time. However, research has been intensified since the 1960s due to the development of IRT and the availability of appropriate computer programs. Among the many test equating procedures, the IRT techniques are generally considered the best. However, only the one parameter model or Rasch model has strong measurement properties. Therefore, in order to examine the achievement level of students over time, it is desirable to apply Rasch model test equating procedures. Hence, in this study of the mathematics achievement and attitudes and views of 13-year-old students towards school and mathematics over time, the horizontal test equating strategy with the concurrent, anchor item equating and common item equating techniques using the Rasch model are best applied.

Item Bias

In the examination of gender differences in the achievement tests and attitudes and descriptive or view scales it is first necessary to consider whether items and statements associated with the tests and scales exhibit item bias.

The issue of bias in a test has become an important research topic in educational measurement since the 1960s. The civil rights movement and the women's rights movements of the 1960s led to the investigation, by those who were interested in fairness and equal rights, of the outcomes of education for certain groups involved in testing (Cole and Moss, 1989). As a result, the possibility of bias in test use has received wide attention from the public and from educational measurement specialists. Therefore, most of the research on bias has been concerned with the fair use of testing instruments in decision-making situations such as employment and admission for tertiary education (Scheuneman, 1979). Moreover, the study of bias has focused not only on gender and social differences, but also the difference in performance of minority groups, however, they are formed and whenever they occur.

Educational measurement researchers and psychologists who are involved in the construction of test instruments have thus become interested in the specific problem of identifying items that seem to be biased prior to the development of the final forms of a test. Because the identified items, can be modified or deleted, such steps, it is argued, would increase the fairness of the final draft of the test. Therefore, researchers have investigated different procedures for assessing bias in test items. In this section the definition of bias and procedures for detecting item bias are discussed.

Definition of Item Bias

'Bias' was defined by Cole and Moss (1989, p. 205) as follows.

> Bias is differential validity of a given interpretation of a test score for any definable, relevant subgroup of test takers.

These researchers and Smith (1994) further, suggest that the procedures can be extended to groups of concern in order to investigate possible differences in validity for these groups. In the 1970s and 1980s, Green (1975), and Benson (1987) also recognised that the problem of bias was basically a question of differential validity between two or more groups.

In the recent literature the terms DIF (differential item function) or DIP (differential item performance) are used instead of item bias. In recent research reports, the term DIF seems to be a more appropriate term to use as a substitute for item bias (Klieme and Stumpf, 1991; Diamond, 1992; Meredith and Millsap, 1992; Raju et al., 1993; Lautenschlager et al., 1994; Mellenbergh, 1994; Nandakumar, 1995; Wainer, 1995). Holland and Thayer (1988), Thissen et al., (1988) and Lautenschlager et al. (1994) have argued that DIF has gradually replaced item bias because DIF is a more accurate and less evaluative term than item bias. Hoijtink and Molenaar (1992) indicated that DIF refers to items that behave differently for different groups of students. Furthermore these authors have argued that "the difference may occur in the values of the parameters characterising the item in the parametric family appropriate for modelling the item or a combination of both" (Hoijtink and Molenaar, 1992, p. 383). DIF allows the identification of test items that do not function similarly for different groups of students. The term *bias* used in this study relates to this meaning of DIF.

Methods of DIF Detection

In the 1980s the DIF detection procedures were improved to provide a better basis for matching on ability (Westers and Kelderman, 1991). DIF detection procedures have been widely accepted that are based on IRT (Wright, Mead and Draba, 1976; Baker, 1977; Lord, 1980; Westers and Kelderman, 1991). For the detection of DIF, IRT has been argued to be the preferred method over classical test theory (Baker, 1977; Lautenschlager et al., 1994). The different procedures which are employed in DIF detection are consequently categorised under the heading of classical test theory or item response theory.

IRT Procedures

Researchers such as Lord (1980), Lautenschlager et al. (1994), Potenza and Dorans (1995), Scheuneman and Bleistein (1997) have argued for the use of IRT techniques for the detection of DIF. Osterlind (1983, p. 54) argued that the significance of the IRT DIF technique over classical test theory procedures arose from it being "the most elegant of all the models discussed to tease out test item bias." Osterlind also claimed that the assumptions of the IRT makes it important for detecting DIF. Furthermore, Scheuneman and Bleistein (1997, p. 746) have pointed out that the characteristics of the IRT which makes it important for examining DIF is that "the estimated parameters of the item response function (IRF) are invariant for different samples drawn from the same population." Osterlind also noted that one of the requirements of the item response theory is the dimensionality of the latent trait. The IRT requires that the items in a test should measure one latent trait or they should be unidimensional items. These unidimensional items should measure the same latent trait in all sub-groups of the population. Lord and Stocking also share similar views with Osterlind (1983). They believe that:

> items in a test that measures a single trait must measure the same trait in all subgroups of the population to which the test is administered. Items that fail to do so are biased for or against a particular sub-group. (Lord and Stocking, 1988, p. 681)

Furthermore these authors argued that theoretically, the item response functions (IRFs) are independent of the group used for calibration, therefore, IRT provided the best method for detecting DIF.

Even if there would seem to be a common understanding among measurement and psychology researchers that item response techniques have an advantage over the classical test theory procedures in detecting DIF, some researchers employ both methods in the same study for different purposes, in order to make comparisons and enhance the precision of detecting DIF (Cohen and Kim, 1993; Rogers and Swaminathan, 1993; Zwick, *et al*., 1994).

IRT techniques for DIF detection are based on comparisons of item-characteristic curves (ICCs), which are estimated separately for each group (Lautenschlager et al., 1994). If an item is unbiased, then the item-characteristic curves for the item as estimated for the different groups of students should be the same. Furthermore, Ellis et al. (1993), and Lautenschlager et al. (1994) have argued that an item is unbiased when the item-characteristic curves estimated separately for the same item for two different groups are the same curves. This means that the item functions equivalently for both groups. When the estimated ICCs of the same item for the two groups differ by more than sampling error, then the item is said to show DIF (Lord, 1980). Since the values of the ICCs represent the probability of a correct response at each level of ability, the ICCs must be the same if the item is measuring the same trait for both groups. If the ICCs are not the same, there is an indication that the item is not functioning in the same way for both groups. Osterlind (1983, p. 67) also shared the same view by arguing that the "bias is inferred any time equated ICCs are not identical". Thus bias is indicated by the size of the area between ICCs. This means that when the equated ICCs are compared for both groups, if a particular item is biased, the two curves would be significantly different. If the item is unbiased, when the ICCs are compared for both groups, the curves would be identical for both groups. Thus the measure of DIF using an IRT is the difference in area between the ICCs. Moreover, the criteria employed for examining whether or not DIF is detected are: item discrimination and difficulty, and the guessing parameter of the item (Osterlind, 1983). If the three parameter model is used, then the three indicators are read from the ICCs of an item which is considered in the comparison of two groups. If any one of the indicators differs considerably between the two groups, then it is possible to conclude that DIF is detected.

Indices of DIF in the Rasch Model

It has been noted that the Rasch model is a one parameter IRT model. Consequently, indicators for the detection of DIF in IRT would also be indicators in the Rasch model. However, the difficulty level of an item is the only criterion available for measuring DIF which is directly obtained from the Rasch model. Consequently, Scheuneman and Bleisten (1994), have argued that it is also necessary to examine the fit of each item for each group to the model (discrimination or infit mean square). The authors did not consider guessing as a criterion, because the advocates of the Rasch model contend that guessing is a characteristic of the examinee, and not of the item. Thus, in the current study guessing is not considered as a criterion for detecting DIF. However, item difficulty and discrimination are considered for detecting DIF, because employing more than one approach enhances the accuracy of detecting DIF.

Item Difficulty. One of the indices used in the detection of DIF between groups is the difficulty level of the item. If the difficulty level of the item is different for different groups and if that difference is substantial, this indicates that the item was more difficult for one group when compared with the other and was not due to the differences in level of achievement between the groups. The difference might arise from differences in racial and ethnic background, sex, age or condition of handicap, which are related to their being members of the two groups.

Therefore, the major indicator for the detection of DIF is the difference in the difficulty level of items between sub-groups of a population. This indicator, as argued by Scheuneman (1979) and Thorndike (1982), identifies items which are unexpectedly easy as well as those unexpectedly hard for a particular sub-group.

Item Discrimination. All the items in the Rasch model are required to be unidimensional. This involves equal discriminating power as shown by the item characteristic curve. Therefore, those items that fit the Rasch model should fall within the predetermined infit mean square range regardless of the groups employed. If the infit mean square value of an item is outside the predetermined acceptable range, such items are showing DIF because the items do not have the same discriminating power for the different sub-groups of the population under investigation.

Computerised DIF Detection

Several computer programs have been developed for handling the detection of DIF. The LOGIST computer program (Wingersky, *et al.*, 1982), FORTRAN 77 computer program (Klieme and Stumpf, 1991), the QUEST program (Adams and Khoo, 1993) and the SIBTEST (Shealy and Stout, 1993) are among the programs currently available. The QUEST computer program (Adams and Khoo, 1993) was selected for use in the present study because of its availability.

The Rasch model DIF procedures available in the QUEST computer program (Adams and Khoo, 1993) include the comparison of item difficulty levels or thresholds between two groups being compared and the consideration of the item fit to the Rasch model for the two groups being compared.

Comparison of Item Difficulty. By employing the QUEST computer program (Adams and Khoo, 1993), items which satisfy the following requirements are identified as DIF.

3. When the difference in level of difficulty of the item between any two groups being compared is outside of a predetermined range. Experience in IEA studies recommends the predetermined range to be between $\pm$0.50. That is:

   $|ad_1 - ad_2| > 0.50$

   where ad_1 = the adjusted item's difficulty level in group1, and

   ad_2 = the adjusted item's difficulty level in group2.

2. When the differences in standardised level of difficulty of the item between any two groups being compared is outside a predetermined range. The range recommended by Adams and Khoo (1993) is $\pm$2.00. Therefore;

   $|st\ (ad_1 - ad_2)| > 2.00$

   where st indicates a standardised difference.

Adams and Khoo (1993) employ this procedure commonly for samples of approximately 400 cases. When the sample size increases the range should also increase. In the current study the sample sizes were over 10,000 students, hence it was necessary to adjust the standardised item difficulty difference. The adjusted standardised item difficulty difference can be calculated by employing the following formula:

$$astd\ st(ad_1 - ad_2)\left(\sqrt{\frac{N}{400}}\right)$$

Where

astd = adjusted standardised difference
N = sample size in the two groups.

The standardised item difficulty difference for the new sample size is adjusted by dividing the parameter by $\sqrt{\frac{N}{400}}$ using the commonly employed sample size 400 as a reference group size.

Vijver and Poortinga (1991) and Hungi (1997) emphasised that items that do not fit the Rasch model should be deleted before identifying DIF. Subsequently, items outside of a pre-defined difficulty level difference should also be discarded. However, items should not be removed from analysis without careful consideration as to why misfit might have occurred.

Comparison of Infit mean square. Adams and Khoo (1993) recommended the infit mean square range to be between 0.77 and 1.30. Items outside of this acceptable range are identified as misfitting items and should be deleted, after consideration of why differential item function might have occurred.

By employing the above DIF detection procedures, the mathematics achievement data and the view and attitude data are examined in Chapter 10 for items bias. With respect to gender in order to examine whether any gender differences might be observed arise from the existence of bias in the test items or in the attitudinal and view statements.

Analysing Data with Complex Samples

Sampling Design

Researchers who are involved in studies that requires the collection of data, must design their samples before undertaking data collection. Three important ideas are associated with sampling design that are employed to facilitate the collection of data and to reduce the size of the errors involved.

1. *Stratification* - the total sample is divided into sub-samples called strata and the random sampling principle is applied independently within each stratum in order to reduce sampling error. The sample estimates are then obtained by combining the information from each stratum. Rosier and Ross (1992) argue that stratification increases accuracy in sample estimates without increasing cost.

2. *Cluster Size* - refers to the number of students within a school or class group who complete a test or questionnaire. Clustering is effective when there is heterogeneity within each cluster, but has substantial effects on sampling errors when the clusters are homogeneous and different from each other.

3. *Three Stage Sampling Designs with different sampling fractions* - can involve areas or regions as the first or primary stage of sampling (PSU), schools as the second stage of sampling (SPU) and students as the third stage of sampling (TSU), with different sampling fractions across the different stages of sampling.

By using these ideas there are different kinds of sampling design available. Two common sampling designs are employed for the two-stage sampling of schools and students.

1. Simple random sample (SRS) selection of schools with a fixed fraction of students sampled within schools.
2. Probability proportional to size (PPS) selection of schools with an equal-size cluster of students sampled within schools.

Sample Size

After deciding which sampling design to use, the next step would be to decide on the sample size, that is, to decide on the number of schools and students that would participate in the study. Rosier and Ross (1992) have suggested that the computation of the size of the sample to be employed should be based on the magnitude of the standard error of the key criterion measures specified for the study. When the purpose of the study is to estimate the population mean for a variable (eg. a test score), the calculation of the size of the sample is based on the magnitude of the standard error of the mean, expressed as a proportion (or percentage) of the standard deviation. When the purpose of the study is to estimate the proportion (or percentage) of a variable in the population (such as item facility), the size of the sample is based on the magnitude of the standard error of the proportion. However, research workers are also interested in the estimation of correlation and regression coefficients, and these require a different approach to determining sample size.

Sample size for the estimation of population mean

Consider the estimation of a population mean, such as student mean scores on tests and subtests. Assume that the study specifies a precision of $\pm$ 10 per cent of a student standard deviation at the 95 per cent confidence level. This level of precision corresponds to 1.96 standard errors; that is, approximately 2 standard errors. Hence, the sample size should be chosen to give a standard error of 5 per cent of a student standard deviation.

The formula to obtain the estimated standard error of the mean for a complex (two-stage) sample is:

$$se\left[\bar{x}(c)\right] = \frac{s}{\sqrt{n^*}}$$ (Rosier and Ross, 1992, p.59)

Where

s = standard deviation for a simple random sample

n^* = the effective sample size

$se\left[\bar{x}(c)\right]$ = standard error of the mean of the complex sample.

Substituting the required value of the standard error of the mean gives:

$$0.05s \leq \frac{s}{\sqrt{n^*}}$$ (Rosier and Ross, 1992, p.73)

Hence

$n^*\surd 400$

In other words, the sample should be designed to give a minimum effective sample size of 400 students. However, in practice researchers have not always applied these procedures when they have determine the sample size for their studies, and have not achieved sample sizes that would give these levels of precision. Moreover, they also have not always taken into consideration the design effect, when they test for significance. Consequently, they might have advanced a wrong conclusion. Young and Fraser (1994, p.863-4) in their study of gender differences in science achievement, reported that

> .. the use of traditional simple random sampling standard errors for the same analyses indicated statistically significant sex differences in biology achievement among 10-year-old students, with boys outperformed by girls, but differences were not found to be significant when sample design effect was incorporated.

Furthermore, these researchers indicated that their findings and those of others have highlighted the importance of taking into consideration the complex nature of sampling designs when calculating statistical significance. Therefore, it is necessary to consider design effects when thinking about significance levels, the size of sampling errors and effective sample sizes.

Design Effects

Design effect is related to the design of the sample included in a study. Peaker (1975) argued that the design effect arises from the sampling design when the sampling is done in two or more stages. Brick et al. (1996, p. 143) defined design effect "as the ratio of the variance under the actual survey design to the variance under simple random sampling given the same sample size for the domain of interest." These researchers have developed a computer program called WesVarPC which is able to calculate the design effect and the standard error of sampling. In the use of the WesVarPC computer program, the numerator of the design effect is the WesVarPC estimate of variance. For a proportion, the variance under simple random sampling is given by the following formula (Brick et al., 1996, p.143).

$$Var(p) = \frac{p(1-p)}{n}$$

where

p is the sample estimate of the population proportion,
n is the sample size on which the estimate of p is based, and
$\mathrm{var}(p)$ is the variance of the estimated proportion.

For the mean of a continuous variable, the variance under simple random sampling is given by the following formula.

$$\mathrm{var}(\bar{y}) = \frac{\sum_{i=1}^{n} w_i (y_1 - \bar{y})^2}{n \sum_{i=1}^{n} w_i} \qquad \text{(Brick, et al., 1996, p. 143)}$$

where

n is the number of observations on which the estimated mean is based,
w_i is the full-sample weight of the *i*-th observation, and
$\bar{y}$ is the estimated mean.

The design effect (DEFF) of a complex sample is defined as

$$\text{DEFF} = \frac{\text{variance of the complex sample (obtained with WesVarPC)}}{\text{variance of the corresponding simple random sample}}$$

Where the simple random sample and the complex sample are considered to be corresponding if they have the same total number of cases as at the second stage of sampling in the complex sample.

The equivalent or effective simple random sample size is given by

$$n^* = \frac{n}{DEFF}$$

Where

n = total number of cases at the second stage of sampling in the complex sample and

n^* = the number of cases in the simple random sample that would be equivalent to or correspond to the complex sample.

Significance Testing with Complex Samples

In any study the question of significance of the findings must be addressed. In an investigation such as this where large samples are employed it becomes relatively easy to obtain statistical significance when large samples have been used. Under these circumstances it is necessary to not only consider issues of practical significance but also to consider the effects of the use of a complex design.

The *t*-test assesses the statistical significance of the difference between two independent sample means for a single dependent variable (Hair et al., 1995, p. 261). The *t* statistic is the ratio of the difference between the sample means to the standard error of the difference. The standard error of the difference is an estimate of the difference between means to be expected because of sampling error. Hair et al. (1995), claim that by forming the ratio of the actual difference between the means to the difference expected due to sampling error, the amount of the actual impact of the treatment that is due to random sampling error is quantified. The following formula can be employed in calculating the *t* statistic between two independent sample means.

$$t\text{statistics} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{se_1^{\,2} + se_2^{\,2}}}$$

where

$\overline{X}_1$ is the mean of group one,

$\overline{X}_2$ is the mean of group two,

se_1 is the standard error of group one, and

se_2 is the standard error of group two.

The value of the standard error of the difference $\left(\sqrt{se_1^2 + se_2^2}\right)$ is calculated by the WesVarPC program to allow for the complex design of the sample, and as a consequence the *t*-statistic can be readily calculated.

The absolute values of the *t* statistic that exceed the critical value of the *t* statistic lead to the rejection of the null hypothesis of no difference between the two groups that are being compared. Commonly a five per cent probability is chosen for significance when the critical value of the *t* statistic is approximately 2.0.

In the present study the *t*-statistic were calculated only for mathematics achievement, and view and attitude scales mean score differences between groups. However, in the PLSPATH analyses no significance testing was undertaken, because the samples employed were not a simple random samples, but a cluster and stratified random sample design, and simple random sample tests of significance would be inappropriate. The WesVarPC program employs a jackknife procedure for this estimation of sampling error that makes allowance for the effects of stratification, cluster design and differential weighting which results for the use of different sampling fractions and response rates across the primary and secondary stages of sampling.

Use of Weighted Scores to allow for Complex Sample Design

All student scores employed in the presented study are weighted scores. Weighted scores were employed in accordance with the differences that arise in a country between the target and the achieved samples as well as to compensate for the differential sampling procedures when combining data across strata used in the sample design of particular country (Postlethwaite, 1992; Rosier and Ross, 1992; 1997). The formula employed to calculate the scores was

$$\text{weighted score} = \frac{N_h}{n_h} x \frac{n}{N} \qquad \text{(Rosier and Ross, 1997, p. 432)}$$

where

N = target population for X country overall

n = achieved sample for X country overall

N_h = target population for stratum h

n_h = achieved sample for stratum h

Effect Size

In this study both the standardised effect size and the magnitude of effect on the calibrated scales are used to examine the level of practical significance of the differences between EMS, FIMS, SIMS and TIMS in mathematics achievements over time and in views and attitudes towards mathematics and schooling. The following formula was employed to calculate an effect size value.

$$\text{Effect size} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{sd_1^2 + sd_2^2}{2}}}$$

where

$\bar{X}_1$ is the mean of group one,

$\bar{X}_2$ is the mean of group two,

sd_1 is the standard deviation of group one, and

sd_2 is the standard deviation of group two.

In this study effect size values less than 0.20 are considered as trivial, while values between 0.20 and 0.50 are considered as small. Furthermore, effect size values between 0.50 and 0.80 are taken as moderate and values above 0.80 are reported as large (Cohen, 1992; Keeves, 1992a).

Conclusion

In this chapter the appropriate methods to be employed in testing the hypothesised models and analyses of data are identified. The confirmatory factor analysis procedure was chosen as an appropriate technique for testing the underlying dimensions of the mathematics test. For the analysis of the students' mathematics achievement test scores and the changes in achievement over time as well as the views and attitudes of students towards mathematics and schooling, the Rasch model was chosen. From the different item response theory the one parameter or the Rasch model was considered to be the most appropriate and the most robust as well as being the easiest to employ.

For the comparisons of achievement over time and across-nations horizontal test equating was used, because for the over time comparison the students were all 13-year-old students, or Year 8 students, while for the across-nations comparison all students were Year 8 students. The equating techniques to be employed were concurrent, anchor item equating and common item difference equating techniques. In addition, PLSPATH was selected as an appropriate procedure for the examination of the hypothesised models of student level factors influencing mathematics achievement.

In next four chapters namely, Chapter 8, 9, 10 and 11 both the mathematics achievement data, the views and attitudes towards mathematics and schooling data, and student background information data are analysed using the appropriate methods identified in this chapter.

8

Changes in Mathematics Achievements Over Time and Across Nations

Over the past three decades researchers have shown considerable interest in the study of student achievement in mathematics at all levels across educational systems. Many important conclusions can be drawn from various research studies about student achievement in mathematics across countries. International studies of mathematics achievement such as the First (Husén, 1967), Second (Robitaille and Travers, 1992) and Third International Mathematics Studies, (Beaton et al. 1996a), the Japan/Illinois Study (Harnisch et al. 1985), Development of Mathematical Thinking Study (Song and Ginsburg, 1987) and Mathematics Achievement in the Elementary Grades: The Michigan Studies (Stigler and Barances, 1988) have all shown that there are substantial differences in mathematics achievement between students across countries.

Stigler et al. (1982) have indicated that children from Japan and Taiwan consistently performed at a higher level than their American counterparts. In 1986 Stevenson et al. (1986) in their study *Mathematics Achievement of Chinese, Japanese and American Children,* reported that American kindergarten children lagged behind the Japanese children in their understanding of mathematics. At the Year 5 level the American children were surpassed by both Chinese and Japanese children. Miura and Okamoto (1989) have suggested that already by Year 1 students in United States and Japan differed in their cognitive representation of number. Furthermore, this difference might have positively influenced the Japanese students' understanding of place value and their subsequent mathematics achievement. Nevertheless, Fuson et al. (1988) found uniformity in the grade placement of topics involving calculation by addition and subtraction in the curricula of China, Japan, the former Soviet Union and Taiwan. Furthermore, the authors reported substantial differences between the placement of topics in these countries and the placement in the United States. More recently in their TIMSS report, Beaton et al. (1996a) showed that, among the 45 participating countries, Singapore was the top-achieving country at both Years 7 and 8 levels, while Colombia, Kuwait and South Africa were found to be low-achieving countries. The issue to be addressed in this study is concerned with the magnitude of the differences between Australia and Ethiopia.

Educational researchers are not only concerned about differences in achievement between educational systems, they are also concerned about changes in achievement over time. Willett (1997, p. 327) argues that by measuring change over time, it is possible to map phenomena at the heart of the educational enterprise. He also argued that education is intended to enhance learning, and to develop changes in achievement, attitudes and values. It is Willett's belief that "only by measuring individual change is it possible to document each person's progress and, consequently, to evaluate the effectiveness of educational systems" (Willett, 1997, p. 327). Therefore, measuring changes in achievement over time is an important tool in finding ways and means to improve the educational system of a country.

Since Australia participated in the 1964, 1978 and 1994 International Mathematics Studies, it is possible to examine the mathematics achievement differences over time across the 30-year time period. The data collected in Ethiopia can also be used to compare the achievement level of the Ethiopian students with their Australian counterparts.

In this chapter the results of the Rasch analyses of the mathematics achievement of the 1964, 1978 and 1994 Australian students who participated in FIMS, SIMS and TIMS and the Ethiopian students who participated in EMS are presented and discussed. The chapter is divided into six sections. The methods employed in the study are presented in the first section of the chapter, while the second section examines the statistical procedures employed in the comparisons of mathematics achievement. The third section considers whether or not the mathematics items fit the Rasch model. The procedures employed for equating the mathematics achievement data between occasions are addressed in the fourth section. Meanwhile, the comparisons of the achievement of EMS, FIMS SIMS and TIMS students are presented in the next section. The last section presents the findings and conclusions drawn from the chapter.

Methods Employed in the Study

In this chapter the mathematics achievements of students over time and across-nations are measured using the Rasch model.

The purposes of these analyses are to:

(a) identify the differences in achievement in mathematics of Australian students between the 1964, 1978 and 1994;

(b) examine the differences in achievement in mathematics between Australian and Ethiopian students; and

(c) examine the differences in achievement in mathematics between government and nongovernment school students in EMS, SIMS and TIMS.

Use of Rasch Model

In Chapter 7 it was argued that the Rasch model should be used to equate student performance in mathematics on a common scale between the FIMS, SIMS, TIMS and EMS data sets.

Unidimensionality

Before the Rasch model could be used to analyse the mathematics test items in the present study, it was important to examine whether or not the items of each test were unidimensional, since the unidimensionality of the test items is one of the requirements for the use of the Rasch model (Hambleton and Cook, 1977; Anderson, 1994). Consequently, confirmatory factor analysis procedures (discussed in Chapter 7) were employed to test the unidimensionality of the mathematics items in FIMS, SIMS and EMS. The results of the confirmatory factor analyses, revealed that the nested model in which the mathematics items were assigned to three specific first-order factors (*Arithmetic, Algebra* and *Geometry*) as well as a general higher order factor, which was labelled as *Mathematics* provided the best fitting model.

In addition, the mathematics items analysed in this section were categorised into three types of cognitive processes, namely computation and verbal processes, lower and higher mental processes, and computation, knowledge, translation and analysis. These three major cognitive processes were also examined using confirmatory factor analysis procedures to investigate whether the items in the mathematics test measured one latent trait or not.

The results of these analyses showed that, the nested model in which the mathematics items were assigned to two (*Computational* and *Verbal Processes*, or *Lower* and *Higher Order Processes*) or four specific first-order factors (*Computation, Knowledge, Translation* and *Analysis*) as well as a general higher order factor, which was labelled as *Mathematics* provided the best fitting models. In addition, no evidence was found to reject the assumption of the existence of one general factor involved in the mathematics tests. In the nested model the *Mathematics* factor extracted more of the total variance than the specific first-order factors. As a result, the unidimensionality of the mathematics test items was confirmed.

It is important to note that researchers involved in analysing the IEA mathematics studies such as Husén (1967), Rosier (1980), Moss (1982), Hanna (1989), Robitaille (1990), Robitaille and Travers (1992), Beaton et al. (1996a) Lock et al. (1996) sometimes considered these mathematics items to comprise, three or four factors, such as arithmetic, algebra, geometry and new mathematics, two or more factors, such as verbal ability and comprehension, higher and lower order mental processes, comprehension, translation, analysis and application, although many, but not all of these authors, also employed a total score in their analyses.

Inability to obtain the item data for the TIMS study until six months before this thesis was to be completed prevented the examination of the TIMS item data using CFA. Moreover, the cluster-based procedure (Adams and Gonzalez, 1996) employed in the construction of the TIMS mathematics tests would seem to preclude the simple use of confirmatory factor analysis to test the unidimensionality of the TIMS data set.

The Statistical Procedures Employed in the Comparison of Mathematics Achievement

In this section the statistical procedures employed in this study in the comparison of achievement in mathematics across countries and over time are presented.

Effect Size

In this study both the standardised effect size and the magnitude of effect on the calibrated scales, referred to here a calibrated effect size, are used to examine the level of practical significance of the differences between EMS, FIMS, SIMS and TIMS in mathematics achievement over time (For further information see Chapter 7).

Growth between Grade Levels

It is possible since the TIMS project tested at two adjacent grades to estimate the gain between the lower grade and the upper grade for the Australian sample and thus to interpret the calibrated effect size in terms of years of mathematics learning at the lower secondary school level. The difference between the lower and the upper grade levels show the growth of achievement level per year in mathematics performance in Australian lower secondary schools. Thus to examine the growth between grade levels, the estimated mean scores difference between the lower and the upper level students was compared using WesVarPC 2.11 computer program (Brick et al. 1997). The results of the comparisons are presented in Table 8.1.

Table 8.1 Comparisons between the Lower and Upper group Levels in TIMS

Levels	Mean	SD	SE	SS	DEFF	ESS	ES	t-value	Sig-L
Lower	520	118.8	7.0	5599	19.44	288.03	-0.30	-3.96	0.01
Upper	557	127.7	6.2	7253	17.36	417.77			

SD = Standard deviation, SE = Standard error SS = Sample size
ESS = Effective sample size ES = Effect size Sig-L = Significant level

The estimated mean score for the lower level was 520 centilogits, while for the upper level students, it was 557 centilogits. The difference was 37 centilogits in favour of the upper level students. Hence, the growth in achievement per year in mathematics performance in Australian lower secondary schools was 37 centilogits. The effect size and t-values were -0.30 and -3.96 respectively. Thus, the mean difference between the lower and the upper grade levels, namely 37 centilogits was equivalent to an effect size of 0.30. Keeves (1992a) indicated that 38 centilogits was found to be equivalent to a year of science learning between the 10 and 14 year-old levels in the Second IEA Science Study in Australia. Therefore, this information allows the differences between the achievement level of the different groups to be interpreted in terms of practical significance as well as statistical significance, and the mean difference between the lower and the upper grade levels in Australian lower secondary schools in 1994 was practically and statistically (at the 0.01 level) significant.

The t -test

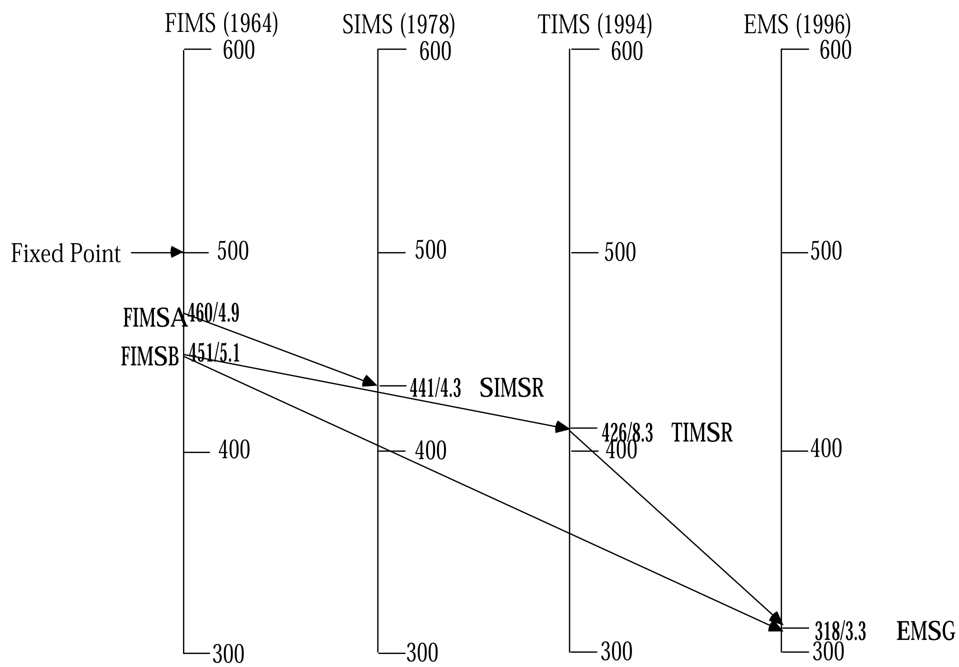
In order to determine the level of statistical significance between the estimated mean scores on the EMS, FIMS, SIMS and TIMS in mathematics achievements a t statistic was calculated (for further information see Chapter 7), which took into account errors from three sources: (a) sampling error, (b) error of calibration and (c) equating error for the comparisons between FIMSB and TIMSR, as well as between EMS and TIMSR. For the other comparisons only the sampling errors were considered, since anchor item and concurrent equating procedures were employed.

The Calculation of Perfect and Zero Scores

The QUEST computer program (Adams and Khoo, 1993) by default does not process cases with perfect and zero scores, because both groups do not provide information for the calibration of the scale. Cases with perfect scores are those cases who provide correct responses for all the items, while cases with zero scores are those cases who provide wrong responses for all items. Hence, in order to include those cases with perfect and zero scores in the calculation of the mean and standard deviation for each student sample, the values of perfect and zero scores were calculated by extrapolation from the logit table produced by the QUEST Computer program. Subsequently, the SPSS 6.1. (Statistical Package for Social Sciences, 1993) computer program was used to calculate the case estimate mean scores and standard deviations with appropriate weights and the WesVarPC 2.11 (Brick et al. 1997) computer program was employed for calculating the standard errors of the mean values, again with appropriate weighting of the data.

Developing a Common Mathematics Scale

The calibration of the mathematics data permitted a scale to be constructed that extended across the four groups, namely EMS, FIMS, SIMS and TIMS students, on the mathematics scale. The fixed point of the scale was set at 500 with one logit, the natural metric of the scale, being set at 100 units. The choice of the fixed point of the scale, namely 500, was an arbitrary value which was necessary to fix the scale, which has been used by several authors in comparative studies (Keeves and Schleicher, 1992; Elley, 1994; Keeves and Kotte, 1996; Lietz, 1996; Tilahun, 1996; Tilahun and Keeves, 1997b). The graphical representation of the mathematics scale constructed in this way for the different sample groups of students in EMS, FIMS, SIMS and TIMS is presented in Figure 8.1, with 100 scale units (centilogits) equivalent to one logit.



Fixed Point - 500, Metric - 100 = 1 logit, 1 unit = 1 centilogit
Values indicated for each occasion are Rasch estimated scores and standard errors of the mean respectively.

Figure 8.1 The Mathematics test scale of government school students in EMS, FIMSA, FIMSB and SIMSR

Calibration & Scoring Procedures employed in the Mathematics Test

The results of the Rasch analyses showed that ignoring the missing data or treating them as wrong had a marked change on the infit mean square values, where there were large numbers of misfitting items. From the analyses, for both calibration and scoring purposes it was decided to treat the missing data as wrong. While all the items in the mathematics test were considered for scoring purposes, for calibration purposes only those items that fitted the Rasch scale were employed. The main justification for the use of these procedures would seem to lie in the greater number of misfitting items when the procedure that involved the ignoring of the missing data was tested with the SIMS data, but not with the EMS data set. Consequently, the procedure that involved treating missing data as wrong was chosen for this study. However, the TIMS data set was not considered for making the decision, because the data could only be obtained for analysis until about six months prior the completion of this study.

Rasch Analysis

Four groups of students namely EMS (1200), FIMS (4320), SIMS (5120) and TIMS (7926) participated in the analyses. The necessary requirement to calibrate a Rasch scale is that the items must fit the unidimensional scale. Items that do not fit the scale must be deleted in calibration. In order to examine whether or not the items fitted the scale, it was also important to evaluate both the item fit statistics and the person fit statistics. The results of these analyses are presented below.

Item Fit Statistics

One of the key item fit statistics is the infit mean square (INFIT MNSQ). The infit mean square measures the consistency of fit of the students to the item characteristic curve for each item with weighted consideration given to those persons close to the 0.5 probability level. The acceptable range of the infit mean squares statistic for each item in this study was taken to be from 0.77 to 1.30 (Adams and Khoo, 1993). Values outside this acceptable range, that is above 1.30 indicate that these items do not discriminate well and below 0.77 the items provide redundant information. Hence, consideration must be given to excluding those items that are outside the range. In calibration, items that do not fit the Rasch model and which are outside of the acceptable range must be deleted from the analysis (Rentz and Bashaw, 1975; Wright and Stone, 1979; Kolen and Whitney, 1981; Smith and Kramer, 1992). The results of the Rasch calibration of the mathematics test across the four occasions are discussed below.

In the EMS data set 37 items had infit mean square values outside the acceptable range and were deleted from the calibration analysis, while in FIMS two (Items 13 and 29) and in SIMS two (Items 21 and 29) items were removed from the calibration analysis due to misfitting the Rasch model. Furthermore in TIMS one item [(Item T1b No 148) together with one item (no 94) that was excluded from the international TIMSS analysis] were removed from the calibration analyses due to the misfitting of these items to the Rasch model. The highest number of items was removed from the Ethiopian data set and the least from the 1994 (TIMS) Australian data file. The reason for the high rate of misfitting items for the EMS data set might be related to the difficulty levels of the anchor items. Table A8.6.9 shows the infit mean square values of the EMS data set analysed by employing the two procedure before equating. In both cases all the items were within the acceptable range, that is between 0.77 and 1.30 (see Appendix 8.6). Consequently, 35 items for EMS, 68 items for FIMS, 70 for SIMS and 156 items for TIMS fitted the Rasch model.

The Fit of Case Estimates

The other way of investigating the fit of the Rasch scale to data is to examine the estimates for each case. The case estimates express the performance level of each student on the total scale. In order to identify whether the cases fit the scale or not, it is important to examine the case OUTFIT mean square statistic (OUTFIT MNSQ) which measures the consistency of the fit of the items to the student characteristic curve for each student, with special consideration given to extreme items. In this study, the general guideline used for interpreting t as a sign of misfit is if t>5 (Wright and Stone, 1979, p. 169). That is, if the OUTFIT MNSQ value of a person has a t value >5, that person does not fit the scale and is deleted from the analysis. However, in this analysis no person was deleted, because the t values for all cases were less than 5.

Equating of Mathematics Achievement between Occasions and over Time

The equating of the mathematics tests requires common items between occasions, that is between EMS, FIMS, SIMS and TIMS.

In this study, the number of common items in the mathematics tests for EMS, FIMS and SIMS data sets was 65. These common items formed approximately 93 per cent of the items for FIMS, and 90 per cent for EMS and SIMS. Thus, the common items in the mathematics test for the three occasions were all above the percentage ranges proposed by Wright and Stone (1979), and Hambleton et al. (1991), which are discussed in Chapter 7.

There were also some items which were common for EMS, FIMS, SIMS and TIMS data sets. Garden and Orpwood (1996, p. 2-2) reported that achievement in TIMSS was intended to be linked with the results of the two earlier IEA studies. Thus, in the TIMS data set there were nine items which were common to the other three occasions. Therefore, it was possible to claim that there were just sufficient numbers of common items to equate the mathematics tests on the three occasions.

Rasch model equating procedures were employed for equating the three data sets. Rentz and Bashaw (1975), Beard and Pettie (1979), Sontag (1984) and Wright (1995) have argued that Rasch model equating procedures are better than other procedures for equating achievement tests. The three types of Rasch model equating procedures, namely concurrent equating, anchor item equating and common item difference equating were all used for equating the four data sets.

Because of the unavailability of the TIMS data set on the expected time frame, it was not possible to employ the concurrent equating procedure for EMS, FIMS, SIMS and TIMS data sets. Thus, a decision was made to score the EMS, FIMS, SIMS and TIMS data sets using the anchor item equating procedure. In order to get good anchor values for the nine common items, it was decided to extract the anchor item values from the concurrent equating employed for calibrating the EMS, FIMS and SIMS data sets. Therefore, concurrent equating for calibration and anchor item equating for scoring were employed for EMS, FIMS and SIMS data sets. Common item difference equating procedure was examined for equating and calibrating the EMS and FIMS data with the TIMS data set. When anchor item equating was employed for calibrating and scoring the TIMS data one of the common items (Item 6 or A6) did not fit the Rasch model and had to be removed from the analysis. This reduced the number of common items between TIMS and the other studies. Thus in order to maintain the nine common items in the process, the equating procedure was changed from anchor item equating to common item difference equating.

Concurrent equating was employed for equating the data sets from EMS, FIMS and SIMS. In this method, the data sets from EMS, FIMS and SIMS were combined into one data set. Hence, the analysis was done with a single data file. Only one misfitting item was deleted at a time so as to avoid dropping some items that might eventually prove to be good fitting items. The acceptable infit mean square values were between 0.77 and 1.30 (Adams and Khoo, 1993). The concurrent equating analyses revealed that among the 65 common items 64 items fitted the Rasch model. Therefore, the threshold values of these 64 items were used as anchor values in the anchor item equating procedures employed in the scoring of the EMS, FIMS and SIMS data sets separately. Among the 64 common items, nine were common to the EMS, FIMS, SIMS and TIMS data sets. The threshold values of these nine items generated in this analysis are presented in Table 8.2 and were used in equating the FIMS and EMS data sets with the TIMS data set.

Table 8.2 Description of the common item difference equating procedure employed in FIMS, SIMS, EMS and TIMS

FIMS, SIMS and EMS		**TIMS**		**TIMS - FIMS**
Item Number	Thresholds	Item Number	Thresholds	Thresholds
12	0.21	K4	0.87	0.66
26	0.21	J14	1.90	1.69
31	-2.38	A6	-0.84	1.54
32	-0.08	R9	1.45	1.53
33	-1.10	Q7	-0.38	0.72
36	-0.82	M7	-0.87	-0.05
38	0.28	G6	1.31	1.03
54	0.27	F7	1.67	1.40
67	0.26	G3	0.47	0.21
Sum / N				8.73 / 9
Mean				0.97

N = number of common items; Equating Constant = 0.97;
Standard deviation of equating constant = 0.59; Standard error of equating constant = 0.197

The design of TIMS was different from FIMS and SIMS in two ways. In the first place, only one mathematics test was administered in both FIMS and SIMS, however, in the 1994 study the test included mathematics and science items and the study was named TIMSS (Third International Mathematics and Science Study). The other difference was that in the first two international studies, the test was designed as one booklet. Every participant used the same test booklet. Whereas in TIMSS, a rotated test design was employed. The test was designed in eight booklets. Garden and Orpwood (1996, p.2-16) explained the arrangement of the test in eight booklets as follows.

> This design called for items to be grouped into “clusters”, which were distributed (or “rotated”) through the test booklets so as to obtain eight booklets of approximately equal difficulty and equivalent content coverage. Some items (the core cluster) appeared in all booklets, some (the focus cluster) in three or four booklets, some (the free-response clusters) in two booklets, and the remainder (the breadth clusters) in one booklet only. In addition, each booklet was designed to contain approximately equal numbers of mathematics and science items.

All in all there were 286 (both mathematics and science) unique items that were distributed across eight booklets for Population 2 (Adams and Gonzalez, 1996, p.3-2). Garden and Orpwood (1996) also reported that the core cluster items (six items for mathematics) were common to all booklets. In addition, the focus cluster and free-response clusters were common to some booklets. Thus, it was possible to equate these eight booklets and report the achievement level in TIMS on a common scale. Hence, among the Rasch model test equating procedures concurrent equating was chosen for equating these eight booklets. Consequently, the concurrent equating procedure was employed for the TIMS data set. The result of the Rasch analysis indicated that only one item was deleted from the analysis. Out of 157 items, 156 of the TIMS test items fitted well the Rasch model. The item which was deleted from the analysis was Item 148 (T1b), whose infit mean square value was below the critical value of 0.77. From this concurrent equating procedure it was possible to obtain the threshold values of the nine common items in TIMS. These threshold values are shown in Table 8.2.

The next step involved the equating of the FIMS data set with the TIMS data set using the common item difference equating procedure. In this method the threshold value of each common item from the concurrent equating run for the combined FIMS, SIMS and EMS mathematics test data set was first subtracted from the threshold value for the item in the TIMS test. Then the differences were summed and divided by the number of anchor test items to obtain a mean difference. Subsequently the mean difference was subtracted from the case estimated mean value on the second test to obtain the adjusted mean value. In addition the standard deviation of the nine difference values and the standard error of the mean were calculated and are recorded in Table 8.2.

Comparisons between Students on Mathematics Test

The comparisons of the performance of students on the mathematics test for the four occasions were undertaken for two different subgroups namely, (a) Year 8 students who participated in the study, and (b) 13-year-old students in both government and nongovernment schools, who participated in the study. All EMS students were Year 8 students, while all SIMS students were 13-year-old students from both government and nongovernment schools. Meanwhile, some of the FIMS students were 13-year-old students, while others were younger and/or older students who were in Year 8. Therefore, for comparison purposes FIMS students were divided into two groups, namely: (a) FIMSA which involved all 13-year-old students, and (b) FIMSB which involved all Year 8 students including 13-year-olds. Thus, FIMSA students' results could be compared with the SIMS government school students' results, because all students were 13-year-olds. In TIMS a decision was made to include only Year 8 government school students, because they were the only group of students who were common to all participating states in FIMS. Thus TIMS Year 8 government school students could be compared with FIMSB and EMS students, because in the three groups students were at different age levels but at the same year level.

The Australian Capital Territory (ACT) and South Australia (SA) participated in the SIMS and TIMS, but not in FIMS and Northern Territory (NT) participated in the TIMS but not in the FIMS and SIMS studies. Consequently, these two territories and South Australia (SA) were excluded from the comparisons between FIMS, SIMS and TIMS. Nongovernment schools were not involved in FIMS, however, they participated in SIMS, TIMS and in EMS. Therefore, for comparability, the nongovernment school students who participated in the EMS, SIMS and TIMS were also excluded from the comparison between FIMSA and SIMS, between EMS and FIMSB, between EMS and TIMS, and between FIMSB and TIMS. Thus the SIMS data set employed for comparison purposes with FIMSA in this section is called SIMSR (SIMS restricted), and the TIMS data sets employed for comparison purposes with EMS and FIMSB is named TIMSR (TIMS restricted).

Comparison between Students on Different Occasions

This section considers two types of comparisons, the first section compares the 13-year-old students between FIMS and SIMS and the Year 8 students between in EMS and FIMSB, EMS and TIMSR, FIMSB and TIMSR. The EMS student group has also been subdivided into government and nongovernment student subgroups in the comparisons. However, the comparisons between EMS and FIMSB, and between EMS and TIMSR involved only government school students. The second comparison is between government and nongovernment school students in EMS, SIMS and TIMS.

Comparisons of Students' Level of Mathematics Achievement between EMS, FIMS, SIMS and TIMS

The first comparison is between all students who participated in EMS, FIMS, SIMS and TIMS. Table 8.3 shows the descriptive statistics for government and nongovernment school students and all students who participated in the mathematics studies on the four occasions. The EMS students' mean score was markedly lower than the mean scores for the other groups (see Table 8.3, Figure 8.2). This indicates that the test was markedly more difficult for the EMS students than for the Australian students. The mean score of the EMS students was approximately 1.80 logits below the 1964 item mean. Robitaille and Travers (1992, p. 691) reported that in the First International Mathematics Study "...all groups of students from all of the participating countries found the tests difficult."

Table 8.3 Descriptive Statistics for Mathematics Achievement of Students for the Four Occasions

	EMS			**FIMS**[a]		
	G	Ng	Total	FIMSA	FIMSB	Total
Mean	318.0	342.0	330.0	460.0	451.0	456.0
standard deviation	52.0	68.0	53.0	96.0	82.0	92.0
Standard error of the mean	3.3	9.2	5.2	4.9	5.1	4.7
design effect	2.4	11.1	8.6	7.7	11.8	11.1
sample size	600	600	1200	2917	3081	4320

	SIMS				**TIMS**			
	G	Ng	Total	R	G	Ng	Total	R
Mean	442.0	478.0	450.0	441.0	427.0	484.0	446.0	426.0
standard deviation	102.0	109.0	104.0	102.0	124.0	113.0	123.1	126.0
standard error of the mean	3.9	9.6	3.9	4.3	7.6	7.3	6.0	8.3
design effect	5.7	8.8	7.0	5.4	17.3	11.5	17.4	16.5
sample size	3989	1131	5120	3038	4648	2744	7392	3786

	Mean differences	**Effect size**	**t-value**	**Significance level**
EMS G vs Ng	-24.0	-0.40	-2.36	<0.05
EMS G vs FIMSB	-142.0	-1.94	-21.83	<0.01
EMSG vs TIMSR	-108.0	-1.12	-4.89[b]	<0.01
SIMSG vs Ng	-36.0	-0.34	-3.47	<0.01
FIMSA vs SIMSR	19.0	0.19	2.91	<0.01
TIMSG vs TIMSNG	-57.0	-0.48	-5.41	<0.01
FIMSB vs TIMSR	25.0	0.24	1.13[b]	NS
Alternative estimation of equating error				
FIMSB vs TIMSR	31.0	0.29	-2.16[c]	<0.05

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; R = Groups of SIMS students comparable with FIMSA and Groups of TIMS students comparable with FIMSB; NS = Not significant

Therefore, the Ethiopian students like their counterparts in many other countries of the world found the test difficult. It would seem very important to undertake further investigation to identify the reasons why the Ethiopians achieved markedly lower than the Australian students. This would help to find ways and means of improving the achievement level of the Ethiopian students in mathematics.

Comparisons between EMS and FIMSB Students

When the mean scores of the EMS government school students and the FIMSB students are compared, the mean score of the EMS students was more than 1.40 logits below the mean scores of the FIMSB students (see Table 8.3, Figures 8.1 and 8.2). This difference indicates how difficult the test was for the Ethiopian government school students. The standard deviation value for EMS (52) was smaller than for FIMSB (82). The standard deviation value indicates that the students' scores showed more variability in FIMSB than in EMS. The standard error was smaller for EMS (3.3) than FIMSB (5.1) and the design effects for FIMSB (11.8) was larger than for EMS which was 2.4. Both the standard error and the design effect values were noticeably higher for FIMSB than EMS students which were consequences of both the stratification and the extent of clustering in the Australian schools. The effect size (-1.94) and t-values were (-21.83) were very large. These values indicate that the mean score difference between EMS government school students and FIMSB students was both practically and statistically significant at the 0.01 level.

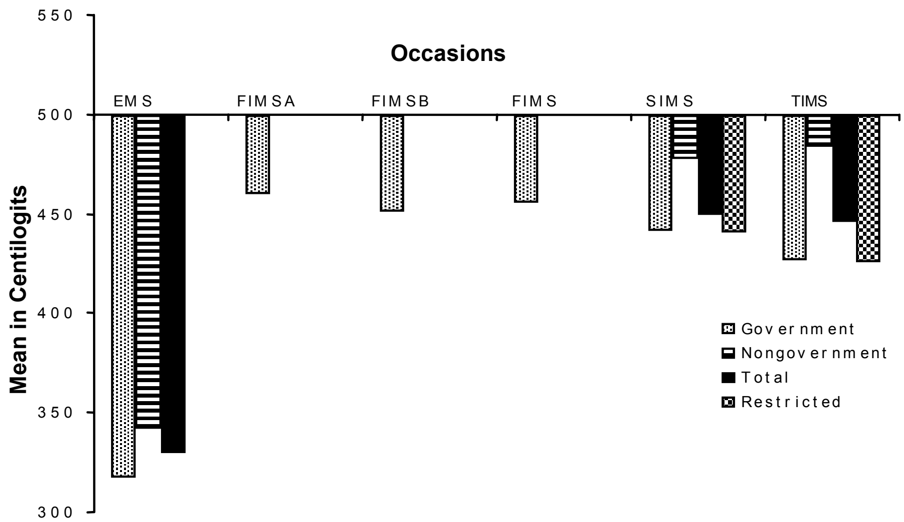


Figure 8.2 Comparison of mean scores of EMS, FIMS, SIMS and TIMS by school type

Comparisons between EMS and TIMS Students

The comparison between the estimated mean score of EMS government school and TIMSR students showed that, the TIMSR students achieved markedly better than the Ethiopians. The mean score difference was 108 centilogits, in favour of the TIMSR students (see Table 8.3, Figures 8.1 and 8.2). The difference was about one logit. The standard deviation, the standard error and design effect values of the TIMSR students were greater than for the EMS government school students. The effect size(-1.12) and t-value (-4.89) were very large. Therefore, the estimated mean score difference between the two groups was practically and statistically significant at the 0.01 level.

Comparisons between FIMSA and SIMSR Students

When the mathematics test estimated mean scores of FIMSA (13-year-old students) and SIMSR (13-year-old students excluding ACT and SA and all nongovernment school students in Australia) students were compared, the FIMSA score was higher than the SIMSR mean score (see Table 8.3, Figures 8.1 and 8.2). The estimated mean score difference between the two occasions was 19 centilogits, the difference was in favour of the 1964 13-year-old Australian students. This revealed that the mathematics achievement of Australian students declined from 1964 to 1978. The differences in standard deviation and standard error values for the two groups were small, while the design effect was slightly larger in 1964 than in 1978. The effect size was small (0.19) and the t-value was 2.91. Hence, the mean difference was statistically significant at the 0.01 level (see Table 8.3, Figures 8.1 and 8.2). Thus in Australia the mathematics achievement level of the 13-year-old students declined over time, between 1964 and 1978, to an extent that represented approximately half of a year of mathematics learning.

Comparisons between FIMSB and TIMSR Students

The next comparison was between FIMSB and TIMSR students. The estimated mean score of the 1964 Australian Year 8 students was 451, while, it was 426 in 1994 for the TIMS restricted sample. The difference was 25 centilogits in favour of the 1964 students (see Table 8.3, Figures 8.1 and 8.2). This difference revealed that the mathematics achievement level of Australian Year 8 students has declined over the last 30 years. The standard deviation, standard error and the design effects were markedly larger in 1994 than in 1964. The effect size was small (0.24) and the t-value was 1.13. While the effect size difference between FIMSB and TIMS is approximately three-quarters of a year of school learning, this difference is not statistically significant as a consequence of the large standard error of the equating constant shown in Table 8.2 and considered to be about of 19.7 centilogits. Because of this extremely large standard error for the equating constant which arose from the use of only nine common items it was considered desirable to undertake alternative procedures to estimate the equating constant and its standard errors. They use the five state subsamples and the nine common items to provide more accurate estimation. With these alternative procedures a mean difference of 31.0 with an effect size of 0.29 (see Table 8.3), or nearly a full year of mathematics learning was obtained which was found to be statistically significant at the five per cent level of significance.

Summary

The results of the comparisons in this section showed that the Ethiopian Year 8 students achieved at a lower level in mathematics, when compared with all groups of Australian students. The investigation also revealed that the mathematics achievement of Australian students declined significantly over time at the 13-year-old level. Moreover, there was not a statistically significant decline at the Year 8 student level. The next comparison shows the differences between government and nongovernment school students' achievement in mathematics on the three occasions.

Comparisons between Government and Nongovernment School Students

The comparison undertaken in this section is between government and nongovernment school students in EMS, SIMS and TIMS, since nongovernment school students did not participate in FIMS. The first comparison is between EMS government and nongovernment school students.

Comparisons between EMS Students

When the mathematics test mean scores of the EMS government and nongovernment school students are compared, the nongovernment school students mean score is higher than that of the government school students (see Table 8.3 and Figure 8.2). The difference was 24 centilogits. The difference was in favour of the nongovernment school students. The design effect of government school students (2.4) was smaller than for nongovernment school students (11.1). The effect size was small (-0. 40) and t-value was -2.36. Therefore, the mean difference between the EMS government and nongovernment school students was practically and statistically significant at the 0.05 level. This shows that in Ethiopia the mathematics performance of the nongovernment school students was higher than that of the government school students.

Comparisons between SIMS Students

The next comparison was between government and nongovernment school students in SIMS. When the estimated mean scores of the SIMS government and nongovernment school students are compared, the nongovernment school students mean score is higher than the government school students' mean score (see Table 8.3 and Figure 8.2). The difference was 36 centilogits in favour of the nongovernment school students. The standard deviation, standard error and the design effects of the nongovernment schools were higher than for the government schools (see Table 8.3). The effect size was small (-0.34) and the t-value was -3.47. Therefore, the mean difference between the SIMS government and nongovernment school students was both practically and statistically significant at the 0.01 level. This shows that like the Ethiopian students in SIMS the mathematics performance of the nongovernment school students was higher than that of the government school students, and approximately equal to a year of learning mathematics in Australian schools.

Comparisons between TIMS Students

The last comparison in this section is between government and nongovernment school students in TIMS. When the estimated mean score of the government and nongovernment school students are compared, the nongovernment school students achieved higher than the government school students (see Table 8.3 and Figure 8.2). The mean score difference was 57 centilogits in favour of the nongovernment school students. The standard deviation, standard error and design effect values of the government school students were larger than for nongovernment schools. The effect size was small (-0.48) and the t-value was -5.41. Therefore, the mean difference between the TIMS government and nongovernment school students was practically and statistically significant at the 0.01 level. This shows that like the Ethiopians and the SIMS students, in TIMS the performance of the nongovernment school students was higher than that of the government school students. Moreover, it should be noted that the difference was approximately about one and a third years of mathematics learning in Australian schools.

It is interesting to observe the same trend on the three occasions. That is in EMS, SIMS and TIMS the nongovernment school students were found to be higher achievers than their government school peers. This shows that the mathematics achievement of students in government schools in both countries is lower than for the nongovernment school students. It is important to note that in both developing and developed countries such as Ethiopia and Australia respectively, nongovernment schools showed higher achievement levels than the government schools. The other important point to be noted is that for the 16-year period between 1978 and 1994, the achievement difference between government and nongovernment school students in Australia would appear to be widening. Hence, further research might be important to investigate why the government school students' achievement level is lower than the nongovernment school students both in Australia and in Ethiopia.

Conclusion

In this chapter differences in mathematics achievement between Ethiopian and Australian students, and between the 1964, 1978 and 1994 Australian students were investigated. The findings of the study are summarised as follows.

1. The Ethiopian students achieved at lower levels than all groups of Australian students.
2. The achievement level of Australian 13-year-old students declined between 1964 and 1978. Moreover, there was a decline of clear significant in practical terms at the Year 8 level between 1964 and 1994.
3. On all occasions nongovernment school students achieved at a higher level than the government school students.

The findings in both comparisons between EMS and FIMSB, EMS and TIMSR revealed that the achievement level of Ethiopian students was markedly lower than that of the Australian students. Meanwhile, the comparison between FIMSA and SIMS, FIMSB and TIMSR showed that the achievement level of Australian students had declined over time. These findings indicated that there is a need to investigate differences in conditions of learning in the two countries. Carroll (1963) has identified five factors that influence school learning. One of the factors identified by Carroll was students' perseverance (motivation and attitude) towards the subject they are learning. Therefore, it is important to examine EMS, FIMS and SIMS students' views and attitudes towards mathematics and schooling. In addition, Carroll (1963) has emphasized the importance of time given to learning and it would seem likely that the decline in achievement in Australia in mathematics recorded over the 30 year period is related to a reduction in time given to the learning of mathematics in Australian schools.

9

Changes in Attitudes Towards School and Mathematics Over Time

Many conclusions can be derived from various research findings about students' attitudes towards school and mathematics. Research studies have shown that the achievement of students in different subjects depended on the level of attitudes of the students towards the subject they were learning and towards schooling and the schools in which they were studying. Research studies in Ethiopia have indicated that students who expressed more favourable attitudes towards schooling and mathematics were higher achievers in mathematics (Derese et al. 1990). Moreover, Australian studies have shown that students who considered that they were good at mathematics said that they liked the subject. While those who performed at a lower level recognised mathematics as a disliked subject (Helfers, 1986).

Australian research findings have also reported that the attitudes of students towards mathematics declined as they moved through school (Fraser, 1979). Many years after Fraser's findings Endalkachew's (1990) and Tadesse's (1993) studies in Ethiopia revealed that students' attitudes towards mathematics and their attitudes towards school and school learning remained low. Nevertheless, from these studies, it would be difficult to conclude that over recent decades the attitudes of students towards mathematics had changed since these studies rarely took place in the same schools and school systems and at the same year levels. However, because of technological developments, curricular reforms and the training of teachers, the attitudes of students might well have changed over time. Knowledge of such changes over time could help educators and mathematics educators, in particular, improve both their curricula and the training of teachers.

In order to generalise with respect to the views and attitudes of students towards schooling and mathematics, it would seem important to analyse data which were collected on different occasions using the same instrument and the same group of students or the same year level of students. Since, Australia participated in the 1964 and 1978 International Mathematics Studies, it was possible to examine the views and attitudes of students towards mathematics and schooling across the 14-year period.

Thus, the purposes of the analyses in this chapter are to:

(a) identify the changes in views and attitudes towards mathematics and schooling of Australian students between the 1964, and 1978;

(b) examine the differences in views and attitudes towards mathematics and schooling between Australian and Ethiopian students; and

(c) examine the differences in views and attitudes towards mathematics and schooling between government and nongovernment school students in EMS and SIMS.

In the 1994 mathematics study, regarding students' views and attitudes towards mathematics and schooling there were insufficient items to make comparisons over time and across countries. Thus, the comparisons of views and attitudes were limited between FIMS and SIMS, and between FIMS and EMS students. Therefore, in this chapter the results of the Rasch analyses of the views and attitudes of the 1964 and 1978 Australian students who participated in FIMS and SIMS and Ethiopian students who participated in EMS are discussed. The chapter is divided into four sections. The methods employed in the study are presented in the first section of the chapter, while the second section examines whether or not the view and attitude items fit the Rasch model. The views and attitudes of the students towards mathematics and schooling in the different studies are compared in section three. The last section of the chapter presents the findings and conclusions drawn from the chapter.

Methods Employed in the Study

In this chapter the students' views and attitudes towards schooling and mathematics over time are measured using the Rasch model.

Use of Rasch model

The Rasch model, which is the most robust of the item response models, is used to equate students' views and attitudes on a common scale between the investigations conducted in FIMS, SIMS and EMS.

Rating Scale analysis

Rating scale analysis was chosen from among the different Rasch models to employ in these analyses. The rating scale procedure was selected for a variety of reasons. On the one hand, rating scale procedures are employed with attitude questionnaires where the responses are either bipolar or unipolar (Wolf, 1994). When responses to the statements are in the form of 'agree', 'disagree' or 'undecided' a scale is bipolar, and when responses are in a form such as 'excellent', 'good', 'fair' or 'poor', the scale is unipolar. The item responses involved in this study were considered to form bipolar scales. On the other hand, the rating scale procedure also "assumes a single underlying dimension for the variable and seeks to scale the item responses in such a way that interval scale data are obtained for the variable formed" (Wolf, 1994, p. 4926). The use of interval scale data has many advantages. One of the advantages is that persons and items are located jointly on the same scale, which is important for measuring change over time (Anderson, 1997, p. 890). Andrich (1997, p. 879) has also argued that the Rasch model allows a separation of the item and person parameters in the estimation, so that no assumptions about the distributions of the person parameters is necessary in estimation. However, such separation is not possible in other models such as the Thurstone Model.

Unidimensionality

In order to employ the Rasch model in analysing both view and attitude scales it was important to examine whether or not these scales were unidimensional since the unidimensionality of items is one of the requirements for the use of the Rasch model (Hambleton and Cook, 1977; Anderson, 1994). If the view and attitude items were found not to be unidimensional, it would not be possible to apply the Rasch procedures to the analysis. Wolf (1997, p. 961) has argued that the Rasch models "assume a single underlying dimension for the variable and seek to scale the data in such a way that the interval scale data are obtained for the variable so formed."

Consequently, a literature search was undertaken to examine whether or not the test developers had examined the unidimensionality of the statements employed in the different scales. The view and attitude statements employed in 1964, 1978 and 1996 were first developed for use in 1964. Wolf (1967) reported that after the items had been developed, they were submitted to a panel of judges for ranking the sets of statements and for placing them along the appropriate scales. A pilot study was also undertaken with a representative sample of about 150 students in each participating country to examine whether the patterns of responses were as expected by the judges' rankings along the view and attitude scales (Wolf, 1967).

Wolf (1967) has reported that the responses of students from the pilot study were initially analysed by employing the Guttman scale analysis procedure. The Guttman scale analysis procedure provides a means of evaluating the unidimensionality of a set of items (Edwards and Kilpatrick, 1948, p. 99). Hence, the first step in the analysis was to compare the position of the items ranked by the judges and the students, and statements were deleted, when differences occurred in the position of the statements. The position of all statements in all view and attitude scales were also compared from country to country, and items were deleted when a marked difference occurred. The third step employed involved the deletion of items because of problems with content and translation across countries.

The coefficients of reproducibility obtained from the Guttman scale analysis were calculated. These coefficients were calculated for the total sets of items and were below 0.90 which was the lower acceptable limit for Guttman scale analysis (Guttman, 1947, p. 452). However, Wolf (1967), reported that the coefficients of reproducibility were increased to the 0.90 level, considered by Guttman to be acceptable, by recalculating the coefficients of reproducibility on the final sets of view and attitude statements.

Therefore, the view and attitude items employed in 1964, 1978 and 1996 were believed to be consistent with the designers' theoretical model of a unidimensional scale of statements and it was anticipated that the items selected would form unidimensional scales, and that the scale scores would be meaningful. Edwards (1957, p. 176) argued that:

> ... if the responses of subjects to the statements were in accord with our theoretical model of a unidimensional scale of statements, we would have confidence in interpreting scores of subjects based upon the statements as also falling along the same unidimensional continuum.

Keeves (1966b, p.108) also reported that "efforts were made to ensure that the statements finally selected formed a unidimensional scale". In addition, Morgensten and Keeves (1994, p. 392) contended that "efforts were made to ensure that the statements finally selected logically formed a unidimensional scale". Therefore, the sets of statements were considered to lie along unidimensional scales without further analyses.

Procedure Used for Equating

The concurrent equating procedure which is probably the strongest and most robust of the current methods of equating several test forms that differ in their threshold levels (Morrison and Fitzpatrick, 1992; Mohandas, 1996) was used in calibrating and equating the attitudes scales employed in 1964, 1978 and 1996.

The Data Processing Procedure

The 1964 Australian and the 1996 Ethiopian Opinion Questionnaires contained 65 view and attitude statements. These 65 view and attitude statements were subdivided into two view and five attitude scales, namely Views about Mathematics Teaching, Views about School and School Learning, Attitudes towards Mathematics as a Process, Attitudes towards Facility of Learning Mathematics, Attitudes towards the Place of Mathematics in Society, Attitudes towards School and School Learning and Attitudes towards Control of the Environment scales (see Chapter 5). From these 65 items, only 36 were used in the 1978 Opinion Questionnaire. Among the 36 items employed in SIMS, Item 22 (Item 49 in FIMS) was assigned in the Attitudes towards the Place of Mathematics in Society scale, while in FIMS it was assigned in Attitudes towards Mathematics as a Process scale. There was no reason stated why the item was moved to the other scale. Therefore, a decision was made to remove the item from further analyses. Thus, the remaining 35 items which were common across the three occasions were taken into consideration for equating purposes.

Among the 64 items there were 25 items, where the response of ‘disagree’ was considered a favourable response. An example of such a reversed statement is:

> 7. My mathematics teacher wants pupils to solve problems only by the procedures he teaches.

For the other items a favourable response was indicated by agreement with the statement. An example of such a statement is:

> 16. My mathematics teacher encourages us to try to find several different methods for solving particular problems.

It should be noted that response categories were presented to the students on an answer sheet in an order with the undecided category last; thus responses were of the form ‘agree’, ‘disagree’ and ‘undecided’. Moreover, missing data or omitted responses were coded ‘0’ in FIMS and ‘9’ in EMS. Subsequently, the missing data or omitted responses in EMS, FIMS and SIMS were recoded as ‘undecided’. In order to process the information meaningfully all the responses to the reversed statement items described above, were recoded from 0123 to 1021 and from 1239 to 1021 for FIMS and EMS respectively. The other items that did not involve a reversed statement, also had to be recoded from 0123 to 1201 and 1239 to 2011 for FIMS and EMS respectively.

The SIMS data set had been previously recoded, however, to allow for the reversed statement items and the recoding required was 1239 to 0121 for all the 36 items. These different recodings (1021), (2011) and (0121) all preserve the order of the original response alternatives, to be consistent with unfavourable: 0, undecided: 1, and favourable: 2. After completing all the above preliminary procedures, the three data sets were combined to form a single data file prior to carrying out concurrent equating.

Treatment of Omits and Non-responses

Students in these analyses tended to omit only one or two items in a scale, and Anderson (1994) has argued that if a student omitted only one item, or less than 20 per cent of the items in a scale, then an omission would seem to show uncertainty in responding and it could be said that assigning a neutral response provided appropriate compensation for the missing data. Under these circumstances, the missing data or the omitted responses were recoded as undecided.

However, if a student failed to provide responses to at least 80 per cent of the items in a scale, Anderson (1994) has recommended the exclusion of such a student from further analysis. Consequently, all students in the Queensland sample were excluded from analysis for the Views about Mathematics Teaching and Views about School and School Learning scales, because most students in this sample failed to provide responses to more than 20 per cent of the items in these two scales. This was the result of students in Queensland being told by the state authorities not to answer certain items that referred to their teachers and schools.

Treatment of Zero and Perfect Scores

The QUEST computer program (Adams and Khoo, 1993) by default does not process cases with perfect and zero scores, because both groups do not provide information about the scale. Cases with perfect scores are those cases who provided favourable responses for all the view and attitude statements, while cases with zero scores are those cases who provided unfavourable responses for all the view and attitude statements. However, in order to include those cases with perfect and zero scores in the calculation of the mean and standard deviation for each scale, the values of perfect and zero scores were calculated by extrapolation from the logit table produced by the computer program. The SPSS 6.1 (Statistical Package for Social Sciences, 1993) computer program was used to calculate the case estimate mean scores and standard deviations and the WesVarPC 2. 11 (Brick et al., 1997) computer program was employed to calculate the jackknife standard error values of the case estimate mean values. Weights were employed in calculating the case estimate mean scores, standard deviations and standard errors for each occasion.

Developing Common View and Attitude scales

The concurrent calibration of the view and attitude data permit scales to be constructed that extend across the three groups, namely EMS, FIMS and SIMS students on the seven view and attitude scales. The fixed point of each scale was set at 500 with one logit, the natural metric of the scale, being set at 100 scale units (centilogits). The mean difficulty level of the items in each scale was the fixed point of that scale, and 500 was taken as the mean of the item difficulty levels for each scale. The seven view and attitude scales constructed in this way for all different sample groups of students in FIMS, SIMS and EMS are presented in Figures 9.1 to 9.7, with 100 scale units (centilogits) being equivalent to one logit.

Rasch Analyses of the View and Attitude Statements

In the previous section the Rasch model was identified as the major procedure for the analyses of the view and attitude statements. In order to employ the Rasch model for a particular data analysis, the statements or the items must fit the model. Consequently, the statements must be examined to determine whether or not they fitted the Rasch model. Therefore, in this section the view and attitude statements are examined to identify those statements that fitted the Rasch model using the QUEST computer program (Adams and Khoo, 1993).

Rasch Analysis

Among the different Rasch models rating scale analysis was chosen for these analyses. The equating procedure employed in this study was concurrent equating which is the most robust of the current horizontal methods of equating (Morrison and Fitzpatrick, 1992; Mohandas, 1996: Hungi, 1997). Three groups of students namely EMS (1200), FIMS (4320) and SIMS (5120) were involved in the analyses. The necessary requirement to employ Rasch scaling is that the items must fit the Rasch scale. Items that do not fit the scale must be deleted. In order to examine whether or not the items fitted the scale, it was important to evaluate both the item fit statistics and the person fit statistics. The results of these analyses are presented below.

Item fit statistics

The seven view and attitude scales were Rasch analysed using the concurrent equating procedure to examine whether or not the items and cases fitted the scale. The results of the analysis showed that 60 out of 64 items had infit mean square coefficients within the acceptable range of 0.77 and 1.30 (Adams and Khoo, 1993). Only four items had coefficients that were outside the acceptable range. Therefore, the four misfitting items were deleted from the final analysis.

The four deleted items were from the scale concerned with Attitudes towards School and School Learning. Four of the seven scales, that is Views about School and School Learning, Attitudes towards Mathematics as a Process, Attitudes towards Facility of Learning Mathematics and Attitudes towards the Place of Mathematics in Society scales fitted well the Rasch model. Since the infit mean square indices from the final analysis of the items in these scales ranged between 0.90 and 1.14, the items fitted well the Rasch model, whereas the infit mean square indices for the other three scales ranged from 0.79 to 1.25. The lower value of 0.79 was near to the lower cutting point 0.77 and the higher value of 1.25 was near to the upper limit 1.30. Hence, the items in these three scales fitted less well when compared with the items in the four good fitting scales.

Case Estimates

The other way of investigating the fit of the scales to data is to examine the estimates for each case. The case estimates express the attitude level of each student on each scale. In order to identify whether the cases fit a scale or not, it is important to examine the case estimate fit t (outfit and infit) statistic. In this study, the general guideline used for interpreting t as a sign of misfit is if $t > 5$ (Wright and Stone, 1979, p. 169). That is, if the case estimate value of a person has a t value >5, that person does not fit the scale and that person should be deleted from the analysis. However, in these analyses no person was deleted, because the t values for all cases were less than 5. Hence, the seven view and attitude scales constructed after the deletion of four non-fitting items were considered to conform to the Rasch model.

Changes of Students' Views and Attitudes towards Mathematics and Schooling over time

This section compares the views and attitudes of EMS, FIMS and SIMS students towards school and mathematics over time. For the comparison of the changes of students' attitudes towards mathematics and schooling over time, the case mean, standard deviation, standard error and design effect of the samples for each occasion, and effect size and t-value were calculated for the compared groups using a weighted data set.

The comparisons of the views and attitudes of students towards mathematics and schooling for the three occasions were undertaken for two different subgroups namely, (a) Year 8 students who participated in the study, and (b) 13-year-old students in government schools, who participated in the study. All EMS students were Year 8 students, while all SIMS students were 13-year-old students. Meanwhile, some of the FIMS students were 13-year-old students, while others were younger and/or older students who were in Year 8. Therefore, for comparison purposes FIMS students were divided into two groups, namely (a) FIMSA which involved all 13-year-old students and (b) FIMSB which included all Year 8 students. Thus, FIMSA students' results could be compared with SIMS government school students, because all students were 13-year-olds, while EMS government school students could be compared with FIMSB students, because in both groups students were at different age levels but at the same year level.

The Australian Capital Territory (ACT) and South Australia (SA) participated in the SIMS, but not in FIMS. Consequently, these two states were excluded from the comparisons between FIMS and SIMS. In 1964, Queensland students did not respond to the two view scale questionnaires, therefore, they were excluded from the comparison between FIMS and SIMS for these two scales. Moreover, for various reasons the nongovernment schools were not involved in FIMS, although, they did participate in SIMS and in EMS. Therefore, for comparability, the nongovernment school students who participated in the EMS and SIMS were also excluded from the comparison between FIMSA and SIMS, and EMS and FIMSB. Thus the SIMS data set employed for comparison purposes with FIMSA in this section is called SIMSR (SIMS restricted).

The first part of this section compares the views and attitudes of students towards mathematics, while the second part considers their views and attitudes towards school and school learning.

Changes in Students' Views and Attitudes towards Mathematics Over Time

In this subsection the changes of students' views and attitudes towards mathematics are compared. The subsection has three main parts. The first part describes the views of students about mathematics teaching, while the second part discusses their attitudes towards mathematics as a process. The third part examines the attitudes of students towards facility of learning mathematics. The last part presents the attitudes of students towards the place of mathematics in society.

Comparison of EMS, FIMS and SIMS students' Views about Mathematics Teaching

One of the four scales used in the comparison of students' views and attitudes towards mathematics was the Views about Mathematics Teaching scale. This scale was developed to measure the views of students about the procedures used by their mathematics teachers' in the teaching of mathematics. The scale ranged from a method that emphasised problem-solving to one that emphasised rote-learning. In order to compare the EMS, FIMS and SIMS students' Views about Mathematics Teaching the case estimated mean scores of the three groups of students were calculated. The standard deviations and the jackknife standard error values were also calculated. The effect size and the t-values were employed to assess the significance level of the comparisons.

Eleven statements formed the Views about Mathematics Teaching scale. When the frequency distribution of students' responses on the three occasions were compared, for Statements 4 and 12 a majority of the students in all four groups responded favourably (see Table 9.1a). However, a majority of the students in FIMS and SIMS provided unfavourable responses for Statements 7, 18 and 19, while the majority of the Ethiopian students provided favourable responses to these statements.

Table 9.1a Frequency Distribution of the FIMS, SIMS and EMS Students' Views about Mathematics Teaching Items

Item No	Response type	FIMSA	FIMSB	FIMS	SIMSR	SIMS	EMS	Total
4	Favourable	73.3	73.6	73.4	68.0	67.3	86.8	
	Undecided	6.6	6.4	6.4	9.3	10.4	4.4	100
	Unfavourable	20.1	20.0	20.2	22.6	22.3	8.8	
6	Favourable	70.7	70.0	70.2	a	a	74.7	
	Undecided	5.6	4.9	5.3			16.1	100
	Unfavourable	23.7	25.0	24.6			9.3	
7	Favourable	31.2	28.9	30.6	38.0	38.4	50.1	
	Undecided	12.6	11.7	12.2	16.7	16.5	21.6	100
	Unfavourable	56.2	59.4	57.2	45.4	45.1	28.3	
12	Favourable	73.1	73.3	73.1	65.1	65.5	81.9	
	Undecided	6.5	5.8	6.1	9.7	9.7	8.2	100
	Unfavourable	20.4	20.9	20.8	25.3	24.8	9.9	
13	Favourable	18.4	16.0	17.3	a	a	41.0	
	Undecided	23.6	23.4	23.3			32.2	100
	Unfavourable	58.1	60.6	59.4			26.8	
15	Favourable	86.1	87.1	86.9	a	a	87.0	
	Undecided	7.2	6.4	6.7			4.5	100
	Unfavourable	6.7	6.5	6.3			8.5	
16	Favourable	48.4	48.1	49.1	41.6	40.5	88.4	
	Undecided	12.0	12.5	12.0	19.5	19.6	4.4	100
	Unfavourable	39.6	39.4	38.9	38.9	39.9	7.2	
17	Favourable	31.5	31.5	33.2	a	a	43.2	
	Undecided	6.2	7.3	6.0			19.9	100
	Unfavourable	62.3	61.1	60.8			36.9	
18	Favourable	34.2	34.0	34.7	25.7	25.3	52.0	
	Undecided	21.4	21.7	21.3	34.3	35.3	14.6	100
	Unfavourable	44.4	44.3	44.0	40.0	39.4	33.4	
19	Favourable	31.2	32.0	32.2	32.9	32.2	68.3	
	Undecided	12.7	12.2	12.5	24.1	23.9	14.3	100
	Unfavourable	56.1	55.8	55.2	43.0	43.8	17.5	
21	Favourable	11.7	11.7	11.8	a	a	25.2	
	Undecided	10.4	10.8	10.6			14.1	100
	Unfavourable	77.8	77.6	77.6			60.7	
	Sample Size	2197	2350	3261	2395	5120	1200	

[a] Items excluded in SIMS

The statements were:

4. My mathematics teacher shows us different ways of solving the same problem.
6. My mathematics teacher does not like pupils to ask questions after he has given an explanation. (D)
7. My mathematics teacher wants pupils to solve problem only by the procedures he teaches. (D)
12. My mathematics teacher expects us to learn how to solve problem by ourselves but helps when we have difficulties.
13. In my mathematics class, pupils who have original ideas get better marks than do pupils who are most careful and neat in their work.
15. My mathematics teacher requires the pupils not only to master the steps in solving problems, but also to understand the reasoning involved.
16. My mathematics teacher encourages us to try to find several different methods for solving particular problems.
17. My mathematics course requires more thinking about the methods of solving problems than memorization of rules and formulae.
18. My mathematics teacher wants us to discover mathematical principles and ideas for ourselves.
19. My mathematics teacher explains the basic ideas; we are expected to develop the methods of solution for ourselves.
21. Most of the problems my mathematics teacher assigns are to give us practice in using a particular rule or formula.(D)

(D) indicates reversed item.

The case mean scores of the four groups of students were above the item mean for the scale which is set to be 500 (see Table 9.1b and Figure 9.1). Thus a majority of the students gave favourable responses to the items. This shows that to some extent the items were such that the students responded favourably.

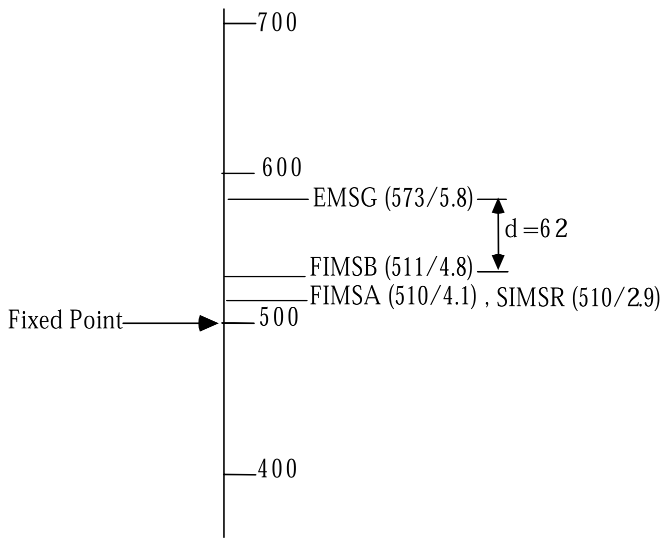
Table 9.1b Descriptive statistics for Views About Mathematics Scale of all students for the three occasions

	EMS			FIMS[a]			SIMS			
	G	Ng	Total	FIMSA	FIMSB	Total	G	Ng	Total	R
mean	573.0	581.0	577.0	510.0	511.0	510.0	510.0	504.0	508.0	510.0
standard deviation	65.0	61.0	63.0	52.0	52.0	52.0	91.0	95.0	92.0	90.0
SE of the mean	5.8	4.2	3.6	4.1	4.8	4.3	2.3	4.1	2.0	2.9
design effect	4.9	2.9	3.8	13.4	19.9	20.0	2.5	2.1	2.4	2.6
sample size	600	600	1200	2197	3078	3261	3989	1131	5120	2395
effect size				**t-value**			**significant level**			
EMS G vs Ng	-0.13			-1.12			NS			
EMS G vs FIMSB	1.05			8.24			<0.01			
SIMSG vs Ng	0.06			1.28			NS			
FIMSA vs SIMSR	0.00			0.00			NS			

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; R = Groups of SIMS students comparable with FIMSA; NS = not significant; SE = standard error

When the case estimate mean scores of the FIMSA (13-year-olds) students were compared with those of 1978 (SIMSR) students, on the Views about Mathematics Teaching scale, the estimated mean scores of both groups of students was the same 510. There was no difference between the two groups of students. This result indicated that both FIMSA and SIMSR students viewed mathematics teaching as emphasising the use problem-solving methods rather than rote-learning. After the introduction of problem solving into Australian schools in the 1970s, it was assumed that the 1978 students would view mathematics teaching as emphasising more the use problem-solving methods rather than rote-learning, therefore the result was not the expected ones since, there was no difference between FIMS and SIMS students' views over the 14-year time period. The design effects (DEFF) of FIMSA and SIMSR samples were 13.4 and 2.6 respectively. The design effect for the 1964 Australian study is extremely large. This is probably a consequence of the very different styles of mathematics teaching in different schools in different parts of Australia at that time which led to a large intra-class correlation coefficient for student groups within schools for this scale. Since the design effect assesses the extent to which school groups differed and was dependent on the intra-class correlation (ρ), where m is the average school group size and ρ is the intra-class correlation] it was to be expected that ρ would be large, if DEFF were large and positive and the values of ρ for FIMS and SIMS were 0.29 and 0.08 respectively.

However, when the case mean scores of FIMSB and EMS government school students are compared on this scale the mean score of the Ethiopian government school students was higher than for FIMSB students (see Table 9.1b, Figures 9.1). The difference between Ethiopian and FIMSB students was 62 centilogits. This shows that the EMS government school students viewed mathematics teaching as emphasising more problem-solving methods rather than rote-learning compared to their Australian peers. The effect size between FIMSB and EMS government school students (1.05) and the t-value (8.24) were large (see Table 9.1b). Thus, the difference was both practically and statistically significant at the 0.01 level.



Fixed Point – 500; Metric - 100 centilogits=1 logit; Values indicated for each occasion in the brackets are Rasch estimated scores and standard errors of the mean respectively; d= differences in views and attitudes between occasions.

Figure 9.1 The Views about Mathematics Teaching scale of Students in EMS, FIMS and SIMS

The reason why the Ethiopian government school students viewed the ways their mathematics teachers taught mathematics more positively than the Australian students might be due to the nature of the educational system of the country. The Ethiopian junior and senior secondary school curricula are centrally planned. Textbooks and teachers' guides are centrally prepared (see Chapter 3). The teachers had the responsibility to implement the centrally prepared syllabuses of the subject they were teaching according to the instructions provided in the teachers' guide.

They had to develop their own annual, semester, unit and daily lesson plans based on the information provided in the teachers' guide. Therefore, the teachers knew what to teach, when to teach, how to teach at the beginning of the year. The information provided by the Ministry of Education covered specific teaching methods such as the teaching of problem solving and this was discussed in the supplementary materials prepared by the Ministry of Education.

> Children learn by doing and mathematical ideas grow out of experience with real objects-by-handling them, arranging them, re-arranging them and comparing them. Therefore a child should be encouraged to do all those things which can help him to discover for himself and so gain a more valuable insight into the basic underlying concepts. (Ministry of Education and Fine Arts, 1967/68, p.102; Ministry of Education, 1981, p. 92)

Therefore, this centralised system of teaching would appear to have helped the Ethiopian students to develop very favourable views about their mathematics teaching.

When the SIMS government and nongovernment school students' estimated mean scores were compared the government school students score was slightly higher than the nongovernment school students' score (see Table 9.1b). However, the difference was not statistically significant, since the effect size and t-values were 0.06 and 1.28 respectively. Moreover, the estimated mean score difference between Ethiopian government and nongovernment school students was only eight centilogits. Unlike SIMS, the difference here was in favour of the nongovernment school students. However, the difference was too small to be considered significant since the effect size (-0.13) was trivial and t-value was only -1.12. Therefore, the difference between government and nongovernment schools was not statistically significant.

The findings can be summarised as follows.

1. The Ethiopian government school students expressed more positive views about mathematics teaching than the 1964 Year 8 Australian students.
2. There was no significant difference between the 1964 and 1978 13-year-old Australian students in their views about mathematics teaching, even though more favourable views might have been expected in the second mathematics study.
3. Significant differences were not found between government and nongovernment school students in Ethiopia and Australia.

Comparison of EMS and FIMS students' Attitudes towards Mathematics as a Process

The second analysis was concerned with the attitudes of students towards mathematics as a process. Only the FIMS and Ethiopian students responded to this scale. Consequently, the comparison was between FIMSB and EMS. The scale originally consisted of eight items, but Item 49 was deleted because it was assigned for different scales in FIMS and SIMS. In this scale low estimated mean scores indicated the degree to which students considered mathematics as fixed and given once and for all time, while high estimated mean scores showed the level to which students considered mathematics as something that was developing, growing, and changing. The estimated mean scores for the students in the two countries were below the mean level of the items that formed the scale. Both groups of students tended to consider that mathematics was a subject that was fixed, and given once and forever (see Table 9.2a, 2b and Figures 9.2). However, when the two groups were compared, the Australian students considered mathematics more as a subject that was developing, growing and changing. It can be seen in Table 9.2a, that the majority of FIMS and EMS students responded in a similar pattern to most of the items on this scale except Items 36, 38 and 62.

An example of such an item is:

> Item 62. Mathematics will change rapidly in the near future.

Table 9.2a Frequency Distributions of EMS and FIMS Students' Attitudes towards Mathematics as a Process Statements

Item number	Response type	FIMSA	FIMSB	FIMS	EMS	Total
23	Favourable	15.0	15.0	15.1	14.0	
	Undecided	5.6	6.4	6.2	11.7	100
	Unfavourable	79.3	78.6	78.8	74.3	
36	Favourable	18.1	17.7	18.2	42.4	
	Undecided	10.9	11.6	11.2	26.4	100
	Unfavourable	71.1	70.7	70.6	31.2	
38	Favourable	43.4	43.0	43.7	16.2	
	Undecided	23.8	24.8	24.0	17.8	100
	Unfavourable	32.7	32.2	32.3	65.9	
42	Favourable	50.1	51.7	50.9	37.0	
	Undecided	20.7	20.4	20.7	29.5	100
	Unfavourable	29.2	28.0	28.5	33.4	
49[d]	Favourable	53.9	54.3	54.7	51.7	
	Undecided	24.9	24.6	24.6	20.4	100
	Unfavourable	21.2	21.1	20.7	27.9	
59	Favourable	38.6	38.2	38.3	52.7	
	Undecided	32.8	34.7	33.9	26.0	100
	Unfavourable	28.6	27.1	27.8	21.3	
62	Favourable	47.2	46.5	46.6	18.8	
	Undecided	26.5	27.6	27.0	56.8	100
	Unfavourable	26.4	25.9	26.4	24.4	
64	Favourable	38.4	36.9	38.0	19.7	
	Undecided	10.9	11.3	11.2	10.0	100
	Unfavourable	50.6	51.8	50.8	70.2	
Sample Size		2917	3081	4320	1200	

The statements were:

23. In mathematics there is always a rule to follow in solving problems. (D)
36. The most important reason for studying arithmetic and secondary school mathematics is that they help people to take care of their own financial affairs. (D)
38. Mathematics helps one to think according to strict rules. (D)
42. Almost all of present-day mathematics was known at least a century ago. (D)
49. Mathematics is a very good field for creative people to enter.
59. There is little place for originality in mathematics. (D)
62. Mathematics will change rapidly in the near future.
64. In the study of mathematics, if the pupil misses a few lessons it is difficult to catch up. (D)

(D) = reversed item; d=Item 49 was deleted as it was assigned for different scales in FIMS & SIMS

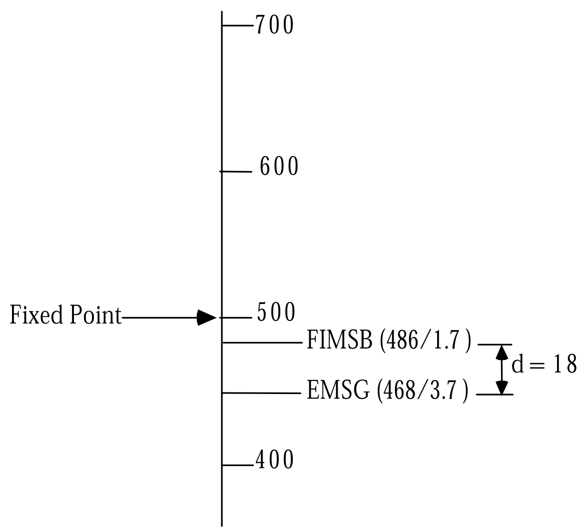
In this statement a response of 'agree' was considered a favourable response and students were expected to respond by agreement. The frequency distribution of the responses of students in Table 9.2a (Item 62) indicates that 57 per cent of the Ethiopian students were uncertain about the statement, whereas 47 per cent of the FIMS students provided a favourable response to the statement. Husén (1967) reported that among the 12 countries which participated in FIMS, those students from the countries that had introduced new mathematics viewed mathematics as more of an open and changing system. Australia was one of the countries that had introduced new mathematics into some of its schools in the 1960s (Owen et al. 1983; Clements, 1989). Hence, at the time of testing some Australian students had experienced changes in their mathematics curricula (see Chapter 3), whereas, there was no change in the mathematics curriculum since the introduction of new mathematics in Ethiopia in the 1970s (see Chapter 3). Therefore, the Ethiopian students had not experienced a change of curriculum in their mathematics education. This might have been the reason why the Australian students expressed more favourable Attitudes towards Mathematics as a Process than the Ethiopian students.

The case mean scores for the two groups of students were very low and were below 500 which was the item mean score. The scores were 486 and 468 for FIMSB and EMS government school students respectively (see Table 9.2b and Figure 9.2). The case mean score difference between the two groups was 18 centilogits, in favour of the 1964 Australian Year 8 students. The design effects of the FIMSB and EMSG samples were 2.7 and 2.6 respectively. The effect size (-0.31) was small and the t-value was -4.42. Thus the difference was practically and statistically significant at the 0.01 level.

Table 9.2b Descriptive statistics for EMS and FIMS students' Attitudes towards Mathematics as a Process Scale

	EMS			FIMS[a]		
	G	Ng	Total	FIMSA	FIMSB	Total
mean	468.0	474.0	471.0	487.0	486.0	486.0
standard deviation	56.0	52.0	54.0	59.0	59.0	59.0
standard error of the mean	3.7	3.3	2.4	2.0	1.7	1.8
design effect	2.6	2.4	2.5	3.5	2.7	4.2
sample size	600	600	1200	2917	3078	4320
effect size		**t-value**			**significant level**	
EMS G vs Ng	-0.11	-1.21			NS	
EMS G vs FIMSB	-0.31	-4.42			<0.01	

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; NS = Not significant



Fixed Point - 500., Metric - 100 centilogits = 1 logit; Values indicated for each occasion in the brackets are Rasch estimated scores and standard errors of the mean respectively; d= differences in views and attitudes between occasions.

Figure 9.2 The Attitudes towards Mathematics as a Process Scale of Students in EMS and FIMS

When the estimated mean scores of Ethiopian government and nongovernment school students are compared, the mean score of the nongovernment school students was slightly higher than that of the government school students. However, the difference was only six centilogits. This difference was too small to be considered significant since the effect size was -0.11 and the t-value was -1.21. Therefore, the difference was not statistically significant.

Thus, the main findings on the Attitudes towards Mathematics as a Process scale may be stated as follows.

1. The Australian students showed more favourable Attitudes towards Mathematics as a Process than the Ethiopian students, but both groups of students expressed relatively unfavourable attitudes to this scale.
2. There was no significant difference between the government and the nongovernment school Ethiopian students in their Attitudes towards Mathematics as a Process.

Comparison of EMS, FIMS and SIMS students Attitudes towards Facility of Learning Mathematics

The third analysis was about the attitudes of students towards the facility of learning mathematics. The EMS, FIMS and SIMS students responded to this scale. Consequently, the comparisons were among the three groups. This scale consisted of seven statements and the scale is considered to measure the attitudes of students towards facility of learning mathematics. The frequency distributions of the students' responses on these seven items are shown in Table 9.3a. The majority of the students on the three occasions responded favourably to all the items. Moreover, there were some differences between groups with respect to those students who responded unfavourably or undecided to the items.

The estimated mean scores of the groups on this scale were higher than for the previous two scales. The mean scores for SIMSR and FIMSA were 643 and 626 respectively, while the EMS government school students' mean score was 581 (see Table 9.3b and Figure 9.3). The estimated mean scores for both groups of Australian students with respect to attitudes concerning the ease of learning mathematics were higher than for the Ethiopian students. The difference in mean scores between the FIMSB and EMS government school students was 49. The effect size (-0.47) was classified as moderate and the t-value was -5.08. Therefore, the difference was practically and statistically significant at the 0.01 level. This difference showed that Australian Year 8 students perceived mathematics as a subject that most students could learn, while the Ethiopian students perceived mathematics as more difficult and a subject to be studied only by more able students.

When FIMSA and SIMSR students were compared on their attitudes about the facility of learning mathematics, the SIMSR students' mean score was higher than for the FIMSA students. The SIMSR students' mean score was higher by 17 centilogits. The effect size was -0.17 and the t-value was -4.16, so the difference was statistically significant at the 0.01 level. This difference indicated that the SIMSR students had more favourable attitudes about the facility of learning mathematics than had the FIMSA students. The introduction of calculators to schools and other related materials might have helped the SIMSR students to think that every one could learn mathematics, compared to the FIMSA students.

When the SIMS government and nongovernment school students mean scores are compared, the mean score of the government school students was higher than the nongovernment school students' mean score. The difference was 13 centilogits, moreover the difference was significant at the 0.10 level. The effect size and the t-values were 0.12 and 1.82 respectively. Since the effect size is trivial it is doubtful whether this difference can be considered as practically or statistically significant.

However, the estimated mean score difference between EMS government and nongovernment school students was 28 centilogits. Unlike SIMS the difference was in favour of the nongovernment school students. The design effect for both groups was 4.0. The effect size was -0.25 and the t-value was -2.21. Therefore, the difference was significant at the 0.05 level.

Table 9.3a Frequency Distribution of FIMS, SIMS and EMS Students' Attitudes towards Facility of Learning Mathematics Statements

Item No	Response type	FIMSA	FIMSB	FIMS	SIMSR	SIMS	EMS	Total
29	Favourable	57.3	57.6	57.3	64.8	63.0	72.6	
	Undecided	9.7	9.9	9.8	13.2	13.6	12.3	100
	Unfavourable	33.0	32.5	32.9	22.0	23.3	15.1	
37	Favourable	83.0	83.9	82.7	85.5	86.2	53.4	
	Undecided	6.5	6.4	6.8	7.8	7.5	24.7	100
	Unfavourable	10.5	9.8	10.5	6.7	6.3	21.9	
48	Favourable	89.3	89.7	89.4	91.4	91.4	84.6	
	Undecided	4.0	3.6	3.9	4.2	4.3	7.3	100
	Unfavourable	6.6	6.6	6.7	4.4	4.3	8.1	
51	Favourable	90.3	89.9	89.7	88.0	88.6	69.4	
	Undecided	4.7	4.9	4.9	6.9	6.6	17.9	100
	Unfavourable	5.1	5.2	5.4	5.1	4.8	12.7	
53	Favourable	66.0	65.8	65.6	56.5	56.8	74.6	
	Undecided	19.8	20.8	20.2	27.0	26.5	13.6	100
	Unfavourable	14.3	13.5	14.2	16.4	16.8	11.8	
57	Favourable	72.6	73.9	72.9	74.7	73.7	80.8	
	Undecided	11.9	11.8	12.0	14.2	14.2	10.3	100
	Unfavourable	15.5	14.3	15.1	11.1	12.1	8.9	
61	Favourable	86.2	86.9	86.3	88.4	88.9	70.5	
	Undecided	4.8	4.4	4.7	5.4	5.3	17.0	100
	Unfavourable	9.0	8.7	9.1	6.2	5.8	12.4	
	Sample Size	2917	3081	4320	3038	5120	1200	

The statements were:
29. Anyone can learn mathematics.
37. Very few people can learn mathematics. (D)
48. Almost anyone can learn mathematics if he is willing to study.
51. Any person of average intelligence can learn to understand a good deal of mathematics.
53. Even complex mathematics can be made understandable and useful to every high school pupil.
57. Almost all pupils can learn complex mathematics if it is properly taught.
61. Only people with a very special talent can learn mathematics. (D)

(D) indicates reversed item.

Table 9.3b Descriptive statistics for FIMS, SIMS and EMS students' Attitudes towards Facility of Learning Mathematics Scale

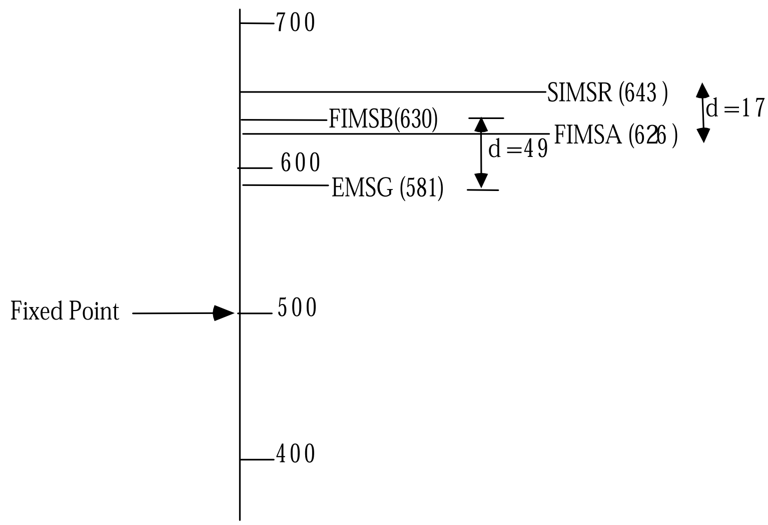
	EMS			FIMS[a]			SIMS			
	G	Ng	Total	FIMSA	FIMSB	Total	G	Ng	Total	R
mean	581.0	609.0	595.0	626.0	630.0	626.0	643.0	630.0	640.0	643.0
standard deviation	113.0	108.0	111.0	94.0	94.0	94.0	107.0	106.0	110.0	105.0
SE of the mean	9.2	8.7	6.6	3.5	2.9	2.9	1.9	6.9	2.2	2.1
design effect	4.0	4.0	4.2	4.0	2.9	4.0	1.3	4.5	2.2	1.2
sample size	600	600	1200	2917	3078	4320	3989	1131	5120	3038
effect size				**t-value**			**significant level**			
EMS G vs Ng	-0.25			-2.21			<0.05			
EMS G vs FIMSB	-0.47			-5.08			<0.01			
SIMSG vs Ng	0.12			1.82			<0.10			
FIMSA vs SIMSR	-0.17			-4.16			<0.01			

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; R = Groups of SIMS students comparable with FIMSA; NS = Not significant; SE = standard error

This study revealed that the attitudes of the Australian students towards the facility of learning mathematics increased over time. Hence, the findings are summarised as follows.

1. SIMSR students had more favourable Attitudes towards the Facility of Learning Mathematics than the FIMSA Students. Thus, students' Attitudes about the Facility of Learning Mathematics in Australia had changed significantly over time, towards seeing mathematics more as a subject that could be learned by all students.
2. Ethiopian government school students were found to have less favourable Attitudes about the Facility of Learning Mathematics than did the Australian students although their mean scores were relatively high.

Significant differences were found between government and nongovernment school students in both countries, the difference in Australia was in favour of government school students although it was only marginally significant, but in favour of nongovernment school students in Ethiopia and was statistically significant.



Fixed Point – 500; Metric - 100 centilogits = 1 logit; Values in the brackets are in centilogits; d= differences in views and attitudes between occasions.

Figure 9.3 The Attitudes about Facility of Learning Mathematics Scale of Students in EMS, FIMS and SIMS

Comparison of FIMS, SIMS and EMS Students' Attitudes towards the Place of Mathematics in Society

The fourth analysis was about the students' Attitudes towards the Place of Mathematics in Society scale. The eight statements included in this scale and the frequency distributions of the students' responses indicated that in five of the eight statements a majority of the respondents on the three occasions responded favourably (see Table 9.4a). A majority of the respondents on each occasion provided a different pattern of responses for three items, namely Items 27, 34 and 45. One of these items was:

> Item 27. More of the most able people should be encouraged to become mathematicians and mathematics teachers.

For this statement a response of 'agree' was considered a favourable response and students were expected to respond by agreement. The frequency distribution of the responses of students indicated that 71 per cent of the Ethiopian students agreed with the statement, while 57 per cent of the FIMS students provided an unfavourable response.

Table 9.4a Frequency Distribution of FIMS, SIMS and EMS students' Attitudes towards the Place of Mathematics in Society Statements

Item No	Response type	FIMSA	FIMSB	FIMS	SIMSR	SIMS	EMS	Total
27	Favourable	28.0	26.0	27.1	a	a	70.7	
	Undecided	16.2	16.0	16.2			15.6	100
	Unfavourable	55.8	58.0	56.7			13.6	
34	Favourable	33.1	32.8	33.1	55.9	56.9	68.5	
	Undecided	50.0	51.1	50.0	13.0	12.7	18.3	100
	Unfavourable	16.9	16.9	16.9	31.1	30.4	13.2	
35	Favourable	77.3	77.2	77.3	70.0	68.7	85.5	
	Undecided	11.0	11.1	11.0	17.7	19.2	6.2	100
	Unfavourable	11.7	11.7	11.7	12.3	12.1	8.4	
39	Favourable	56.0	57.2	56.8	51.5	51.7	59.3	
	Undecided	12.9	12.6	12.6	19.0	19.2	23.6	100
	Unfavourable	31.1	30.2	30.6	29.5	29.1	17.0	
45	Favourable	50.2	50.2	50.0	32.4	31.5	27.4	
	Undecided	24.1	26.1	26.2	40.4	30.5	35.8	100
	Unfavourable	25.7	23.6	23.9	27.2	28.0	36.8	
47	Favourable	77.4	78.2	77.9	76.1	74.1	78.4	
	Undecided	8.0	7.6	7.7	11.9	12.8	11.5	100
	Unfavourable	14.6	14.2	14.4	12.1	13.1	10.0	
50	Favourable	57.3	59.4	58.7	59.6	59.1	58.4	
	Undecided	12.8	13.1	13.0	16.4	16.0	23.3	100
	Unfavourable	29.9	27.5	28.3	24.1	24.9	18.2	
54	Favourable	69.3	69.1	68.6	54.1	52.7	69.4	
	Undecided	15.5	15.1	15.9	27.3	28.4	21.3	100
	Unfavourable	15.2	15.8	15.5	18.6	18.9	9.3	
	Sample Size	2917	3081	4320	3038	5120	1200	

a item excluded from SIMS.

The statements of this scale were:

27. More of the most able people should be encouraged to become mathematicians and mathematics teachers.
34. Outside of science and engineering, there is little need for mathematics (algebra, geometry, etc.) in most jobs. (D)
35. Mathematics is of great importance to a country's development.
39. Mathematics (algebra, geometry, etc.) is not useful for the problems of everyday life. (D)
45. A thorough knowledge of advanced mathematics is the key to an understanding of our world in the twentieth century.
47. It is important to know mathematics (algebra, geometry, etc.) in order to get a good job.
50. Unless one is planning to become a mathematician or scientist, the study of advanced mathematics is not very important.
54. In the near future most jobs will require a knowledge of advanced mathematics.

(D) indicates reversed item.

The results with respect to the students' attitudes towards the place of mathematics in society showed that the Ethiopian students had the higher mean score as compared with the 1964 Year 8 Australian students (see Table 9.4b and Figure 9.4). The differences was 24 centilogits. The design effects of the FIMSB, and EMSG samples were 2.9 and 2.3. The effect size (0.25) was small and the t-value was 3.17. Hence, the difference was significant at the 0.01 level. This indicates that the Ethiopian students believed that mathematics had an important and vital role in society, while the Australian students believed that the role of mathematics in society was of lesser value. Mathematics had a higher place in Ethiopian society than other subjects taught in schools. In Ethiopia it was a compulsory subject at all levels of education (Tilahun, 1995). Mathematics was also a compulsory subject for entry to higher educational institutions. Tilahun (1994, p. 44) indicated the importance of mathematics as follows:

Each candidate for ESLCE is awarded a certificate which lists the subjects on which a grade D or better was obtained. . . . The Grade Point Average (G. P. A.) of candidates is calculated from his/her best five subjects results including English and Mathematics and is used in selection for entry to higher education.

In addition, most employers required knowledge and skills in mathematics when they sought to recruit workers. This practice might well have helped Ethiopian students to develop more favourable Attitudes towards the Place of Mathematics in Society.

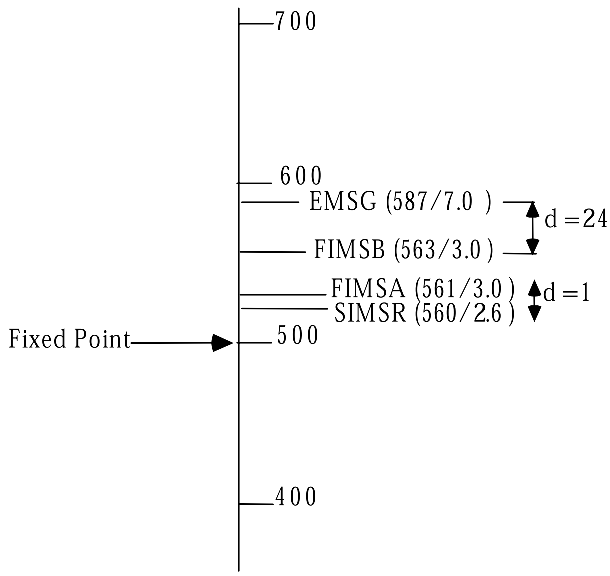
Table 9.4b Descriptive statistics FIMS, SIMS and EMS students' Attitudes towards the Place of Mathematics in Society Scale

	EMS			**FIMS**[a]			**SIMS**			
	G	Ng	Total	FIMSA	FIMSB	Total	G	Ng	Total	R
mean	587.0	586.0	586.0	561.0	563.0	561.0	558.0	557.0	558.0	560.0
standard deviation	113.0	108.0	110.0	78.0	75.0	76.0	102.0	103.0	102.0	103.0
SE of the mean	7.0	7.3	4.9	3.0	3.0	2.7	2.3	4.6	2.1	2.6
design effect	2.3	2.7	2.4	3.8	2.9	5.6	2.1	2.3	2.1	1.9
sample size	600	600	1200	2917	3078	4320	3989	1131	5120	3038
effect size				**t-value**			**significant level**			
EMS G vs Ng	0.01			0.10			NS			
EMS G vs FIMSB	0.25			3.17			<0.01			
SIMSG vs Ng	0.01			0.19			NS			
FIMSA vs SIMSR	0.00			0.00			NS			

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; R = Groups of SIMS students comparable with FIMSA; N= Not significant; SE = standard error

Meanwhile, when the mean score of the 1964 13-year-old Australian students was compared with that of the 1978 students, the mean scores difference was only one centilogit, in favour of the 1964 13-year-olds. This difference was too small to be considered significant. This shows that the role of mathematics in Australian society could be considered to be the same over the 14 year period. Like the Views about Mathematics scale, here the expectation was that the attitudes of students towards the importance of mathematics for the society would be more favourable in 1978 than in 1964. However, the result did not reach this level of expectation. Therefore, it might be of interest to examine curriculum statements in a search for the reasons the 1964 and 1978 students expressed similar Attitudes towards the role of Mathematics in Society.

The estimated mean score difference between the SIMS government and nongovernment school students was one centilogit, the difference was in favour of the government school students. However, this difference was not significant since the effect size (0.01) and t-value (0.19) were trivial. The design effects were 2.1 and 2.3 respectively. Meanwhile, the mean score difference between EMS government and nongovernment school students was also one centilogit, like the SIMS data the difference was in favour of the government school students. The design effects for the government school students and for nongovernment school students were 2.3 and 2.7 respectively. The effect size (0.01) and the t-value (0.10) were also trivial. Therefore, the difference was not statistically significant.



Fixed Point – 500; Metric - 100 centilogits = 1 logit; Values indicated for each occasion in the brackets are Rasch estimated scores and standard errors of the mean respectively;
d= differences in views and attitudes between occasions.

Figure 9.4 The Attitudes towards the Place Mathematics in Society Scale of Students in EMS, FIMS and SIMS

The summary of the findings may be stated as follows.

1. The Ethiopian students had more favourable Attitudes about the Place of Mathematics in Society than the Year 8 Australian students.
2. There was no significant difference between the 1964 and 1978 Australian 13-year-old students in their Attitudes towards the Place of Mathematics in Society.
3. Significant differences were not found between government and nongovernment school students in either country.

Changes of Students' Views and Attitudes towards School over time

In this subsection the changes of students' views and attitudes towards school and school learning are compared. The subsection has three main parts. The first part describes the views of students about school and school learning, while the second part discusses their attitudes towards school and school learning. The third part examines the attitudes of students towards man's ability to control the environment.

Comparison of FIMS and EMS students' Views about School and School Learning

The first analysis involved the Views about School and School Learning scale. The scale assesses the views of the students about the teaching-learning process in their respective schools. The items in this scale were not administered in 1978. Therefore, the comparison was between the EMS and FIMS students. Eleven statements formed this scale and the frequency distributions of the responses for each of the statements are provided in Table 9.5a. The mean score of the FIMSB students was 503, while, for the EMS government school students the score was 503 (see Table 9.5b, Figure 9.5).

This shows that the FIMSB students' mean scores was the same as that of the Ethiopian students.

Table 9.5a Frequency Distribution of FIMS and EMS Students' Views about School and School Learning Statements

Item number	Response type	FIMSA	FIMSB	FIMS	EMS	Total
1	Favourable	23.3	23.7	23.7	48.2	
	Undecided	10.3	11.1	10.8	30.5	100
	Unfavourable	66.5	65.2	65.5	21.3	
2	Favourable	38.6	36.6	37.4	12.3	
	Undecided	16.7	16.6	16.6	6.0	100
	Unfavourable	44.8	46.9	45.9	81.7	
3	Favourable	24.6	23.5	23.7	15.9	
	Undecided	7.7	8.4	8.0	11.6	100
	Unfavourable	67.7	68.1	68.2	72.5	
5	Favourable	46.9	45.9	46.5	24.9	
	Undecided	13.4	13.5	13.5	14.4	100
	Unfavourable	39.7	40.6	40.0	60.7	
8	Favourable	55.9	54.2	54.9	70.3	
	Undecided	15.0	15.3	15.0	8.3	100
	Unfavourable	29.1	30.4	30.0	21.5	
9	Favourable	78.4	79.5	78.7	72.5	
	Undecided	9.9	9.2	9.8	10.6	100
	Unfavourable	11.7	11.3	11.5	16.9	
10	Favourable	34.4	33.4	34.2	17.2	
	Undecided	16.5	16.1	15.8	17.5	100
	Unfavourable	49.1	50.5	50.0	65.3	
11	Favourable	54.0	52.3	53.2	69.2	
	Undecided	15.8	15.8	15.7	14.9	100
	Unfavourable	30.2	31.9	31.1	15.9	
14	Favourable	54.6	54.3	54.6	12.3	
	Undecided	12.2	12.7	12.3	10.3	100
	Unfavourable	33.3	33.0	33.1	77.5	
20	Favourable	61.2	59.3	60.2	73.5	
	Undecided	8.6	8.6	8.6	9.0	100
	Unfavourable	30.2	32.1	31.2	17.4	
22	Favourable	31.7	31.9	31.8	71.6	
	Undecided	16.8	16.3	16.8	12.0	100
	Unfavourable	51.5	51.9	51.4	16.4	
Sample size		2197	2350	3261	1200	

The statements of this scale were:
1. Most school work is the memorizing of information. (D)
2. In our school we get a great deal of practice & drill until we are almost perfect in our learning. (D)
3. The pupils spend most of their class time listening to the teachers and taking notes. (D)
5. Our teachers want us to do most of our learning from the textbook which is used in the course. (D)
8. We are expected to learn and discover many ideas for ourselves.
9. We are expected to develop a thorough understanding of ideas and not just to memorize information.
10. Our teachers believe in strict discipline and each pupil does exactly what he is told to do. (D)
11. Pupils are encouraged to devise their own projects or experiments in order to learn on their own.
14. Most of our classroom work is listening to the teacher. (D)
20. We do not use just one textbook for most of our subjects. Various sources and books from which we can learn are suggested to us.
22. Much of our classroom work is discussing ideas and problems with the teacher and the other pupils.

(D) indicates reversed item.

Moreover, the mean score difference between EMS government and nongovernment school students was only one centilogit and this slight difference was in favour of the nongovernment school students. The effect size (-0.03) and the t-vale (0.32) were trivial. Therefore, the difference was not statistically significant.

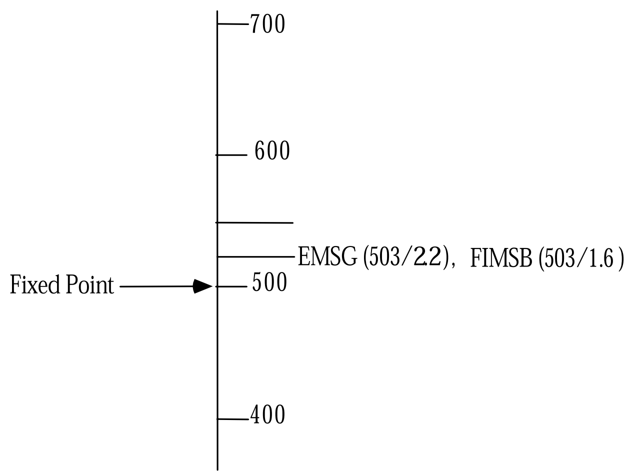
Thus the summary of the findings are as follows.

1. There was no statistically significant difference between the Australian and Ethiopian students in their views about their schools and school learning.
2. There was no significant difference between EMS government and nongovernment school students' views about their schools and school learning.

Table 9.5b Descriptive statistics for FIMS and EMS Students' Views about School and School Learning Scale

	EMS			FIMS[a]		
	G	Ng	Total	FIMSA	FIMSB	Total
Mean	503.0	504.0	503.0	507.0	503.0	504.0
standard deviation	33.0	31.0	32.0	43	42.0	45.0
standard error of the mean	2.2	2.2	1.5	1.7	1.6	2.1
design effect	2.6	2.8	2.6	3.4	3.5	6.0
sample size	600	600	1200	2197	3078	3261
effect size		**t-value**		**significant level**		
EMS G vs Ng	-0.03	-0.32		NS		
EMS G vs FIMSB	0.00	0.00		NS		

a = All students who participated in FIMS were from Government schools;
G = Government school students; Ng = Nongovernment school students;
R = Groups of SIMS students comparable with FIMSA; N = not significant



Fixed Point – 500; Metric - 100 centilogits = 1 logit; Values indicated for each occasion in the brackets are Rasch estimated scores and standard errors of the mean respectively.

Figure 9.5 Views about School and School Learning Scale in EMS and FIMS

Comparison of EMS, FIMS and SIMS Students' Attitudes towards School and School Learning

The second analysis in this section was on the students' Attitudes towards School and School Learning scale. The frequency distribution of the students' responses on the Attitudes towards School and School Learning scale indicated that the majority of the respondents on all the three occasions provided favourable responses to most of the statements (see Table 9.6a). However, the Australian and the Ethiopian students responded differently to three items, namely Items 44, 52 and 58. One of these items was:

Item 52. The most enjoyable part of my life is the time I spend in school.

Table 9.6a Frequency Distribution of FIMS, SIMS and EMS Students' Attitudes towards School and School Learning Statements

Item No	Response type	FIMSA	FIMSB	FIMS	SIMSR	SIMS	EMS	Total
24	Favourable	74.5	74.5	74.1	72.5	72.6	79.3	
	Undecided	10.8	10.7	10.8	16.7	16.6	11.9	100
	Unfavourable	14.1	14.7	15.1	10.8	10.8	8.9	
30[d]	Favourable	80.0	81.6	80.7	a	a	74.5	
	Undecided	5.9	5.4	5.7			14.2	100
	Unfavourable	14.1	13.1	13.6			11.3	
32	Favourable	73.9	74.3	74.0	68.5	70.4	70.4	
	Undecided	10.9	10.8	10.9	18.1	17.2	16.4	100
	Unfavourable	15.2	14.9	15.1	13.4	12.4	13.2	
41	Favourable	67.4	67.4	67.6	58.8	58.8	71.1	
	Undecided	10.4	10.8	10.3	16.9	17.4	17.9	100
	Unfavourable	22.2	21.8	22.1	24.3	23.8	10.9	
44	Favourable	32.3	30.8	31.8	18.8	17.6	74.2	
	Undecided	16.2	17.3	16.4	27.0	27.4	15.0	100
	Unfavourable	51.5	51.8	51.8	54.3	55.0	10.9	
46[d]	Favourable	72.8	70.5	71.8	15.3	15.2	42.1	
	Undecided	10.5	11.6	11.1	15.7	15.4	33.2	100
	Unfavourable	16.7	17.8	17.2	69.1	69.4	24.6	
52	Favourable	26.7	25.3	25.9	11.7	11.1	84.9	
	Undecided	20.6	21.2	20.8	24.3	24.7	8.0	100
	Unfavourable	52.7	53.5	53.2	64.0	64.3	7.1	
58	Favourable	30.3	28.8	29.2	18.9	18.6	74.4	
	Undecided	11.2	11.3	11.6	17.2	17.4	12.2	100
	Unfavourable	58.5	59.9	59.3	63.9	64.0	7.1	
60	Favourable	71.8	71.9	71.4	59.5	58.1	89.9	
	Undecided	13.4	14.0	14.5	24.2	25.0	3.0	100
	Unfavourable	14.8	14.1	14.1	16.3	16.9	7.1	
63[d]	Favourable	79.9	78.9	79.3	a	a	84.6	
	Undecided	10.1	10.4	10.3			6.1	100
	Unfavourable	10.0	10.7	10.4			9.3	
65[d]	Favourable	66.4	66.3	66.1	54.3	55.2	92.5	
	Undecided	15.5	14.7	14.8	26.6	26.2	3.0	100
	Unfavourable	18.1	19.0	19.1	19.1	18.7	4.5	
	Sample Size	2917	3081	4320	3038	5120	1200	

[a] Items excluded in SIMS
The statements included in this scale were:
24. I generally like my school work.
30. Most school learning has little value for a person. (D)
32. I disliked school and will leave just as soon as possible. (D)
41. I am bored most of the time in school. (D)
44. I enjoy everything about school.
46. School is not very enjoyable, but I can see value in getting a good education.
52. The most enjoyable part of my life is the time I spend in school.
58. I like all school subjects.
60. I enjoy most of my school work and want to get as much additional education as possible.
63. Although school is difficult, I want as much education as I can get.
65. I find school interesting and challenging.

(D) indicates reversed item.
d= Statements 30, 46, 63 and 65 did not fit the Rasch model, thus they were dropped from the analyses.

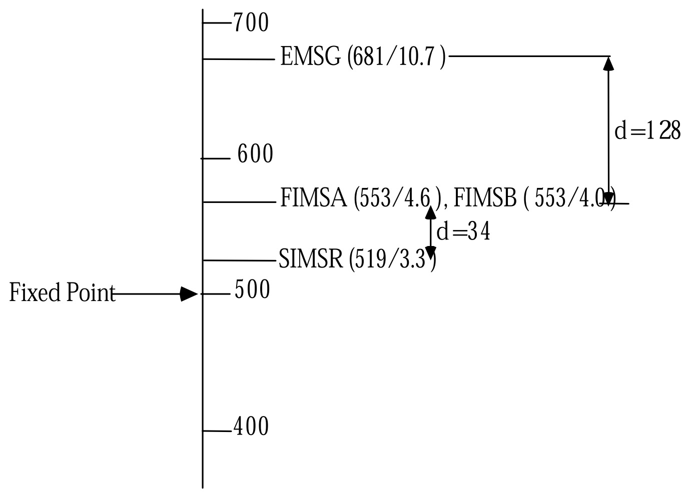
For this item a response of 'agree' was considered a favourable response and students were expected to respond by agreement. The frequency distribution showed that 53 and 64 per cent of the FIMS and SIMS students provided unfavourable responses respectively, while 85 per cent of the Ethiopian students responded favourably (see Table 9.6a). This difference in attitude between Australian and Ethiopian students warrants further consideration. In this attitude scale, the mean score of the Ethiopian government school students was much higher than that of the 1964 Year 8 Australian students. The difference was 1.28 logits (see Table 9.6b and Figure 9.6). The design effects were 3.9 and 2.6 for EMSG and FIMSB respectively. The effect size (0.94) was large and the t-value was 11.2. This shows that the Ethiopian students' liking of school was much higher than that of the Australian students.

Table 9.6b Descriptive Statistics for FIMS, SIMS and EMS Students' Attitudes towards School and School Learning Statements

	EMS			FIMS[a]		SIMS				
	G	Ng	Total	FIMSA	FIMSB	Total	G	Ng	Total	R
mean	681.0	692.0	686.0	553.0	553.0	553.0	518.0	515.0	518.0	519.0
Standard deviation	133.0	113.0	124.0	139.0	139.0	138.0	114.0	115.0	115.0	114.0
SE of the mean	10.7	6.5	6.1	4.6	4.0	3.9	3.0	7.2	2.8	3.3
design effect	3.9	2.0	3.0	3.3	2.6	3.4	2.7	4.4	3.1	2.5
sample size	600	600	1200	2917	3078	4320	3989	1131	5120	3038

effect size		**t-value**	**significant level**
EMS G vs Ng	-0.09	-0.87	NS
EMS G vs FIMSB	0.94	11.21	<0.01
SIMSG vs Ng	0.03	0.38	NS
FIMSA vs SIMSR	-0.27	-6.01	<0.01

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; R = Groups of SIMS students comparable with FIMSA; NS = Not significant; SE = standard error



Fixed Point – 500; Metric - 100 centilogits = 1 logit; Values indicated for each occasion in the brackets are Rasch estimated scores and standard errors of the mean respectively; d= differences in views and attitudes between occasions.

Figure 9.6 The Attitudes towards School and School Learning Scale of Students in EMS, FIMS and SIMS

The participation rates of students in primary, junior and senior secondary schools in Ethiopia in 1996 were 29, 20 and 6.5 per cent (Ministry of Education, 1996) respectively. In a country where 80 per cent of the Year 8 children are out of school, it is not surprising that those students who had the chance to attend school expressed highly favourable attitudes. Moreover, this finding was consistent with findings presented by Walker (1976), who reported that students from developing countries expressed more favourable attitudes towards school and school learning. Furthermore, Kotte (1992) in his analysis of the attitudes towards school and school learning scale administered in the Second International Science Study reported that among the participating countries students in Thailand scored higher than the other participating countries. Kotte argued that this result must be seen in relation to the economic situation in the countries with Thailand being the only developing country of the ten under survey. In Thailand, schooling was still understood to be a privilege as not every 14-year-old had the chance to attend school. Students perceived school and school learning as a great opportunity for their future careers. Therefore, the findings recorded here were to be expected.

On the other hand, when the two Australian groups were compared the mean score of the FIMSA students was noticeably higher than that of the SIMSR students. The mean score difference was 34 centilogits. The effect size was -0.27 and the t-value was -6.01. The difference was significant at the 0.01 level. This result reveals that the attitudes of students towards school and school learning had decreased over time, inspite of the massive input of financial resources into Australian schools between 1964 and 1978. The decrease was 34 centilogits on the scale, which is the largest when compared with the other scale mean differences over time. The students in 1978 were expected to like schooling more than the 1964 students, because governments had spent more money on education during the 1970s than in the 1960s. Furthermore, modern technological equipments such as calculators had been introduced into some schools. However, the investment by the government and the introduction of new technology into schools did not contribute to students liking their schools more. Therefore, further investigation is needed to find out the reason for the decline of students' attitudes towards school and school learning in Australia over the period from 1964 and to 1978.

When the government and nongovernment school SIMS students' estimated mean scores were compared the mean score of the government school students was slightly higher. The effect size (0.03) and the t-value (0.38) were trivial. Hence, the difference was not statistically significant.The estimated mean score difference between the Ethiopian government and nongovernment school students was 11 centilogits. Unlike the SIMS students the difference was in favour of the nongovernment school students. However, the difference was not statistically significant.

A summary of the findings is presented below.

1. The Ethiopian students expressed more favourable attitudes towards school and school learning than the 1964 Year 8 Australian students.
2. The 1964 Australian 13-year-old students showed more favourable attitudes towards school and school learning than did the 1978 students, and the attitudes of students towards school and school learning in Australian schools had decreased noticeably over time.
3. There was no significant difference in attitudes towards school and school learning between government and nongovernment school students in Ethiopia and Australia.

Comparison of EMS, FIMS and SIMS Students' Attitudes towards Control of the Environment

Table 9.7a presents the responses of the students on the Attitudes towards Control of the Environment scale. The table shows that the responses of the majority of the FIMS and EMS students were favourable except on Items 25, 33 and 43. Items 25 and 43 were excluded from SIMS. On Items 25 and 33 a majority of the Ethiopian students provided undecided and unfavourable responses respectively and favourable responses for Item 43, whereas the FIMS students responded favourably for Items 25 and 33 and unfavourably for Item 43. One of these items was:

> Item 25. It should be possible to eliminate war once and for all.

Table 9.7a Comparison of FIMS, SIMS and EMS Students' Attitudes towards Control of the Environment Scale

Item No	Response type	FIMSA	FIMSB	FIMS	SIMSR	SIMS	EMS	Total
25	Favourable	51.0	52.3	51.7	A	a	17.3	
	Undecided	25.3	25.7	25.4			48.2	100
	Unfavourable	23.7	22.0	23.0			34.4	
26	Favourable	72.3	74.1	73.3	A	a	70.1	
	Undecided	9.8	9.0	9.8			18.1	100
	Unfavourable	18.0	16.9	16.9			11.8	
28	Favourable	72.4	72.1	72.0	50.6	49.8	63.4	
	Undecided	12.0	12.9	12.4	27.5	28.0	11.3	100
	Unfavourable	15.7	15.1	15.6	21.9	22.2	11.3	
31	Favourable	56.1	55.4	55.7	33.6	34.2	81.1	
	Undecided	26.1	26.8	26.4	40.8	40.7	8.7	100
	Unfavourable	17.8	17.7	17.9	25.6	25.1	8.7	
33	Favourable	44.6	44.9	44.5	29.5	28.5	19.4	
	Undecided	25.0	24.5	25.0	31.6	32.9	15.5	100
	Unfavourable	30.4	30.6	30.6	39.0	38.6	51.5	
40	Favourable	59.2	59.9	59.3	34.6	35.1	61.7	
	Undecided	21.7	21.3	21.6	36.2	37.0	15.6	100
	Unfavourable	19.1	18.8	19.0	29.3	27.9	15.6	
43	Favourable	31.5	31.2	31.6	A	a	51.7	
	Undecided	21.6	22.6	22.1			32.8	100
	Unfavourable	46.9	46.2	46.3			32.8	
55	Favourable	78.1	76.8	77.2	74.6	73.4	94.8	
	Undecided	8.2	8.8	8.5	11.5	12.1	4.2	100
	Unfavourable	13.7	14.4	14.3	13.9	14.5	4.2	
56	Favourable	39.3	40.1	39.7	23.9	22.6	53.4	
	Undecided	25.8	25.1	25.4	33.8	32.6	8.0	100
	Unfavourable	34.8	34.8	34.9	42.4	44.8	8.0	
	Sample Size	2917	3081	4320	3038	5120	1200	

a=Items excluded in SIMS

The statements included in this scale were:

25. It should be possible to eliminate war once and for all.
26. Success depends to a large part on luck and fate. (D)
28. Some day most of the mysteries of the world will be revealed by science.
31. By improving industrial and agricultural methods, poverty can be eliminated in the world.
33. With increased medical knowledge, it should be possible to lengthen the average life span to 100 years or more.
40. Someday the deserts will be converted into good farming land by the application of engineering and science.
43. Education can only help people develop their natural abilities; it cannot change people in any fundamental way. (D)
55. With hard work anyone can succeed.
56. Almost every present human problem will be solved in the future.

(D) indicates reversed item.

For the above statement only 17 per cent of the Ethiopian students provided favourable responses, while 48 per cent of students were uncertain about the statement and provided an 'undecided' response. Meanwhile 52 per cent of the FIMS students provided favourable responses. This indicates that the Australian students believed that war could be eliminated once and for all, while the Ethiopian were very uncertain. The recent civil war and the present ethnic conflict in Ethiopia might be the cause of their uncertainty. The mean score difference between FIMSB and EMS government school students was six centilogits, and the effect size (-0.07) and t-value (-0.92) were trivial. Hence, the differences between FIMSB and EMS students' mean scores were not statistically significant.

However, the mean score difference between FIMSA and SIMSR was 47 centilogits, in favour of the 1964 13-year-olds. The effect size (0.61) was medium and the t-value (17.04) was large. Therefore, the difference between the FIMSA and SIMSR students was practically and statistically significant at the 0.01 level. This shows that the attitudes of Australian students towards control of the environment had decreased over time. This scale dealt with the relationship between man and the environment, concerning the extent to which man was perceived to have effective control over the environment. High mean scores indicated that the students perceived that man could have control over the natural and social environment. Because of technological development it was expected that the SIMS students would have more favourable attitudes towards control of the environment than the FIMS students, but the result was different from the expectation. Consequently, further study seems very important in order to examine the attitudes of students over time towards control of the environment.

When the mean scores of the SIMS government and nongovernment school students were compared, the government students mean was higher than the mean for the nongovernment school students by nine centilogits and the effect size (0.11) was small, and the t-value was 2.77, and the difference was statistically significant at the 0.01 level (see Table 9.7b and Figure 9.7).

Table 9.7b Descriptive Statistics for FIMS, SIMS and EMS Students' Attitudes towards Control of the Environment Scale

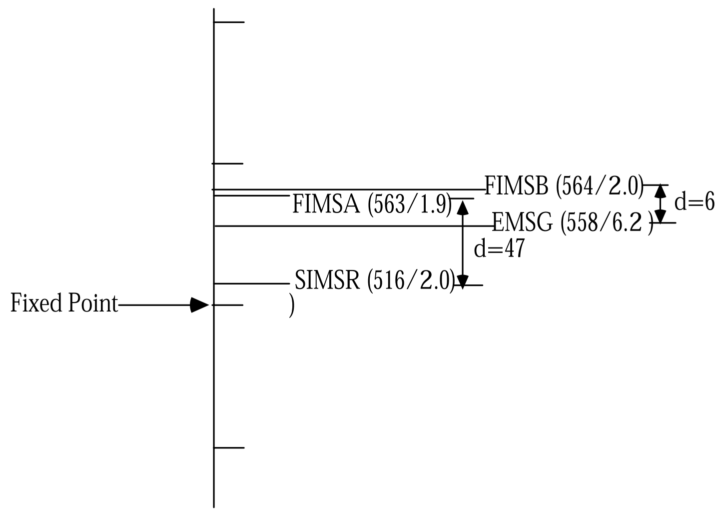
	EMS			**FIMS[a]**		**SIMS**				
	G	Ng	Total	FIMSA	FIMSB	Total	G	Ng	Total	R
mean	558.0	560.0	559.0	563.0	564.0	563.0	516.0	507.0	514.0	516.0
Standard deviation	92.0	89.0	90.0	71.0	72.0	72.0	82.0	82.0	84.0	82.0
SE of the mean	6.2	4.3	3.7	1.9	2.0	1.8	1.8	2.7	1.5	2.0
design effect	2.8	1.4	2.0	2.1	2.4	2.6	1.9	1.2	1.8	1.7
sample size	600	600	1200	2917	3078	4320	3989	1131	5120	3038

effect size		**t-value**	**significant level**
EMS G vs Ng	-0.02	-0.27	NS
EMS G vs FIMSB	-0.07	-0.92	NS
SIMSG vs Ng	0.11	2.77	<0.01
FIMSA vs SIMSR	0.61	17.04	<0.01

a = All students who participated in FIMS were from Government schools; G = Government school students; Ng = Nongovernment school students; R = Groups of SIMS students comparable with FIMSA; NS = not significant; SE = standard error

When the mean scores of the EMS government and nongovernment school students were compared, the nongovernment students' mean score was higher than the government school students by two centilogits and the effect size (-0.02), and the t-value (-0.27) were trivial, and the difference was not statistically significant. Hence, the summary of the findings for the Control of the Environment scale are as follows.

1. The 1964 Australian 13-year-old students expressed more favourable Attitudes towards Control of the Environment than the SIMS students, and the Australian students' Attitudes towards Control of the Environment had declined markedly over time.
2. There was no significant difference between the FIMSB and EMS government school students on their Attitudes towards Control of the Environment scale.
3. The SIMS government school students expressed more favourable Attitudes towards Control of the Environment than the nongovernment school students.
4. There was no significant difference between EMS government and nongovernment school students in their Attitudes towards Control of the Environment.



Fixed Point – 500; Metric - 100 centilogits=1 logit; Values indicated for each occasion in the brackets are Rasch estimated scores and standard errors of the mean respectively;
d= differences in views and attitudes between occasions.

Figure 9.7 The Attitudes towards Control of the Environment Scale of Students in EMS, FIMS and SIMS

Conclusion

The data on the views and attitudes of the EMS, FIMS and SIMS students were equated using the Quest computer program, and the analysis of the concurrent equating showed that 60 of the 64 items fitted the Rasch model. From the results of the Rasch model analysis the FIMS, SIMS and EMS students' views and attitudes towards mathematics and schooling on the seven scales were compared. It should be noted, however, that the SIMS students did not respond to all items and could not be compared to other groups on some scales. However, where only certain items were dropped from within a particular scale, the Rasch scoring procedure permits meaningful scores to be calculated that are independent of the items to which the students responded. The common scales developed for the seven view and attitude scales indicated differences between the groups in their views and Attitudes towards mathematics and schooling.

The results of the analyses of the EMS, FIMS and SIMS data sets indicated that the attitudes of Australian students about the facility of learning mathematics had improved over time, whereas, their attitudes towards school and school learning and control of the environment had declined over time. However, there was no significant change observed between 1964 and 1978 13-year-old Australian students in their views about the teaching of mathematics and in their attitudes towards the place of mathematics in the society. Meanwhile, the views and the attitudes of the Ethiopian students towards mathematics teaching, the place of mathematics in society and attitudes towards school and school learning were more favourable than those of the Australian students. Moreover, the Ethiopian students' attitudes towards mathematics as a process and facility of learning mathematics were low.

A comparison was also undertaken between the government and nongovernment school students in EMS and SIMS. All FIMS students were from government schools, and they were excluded from this comparison. The result of the analysis showed that for SIMS students for all the five view and attitude scales the mean scores of the government school students were higher than those of the nongovernment school students. However, the differences were significant only on Attitudes towards the Facility of Learning Mathematics and Attitudes towards Control of the Environment scales. The government school students were more aware about the importance of school and schooling and were more socially aware of the environmental problems than the nongovernment school students. Whereas, for the Ethiopian students except for Attitudes towards the Place of Mathematics in Society scale the mean scores of the nongovernment school students were higher than those of the government school students. Nevertheless, the difference was significant only on Attitudes towards the Facility of Learning Mathematics scale.

In SIMS the government school students expressed more favourable views and attitudes towards mathematics and school learning than did the nongovernment school students, while in EMS the opposite was true. This shows that the findings in EMS are very different from the findings in SIMS. In SIMS, students in the government schools might have had all kinds of support, which helped them to develop more positive attitudes towards mathematics and schooling. However, in Ethiopia, the students in government schools did not have all the necessary facilities in their schools.

Moreover the students in Ethiopian government schools have the chance to work in schools for only four hours in a day. While students in nongovernment schools have the chance to stay in school for the whole day. This finding is consistent with the findings of Endalkachew (1990) and Tadesse (1993). However, their studies were concentrated on senior secondary school students. Tadesse argued that senior secondary school students had a low level of attitudes towards learning, because the school administration failed to satisfy the needs of the students. Tadesse contended that this led the students to be disinterested in learning and eventually to develop negative attitudes towards school and school learning. Therefore, the differences in the findings between EMS and SIMS students can be considered to have been expected.

10

Sex Differences in Mathematics Achievement, Views and Attitudes Towards Mathematics and Schooling

In the previous two chapters the mathematics achievement levels of students and their views and attitudes towards mathematics and schooling are examined. This chapter investigates the differences in mathematics achievement levels, views and attitudes towards mathematics and schooling between boys and girls on different occasions. Thus, the first part examines sex differences in mathematics achievement on the different occasions, while sex differences in views and attitudes towards mathematics and schooling are investigated in the second part of the chapter.

Sex Differences in Mathematics Achievement

Public opinion in Australia suggests that in Australian schools boys and girls do not have the same performance level in mathematics. The findings indicate that boys perform better than girls in mathematics. However, the differences in achievement between the two sexes only begin to emerge after the lower secondary school level (Carss, 1980; Leder, 1980, 1985, 1989a, 1990; Moss, 1982; Anderson, 1989; Willis, 1989; Ainley et al. 1990; Leder and Forgasz; 1991). A difference between boys and girls in mathematics achievement is also observed outside Australia (Keeves and Kotte, 1994). In Ethiopia research studies have indicated that at all educational levels in both rural and urban areas of the country the achievement of girls in mathematics is much lower than that of boys (Seyoum, 1986; Anbessu and Junge, 1988; Atsede and Kebede, 1988; Tsion, 1990; Assefa, 1991; Gennet, 1991a; Behutiye and Wagner, 1993, 1995; Seleshi, 1995; Yelfign et al., 1995). Thus, in this section the comparisons of the sex differences in mathematics achievement on the four occasions are presented. The strength of these comparisons lies in the use of large, carefully drawn random samples with estimates of sampling and measurement error from which sound generalisations can be derived.

The main purposes of this section are to:

(a) investigate sex differences in achievement in mathematics in 1964, 1978 and 1994 in Australian lower secondary schools,

(b) examine sex differences in achievement in mathematics in Ethiopian lower secondary schools, and

(c) identify the mathematics test items that are biased against boys or girls for the 1964, 1978 and 1994 Australian and 1996 Ethiopian data sets.

Comparison of Sex Difference in EMS, FIMS, SIMS and TIMS

In this section the mathematics achievement levels of boys and girls in EMS, FIMSA, FIMSB, SIMS and TIMS are compared. The first part of this section examines whether or not the mathematics test items show bias against boys or girls. While the next part compares differences in mathematics achievement between the two sexes on the four occasions. These comparisons for Australia in 1964 and 1978, differ from those carried out in previous studies (Keeves, 1968; Moss, 1982) in that proper account can now be taken of the complex design of the samples employed in testing for statistical significance. However, in testing for significant differences, while multiple comparisons are involved no use is made of the Bonferroni Adjustment (Finn, 1997), because the thrust of the comparisons is more toward the detection of no differences, than towards the detection of highly significant ones.

Detection of Item Bias in Mathematics Test Items

The items in the mathematics achievement scale were examined to determine whether the items showed item bias or differential item functioning (DIF) between male and female students. The indices used for the detection of DIF are differences in the difficulty levels, standardised differences in the difficulty levels of items, between male and female students and the infit mean square values of the male and female students for each item for the four occasions.

The analyses identified the items whose differences in level of difficulty indices between male and female students fell outside of the range ± 0.50 and items whose differences in standardised level of difficulty indices between the boys and the girls fell outside of the specified ranges for EMS (± 3.46), FIMS (± 6.57), SIMS (± 7.16), and TIMS were taken as criteria for measuring DIF. The other measure was the presence of an infit mean square which was outside the range between 0.77 and 1.30. Overall 17 items were found as suspect items for DIF. Two items for FIMS and 15 items for TIMS data sets were identified as suspect items when using the level of item difficulty indices between male and female students that fell outside of the range ± 0.50 and adjusted standardised differences outside the specified ranges for the two occasions. Among the 15 suspect items for TIMS data set, Items 51 and 122 were also found as suspect items when infit mean square values outside the range between 0.77 and 1.30 was considered. Hence, it seems possible to conclude that Items 13 and 53 in FIMS and 15 items in TIMS were not functioning equivalently for both male and female students on the mathematics scale. The two suspect items in FIMS were functioning in favour of male students. Whereas, in TIMS seven items were in favour of male students (Items 4, 13, 51, 72, 89, 105 and 116) and the remaining eight items were in favour of female students (Items 28, 55, 68, 117, 122, 133, 139 and 144).

However, all items in EMS and SIMS were functioning equivalently for both male and female students on the mathematics scale.

Out of the 17 biased items nine were in favour of boys and eight items were in favour of girls. Among the nine items that favoured boys, some were geometry and some were measurement items, while the items that favoured girls were related to arithmetic. Previous Australian research findings showed that boys were likely to achieve better in geometry and measurement than girls, while girls were more likely to perform better than boys in arithmetic (Bourke et al., 1981; Moss, 1982; Crawford, 1988; Willis, 1989; Bishop and Clements, 1994). This might be the reason why these items were functioning differently for boys and girls. However, it might not be important to remove these 17 items from the analyses, due to being suspected for DIF, because the items did not show a sign of DIF for all groups. Therefore, the items were not deleted from further analyses.

Sex Difference in EMS, FIMS, SIMS and TIMS

Table 10.1 shows the descriptive statistics of government and nongovernment school and all students who participated in the mathematics tests on the four occasions by sex. The first comparison to be discussed is between Ethiopian Students.

Comparisons of Sex Differences between Ethiopian Students

Three different comparison were considered. The first comparison was between male and female students in government schools. The estimated mean score for the boys was higher than for the girls. The difference was 13 centilogits in favour of the boys (see Table 10.1 and Figure 10.1). The standard deviation and standard error values were slightly higher for the boys, while the design effect was higher for the girls than the boys. The effect size (0.25) and t-values (2.12) showed that the difference was both practically and statistically significant at the 0.05 level. This indicates that boys in government schools were higher achievers in mathematics than their female classmates.

The second comparison was between male and female students in nongovernment schools. The mean score difference between the two sexes was 13 centilogits. The standard deviation value of the boys was higher than that of the girls. Like their government school peers the mean score difference was in favour of the boys. However, the effect size (0.19) and the t-value (0.83) were trivial and the difference was not statistically significant. Nevertheless, the achievement level of the boys was higher than that of the girls by approximately one-third of a year of mathematics learning in Australia.

The third comparison was between boys and girls in all schools. From Table 10.1, it can be seen that the case estimate mean score of EMS male students was 338, while their standard deviation and the standard error values were 66 and 6.1 respectively. Meanwhile, the case estimate mean score of female students was 324 and the standard deviation and the standard error values were 58 and 6.5 respectively. The effect size (0.23) and t-value (1.57) showed that the estimated mean score difference between the two sexes was not statistically significant. However, it would seem important to observe that the achievement level of the boys was higher than that of the girls by approximately two-thirds of a year of mathematics learning in Australia.

The result of the three comparisons showed that the estimated mean score of the boys was higher than that of the girls and in Ethiopian schools the boys achieved better in mathematics than the girls. These findings served to confirm the findings of the

Ethiopian researchers such as Atsede and Kebede (1988), Gennet (1991a, 1991b), Behutiye and Wagner (1993, 1995), Seleshi (1995) and Sewnet (1995). However, the difference in the present study was statistically significant only for government school students. The findings were not significant in nongovernment schools and in all schools. Atsede and Kebede (1988), Gennet (1991a, 1991b), and Sewnet (1995) did not report whether or not the differences between the two sexes were practically or statistically significant, but Behutiye and Wagner (1993) reported a significant difference at the Grade 5 level and in their 1995 study, they reported significant differences from Grades 2 to 5 (Behutiye and Wagner, 1995). In addition Seleshi (1995) reported a significant difference from Grades 8 to 11.

Table 10.1 Descriptive statistics for mathematics achievement of all students for the three occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	325.0	312.0	349.0	336.0	338.0	324.0
Standard Deviation	58.0	46.0	70.0	65.0	66.0	58.0
Jackknife Standard Error	4.8	3.8	10.4	11.8	6.1	6.5
Design Effect	1.8	2.2	6.0	10.8	4.7	8.3
Sample Size	268	332	272	328	540	660

FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	462.0	457.0	455.0	448.0	459.0	453.0
Standard Deviation	101.0	90.0	87.0	76.0	98.0	87.0
Jackknife Standard Error	9.7	7.3	10.1	7.5	9.3	6.5
Design Effect	14.9	9.6	17.3	11.2	20.5	11.4
Sample Size	1530	1386	1619	1462	2275	2044

SIMS	**G**		**Ng**		**Total**		**R**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	442.0	443.0	488.0	464.0	453.0	448.0	440.0	443.0
Standard Deviation	105.0	100.0	117.0	97.0	110.0	99.0	105.0	99.0
Jackknife Standard Error	4.2	5.1	14.9	9.6	5.2	4.6	4.7	5.8
Design Effect	3.4	5.0	9.5	5.4	6.2	5.2	3.2	4.8
Sample Size	2095	1894	580	551	2675	2445	1614	1424

TIMS	**G**		**Ng**		**Total**		**R**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	423.0	431.0	488.0	474.0	443.0	448.0	422.0	431.0
Standard Deviation	125.0	124.0	117.0	110.0	126.0	121.0	126.0	126.0
Jackknife Standard Error	7.7	8.4	10.2	8.4	7.0	6.4	8.4	9.2
Design Effect	9.4	9.9	8.5	9.6	11.1	10.7	8.9	9.4
Sample Size	2479	2168	1111	1633	3590	3801	2030	1755

EMS G Male vs Female	13	0.25	2.12	<0.05
EMS NG Male vs Female	13	0.19	0.83	NS
EMS Male vs Female	14	0.23	1.57	NS
FIMSA Male vs Female	5	0.05	0.41	NS
FIMSB Male vs Female	7	0.09	0.56	NS
FIMS Male vs Female	6	0.06	0.53	NS
SIMS G Male vs Female	-1	-0.01	-0.15	NS
SIMS NG Male vs Female	24	0.22	1.35	NS
SIMS Male vs Female	5	0.05	0.72	NS
SIMSR Male vs Female	-3	-0.03	-0.40	NS
TIMS G Male vs Female	-8	-0.06	-0.70	NS
TIMS NG Male vs Female	14	0.12	1.06	NS
TIMS Male vs Female	-5	-0.04	-0.52	NS
TIMSR Male vs Female	-9	-0.07	-0.72	NS

G=Government school students; NG=Nongovernment school students; R=Students in SIMS which are comparable with FIMSA and students in TIMS which are comparable with FIMSB; NS=Not Significant

However, the present study did not fully support the significant differences reported by these earlier research workers. If the sex difference in favour of boys was significant at the Grade 2 level, the expectation is that at Grade 8 the sex difference would be highly significant. However, the sex difference at the Grade 8 level in the present study shows that only a marginally significant difference emerged in government schools. The findings reported from cross-national research would suggest that clear sex differences in mathematics achievement are not found in the lower secondary schools of most countries (Keeves and Kotte, 1994). Thus, this study would seem to show that sex differences in mathematics achievement would begin to emerge in Ethiopian junior secondary schools. A statistically significant difference was found only in government schools, although the size of the recorded difference would seem to be of practical significant for all schools. Thus, further investigation would seem important to locate the level at which sex differences begin to emerge in Ethiopian schools.

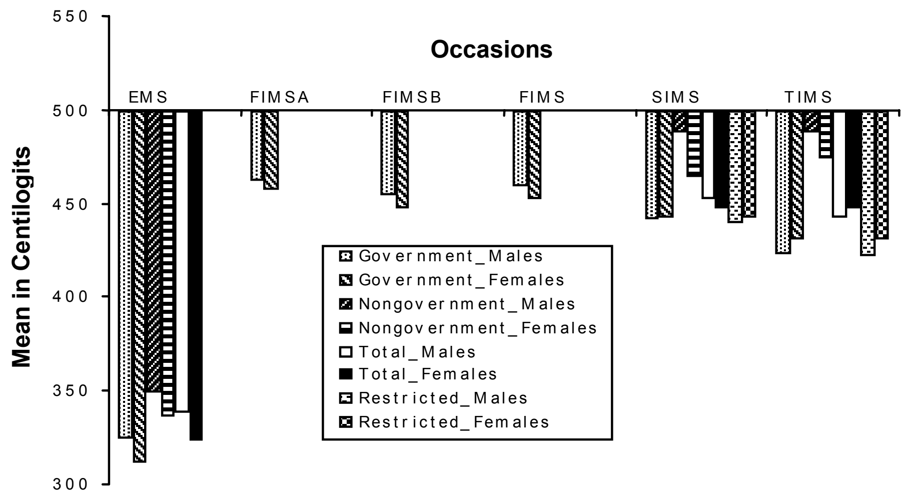


Figure 10.1 The Mathematics test scale of Government and Nongovernment school students in EMS, FIMS, SIMS and TIMS by sex

Comparisons of Sex Differences between 1964 Australian Students

Table 10.1 and Figure 10.1 show the three comparisons which were considered in FIMS. The estimated mean score differences on the three comparisons were in favour of boys. This shows that boys achieved higher than girls. However, the effect size and t-values were trivial, thus, in all the comparisons the differences were neither practically nor statistically significant.

Comparisons of Sex Differences between 1978 Australian Students

In the SIMS data set four comparisons were undertaken. The comparisons were between boys and girls in government schools, nongovernment schools, in all schools, and the restricted sample of schools.

The mean score differences between the two sexes in government schools (both Government and Restricted) indicated that girls achieved at a higher level than boys. However, the differences were neither practically nor statistically significant since the effect size and t-values were too small to be considered significant.

Furthermore, the estimated mean score difference between male and female nongovernment school students in SIMS was in favour of the boys, although the effect size (0.22) and t-value (1.35) were small. Thus, the difference was not statistically significant. Nevertheless, the achievement level of the boys was higher than that of the girls by approximately two-thirds of a year of mathematics learning in Australian schools.

The last comparison in SIMS was between boys and girls in all schools. The mean score difference was in favour of the boys. However, the effect size (0.05) and t-value (0.72) were very small. Hence, the difference between boys and girls in 1978 was not practically or statistically significant across Australian schools.

Out of the four comparisons, only in government schools was the achievement of the girls slightly higher than that of the boys, in the remaining comparisons boys achieved better than girls. These findings were not similar to the findings in FIMS. In the 1964 data set all the differences were in favour of boys. This might indicate that on the 14-year period there was a small shift in achievement level differences between boys and girls. However, all the differences were not significant. The absence of sex difference at the 13-year-old students in SIMS is supported by previous studies. Hanna and Kundiger (1986) reported that most sex differences were not statistically significant at the 0.01 level. Ethington (1990) and Xin Ma (1995) claimed that the sex differences were weak. The next comparisons might provide evidence as to whether the shift continued for the 30-year period or not, since the remaining comparisons are between boys and girls in TIMS.

Comparisons of Sex Differences between 1994 Australian Students

The estimated mean score differences between the two sexes in government schools (both Government and Restricted) were in favour of girls (see Table 10.1 and Figure 10.1). However, the effect size and t-values were very small and the mean differences were neither practically nor statistically significant.

The other comparison was between the two sexes in nongovernment schools. Unlike the government school students the difference was in favour of boys. However, the difference was not statistically significant although the effect size was small (0.12) the t-values was trivial, because the design effects were large.

The last comparison in TIMS considered all students (government and nongovernment together). The mean score difference was in favour of girls. However, the effect size (-0.04) and t-values (-0.52) were too small to be consider significant. Consequently the mean difference was neither practically nor statistically significant.

Three of the four different comparisons in TIMS between boys and girls in mathematics achievement revealed that the mathematics achievement levels of the girls were slightly better than that of the boys. When these findings are compared with the findings in FIMS and SIMS, the achievement level of girls changed relative to those of the boys. In 1964 the differences were in favour of the boys, while in 1978 the differences were in favour of the girls only in government schools. Furthermore, in 1994 the differences, except in nongovernment schools, were in favour of the girls. However, the differences were not found to be statistically significant. These insignificant differences might suggest that a sex difference is starting to emerge at this stage in favour of girls in contrast to the findings of Keeves (1972), Carss (1980), Moss (1982), Anderson (1989), Ainley et al. (1990), Leder and Forgasz (1991). These researchers have argued that sex difference in mathematics in Australian schools had started to emerge at the junior secondary school stage in favour of boys. The changes might have been a result of the implementation of different government policies to increase the participation and the mathematics achievement level of girls by States and Federal Governments. Alternatively, it is possible that the level of performance of the boys has declined more than that of the girls over time since there is a general decline in achievement of all Australian students over the 30-year period. Tilahun and Keeves (2001) reported that the decline in mathematics achievement over this 30 year period for boys is equivalent to nearly one year of mathematics learning, while the drop for girls is only approximately equivalent to half a year of mathematics learning.

Sex Differences in Views and Attitudes towards Mathematics and Schooling

Research studies in Australia have revealed that boys and girls do not have the same attitudes towards school subjects and schooling. The findings of previous research have indicated that, girls have more favourable attitudes towards school than boys (Keeves, 1972). Whereas boys show more favourable attitudes towards mathematics (Keeves, 1972; Fraser, 1980; Leder, 1989b) than girls. Similar results were observed in Ethiopia. Derese et al. (1990), and Seleshi (1995) have argued that boys expressed more positive attitudes towards mathematics than girls. However, these studies were conducted at different times using different instruments and different samples, so it would seem difficult to generalise about the attitudes of boys and girls towards mathematics and schooling. In order to generalise about the attitudes of boys and girls towards schooling and mathematics, it would seem meaningful to analyse data which were collected on different occasions using the same instrument and for the same age group or at the same year level of students. Since, Australia participated in the 1964 and 1978 International Mathematics Studies, it was possible to examine the views and attitudes of boys and girls towards mathematics and schooling over the 14-year time period. The data collected in Ethiopia could also be used to compare the attitudes of the two sexes in Ethiopian lower secondary schools.

Hence, in this section the results of the analyses of the sex differences in the seven view and attitude scale scores for EMS, FIMS and SIMS students are discussed. The first part of the analysis examines whether or not the view and attitude statements show DIF between boys and girls. The second section compares sex differences in views and attitudes towards mathematics and schooling on the three occasions.

Detection of DIF in View and Attitude Scales

The seven view and attitude scales were examined to determine whether the items showed DIF between male and female students. The indices used for the detection of DIF were differences and standardised differences in the levels of difficulty of each item between male and female students and the infit mean square values of the male and female students for each item (see Chapter 7 for the detail about DIF

The analyses showed that the items whose differences in adjusted standardised level of difficulty indices between the boys and the girls fell outside of the range $\pm$9.97 for Views about Mathematics Learning scale, $\pm$6.68 for the Views about School and School Learning scale, $\pm$7.43 for Attitudes towards Mathematics as a Process scale and $\pm$10.32 for Attitudes towards Facility of Learning Mathematics, Attitudes towards the Place of Mathematics in Society, Attitudes towards School and School Learning and Attitudes towards Control of the Environment scales. These values were taken as a criteria for measuring DIF. The other test was whether the infit mean square values fell outside the range between 0.77 and 1.30. However, no item was found as a suspect item for DIF. Hence, it would seem possible to conclude that all items were functioning equivalently for both male and female students on all the seven view and attitude scales.

Sex Differences in Attitudes towards Mathematics and School Over Time

The result of the detection of item bias showed that no item was suspect for DIF. Hence, after confirming that there were no view and attitude statements which indicated DIF, it was meaningful to examine the data set, to determine whether or not sex differences appeared on the seven view and attitude scales in EMS, FIMS and SIMS. Therefore, in this section the views and attitudes of boys and girls towards mathematics and schooling are compared on the three occasions in order to investigate whether sex differences in attitudes towards mathematics and schooling are beginning to emerge at the junior secondary school level in Australia and Ethiopia.

Sex Differences in Views and Attitudes towards Mathematics

The views and attitudes of boys and girls are compared on four view and attitude towards mathematics scales. The first scale was Views about Mathematics Teaching, and the second scale was Attitudes towards Mathematics as a Process. The third scale was Attitudes towards Facility of Learning Mathematics and the fourth scale was Attitudes towards the Place of Mathematics in Society.

Sex Differences in Views about Mathematics Teaching Scale

Table 10.2 shows the descriptive statistics with respect to sex of student of government and nongovernment school students who responded for the Views about Mathematics Teaching statements, and Figure 10.2 compares the estimated mean scores of these students on the three occasions.

The estimated mean scores of EMS total male and female students were 572 and 581, respectively (see Table 10.2 and Figure 10.2). The estimated mean score of the female students was higher by nine centilogits than that of the male students. The effect size (-0.14) and the t-value (-1.61) revealed that the mean difference between male and female EMS students was neither practically nor statistically significant.

When the government school male and female students' estimated mean scores were compared, the mean score of the females was higher than that of their male classmates. However, the difference was not statistically significant. The effect size and t-value were -0.05 and -0.34 respectively. Meanwhile, the estimated mean score difference between nongovernment school male and female students was 14 centilogits in favour of the females. The effect size and t-value were -0.23 and -2.01 respectively. Consequently, the difference was marginally significant at the 0.05 level. The EMS female students tended to view mathematics more as something that was changing, developing and growing, while their male classmates considered mathematics as more fixed and given once and for all time. However, the significance difference was only between nongovernment school students.

When the case estimated mean scores of FIMS male and female students were considered, the case estimated mean score of the females was higher than that of their male counterparts. The mean scores of the male and female students were 507 and 518 respectively (see Table 10.2 and Figure 10.2). The effect size and t-value were -0.21 and -1.54 respectively. Consequently, the difference between male and female students in Views about Mathematics Teaching scale was not statistically significant. When the mean scores of boys and girls at the 13-year-old level (FIMSA) and all students at the Year 8 level (FIMSB) were compared the mean differences in both cases were in favour of the girls. The effect size and t-value for FIMSA were -0.17 and -1.31 respectively. Thus, the difference was not statistically significant.

Meanwhile, the effect size and t-value for FIMSB were -0.19 and -1.26 respectively, hence, this difference was also not statistically significant.

The estimated mean score of government school male and female students in SIMS was also considered. Unlike the EMS and FIMS students the difference here was in favour of the boys. The difference was five centilogits. The effect size (0.05) and the t-value (1.18) were very small. Hence, the difference was not statistically significant. The difference in estimated mean scores between the two sexes in nongovernment schools in SIMS was also considered. The mean difference was two centilogits in favour of the girls. However, the effect size (-0.02) and the t-value (-0.24) were trivial, so the difference was neither practically nor statistically significant. In addition, when the mean scores of the two sexes in SIMS total were compared, the difference was three centilogits in favour of the boys. The effect size and t-values were very small (see Table 10.2), and the difference was neither practically nor statistically significant.

Table 10.2 Descriptive statistics for Views about Mathematics Teaching scale of students on the three occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	571.0	574.0	573.0	587.0	572.0	581.0
Standard Deviation	64.0	66.0	62.0	59.0	63.0	63.0
Jackknife Standard Error	6.5	6.1	5.4	4.4	4.1	3.8
Design Effect	2.8	2.8	2.1	1.8	2.3	2.5
Sample Size	268	332	272	328	540	660
FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	506.0	515.0	507.0	517.0	507.0	518.0
Standard Deviation	52.0	53.0	52.0	53.0	51.0	53.0
Jackknife Standard Error	4.5	5.2	4.7	6.4	4.4	5.6
Design Effect	9.3	10.4	10.3	16.5	13.0	17.8
Sample Size	1144	1042	1248	1102	1726	1534

SIMS	**G**		**Ng**		**Total**		**R**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	512.0	507.0	503.0	505.0	510.0	507.0	512.0	506.0
Standard Deviation	88.0	94.0	94.0	97.0	90.0	95.0	87.0	94.0
Jackknife Standard Error	2.5	3.4	5.9	5.7	2.3	2.9	3.0	4.6
Design Effect	1.6	2.5	2.3	1.9	1.8	2.3	1.6	2.6
Sample Size	2095	1894	580	551	2675	2445	1302	1093

	Mean difference	**Effect size**	**t-value**	**Significance level**
EMS G Male vs Female	-3.0	-0.05	-0.34	NS
EMS NG Male vs Female	-14.0	-0.23	-2.01	<0.05
EMS Male vs Female	-9.0	-0.14	-1.61	NS
FIMSA Male vs Female	-9.0	-0.17	-1.31	NS
FIMSB Male vs Female	-10.0	-0.19	-1.26	NS
FIMS Male vs Female	-11.0	-0.21	-1.54	NS
SIMS G Male vs Female	5.0	0.05	1.18	NS
SIMS NG Male vs Female	-2.0	-0.02	-0.24	NS
SIMS Male vs Female	3.0	0.03	0.81	NS
SIMSR Male vs Female	6.0	0.07	1.09	NS

G= government school students; NG = Nongovernment school students; R = Those group of students in SIMS which are comparable with FIMSA; NS = Not significant

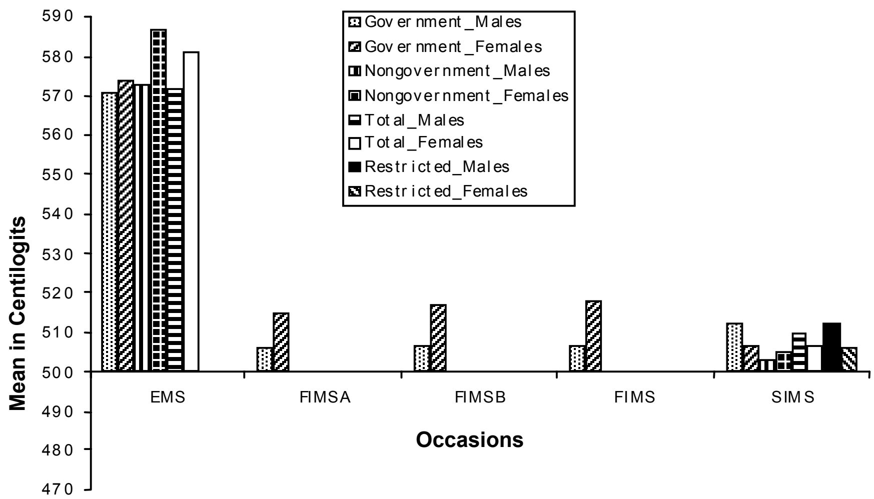


Figure 10.2 Sex differences on Views about Mathematics Teaching on three occasions

Thus the findings can be summarised as follows.

1 EMS female students showed more positive views about mathematics teaching than their male classmates. However the difference was only statistically significant for the nongovernment school students.

2. Sex differences were not found between FIMS male and female students, with respect to their views about mathematics teaching.

3. There was no significant differences between the male and female 13-year-old Australian students in 1978 with respect to their views about mathematics teaching.

Sex Differences in Attitudes towards Mathematics as a Process

Table 10.3 shows the descriptive statistics of government and nongovernment school students who responded for the Attitudes towards Mathematics as a Process statements, and Figure 10.3 compares the estimated mean scores of the EMS and FIMS students. Because this scale was excluded from SIMS, the comparisons were only available for EMS and FIMS students.

On the Attitudes towards Mathematics as a Process scale Ethiopian female students expressed less favourable attitudes than their male classmates. In government schools the mean score difference between male and female students was seven centilogits, and the difference was in favour of the male students. The effect size (0.13) and t-value (1.18) were very small. Thus this difference was not significant.

Meanwhile in nongovernment schools the mean difference was 11 centilogits in favour of the males. The effect size (0.21) was small and t-value was 2.01. Hence, the difference was statistically significant at the 0.05 level. When the overall EMS students' mean scores were considered, the sex difference was nine centilogits. The difference was also in favour of the males. The effect size (0.17) was small and t-value was 2.23 (see Table 10.3 and Figure 10.3). Thus this difference was significant at the $p<0.05$ level. Consequently, the male students held attitudes that mathematics was a subject that was still in a process of change and development and involved the understanding of mathematical phenomena rather than mechanical application of formulae and rules, compared to the female students who held attitudes that mathematics was more a fixed, formal system governed by rigid and unchanging rules which they had to master. However, even the attitudes of the male students were not clearly favourable in this respect since the estimated mean score was lower than the mean level of the items (see Table 10.3 and Figure 10.3).

Table 10.3 Descriptive statistics for Attitudes towards Mathematics as a Process scale of students on the two occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	472.0	465.0	480.0	469.0	476.0	467.0
Standard Deviation	53.0	58.0	52.0	52.0	53.0	55.0
Jackknife Standard Error	4.0	4.4	4.3	3.4	3.0	2.7
Design Effect	1.5	1.9	1.9	1.5	1.7	1.6
Sample Size	268	332	272	328	540	660
FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	485.0	490.0	483.0	490.0	485.0	490.0
Standard Deviation	61.0	58.0	61.0	57.0	60.0	57.0
Jackknife Standard Error	2.5	2.6	1.7	2.8	1.9	2.7
Design Effect	2.6	2.9	1.3	3.6	2.4	4.7
Sample Size	1530	1386	1619	1462	2275	2044

	Mean difference	**Effect size**	**t-value**	**Significance Level**
EMS G Male vs Female	7.0	0.13	1.18	NS
EMS NG Male vs Female	11.0	0.21	2.01	<0.05
EMS Male vs Female	9.0	0.17	2.23	<0.05
FIMSA Male vs Female	-5.0	-0.08	-1.39	NS
FIMSB Male vs Female	-7.0	-0.12	-2.14	<0.05
FIMS Male vs Female	-5.0	-0.09	-1.51	NS

G= government school students; NG = Nongovernment school students; NS = Not significant

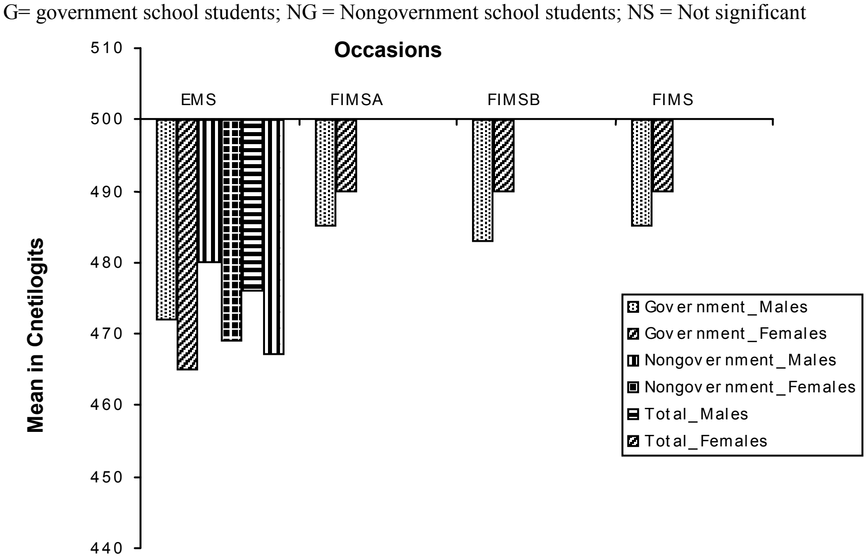


Figure 10.3 Sex differences in Attitudes towards Mathematics as a Process on two occasions

The mean difference between 13-year-old male and female students in FIMSA was five centilogits in favour of the girls. The effect size and t-values were -0.08 and -1.39

respectively. Consequently, the difference was not statistically significant. Meanwhile, when the mean difference between Year 8 male and female students in FIMSB was considered the difference was seven centilogits in favour of the girls. The effect size and t-values were -0.12 and -2.14 respectively. Hence, the difference was statistically significant at the 0.05 level. Furthermore, the mean difference between the total FIMS male and female students in their attitudes towards mathematics as a process was five centilogits (see Table 10.3 and Figure 10.3). The effect size was -0.09 and the t-value was -1.51. Consequently, this difference was neither practically nor statistically significant. Hence, there was no sex difference in attitudes towards mathematics as a process for FIMS students as a combined group.

The findings may be summarised as follows.

1. Ethiopian male students had more favourable attitudes towards mathematics as a process than the female students. However, the difference was not significant for government school students, but was significant for the nongovernment school students and the total group of students.
2. Year 8 female Australian students expressed more favourable attitudes towards mathematics as a process than the male students.

Sex Differences in Attitudes towards Facility of Learning Mathematics

Table 10.4 shows the descriptive statistics for the sex differences for government and nongovernment school students who responded for the Attitudes towards Facility of Learning Mathematics scale, and Figure 10.4 compares the estimated mean scores of the male and female students.

On this scale all groups, the EMS, FIMS and SIMS students, gave their responses, hence the comparisons were for the three groups of students. The scale was considered as a measure of students' attitudes towards the facility or difficulty of learning mathematics. The estimated mean scores of Ethiopian government school male and female students were 584 and 579 (see Table 10.4 and Figure 10.4) respectively. The mean difference between the two sexes was five centilogits, in favour of the boys. The effect size (0.04) and the t-value (0.34) were trivial and this difference was not significant. Whereas the mean score difference between male and female nongovernment school EMS students was in favour of females and the mean difference was five centilogits. However, this difference was also not statistically significant. Furthermore, the estimated mean score difference between the total EMS male and female students groups was 0.00. Hence, the difference was not significant. Thus, there was no significant difference between Ethiopian male and female students in their attitudes towards the difficulty of learning mathematics. The findings are not consistent with the findings of Derese et al. (1990), and Seleshi (1995). These researchers have argued that boys expressed more favourable attitudes towards mathematics than their female peers. However, in the present study the differences were not statistically significant, when attitudes towards the difficulty or facility of learning mathematics was considered.

Furthermore, male and female 13-year-old students in FIMSA had mean estimated scores of 636 and 629 respectively. Thus the mean difference between the two sexes was seven centilogits. The effect size was 0.07 and the t-value was 1.13, and the difference was neither practically nor statistically significant. Whereas the estimated mean score difference between Year 8 FIMSB male and female students was nine centilogits. The effect size and t-values were 0.10 and 1.73 respectively. Consequently, the difference was statistically significant at the ten per cent level. Meanwhile, when the mean difference of all male and female students in FIMS was considered, the estimated mean score of the males was slightly higher than that of the females. The difference was seven centilogits. This difference was not statistically significant. Thus, it would seem possible to conclude, that in 1964 boys expressed more positive attitudes towards the facility of learning mathematics, but a statistically significance difference was identified only between Year 8 students. The findings are not consistent with the findings of Keeves (1972), Fraser (1980) and Leder (1989b). These researchers have argued that boys expressed more favourable attitudes towards mathematics than their female peers. However, in the present study the difference was significant only at the Year 8 level, when attitudes towards the difficulty or facility of learning mathematics were considered.

Table 10.4 Descriptive statistics for Attitudes about the Facility of Mathematics scale of students on the three occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	584.0	579.0	606.0	611.0	595.0	595.0
Standard Deviation	115.0	111.0	113.0	103.0	115.0	108.0
Jackknife Standard Error	12.1	8.2	10.5	9.2	8.0	6.6
Design Effect	3.0	1.8	2.4	2.6	2.6	2.5
Sample Size	268	332	272	328	540	660

FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	636.0	629.0	634.0	625.0	629	622.0
Standard Deviation	94.0	93.0	95.0	93.0	95	94.0
Jackknife Standard Error	3.7	5.0	2.6	4.5	2.4	4.5
Design Effect	2.3	4.0	1.2	3.4	1.5	4.7
Sample Size	1530	1386	1619	1462	2275	2044

SIMS	**G**		**Ng**		**Total**		**R**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	641.0	645.0	622.0	641.0	637.0	644.0	641.0	646.0
Standard Deviation	106.0	105.0	109.0	111.0	107.0	106.0	106.0	105.0
Jackknife Standard Error	2.5	3.1	9.6	8.3	3.1	3.0	2.7	3.4
Design Effect	1.1	1.6	4.6	3.1	2.2	1.9	1.1	1.5
Sample Size	2095	1894	580	551	2675	2445	1614	1424

	Mean difference	**Effect size**	**t-value**	**Significance Level**
EMS G Male vs Female	5.0	0.04	0.34	NS
EMS NG Male vs Female	-5.0	-0.05	-0.36	NS
EMS Male vs Female	0.0	0.00	0.00	NS
FIMSA Male vs Female	7.0	0.07	1.13	NS
FIMSB Male vs Female	9.0	0.10	1.73	<0.10
FIMS Male vs Female	7.0	0.07	1.37	NS
SIMS G Male vs Female	-4.0	-0.04	-1.00	NS
SIMS NG Male vs Female	-19.0	-0.17	-1.50	NS
SIMS Male vs Female	-7.0	-0.07	-1.62	NS
SIMSR Male vs Female	-5.0	-0.06	-1.16	NS

G= government school students; NG = Nongovernment school students; R = Those group of students in SIMS which are comparable with FIMSA; NS = Not Significant

In SIMS the estimated mean scores of government school male and female students were 641 and 645 respectively, and the mean difference between two sexes was four centilogits. The effect size (-0.04) and the t-value (-1.00) were trivial and the difference between the two sexes was not significant. The same was true when the mean estimated scores of nongovernment school males and females were compared, while, the difference was 19 centilogits, the effect size (-0.17) and the t-value (-1.50) were too small to be found both practically and statistically significant partly because of the size of the design effects involved. Nevertheless, the somewhat larger effect size, while still less than 0.20, indicated that girls in the nongovernment schools would seem to hold more favourable attitudes towards the facility of learning mathematics.

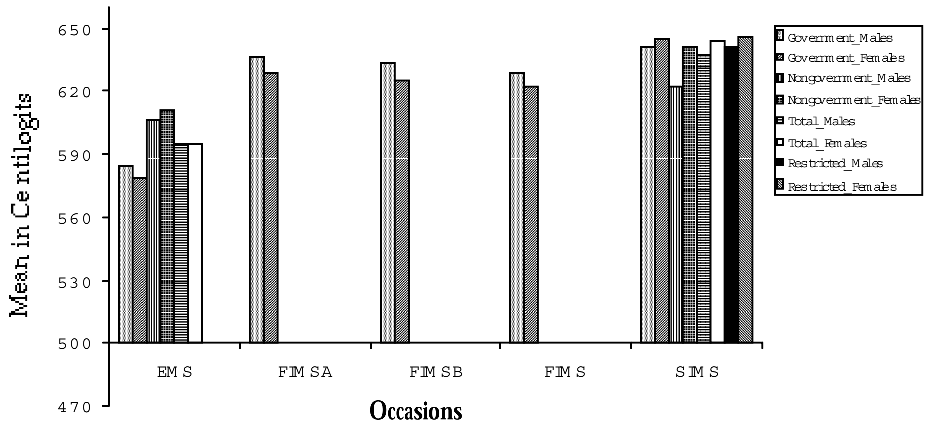


Figure 10.4 Sex differences in Attitudes towards Facility of Learning Mathematics on three occasions

When the estimated mean scores of SIMSR males and females are compared, the difference was five centilogits also in favour of the females. The effect size (-0.06) and the t-value (-1.16) were very small and the difference between the two sexes was not significant. Furthermore, when the total SIMS male and female students estimated mean scores were compared, the mean score of the females' attitudes was higher by seven centilogits than the males' attitudes. The effect size (-0.07) and the t-value (-1.62) were too small and the difference between the two sexes was not significant. Consequently, there was no sex difference identified between SIMS students for their attitudes towards the facility of learning mathematics. The findings in SIMS was consistent with the findings in FIMSA (13-year-olds). On both occasions, significant differences were not found between the two sexes. However, it is important to observe that in 1964 in Australia the mean score of boys was generally higher than that of the girls, whereas in 1978, the mean score of the girls was generally higher than that of the boys. This suggests a shift in attitudes towards the difficulty of learning mathematics from boys to girls over the 14-year period, but the shift did not involve statistically significant differences.

A summary of the findings is as follows.

1. There was no significant sex difference between Ethiopian students in their attitudes towards facility of learning mathematics.
2. FIMS Year 8 male students expressed more favourable attitudes towards facility of learning mathematics than did the female students.
3. Sex differences were not found between SIMS students in their attitudes towards facility of learning mathematics.

Sex Differences in Attitudes towards the Place of Mathematics in Society

EMS, FIMS and SIMS students responded to this scale. Table 10.5 and Figure 10.5 show that the estimated mean scores of both sexes on the three occasions were above the mean threshold level of the items. This suggests that all groups of students expressed positive attitudes towards the place of mathematics in society. The first comparison was between the EMS government school male and female students. The estimated mean score for the boys was 588 and for the girls was 586. The mean difference indicated that the males had marginally higher mean scores than the female students (see Table 10.5 and Figure 10.5). But, this difference was not statistically significant. Meanwhile, the estimated mean score difference between EMS nongovernment school female and male students revealed that the females' mean score was higher than that of the male students by 14 centilogits (see Table 10.5 and Figure 10.5). The effect size (-0.13) and t-value (-1.06) were too small to be considered practically or statistically significant. Furthermore, when the total EMS male and female students estimated mean scores were considered, the mean score of the female students was higher than that of the male students. However, the effect size (-0.05) and t-value (-0.67) were very small. Consequently, it could be concluded that there were no sex differences in Attitudes towards the Place of Mathematics in Society among Ethiopian students.

Table 10.5 Descriptive statistics for Attitudes towards the Place of Mathematics in Society scale of students on the three occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	588.0	586.0	578.0	592.0	583.0	589.0
Standard Deviation	119.0	107.0	120.0	97.0	119.0	102.0
Jackknife Standard Error	10.4	7.6	11.3	6.8	7.5	5.0
Design Effect	2.0	1.7	2.8	1.6	2.1	1.6
Sample Size	268	332	272	328	540	660
FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	573.0	552.0	571.0	553.0	569.0	552.0
Standard Deviation	75.0	79.0	74.0	76.0	74.0	78.0
Jackknife Standard Error	3.2	3.5	3.5	3.6	2.9	3.5
Design Effect	2.8	2.5	3.6	3.3	3.4	4.0
Sample Size	1530	1386	1619	1462	2275	2044

SIMS	**G**		**Ng**		**Total**		**R**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	564.0	552.0	554.0	560.0	562.0	554.0	566.0	554.0
Standard Deviation	102.0	103.0	104.0	100.0	102.0	102.0	102.0	103.0
Jackknife Standard Error	3.0	3.4	7.0	5.1	2.8	2.9	3.3	3.8
Design Effect	1.8	2.1	2.6	1.4	2.0	2.0	1.7	2.0
Sample Size	2095	1894	580	551	2675	2445	1614	1424

	Mean difference	**Effect size**	**t-value**	**Significance Level**
EMS G Male vs Female	2.0	0.02	0.16	NS
EMS NG Male vs Female	-14.0	-0.13	-1.06	NS
EMS Male vs Female	-6.0	-0.05	-0.67	NS
FIMSA Male vs Female	21.0	0.27	4.43	<0.01
FIMSB Male vs Female	18.0	0.24	3.58	<0.01
FIMS Male vs Female	17.0	0.22	3.74	<0.01
SIMS G Male vs Female	12.0	0.12	2.65	<0.01
SIMS NG Male vs Female	-6.0	-0.06	-0.69	NS
SIMS Male vs Female	8.0	0.08	1.98	<0.05
SIMSR Male vs Female	12.0	0.12	2.38	<0.05

G= government school students; NG = Nongovernment school students; R = Those group of students in SIMS which are comparable with FIMSA; NS = Not Significant

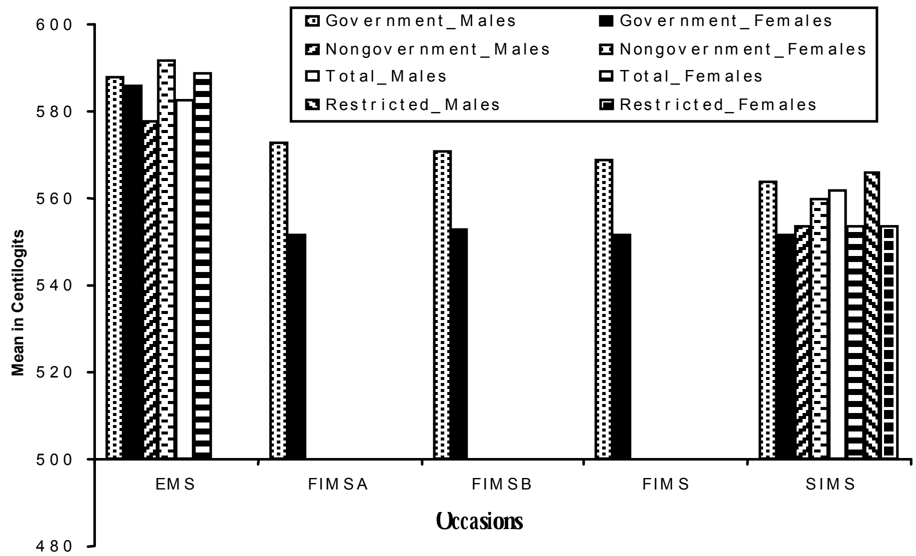


Figure 10.5 Sex differences in Attitudes towards the Place of Mathematics in Society on the three occasions

The comparison between FIMSA male and female students revealed that the estimated mean score of male students was higher by 21 centilogits than that of the female students and that difference was significant at the p<0.01 level (see Table 10.5 and Figure 10.5). Similar results were observed between FIMSB male and female students. The comparison between FIMSB female and male students indicated that the estimated mean score of the males was higher than the female students' mean score by 18 centilogits, and the difference was statistically significant at the p<0.01 level. Furthermore, the comparison between the total FIMS male and female students revealed that the estimated mean score of male students was higher by 17 centilogits than that of the female students and that difference was also significant at the p<0.01 level. Therefore, in FIMS boys expressed more favourable Attitudes towards the Place of Mathematics in Society than their female counterparts.

The estimated mean scores for the SIMS government school male and female students were 564 and 552 respectively. The difference was only 12 centilogits in favour of the male students (see Table 10.5 and Figure 10.5). However, this difference was statistically significant at the p<0.01 level. Similar results were observed between SIMSR male and female students. Their mean difference was 12 centilogits, and the effect size and t-value were 0.12 and 2.38 respectively. Hence, the difference was statistically significant at the p<0.05 level. However, the mean difference between nongovernment school male and female students was six centilogits in favour of female students. The effect size (-0.06) and t-value (-0.69) were trivial. Hence, the difference was not statistically significant.

Meanwhile, the comparison between the total SIMS male and female students revealed that the mean estimated score of male students was higher by eight centilogits than that of the female students and that difference was significant at the 0.05 level. Consequently, in SIMS, boys expressed more favourable Attitudes towards the Place of Mathematics in Society scale than their female counterparts, except for students in nongovernment schools.

The important point to be observed here was that both in 1964 and 1978 at the 13-year-olds level in Australia boys expressed more favourable attitudes than the girls and also at the Year 8 level boys expressed more positive attitudes than girls in FIMSB and EMS government schools. However, the differences were not statistically significant in Ethiopia.

A summary of the findings are as follows.

1. There were no significant sex differences between Ethiopian male and female students in attitudes towards the place of mathematics in society.
2. Male FIMS students expressed more favourable attitudes towards the place of mathematics in society than female students.
3. In SIMS male students expressed more favourable attitudes towards the place of mathematics in society than female students, except in nongovernment schools.

Sex Differences in Views and Attitudes towards School and School Learning

The views and attitudes of female and male students towards school and school learning are compared on three view and attitude scales. The first scale was Views about School and School Learning, the second scale was Attitudes towards School and School Learning. While, the third scale was Attitudes towards Control of the Environment scale.

Sex Differences in Views about School and School Learning

Table 10.6 shows the descriptive statistics for government and nongovernment school students who responded for the Views about School and School Learning statements, and Figure 10.6 compares the estimated mean scores of EMS and FIMS students. Because the students in SIMS did not respond to this scale, the comparisons were only available for EMS and FIMS students.

Table 10.6 Descriptive statistics for Views about School and School Learning scale of students on the two occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	505.0	501.0	503.0	504.0	504.0	503.0
Standard Deviation	34.0	32.0	32.0	31.0	33.0	31.0
Jackknife Standard Error	4.0	2.5	2.3	2.6	1.8	1.8
Design Effect	1.8	2.1	1.4	2.4	1.5	2.2
Sample Size	268	332	272	328	540	660
FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	505.0	509.0	502.0	505.0	503.0	507.0
Standard Deviation	44.0	39.0	41.0	43.0	43.0	43.0
Jackknife Standard Error	2.8	1.8	2.6	2.0	2.6	1.5
Design Effect	4.3	1.8	4.9	2.5	6.2	2.0
Sample Size	1144	1042	1248	1102	1726	1534

	Mean difference	**Effect size**	**t-value**	**Significance Level**
EMS G Male vs Female	4.0	0.12	0.85	NS
EMS NG Male vs Female	-1.0	-0.03	-0.29	NS
EMS Male vs Female	1.0	0.03	0.39	NS
FIMSA Male vs Female	-4.0	-0.10	-1.20	NS
FIMSB Male vs Female	-3.0	-0.07	-0.91	NS
FIMS Male vs Female	-4.0	-0.09	-1.33	NS

G= government school students; NG = Nongovernment school students; NS = Not significant

The mean difference between government school male and female students in EMS was four centilogits (see Table 10.6 and Figure 10.6). The difference was in favour of boys. However, the effect size (0.12) and the t-value (0.85) were very small and the difference was not statistically significant. Moreover, the estimated mean score difference between the two sexes in Ethiopian nongovernment schools was in favour of the girls. The difference was one centilogit. Thus, the effect size (-0.03) and the t-value (-0.29) were too small to be considered significant. In addition, the mean score difference between boys and girls in EMS total was only one centilogit in favour of boys. Consequently, this difference was not statistically significant. Therefore, in this analysis it can be concluded that in Ethiopian junior secondary schools boys and girls expressed similar views about their schools and school learning.

When the FIMSA 13-year-old male students' mean score was compared with the estimated mean score of their female peers, the latter scored higher than the former. The mean score difference between the two sexes was four centilogits. However, the effect size (-0.10) and t-value (-1.20) were trivial and the difference was not significant.

In addition, when the mean score difference between the two sexes in Year 8 FIMSB was compared, the mean score of the girls was higher than that of the boys. Like the 13-year-olds, the effect size (-0.07) and t-value (-0.91) were very small and the difference was not statistically significant. Furthermore, when the mean score difference between boys and girls in the FIMS total were compared, the difference was four centilogits in favour of the girls. However, the effect size (-0.09) and t-value (-1.33) were not statistically significant. Hence, in FIMS, there were no significant differences between boys and girls in their views about their schools and school learning.

Thus the findings in this section are summarised below.

1. Sex difference did not exist between Ethiopian Year 8 students in their views about school and school learning.
2. In Australian lower secondary schools both girls and boys expressed similar views about their schools and school learning in 1964.

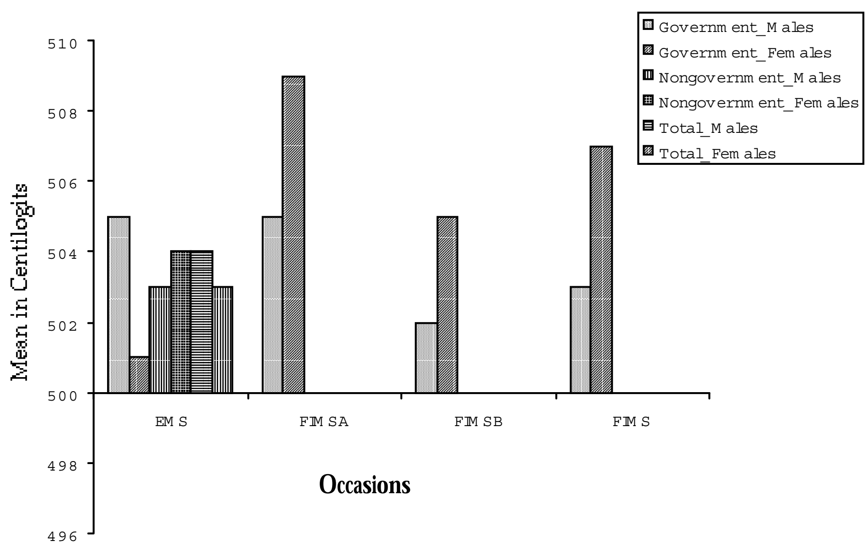


Figure 10.6 Sex differences in views about School and School learning on the two occasions

Sex Differences in Attitudes towards School and School Learning

Table 10.7 shows the descriptive statistics of government and nongovernment school students who responded to the Attitudes towards School and School Learning statements. In addition, Figure 10.7 compares the estimated mean scores of EMS, FIMS and SIMS students since all EMS, FIMS and SIMS students responded to statements on this scale.

Table 10.7 Descriptive statistics for Attitudes towards School and School Learning of students on the three occasions by sex

EMS	**G**		**Ng**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	675.0	685.0	689.0	694.0	682.0	690.0
Standard Deviation	132.0	135.0	113.0	113.0	123.0	124.0
Jackknife Standard Error	11.8	12.2	8.0	6.5	7.0	6.8
Design Effect	2.2	2.8	1.4	1.2	1.8	2.0
Sample Size	268	332	272	328	540	660
FIMS	**FIMSA**		**FIMSB**		**Total**	
	Male	Female	Male	Female	Male	Female
Mean	536.0	578.0	534.0	574.0	535.0	570.0
Standard Deviation	135.0	141.0	135.0	140.0	134.0	141.0
Jackknife Standard Error	5.3	8.8	6.4	6.1	4.8	6.9
Design Effect	2.5	5.3	3.6	2.8	2.9	4.9
Sample Size	1530	1386	1619	1462	2275	2044

SIMS	**G**		**Ng**		**Total**		**R**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	509.0	529.0	508.0	523.0	509.0	528.0	510.0	529.0
Standard Deviation	113.0	115.0	111.0	121.0	113.0	116.0	113.0	116.0
Jackknife Standard Error	3.7	3.7	9.3	11.1	3.5	3.7	4.1	4.1
Design Effect	2.5	2.0	4.1	4.6	2.6	2.5	2.1	1.8
Sample Size	2095	1894	580	551	2675	2445	1614	1424

	Mean difference	**Effect size**	**t-value**	**Significance Level**
EMS G Male vs Female	-10.0	-0.07	-0.59	NS
EMS NG Male vs Female	-5.0	-0.04	-0.49	NS
EMS Male vs Female	-8.0	-0.06	-0.82	NS
FIMSA Male vs Female	-42.0	-0.30	-4.09	<0.01
FIMSB Male vs Female	-40.0	-0.29	-8.84	<0.01
FIMS Male vs Female	-35.0	-0.25	-4.16	<0.01
SIMS G Male vs Female	-20.0	-0.18	-3.82	<0.01
SIMS NG Male vs Female	-15.0	-0.13	-1.04	NS
SIMS Male vs Female	-19.0	-0.17	-3.73	<0.01
SIMSR Male vs Female	-19.0	-0.17	-3.28	<0.01

G= government school students; NG = Nongovernment school students; R = Those group of students in SIMS which are comparable with FIMSA; NS = Not Significant

The estimated mean scores of the Ethiopian government school male and female students on attitudes towards school and school learning were 675 and 685 respectively. The mean difference between the two groups was ten centilogits in favour of girls. The effect size (-0.07) and t-value (-0.59) were very small. Consequently the difference was not significant (see Table 10.7 and Figure 10.7). This indicated that there was no sex difference found between Ethiopian government school students in their attitudes towards school and school learning.

Furthermore, the estimated mean score difference between the Ethiopian nongovernment school boys and girls on attitudes towards school and school learning was five centilogits and the effect size (-0.04) and t-value (-0.49) were very small (see Table 10.7 and Figure 10.7). Like the government school students the mean score of the girls was slightly higher than that of the boys but the difference was not statistically significant. This showed that there was no sex difference between Ethiopian nongovernment school students in their attitudes towards school and school learning.

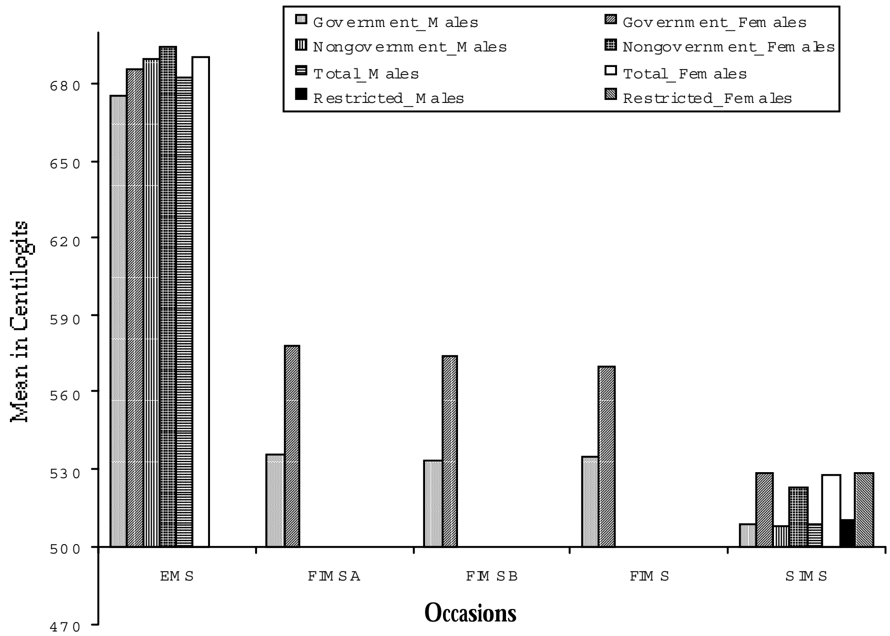


Figure 10.7 Sex differences in Attitudes towards School and School learning on the three occasions

When the estimated mean score difference between the total Ethiopian male and female students was calculated, the mean score of the female students was also slightly higher than the male students, but, the difference was not statistically significant. Hence, in all three comparisons, the mean scores of the girls were slightly higher than that of the boys but all differences were not statistically significant. Therefore, it would seem possible to conclude that there was no significant sex differences found between Ethiopian students in their attitudes toward school and school learning.

A comparison was also undertaken between FIMS male and female students' attitudes towards school and school learning. When the 13-year-old male students mean score was compared with the mean score of their female peers, the latter mean score was markedly higher than the former. The difference was 42 centilogits. This indicated that 13-year-old female students showed more enthusiasm for school and the experiences their schools provided than their male counterparts. Thus, the difference was statistically significant at the $p<0.01$ level (see Table 10.7 and Figure 10.7).

In addition, the estimated mean score difference between FIMSB male and female students' attitudes towards school and school learning was also considered. The females' estimated mean score was markedly higher than that of the male students, and the difference was 40 centilogits. This difference was statistically significant at the $p>0.01$ level. This significant difference showed that the FIMS Year 8 female students expressed more favourable attitudes towards school and school learning than their male classmates. Furthermore, the estimated mean scores of the female students in FIMS total was also markedly higher than those of their male peers. The difference was 35 centilogits. Hence, the difference was significant at the $p<0.01$ level. Consequently, the findings in FIMS were in a similar direction to the findings in EMS, with on both occasions, the female students scoring higher than their male peers. However, the differences in EMS were not statistically significant, while those for the Australian 1964 samples were strongly significant.

A comparison was also undertaken between SIMS government school male and female students' attitudes towards school and school learning. The difference between the two mean scores was 20 centilogits in favour of the female students. This difference was also statistically significant at the 0.01 level. Furthermore, when the two mean scores in nongovernment schools were compared, the mean score of the female students was higher than that of the male students' mean score by 15 centilogits. However, the difference was not statistically significant. Meanwhile, when the male and female student mean scores in SIMSR were compared, like that in the government schools the mean scores of the female students was higher than that of the male students by 19 centilogits. This difference was statistically significant at the 0.01 level. Furthermore, the estimated mean scores of all SIMS female students was also higher than that of their male peers. This difference was 19 centilogits, and the difference was significant at the 0.01 level.

In this analysis the interesting finding was that in all the comparisons, female students expressed more favourable attitudes towards school and school learning than their male peers. These findings were consistent with Keeves (1972), Keeves and Kotte (1992), Kotte (1992). Keeves (1972) pointed out that at Years 6 and 7 girls in the Australian Capital Territory showed more favourable attitudes to school than boys. Furthermore, Keeves and Kotte (1992, p. 156) in their study *Disparities between the Sexes in Science Education: 1970-84*, reported that " with quite remarkable cross-country consistency, girls hold more favourable attitudes to school and school learning." In addition, Kotte (1992) in his analysis of the attitudes towards school and school learning scale administered in the Second International Science Study reported that female students showed more favourable attitudes towards school and school learning. Furthermore, this researcher indicated that "across all countries and at each of the levels tested girls enjoyed being at school more strongly than their male classmates" (Kotte, 1992, p. 113).

The findings are summarised below.

1. There was no significant sex difference between Year 8 Ethiopian students' attitudes towards their school and school learning.
2. In 1964, female Australian students indicated more favourable attitudes towards school and school learning than male students.
3. Female Australian students in 1978 also indicated more favourable attitudes towards school and school learning than male students, however, the difference was not significant between nongovernment school students.

Sex Difference in Attitudes towards Control of the Environment

The last scale involved in close comparisons was Attitudes towards Control of the Environment. Table 10.8 shows the descriptive statistics of government and nongovernment school students who responded to the Attitudes towards Control of the Environment scale. In addition Figure, 10.8 compares the estimated mean scores of EMS, FIMS and SIMS students since all EMS, FIMS and SIMS students responded to statements on this scale.

When the estimated mean scores of EMS government school male and female students were compared, boys scored higher than girls by nine centilogits (see Table 10.8 and

Figure 10.8). However, this difference was not statistically significant. On the other hand, when the nongovernment school male and female students' mean scores were compared, the females scored higher than the males, but the difference was not statistically significant. Meanwhile, comparison between the mean scores of the two sexes for EMS total showed that the mean score of the girls was higher than that of the boys, and the difference was not statistically significant. Thus, it can be concluded that in the EMS study boys and girls expressed similar attitudes towards man's ability to control the environment.

Table 10.8 Descriptive statistics for Attitudes towards Control of the Environment of students on the three occasions by sex

EMS	**G**		**Ng**		**Total**			
	Male	Female	Male	Female	Male	Female		
Mean	563.0	554.0	554.0	565.0	558.0	560.0		
Standard Deviation	92.0	91.0	91.0	88.0	91.0	90.0		
Jackknife Standard Error	8.3	6.3	4.6	5.7	4.7	4.2		
Design Effect	2.2	1.6	0.7	1.4	1.4	1.5		
Sample Size	268	332	272	328	540	660		
FIMS	**FIMSA**		**FIMSB**		**Total**			
	Male	Female	Male	Female	Male	Female		
Mean	568.0	557.0	570.0	557.0	570.0	556.0		
Standard Deviation	71.0	69.0	73.0	70.0	141.0	69.0		
Jackknife Standard Error	2.1	2.4	2.6	3.1	1.9	2.5		
Design Effect	1.3	1.8	2.0	2.8	1.5	2.7		
Sample Size	1530	1386	1619	1462	2275	2044		
SIMS	**G**		**Ng**		**Total**		**SIMSR**	
	Male	Female	Male	Female	Male	Female	Male	Female
Mean	520	511.0	507.0	508.0	517.0	510.0	520.0	511.0
Standard Deviation	83	80.0	85.0	83.0	83.0	81.0	83.0	80.0
Jackknife Standard Error	2.3	2.3	5.7	4.3	2.1	2.0	2.6	2.6
Design Effect	1.7	1.6	3.5	1.5	1.6	1.6	1.6	1.5
Sample Size	2095	1894	580	551	2675	2445	1614	1424
	Mean difference		**Effect size**		**t-value**		**Significance Level**	
EMS G Male vs Female	9.0		0.10		0.86		NS	
EMS NG Male vs Female	-11.0		-0.12		-1.49		NS	
EMS Male vs Female	-2.0		-0.02		-0.32		NS	
FIMSA Male vs Female	9.0		0.13		2.82		<0.01	
FIMSB Male vs Female	13.0		0.18		3.58		<0.01	
FIMS Male vs Female	14.0		0.13		3.14		<0.01	
SIMS G Male vs Female	9.0		0.11		2.77		<0.01	
SIMS NG Male vs Female	-1.0		-0.01		-0.14		NS	
SIMS Male vs Female	7.0		0.09		2.41		<0.05	
SIMSR Male vs Female	9.0		0.11		2.45		<0.05	

G= government school students; NG = Nongovernment school students; R = Those group of students in SIMS which are comparable with FIMSA; NS = Not Significant

When the FIMS male and female students were considered, in all the three groups (FIMSA, FIMSB and the total) unlike the EMS students, the male students scored higher than their female peers. The differences were for FIMSA nine, for FIMSB 13, and 14 centilogits for the total, and all the differences were statistically significant at the $p<0.01$ level.

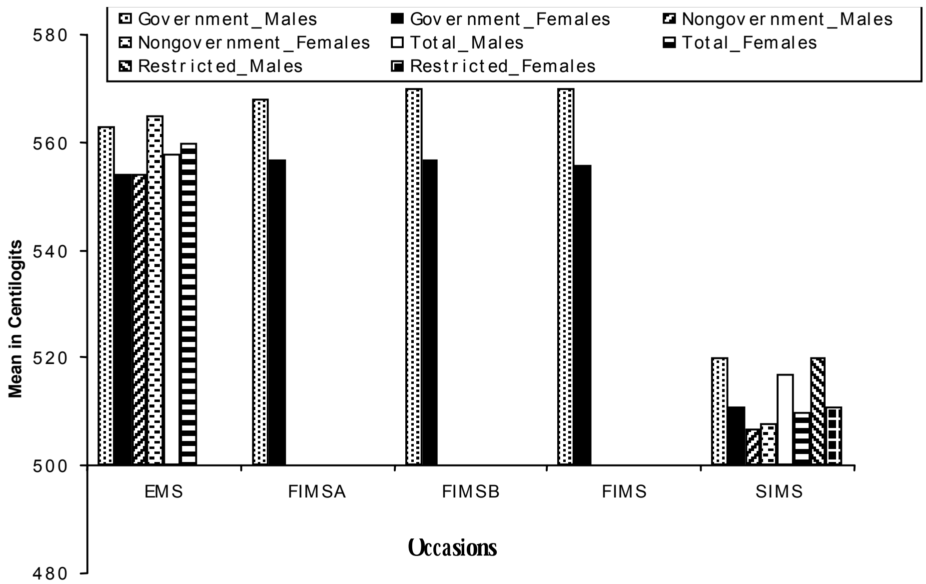


Figure 10.8 Sex differences in Attitudes towards Control of the Environment on the three occasions

In the comparison between the estimated mean scores of the two sexes in the SIMS government schools, the boys also scored at a higher level than the female students by nine centilogits (see Table 10.8 and Figure 10.8). Thus, the difference was statistically significant at the 0.01 level. Similar results were found in the comparison between boys' and girls' mean scores in SIMSR. The boys' score was higher than their female counterparts' by nine centilogits, and this difference was significant at the 0.05 level.

In SIMS the nongovernment school female students' mean score was higher than that of their male peers, however, the difference was not statistically significant. Meanwhile, the mean score difference between the total SIMS male and female students was seven centilogits in favour of male students. Furthermore, this difference was statistically significant at the 0.05 level. Consequently, it can be concluded that in SIMS with the exception of the nongovernment school sample, the male students expressed more favourable attitudes towards man's ability to control the environment than did their female peers.

There are three important observations to be made. In both FIMS and SIMS government schools the boys expressed more favourable attitudes towards control of the environment than their female peers. Whereas in EMS and SIMS nongovernment schools female students expressed more favourable attitudes towards control of the environment scale than their male counter parts, however, the differences were not statistically significant on both occasions. In addition at the Year 8 level, both in Australia and Ethiopia boys expressed more positive attitudes than their female classmates. However, the difference was significant only for Australian students.

The findings may be summarised in the following terms.

1. A significant sex difference was not detected for Ethiopian students' attitudes towards man's role to control the environment.
2. Male Australian students in 1964 expressed more favourable attitudes towards the ability and role of man to control and change the environment than their female peers.

3. In 1978 Australian government school male 13-year-old students expressed more favourable attitudes towards the ability and role of man to control and change the environment than their female counterparts.

Conclusion

The first part of this chapter discussed the analysis of sex differences in mathematics achievement between boys and girls for the four data sets. The analysis showed no significant difference in mathematics achievement between boys and girls in Australian lower secondary school over the past 30-year period. Whereas a marginally significant difference was found between the two sexes in Ethiopian government schools. The other major analysis in this chapter was concerned with sex differences in views and attitudes towards mathematics and schooling. One of the strong findings is the more favourable attitudes of girls to school and school learning compared to boys, as a consequence, it is not surprising that retention rate for girls now exceed that of boys (Leder, 1990; Leder and Forgasz, 1992). The Australian Bureau of Statistics (1997, p. 6) reported the 1996 retention rates in Australia schools "as in previous years, the retention rate for female students (77%) was higher than the corresponding rate for males (65.9%)." The results of the comparisons of sex differences on the seven view and attitude scales are presented in Table 10.9.

Table 10.9 Effect size for sex differences in views and attitudes towards mathematics and schooling in EMS, FIMS and SIMS[a]

Scale	EMS			FIMS			SIMS				N	Sign		Mean
	G	NG	Total	FIMSA	FIMSB	Total	G	NG	Total	R	items	+	-	
Viewmath	-0.05	-0.23	-0.14	-0.17	-0.19	-0.21	0.05	-0.02	0.03	0.07	11/6	3	7	-0.09
Mathpro	0.13	0.21	0.17	-0.08	-0.12	-0.09	nc				8	3	3	0.04
Facimaths	0.04	-0.05	0.00	0.07	0.10	0.07	-0.04	-0.17	-0.07	-0.06	7	5	5	-0.01
Mathsoc	0.02	-0.13	-0.05	0.27	0.24	0.22	0.12	-0.06	0.08	0.12	8	7	3	0.08
Viewsch	0.12	-0.03	-0.03	-0.10	-0.07	-0.09	nc				11	1	5	-0.03
Schenjoy	-0.07	-0.04	-0.06	-0.30	-0.29	-0.25	-0.18	-0.13	-0.17	-0.17	11/9	0	10	-0.17
Contrenv	0.10	0.12	-0.02	0.13	0.18	0.13	0.11	-0.01	0.09	0.11	9/6	8	2	0.09

Viewmath = Views about Mathematics Teaching scale, Mathpro = Attitudes towards Mathematics as a Process scale; Facimaths = Attitudes towards Facility of Learning Mathematics scale; Mathsoc = Attitudes towards the Place of Mathematics in Society scale; Viewsch = Views about School and School Learning scale; Schenjoy = Attitudes towards School and School Learning scale; Contrenv = Attitudes towards Control of the Environment Scale; nc = Not conducted; The number of statements administered in SIMS for Views about Mathematics Teaching and Attitudes towards School and School Learning scales were only six and nine respectively; [a] Statistically significant sex differences are in bold; G = Government; NG = Nongovernment; R = Restricted

The expected significance by chance for multiple comparisons of sex differences on the seven view and attitude scales was only five per cent, however, the significances in Table 10.9 increased to 37 per cent. Among these significant differences 29 per cent were from the three attitude scales, namely Attitudes towards the Place of Mathematics in Society, Attitudes towards School and School Learning, and Attitudes towards Control of the Environment scales. In Attitudes towards the Place of Mathematics in Society, and Attitudes towards Control of the Environment scales in FIMS and SIMS data sets boys expressed more favourable attitudes than their female counterparts. Meanwhile, in Attitudes towards School and School Learning scale females expressed more favourable attitudes than their male colleagues. In the Attitudes towards Mathematics as a Process scale on all the comparisons male Ethiopian students showed more favourable attitudes than their female classmates, whereas, in Australia girls indicated more favourable attitudes than their male counterparts.

The findings in Chapter 8 showed significant differences in mathematics achievement between countries and occasions. However, the results in this chapter revealed that there was no sex difference in mathematics achievement on the different occasions. In addition, no significant sex differences were found in students' attitudes towards the difficulty of learning mathematics (except for FIMSB) and their views of the ways their mathematics teachers taught mathematics. Thus, these sex differences in mathematics, and views and attitudes towards mathematics are unlikely to be explanations for differences in achievement between occasions. Moreover, these findings indicate a need for more detailed investigations into differences in the conditions of learning in the two countries. Carroll (1963) in his model of school learning, has identified five factors that influence school learning. One of the factors that he proposed was perseverance, which is related to attitudes towards school and mathematics learning. From Carroll's model of school learnings, models of student level factors that influence mathematics achievement are developed in Chapter 6 and include attitudes as well as sex of student as mediating and antecedent variables respectively. The examination of these models demands the use of multivariate analysis, thus in the next chapter these models of student level factors that influence the mathematics achievement level of Australian and Ethiopian students are investigated using partial least square path analyses procedures.

11

Student Level Factors that Influences Mathematics Achievements

Research findings have indicated that there are substantial differences between students in their achievement in school mathematics, and the factors that influence the achievement levels of students need to be investigated. From the findings of previous research, models of the student level factors influencing mathematics achievement at the 13-year-old level and at the Year 8 level were developed (see Chapter 6). PLSPATH was chosen as an appropriate multivariate technique to investigate the hypothesised models described in Chapter 6. This chapter discusses the results obtained when the hypothesised models were tested using PLSPATH in the analysis of data from the First, Second and Third International Mathematics Studies conducted in Australia and the Ethiopian Mathematics Study.

Therefore, the purposes of this Chapter are to:

(a) examine the student level factors influencing student learning in mathematics in Australia and Ethiopia;

(b) investigate the differences and the similarities in the student level factors that influence the learning of mathematics in the two countries;

(c) compare the differences between 1964 and 1978 in the student level factors that influence the learning of mathematics at the 13-year-old level in Australia; and

(d) compare the differences between 1964 and 1994 in the student level factors that influence the learning of mathematics at the Year 8 level in Australia.

It is necessary to separate the comparisons between 1964 and 1978 from those between 1964 and 1994 because of differences in the nature of the sub-populations under investigation.

The first section describes the results of the EMS data set. In the second and third sections the FIMS data analyses are discussed. The SIMS results of PLSPATH analyses are presented in the fourth section. The fifth section describes the results of analyses with the TIMS data set. The final section considers the similarities and differences between the five analyses that were undertaken.

In this analysis weighted scores were employed in compensating for the differential sampling procedures and student losses when combining data across strata for all the different occasions (see Chapter 7).

Results of the EMS Data Set

Tables 11.1a and 11.1b show the outer model and the inner model results for the EMS data set. Eleven LVs and 25 MVs were included in the model. Thus the results of the PLSPATH analyses are presented in two parts. The first part gives the results for the outer model and the second part discusses the inner model.

Outer Model Results

In the following discussion the weights and the factor loadings, the communality, redundancy and the tolerance values of each MV within a construct are discussed with respect to the LV to which it contributes. The weights (ßs) are considered significant if their values are ß≥0.07, while the factor loadings (ls) are regarded as being significant if l ≥0.30. The criterion employed for measuring the strength of the outer model is the average of the communalities of the MVs (Falk, 1987). Furthermore Falk has argued that the higher the average of the communalities the better the outer model, and an average communalities value of 0.30 would generally be considered too low.

The other important point to be considered in this study is that, no significance testing was undertaken, because the samples employed were not simple random samples, but a cluster and stratified random sample design, and simple random sample tests of significance are clearly inappropriate (Sellin and Keeves, 1997).

Home Background (Homeback)

Table 11.1a shows that *Homeback* was constructed in an outward mode and reflected in six MVs. In the hypothetical model developed in Chapter 6, it was assumed to be formed in the inward mode, however, during the analysis, it was changed to an outward mode. Since this LV had as many as six observed or MVs the outward mode was chosen to avoid problems of multicollinearity. One MV, namely *Mocc* was deleted from the analysis, since the factor loading (for outward mode) of this variable was 0.14 which was below the critical value of 0.30 (Campbell, 1996). This indicated that *Mocc* failed to contribute in a meaningful way to the latent variable *Home Background*. The other MVs reflecting *Homeback* were *Focc, Fed* and *Med*. The factor loadings for *Focc* was 0.39, while for *Fed* and *Med*, they were 0.84. This shows that the educational back grounds of both the father and the mother were highly important for the development of the construct.

Homebook was another MV that contributed to *Homeback*. Its loading was 0.71. The variable that contributed least to the LV was *Siblings* (number of siblings), and its loading was -0.31. While, it did not contribute as strongly as did the other MVs, this factor could be taken as a measure of family size, and the marked negative effect of a larger family should be noted.

Hence, apart from *Mocc* all other MVs were well suited to reflect a student’s home background or socioeconomic status. The communalities in Table 11.1a show that all the MVs contributed to this construct. Communality is defined as the squared loadings for a MV. It is a measure of the explained variance of a particular MV with respect to the LV it reflects (Sellin, 1990). Consequently, the variance explained by each MV was from 10 per cent by *Siblings* to over 70 per cent by *Fed*.

Table 11.1a EMS Outer Model Results

Variable		Weight/Loading[a]	Communality	Redundancy	Tolerance
Homeback[o]	*Focc*	.39	.15	.00	.09
	Mocc	Deleted			
	Fed	.84	.71	.00	.43
	Med	.84	.70	.00	.41
	Homebook	.71	.50	.00	.21
	Siblings	-.31	.10	.00	.05
Gender[u]	*Sex*	1.00	1.00	.00	.00
Studage[u]	*Age*	1.00	1.00	.00	.00
Province[i]	*Homelan*	.22	.05	.00	.002
	Fprov	.50	.37	.00	.30
	Mprov	.10	.23	.00	.31
	Prov	.74	.65	.00	.03
Classize[u]	*Clssize*	1.00	1.00	.01	.00
Views[o]	*Viewmath*	1.00	1.00	.01	.00
	Viewsch	Deleted			
Values[o]	*Mathinso*	.80	.65	.05	.07
	Contrenv	.79	.62	.05	.07
Motivation[i]	*Attitsch*	1.00	1.00	.22	.00
	Hmwall	Deleted			
Timlearn[i]	*Hourmath*	.98	.97	.15	.0004
	Hourmhmw	.18	.04	.01	.0004
Attitude[o]	*Likemath*	.52	.27	.06	.10
	Mathmark	.47	.22	.05	.09
	Diffmath	.83	.69	.15	.01
Mathachi[u]	*Rasch score*	1.00	1.00	.12	.00
Mean Communalities			0.59		

i = Inward mode; o = Outward mode; u = Unity; [a] = weight for inward mode and loading for outward mode

Kotte (1992) has defined redundancy as the squared correlations between a particular MV and the set of LVs linked indirectly through inner model relationships. Kotte argues that a high redundancy value can be taken as a possible misplacement of that MV in relation to the LVs which predict it. It can be seen from the table, that the redundancy values for each of the five MVs forming *Homeback* are 0.00, because *Homeback* operates as an antecedent and is not predicted by any other LVs in the model.

Table 11.1a also gives the tolerance values for the MV. The tolerances show that the five MVs contribute to the formation of the LV *Homeback*. The tolerance value is the squared multiple correlation of a particular MV with all remaining MVs in the set. The tolerance values provide information about possible multicollinearity within a block of MVs. High values (≥0.50) indicate possible multicollinearity and the variables involved should be considered with caution particularly in the inward mode (Sellin, 1990). However, here the highest tolerance value is only 0.43, and it seems safe to say that there is little multicollinearity, which would confound estimated relationships.

Gender

The sex of the student was considered to indicate *Gender*. Thus, this LV comprised a single MV.

Student Age (Studage)

The age of the student in years was taken as an index of *Studage*. Hence, this LV comprised just a single MV.

Province

Four MVs, namely, *Fprov* (Father 's Province of birth), *Mprov* (Mother's Province of birth), *Prov* (Student's Province of birth) and *Homelan* (Language spoken at home) were hypothesised to form this LV. Because this LV was in an inward mode the weights for the MVs were 0.50, 0.10, 0.74 and 0.22 respectively. *Fprov* and *Prov* contributed more to the formation of the LV *Province* than did the other two MVs.

Class Size (Classize)

The number of students in class was taken as the index of *Classize*. Hence, this LV comprised just a single MV called *clssize*.

Views about Mathematics (Views)

This construct was in the outward mode and two MVs, namely, *Viewmath* and *Viewsch* reflected this construct. The former was a scale measuring the students' *Views* about the way mathematics teachers taught mathematics and the latter was a scale measuring the students' *Views* about their schools and school learning. However, the latter was dropped from further analysis because the loading was below the critical value of 0.30. The interesting point to be observed is that the *Views* of the students about the climate of their schools and school learning did not correlate substantially with their views about mathematics teaching (r=0.10), and the only contributing factor was their *Views* about the methods their mathematics teachers used to teach them. This indicated that the students' views about the overall climate of the school and school learning did not have an effect on their learning of mathematics. Instead only their views about the ways and methods employed by their mathematics teachers had an effect on their learning of mathematics. This means that the students at this level did not relate the general climate of their schools to their learning in a particular subject area, namely mathematics. Consequently, this LV involved only one MV namely, *Viewmath.*

Values about Mathematics (Values)

Mathinso (a scale measuring students' Attitudes towards the Place of Mathematics in Society) and *Contrenv* (a scale measuring students' Attitude towards Control of the Environment) reflected this LV. The LV was constructed in the outward mode. It can be seen in Table 11.1a that these two MVs had high loadings (>0.70) on the latent variable *Values*.

Motivation towards Mathematics (Motivation)

Two MVs, namely, *Attitsch* (a scale measuring students' Attitudes towards School and School Learning) and *Hmwall* (Hours taken by the student to do all his/her homework) formed this LV in the inward mode. The former MV involves the students' attitudes towards their schools and school learning and the latter involves the time taken by a student to do all his or her homework. However, the latter was dropped from further analysis because the weight fell below the critical value (0.07). It is important to observe here that the time taken by students to do their homework in all subjects did not contribute for the formation of the LV *Motivation.* A majority of the students (62 %) spent about 12 hours a week on all their homework. This means that the students were spending substantial time each night on homework. However, the brighter students might not have done as much homework, when compared with the weaker students. Thus homework might have had a remedial purpose.

Consequently, this LV was formed by only one MV namely, *Attitsch* and its loading and communality were 1.00, while the redundancy and the tolerance values were 0.22 and 0.00 respectively (see Table 11.1a). In the hypothetical model developed in Chapter 6, *Motivation* was assumed to be in an outward mode, however, in the analyses, it was changed to an inward mode, in order to get better estimations from the MVs. Moreover, in both estimations, *Hmwall* has been removed from the analyses. Thus the LV formed with only one MV and its estimation became equal to unity, irrespective of whether such single MV blocks are specified as inward or outward mode (Sellin, 1992, p. 404).

Time in Learning (Timlearn)

The MVs *Hourmath* and *Hourmhmw* formed this LV in the inward mode. *Hourmath* was the number of periods that the student had in a week to learn mathematics, while *Hourmhmw* was the number of hours taken by the student to do his or her mathematics homeworks in a week. It can be seen in Table 11.1a that the former which provided the number of hours assigned for the student to learn mathematics had a noticeably high weight (0.98), while the latter had a weight of 0.18.

Attitudes towards Mathematics (Attitude)

Three MVs namely, *Likemath*, *Mathmark,* and *Diffmath* contributed to reflect this construct. The analysis showed that *Diffmath* (0.83) (a scale measuring students' Attitudes towards Facility of Learning Mathematics) was the strongest contributor to reflecting the LV *Attitude*, while *Likemath* and *Mathmark* had effects given by the loadings of 0.52 and 0.47 respectively.

Mathematics Achievement (Mathachi)

The Rasch analysed mathematics test score of students was considered to indicate this construct. Thus, this LV comprised a single MV.

In summary, for the outer model, among the 25 hypothesised MVs that might contribute to the 11 constructs, three MVs were removed from further analysis because of lack of fit to the path model and the remaining 22 MVs contributed to the 11 LVs. The average of the communalities of the MVs which is considered a measure of the strength of the outer model was 0.59. The next section presents the results for the inner model.

Inner Model Results

The results for the outer model are discussed in the previous section, while in this section the results of the inner model are presented. Table 11.1b shows the beta (β), correlation and tolerance values and Table 11.1c indicates the direct, indirect and total effects, correlations, fit and R^2. There are 11 LVs in the inner model, and the results obtained from the analyses of these LVs are presented in Figure 11.1.

Among the 11 LVs, *Homeback, Gender, Province* and *Studage* were considered antecedents, which means that they were not influenced by any other LV. Therefore, the discussion in this section examines only those seven LVs that were assumed to be influenced by an other LV in the hypothesised model. The criterion employed for measuring the strength of the inner model is the average multiple R^2 for the endogenous variables in the model (Falk, 1987). He indicated that the larger the average multiple R^2 the better the inner model. The size of a β coefficient was considered to be sufficient magnitude or importance and significance in the path model if it was equal or exceeded 0.07, which was associated with the explanation of approximately 0.5 per cent of the variance in the criterion latent variable.

Class Size (Classize)

This LV provided information about the number of students in a mathematics class and this information was obtained from each student. Four variables were hypothesised to influence this construct, however, only *Homeback* (0.08) was found to influence it, and the influence was not strong. Table 11.1b and Figure 11.1 show that the direct effect of *Homeback* was 0.08 and there was no any other factor which had an indirect effect on this LV. The correlation and the tolerance values were 0.08 and 0.00 respectively. The explained variance (R^2=0.01) for the LV *Classize* was very small. R^2 shows the variance of a construct explained when all significant variables which influence the criterion are included in the model. Thus, the larger the R^2 the more variance explained. This means that those students from higher socioeconomic status backgrounds were in larger classes for mathematics.

Table 11.1b EMS Inner Model statistics

Variable	Beta	Correlation	Tolerance	R^2
Classize				0.01
Homeback	.08	.08	.00	
Views				0.01
Homeback	.09	.09	.0003	
Gender	.07	.07	.0003	
Values				0.08
Province	-.08	-.08	.0003	
Views	.27	.27	.0003	
Motivation				0.22
Views	.23	.33	.07	
Values	.35	.41	.07	
Timlearn				0.16
Homeback	.23	.25	.01	
Classize	.31	.32	.01	
Attitude				0.21
Views	.10	.25	.13	
Values	.20	.35	.19	
Motivation	.29	.40	.22	
Mathachi				0.12
Homeback	.14	.19	.10	
Gender	-.08	-.10	.01	
Studage	-.12	-.16	.05	
Classize	-.18	-.14	.11	
Timlearn	.11	.11	.16	
Attitude	.16	.20	.02	
Mean R^2				0.12

Views about Mathematics (Views)

The *Views* of students about mathematics were hypothesised to be influenced by five LVs. Among these factors only two latent variables influenced this construct (see Tables 11.1b and 11.1c and Figure 11.1). The factors that influenced *Views* were *Gender* and *Homeback.* Girls (0.07) expressed stronger *Views* about mathematics than boys, and students from higher socioeconomic status (0.09) indicated stronger *Views* about mathematics than their classmates from lower socioeconomic backgrounds (see Tables 11.1b and 11.1c and Figure 11.1). It can be seen from Table 11.1c that there were no indirect effects associated with this particular LV. The correlation and the tolerance for *Homeback* can be seen in Table 11.1b. The values were 0.09 and 0.0003 respectively and were 0.07 and 0.0003 for *Gender* respectively. The variance explained (R^2=0.01) for the construct was very small. Therefore, from these results it would seem possible to conclude that student's *Views* about mathematics were influenced by their *Gender* and *Home Background,* but with considerable unexplained variance.

Values about Mathematics (Values)

The values of students about mathematics were hypothesised to be influenced by six LVs, namely *Homeback, Gender, Province, Studage, Classize* and *Views*. Among these factors only two influenced this construct directly (see Table 11.1b and 11.1c and Figure 11.1). The factors that influenced *Values* directly were *Province*: [those students whose homes were outside of Addis Ababa expressed stronger values about mathematics than their classmates from Addis Ababa (see Tables 11.1b and 11.1c and Figure 11.1)]. In addition, the effects of their *Views* were found to be significant. These results show that students who indicated strong *Views* valued mathematics more than those students who expressed weak *Views* about mathematics. Table 11.1c also shows that *Homeback* (0.03) and *Gender* (0.02) indirectly influenced this particular LV. The indirect effects indicate that those students from higher socioeconomic status back grounds and female students indirectly expressed stronger values about mathematics than their classmates from lower socioeconomic status backgrounds and male students. These two factors indirectly influenced the values of students through their *Views,* which was a mediating variable. Tables 11.1b and 11.1c show the correlations and the tolerance values. The correlations for *Homeback, Gender, Province* and *Views* were 0.03, 0.02, -0.08 and 0.27 respectively, while the tolerance values were provided only for the variables that influenced the construct directly, and, the tolerances for *Province* and *Views* were 0.003 each. The variance explained (R^2=0.08) of the construct was small. Thus, *Values* about mathematics was influenced directly by the LVs *Province* and *Views* about mathematics, and only indirectly by *Homeback* and *Gender*.

Motivation towards Mathematics (Motivation)

The *Motivation* level of students towards mathematics was hypothesised to be influenced by seven LVs, namely *Homeback, Gender, Province, Studage, Classize, Views* and *Values*. *Views* showed a direct (0.23) and an indirect (0.09) influence on this construct. This means that those students who held stronger *Views* towards mathematics also expressed stronger *Motivation.* Meanwhile, *Values* (0.35) showed only a direct influence on this construct, which indicates that students who expressed stronger values were more highly motivated towards mathematics than those students who expressed weaker values about mathematics. Three other factors only influenced the LV *Motivation* indirectly (see Tables 11.1b and 11.1c and Figure 11.1). The factors that had indirect influence on *Motivation* were *Homeback* (0.03), *Gender* (0.02) and *Province* (-0.03). The indirect effects show that those students from higher socioeconomic status backgrounds expressed slightly stronger motivation towards mathematics learning than those students from lower socioeconomic backgrounds. While, the indirect effect of *Gender* demonstrated that female students expressed slightly stronger motivation towards mathematics than their male peers with the effect operating largely through their views. Meanwhile, the effect of *Province* on

Motivation revealed that those students whose background was not in Addis Ababa indicated slightly stronger motivation than students whose background was in Addis Ababa. These three factors indirectly influenced the motivation of students through *Values*. The correlations for *Homeback, Gender, Province, Views* and *Values* with *Motivation* were 0.02, 0.03, -0.03, 0.33 and 0.41 respectively. Meanwhile, the variance explained (R^2=0.22) for the construct was medium. Therefore, these results revealed that students' *Motivation* towards mathematics was directly and largely influenced by their *Values* and their *Views* (directly and indirectly) and only indirectly by *Homeback, Gender,* and *Province*.

Time in Learning (Timlearn)

Eight LVs, namely *Homeback, Gender, Province, Studage, Classize, Views, Values* and *Motivation* were factors which were expected to influence *Timlearn*. However, the results of the analysis revealed that only two of these factors influenced this construct (see Tables 11.1b and 11.1c and Figure 11.1). The factors that had a direct influence on *Timlearn* were *Homeback* (0.23)[those students from higher socioeconomic status background spent more time in learning mathematics than students from lower socioeconomic background (see Tables 11.1b and 10.1c and Figure 11.1)] and *Classize* (0.31) [students from larger class groups spent more time in learning mathematics than students from smaller class groups]. Table 11.1c indicates that *Homeback (*0.02), also influenced *Time in Learning* indirectly through *Classize*. Thus the total effect of *Homeback* was 0.25 while the total effect of *Classize* was 0.31. The correlations for *Homeback*, and *Classize* were 0.25 and 0.32 respectively. The R^2(0.16) value for this construct was medium. Therefore, these results revealed that the time in which students were involved in learning mathematics was influenced largely by their home background and the number of students in the class. This was a very interesting observation, because the latest figures from the Ministry of Education in Ethiopia showed that the number of students in Year 8 was stated in the following terms "Addis Ababa was the highest with 74 pupils per class" (Ministry of Education, 1996, p. 1). Consequently, it is difficult to say that optimal learning is taking place in a class size of 74 students. The teacher would not even have time to check whether the students had done their assignments. It is very important to recognise that the figure revealed by the Ministry of Education, that is 74 students in a class, is an average, so the reader must bear in mind that there are classes in Addis Ababa with more than 74 students.

Attitudes towards Mathematics (Attitude)

The attitudes of students towards mathematics were hypothesised to be influenced by nine LVs, namely *Homeback, Gender, Province, Studage, Classize, Views, Values, Motivation* and *Timlearn*. The result of the PLSPATH analysis demonstrated that two of these factors showed both direct and indirect influences on *Attitude*, and one factor influenced this criterion variable only directly (see Tables 11.1b and 11.1c and Figure 11.1), while the remaining factors did not have any recognisable effect on the construct. The factors that showed both direct and indirect effects on *Attitude* were *Views* (total effect =0.25) [those students who showed stronger views about mathematics also showed more positive attitudes towards the subject (see Tables 11.1b and 11.1c and Figure 11.1)] and *Values* (total effects=0.30) [students who expressed stronger values also demonstrated more positive attitudes towards mathematics than those students who showed weaker values]. *Motivation* influenced the criterion variable only directly (0.29). Those students who demonstrated higher motivation towards mathematics also showed more positive attitudes towards mathematics. *Homeback* (0.02), *Gender* (0.02) and *Province* (-0.02) weakly influenced indirectly this particular LV. The indirect effects of *Homeback* indicated that those students from higher socioeconomic status backgrounds had more positive attitudes towards mathematics than students from lower socioeconomic backgrounds. The indirect effect of *Gender* indicated that female students showed more positive attitudes towards mathematics than their male peers. Meanwhile, the effect of *Province* on *Attitude* revealed that those students whose homes were not in Addis Ababa indicated more positive attitudes than students from Addis Ababa. These three factors indirectly influenced the *Attitude* of students through *Values, Views* and *Motivation.*

Table 11.1b and 11.1c show the correlations and the tolerance values for the LVs. The correlations for *Homeback*, *Gender*, *Province, Views*, *Values* and *Motivation* were 0.08, -0.07, -0.02, 0.25, 0.35 and 0.40 respectively. The variance explained (R^2=0.21) for this construct was medium. Therefore, this analysis showed that students' attitudes towards mathematics were directly and indirectly influenced by their *Views*, and *Values,* directly by their *Motivation* towards mathematics and indirectly by the *Homeback*, *Gender* and the *Province* of the students.

Table 11.1c EMS Inner Model Effects (On)

Variable	Direct	Indirect	Total	Correlation	Fit	R^2
Classize						.01
Homeback	.08	-	.08	.08	-	
Views						.01
Homeback	.09	-	.09	.09	-	
Gender	.07	-	.07	.07	-	
Values						.08
Homeback	-	.03	.03	.03	.002	
Gender	-	.02	.02	.02	.01	
Province	-.08	-	-.08	-.08	-	
Views	.27	-	.27	.27	-	
Motivation						.22
Homeback	-	.03	.03	.02	-.01	
Gender	-	.02	.02	.03	.01	
Province	-	-.03	-.03	-.03	-.004	
Views	.23	.09	.33	.33	-	
Values	.35	-	.35	.41	-	
Timlearn						.16
Homeback	.23	.0237	.25	.25	-	
Classize	.31	-	.31	.32	-	
Attitude						.21
Homeback	-	.02	.02	.08	.06	
Gender	-	.02	.02	-.07	-.09	
Province	-	-.02	-.02	-.02	.003	
Views	.10	.15	.25	.25	-	
Values	.20	.10	.30	.35	-	
Motivation	.29	-	.29	.40	-	
Mathachi						.12
Homeback	.14	.02	.16	.19	-	
Gender	.08	.003	-.08	-.10	-	
Studage	.12	-	-.12	-.16	-	
Province	-	-.004	-.004	-.01	-.01	
Classize	.18	.04	-.15	-.14	-	
Views	-	.04	.04	.05	-.001	
Values	-	.05	.05	.10	.04	
Motivation	-	.05	.05	.02	-.05	
Timlearn	.11	-	.11	.11	-	
Attitude	.16	-	.16	.20	-	
Mean R^2						.12

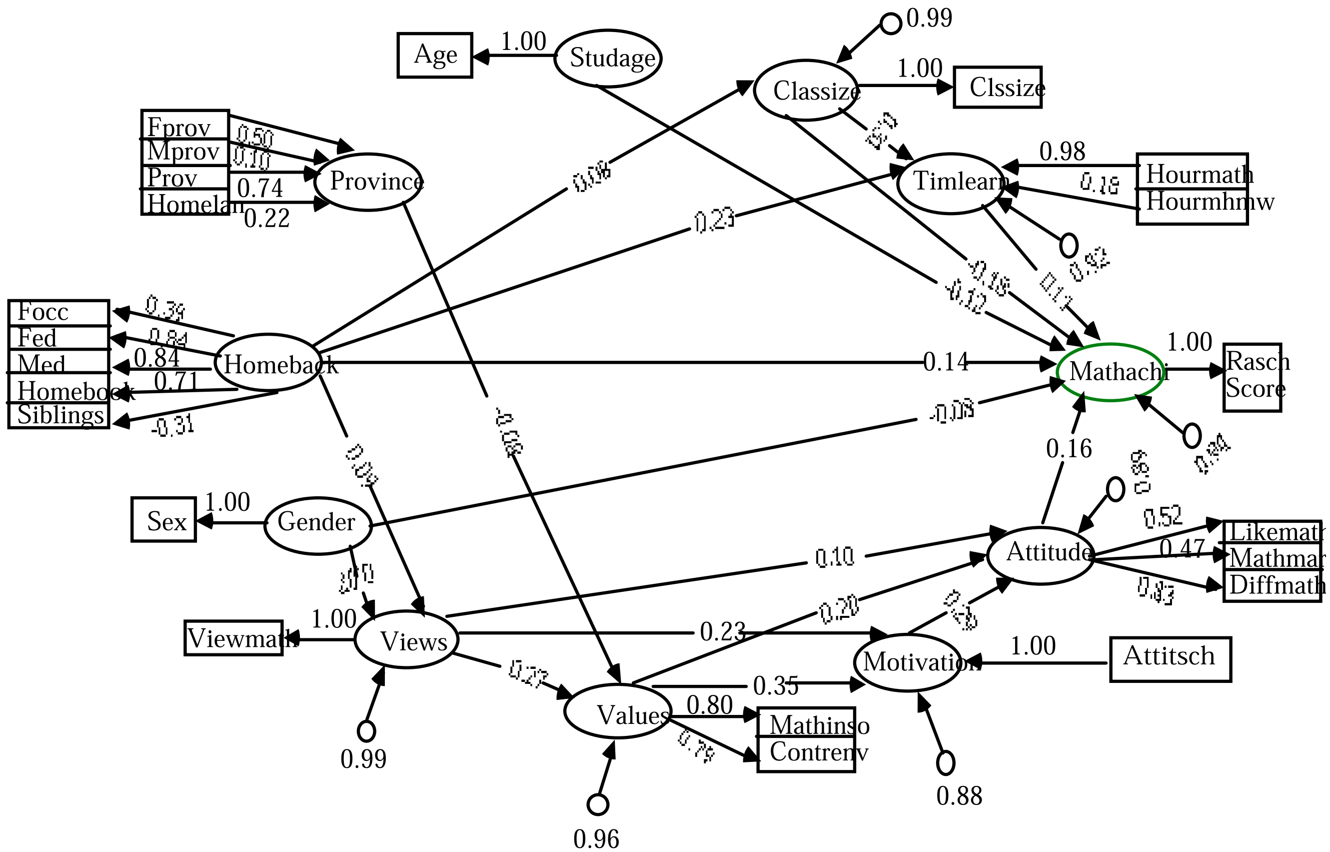


Figure 11.1 EMS - Student level factors influencing Mathematics Achievement

Mathematics Achievement (Mathachi)

The mathematics achievement level of students was hypothesised to be influenced by ten LVs. The result of the PLSPATH analysis demonstrated that three of these factors influenced the outcome measure both directly and indirectly, three other factors influenced *Mathachi* directly, while four factors influenced this LV only indirectly (see Tables 11.1b and 11.1c and Figure 11.1).

Direct Effects

The factors that had direct influence on *Mathachi* are discussed in greater detail below.

Home Background (Homeback)

This factor influenced *Mathachi* directly (0.14) and indirectly (0.02) through *Attitude*. The total effect was 0.16 (direct effect = 0.14 and indirect effect = 0.02), while the correlation was 0.19. The analysis showed that students from a family of higher socioeconomic status background had a higher achievement level in mathematics than their classmates from a lower socioeconomic background. This observation was consistent with previous studies undertaken in Ethiopia (Damtew, 1972; Berhanu, 1986; Endalkachew, 1990; Teshome, 1993; Behutiye and Wagner, 1995). However, this analysis is the first that takes the influence of other factors into account.

Gender

Gender is another factor that influenced *Mathematics Achievement* directly and indirectly. The direct effect was -0.08 and the indirect effect was 0.003, while the total effect remained as -0.08. Thus *Gender* negatively influenced mathematics achievement at the Year 8 level in Addis Ababa. The negative sign indicated that boys were higher achievers in mathematics than girls. The correlation between *Gender* and *Mathematics achievement* was -0.10. The evidence presented above shows that boys were higher achievers in mathematics than girls. This finding is consistent with the findings of other research studies conducted in Ethiopia. These research findings have revealed that at all educational levels in both rural and urban areas of Ethiopia the achievement level of girls in mathematics was much lower than that of boys (Seyoum, 1986; Anbessu and Junge, 1988; Atsede and Kebede, 1988; Ademe and Gebre, 1990; Tsion, 1990; Assefa, 1991; Gennet, 1991a; Behutiye and Wagner, 1993, 1995; Seleshi, 1995; Yelfign et al., 1995). Again, however, this is the first analysis that makes allowance for the influence of other factors on the relationship between *Gender* and *Mathematics Achievement.* The reason why girls were lower achievers in Ethiopia has been argued to be related to the involvement of female students in home activities. In the Ethiopian family girls were considered fully responsible to manage and take care of the house whenever both parents were away. In general, girls in the Ethiopian society were more involved in a larger number of household activities compared to boys. Thus, girls had less time for their studies and play activities during out of school hours, which might have had a negative effect on their mathematics achievement. Research findings have contended that in Ethiopian society the achievement level of girls was generally lower than that of boys, because of the greater involvement of the girls in household activities than the boys (Damtew, 1972; Derese et al., 1990; Gennet, 1991a; Daniel, 1995). However, there was no marked indirect effect of *Gender* on Mathematics Achievement operating through *Time in Learning*. Moreover, this analysis showed, that the indirect effects although small, that operated indirectly through *Views* and *Attitude* enhanced the performance of girls which would seem contrary to commonly held explanations which maintained that the lower achievement of girls in mathematics was related to their less favourable attitudes towards the subject. The effects of gender in the learning of mathematics are clearly more complex than is commonly assumed.

Student Age (Studage)

The other LV that influenced mathematics achievement of students in Ethiopia negatively was their age level. This variable influenced mathematics achievement only directly, and the effect was -0.12. The negative sign demonstrated that younger Year 8 students performed better than older students. In other words, younger Year 8 students were higher achievers in mathematics than older students. The correlation of *Studage* with *Mathematics Achievement* was -0.16.

Class Size (Classize)

Class size was another LV that influenced the *mathematics achievement* of students in Ethiopia. This variable influenced mathematics achievement directly (-0.18) and indirectly (0.04) through *Timlearn*. The total effect was -0.15. The indirect effect was positive while the direct and the total effects were negative. The positive sign of the indirect path showed that students in larger classes had more time to learn mathematics than students in smaller classes, because here *Classize* was mediated by *Timlearn,* and the value was positive. However, the negative sign for the direct and total effect indicated that students from smaller class groups were higher achievers in mathematics than students from larger class groups after other factors had been taken into account. This observation indicated that the number of students in a mathematics class influenced the mathematics achievement level of students. This finding was consistent with the findings of Evaluative Research of the General Education System in Ethiopia (1986).

Time in Learning (Timlearn)

Time in Learning was one of the LVs that influenced the achievement of mathematics students in Ethiopia. This variable influenced mathematics achievement directly (0.11). This observation suggested that students who spent more time in learning mathematics were higher achievers in mathematics than those students who spent less time in learning mathematics. This indicates that in Ethiopia, students who want to improve their achievement level in mathematics need to spend more time in learning mathematics.

Attitude towards Mathematics (Attitude)

This LV influenced Achievement in Mathematics only directly and its effect was 0.16. The evidence showed that students who demonstrated more positive attitudes towards mathematics were likely to achieve at a higher level in mathematics than their classmates who expressed less favourable attitudes. This observation was also consistent with previous studies undertaken in Ethiopia (Endalkachew, 1990; Tadesse, 1993; Seleshi, 1995).

Indirect Effects

Province, *Views*, *Values* and *Motivation* also influenced mathematics achievement of students indirectly (see Table 11.1c). The indirect effect of *Province* was -0.004, the negative sign indicating that students whose family background was outside of Addis Ababa had slightly higher achievement levels than students whose family background was from Addis Ababa. However, the effect was very small. This indirect influence was mediated through *Values* and *Attitudes*. The other factors that indirectly influenced Mathematics Achievement which were mediated by *Attitudes* were *Views* (0.04), *Values* (0.05) and *Motivation* (0.05). The indirect effects of these LVs namely

Motivation, Views and *Values* indicated that students whose *Motivation*, *Views* and *Values* about mathematics were likely to achieve better in the subject than other students. The variance explained for this outcome variable was 0.12.

Conclusion

In this section 25 MVs and 10 LVs were hypothesised to influence the mathematics achievement levels of Ethiopian students. Three out of 25 MVs were removed from further analysis, because these variables did not contribute to the outer model.

The results of the analyses showed that among the 10 LVs hypothesised to influence the mathematics achievements of Ethiopian students, six of them proved to be student level factors that influenced mathematics achievement substantially, at the Year 8 level in Ethiopia. These student level factors that influenced mathematics achievement were *Home Background, Gender, Student Age, Class Size, Time in Learning* and *Attitudes towards Mathematics*, while *Province, Views, Values* and *Motivation* influenced Mathematics Achievement only indirectly. The mean of the R^2 values of the endogenous variables was 0.12. While the proportion of variance explained for the criterion variable of Mathematics Achievement (0.12) was not large, important factors have been identified as influencing level of achievement in this subject.

The relatively low proportion of variance explained was due to several possible reasons. First, the criterion test was very difficult for Ethiopian students and as a consequence the variance of the criterion was greatly reduced, leaving less variance to be explained. Secondly, only approximately 20 per cent of the age group remained at school to the Year 8 level in Ethiopia in 1996, although the proportion must be expected to be slightly higher in Addis Ababa. As a consequence considerable selection had occurred in the sample under survey, and this greatly reduced the variance associated with several of the predictor variables, such as those involving a student's home background and attitudes. With reduced variance in the predictors there was likely to be less variance in the criterion explained by the regression analyses. Finally, no variables involving the ability of a student were included in the analyses, and it would seem likely that such variables would be strong predictors of the criterion variable of achievement in mathematics. A greater proportion of variance explained could thus be obtained through the use of a more appropriate criterion measure and through the inclusion of a predictor variable of student ability in the model being tested.

Results of the FIMSA Data Set

In Chapter 6 it was proposed that the FIMS data set should be divided into two groups, namely, all 13-year-old students as FIMSA data set and all Year 8 students as FIMSB data set. The main purpose of dividing the data was to investigate if there were different factors that influenced student's mathematics achievement levels between Year 8 and 13-years-old students. In other words, to examine the effects of the student's age and year level differences by analysing these data sets separately. Therefore, in this section the results of the 13-year-old students, that is FIMSA, and in the next section the Year 8 students, that is FIMSB data set, are discussed.

Tables 11.2a and 11.2b show the outer model and the inner model results for the FIMSA data set. Twelve LVs and 26 MVs were included in the model, and the results of the PLSPATH analyses are discussed in the following two parts. The first part addresses the results of the outer model and the second part considers the inner model.

Outer Model Results

Table 11.2a shows the weights and the factor loadings, the communalities, redundancies and the tolerance values of each MV within a construct, they are discussed with respect to the LV to which they contribute.

Home Background (Homeback)

Table 11.2a shows that *Homeback* was formed in an inward mode by three MVs, namely *Focc, Fed* and *Med*. The weights for these three MVs forming this antecedent construct, were 0.49, 0.47 and 0.35 respectively. It is of interest to observe that both *Focc* and *Fed* contributed more to the formation of the LV *Homeback* than *Med*. However, all were highly significant for the development of this construct. The communalities recorded in Table 11.2a show that all the MVs contributed to this construct. The redundancy values for each MV forming *Homeback* were 0.00, because *Homeback* operated as an antecedent LV and was not predicted by any other LV in the model.

Gender

The sex of the student was considered to indicate *Gender*. Thus, this LV comprised a single MV.

Year Level (Yearlevel)

Three variables *Year7, Year8,* and *Year9* formed this antecedent construct. This LV was formed in the inward mode. It can be observed from Table 11.2a that the MV *Year9* had a high weight (0.80) on the LV *Yearlevel*. This showed that *Year9* was the strongest contributor for the LV *Yearlevel.* MV *Year8* was treated as a dummy variable[1] with zero weight. *Year7* (-0.44) also contributed to this construct with a negative weight as might be expected.

Class Size (Classize)

The number of students in a mathematics class was taken to indicate *Classize*. Hence, this LV comprised just a single MV called *clssize*.

Views about Mathematics (Views)

Two MVs, namely, *Viewmath* and *Viewsch* reflected this outward mode construct. However, the latter was removed from further analysis because its loading was below the critical value of 0.30. As this latent variable consisted of only a single manifest variable it was estimated using unity mode with its loading being assigned as 1.0. The interesting point to be observed is that the views of students about their schools and school learning showed very low correlation (0.15) with their views about mathematics. This result was consistent with the result for EMS data set discussed in the previous section. This indicated that the students' views about the overall climate of the school and school learning had little effect on their views about mathematics. The only influence was their views about the methods their mathematics teachers employed to teach them, which is what the observed variable sought to measure.

---

[1] Independent variable employed to account for the effect that different levels of a nonmetric variable have in predicting the criterion latent variable

Table 11.2a FIMSA-13-year-olds-Outer Model Results

Variable		Weight/Loading	Communality	Redundancy	Tolerance
Homeback[i]	*Focc*	.49	.62	.00	.22
	Fed	.47	.70	.00	.30
	Med	.35	.40	.00	.15
Gender[u]	*Sex*	1.00	1.00	.00	.00
Yearlevel[i]	*Year7*	-.44	.42	.01	.07
	Year8	Dummy variable			
	Year9	.80	.82	.01	.07
Classize[u]	*Clssize*	1.00	1.00	.01	.00
Views[o]	*Viewmath*	1.00	1.00	.02	.00
	Viewsch	Deleted			
Values[o]	*Mathinso*	.85	.72	.03	.05
	Contrenv	.71	.51	.02	.05
Motivation[i]	*Hmwall*	.82	.74	.09	.01
	Attitsch	.51	.34	.04	.01
Timlearn[o]	*Hourmath*	.54	.29	.07	.0001
	Hourmhmw	.84	.70	.17	.0001
Aspiration[o]	*Exptedu*	.91	.83	.14	.66
	Desiredu	.91	.83	.14	.66
	Exptocc	.48	.23	.04	.88
	Desirocc	.51	.30	.05	.88
	Futmath[o]				
	Expmorma	.62	.38	.07	.11
	Wishmorm	.95	.90	.16	.11
Attitude[o]	*Belima*	.69	.47	.05	.15
	Besubma	.75	.56	.06	.15
	Diffmath	.58	.33	.04	.01
Mathachi[u]	*Rasch score*	1.00	1.00	.40	.00
Mean Communality		0.63			

i = inward mode; o = outward mode; u = unity mode

Values about Mathematics (Values)

Two MVs, *Mathinso* and *Contrenv* were combined to reflect this LV which was constructed in the outward mode. The factor loadings indicate that both MVs contributed much ($\lambda > 0.70$) to the formation of the LV Values. However, *Mathinso* (0.85) was the higher contributor in the reflection of this LV. The communality values for *Mathinso* and *Contrenv* were 0.72 and 0.51 respectively. Therefore, these two MVs would seem strong contributors to the LV Values.

Motivation towards Mathematics (Motivation)

Two MVs, namely, *Attitsch* and *Hmwall* were selected to form this LV in the inward mode. In the hypothetical model developed in Chapter 6, it was assumed to be in an outward mode, however, in the analyses, it was changed to an inward mode, in order to get better estimations from the MVs. Preliminary exploratory PLSPATH analysis suggested that both MVs contributed to the formation of this LV. The weights for the two MVs were 0.51 and 0.82 respectively. This indicated that *Hmwall* contributed more to the formation of this construct than did *Attitsch*. It is important to point out that unlike the Ethiopian students the time taken by the Australian students to do their homework in all subjects contributed more to the LV Motivation than their assessed attitudes towards school and school learning, suggesting that their behaviours were a stronger indicator of motivation than expressed attitudes. The communality values for *Hmwall* and *Contrenv* were 0.74 and 0.34 respectively.

Time in Learning (Timlearn)

The MVs *Hourmath* and *Hourmhmw* reflected this LV in the outward mode. In the hypothetical model developed in Chapter 6, it was assumed to be in an inward mode, however, in the analyses, it was changed to an outward mode, in order to get better estimations from the MVs. It can be seen in Table 11.2a that the latter (λ=0.84) was a noticeably higher contributor in the reflection of this construct than the former (λ=0.54). The communality values were 0.29 and 0.70 respectively.

Aspiration about Mathematics (Aspiration)

Four MVs, namely *Exptedu, Desiredu, Exptocc* and *Desirocc* were selected to reflect this outward mode LV. *Exptedu* involved the educational level which students expected to complete, while *Desiredu* indicated the educational level which students wished to complete. Meanwhile, *Expocc* showed the occupational status which students expected to obtain and *Desirocc* showed the occupational level at which students wished to work in their future career. The loadings indicated that both *Exptedu* and *Desiredu* were the strongest variates reflecting this LV, while the weaker variate was *Exptocc* (0.48). Table 11.2a shows that the maximum communality values were 0.83 (*Exptedu, Desiredu*) while the minimum value was 0.23 (*Expocc*).

Future Mathematics (Futmath)

Expmorma (0.62) and *Wishmorm* (0.95) reflected the construct *Futmath*, which was constructed in the outward mode. The former variable involved students' expectations to take more mathematics courses, while the latter indicated the students' wishes to take more mathematics courses. Both MVs contributed to the reflection of the LV, however, *Wishmorm* was the strongest contributor, (0.95).

Attitudes towards Mathematics (Attitude)

Three MVs namely, *Belima, Besuma* and *Diffmath* were combined to reflect this outward mode LV *Attitude*. In *Belima* students indicated whether mathematics was their best liked subject or not. Furthermore in *Besuma,* they expressed whether their mathematics test and assignment results were for their best subject. While in *Diffmath* they reflected their perceived ease of learning mathematics. The loadings show that the three MVs combined well to reflect this construct. The analysis shows that MV *Besuma* (0.75) was the strongest contributor in reflecting the construct than the other two, while the least contributor was *Diffmath* (0.58).

Mathematics Achievement (Mathachi)

This LV consists of a single MV, namely *Rasch score,* because it was in unity mode the loading and the communality were both one.

In summary, for the outer model, among the 26 hypothesised MVs that contributed to the 12 constructs, only one MV, namely *Viewsch* was removed from further analysis, because it misfitted the path model and the remaining 25 MVs contributed towards the 12 LVs. The average of the communalities of the MVs was 0.63 which indicated that the model was a sound model. The next section presents the results of the analysis of the inner model.

Inner Model Results

The results of the outer model are discussed in previous section, while in this section the results of the inner model are described. Table 11.2b shows the beta (ß), correlation and tolerance coefficients and Table 11.2c gives the direct, indirect and total effects, correlations, fit and R^2. There are 12 LVs in the inner model, and the results of the analyses of these LVs are presented in Figure 11.2.

Among the 12 LVs, *Homeback* and *Gender* were exogenous LVs, which meant that they were not influenced by any other LV. Thus, the discussion in this section considers only those ten endogenous LVs that were assumed to be influenced by another LV in the hypothesised model.

Year Level (Yearlevel)

Two LVs were hypothesised to influence this construct, however, the result of the analysis showed that only *Homeback* (0.11) influenced *Yearlevel* directly. No other factor influenced *Yearlevel* directly or indirectly. *Gender* did not play any role either directly or indirectly in influencing this LV. The R^2 (0.01) value of this LV was very small. This result revealed that those students from higher socioeconomic status backgrounds were in a higher year level than students from lower socioeconomic status backgrounds.

Class Size (Classize)

Three factors namely, *Homeback, Gender* and *Yearlevel* were hypothesised to influence this construct. However, only *Homeback* (0.11) influenced the LV *Classize* (see Figure 11.2). This meant that those students from higher socioeconomic status backgrounds were in larger class groups for mathematics. Like *Yearlevel* the explained variance (R^2=0.01) for this construct was very small.

Views about Mathematics (Views)

Four LVs were hypothesised to influence this LV. Among these factors only two LVs namely *Gender* (0.09) and *Classize* (-0.09) influenced this LV directly, while *Homeback* (-0.01) acted indirectly (see Tables 11.2b and 11.2c and Figure 11.2). The variance explained (R^2=0.02) of the construct was small. The effect of the LV *Gender* showed that female students expressed stronger *Views* about mathematics than male students. The value for *Classize* was negative which indicated that students from small class groups expressed stronger *Views* about mathematics than students from large class groups (see Tables 11.2b and 11.2c and Figure 11.2). Moreover, the indirect effect of *Homeback* (-0.01) revealed that students from a lower socioeconomic status background indirectly expressed stronger *Views* about mathematics than students from a higher socioeconomic background. The R^2 (0.02) value for this construct was small. Therefore, from the analysis it was possible to conclude that students' *Views* about mathematics were influenced by *Gender* and *Classize*. It is also important to point out that *Home background* influenced the students' *Views* about mathematics indirectly through the mediating variable, *Classize*.

Values about Mathematics (Values)

Five LVs were hypothesised to influence students' *Values* about mathematics. The result of the PLSPATH analyses revealed that *Gender* (total effect = -0.14) and *Classize* (total effect = 0.08) influenced this criterion variable both directly and indirectly, while *Views* (0.12) influenced this construct directly and *Homeback* (0.01) only indirectly (see Tables 11.2b and 11.2c and Figure 11.1). The findings indicated that:

(a) boys expressed stronger values about mathematics than girls;

(b) students in larger class groups expressed stronger values about mathematics than students in smaller class groups; and

(c) students who expressed stronger *Views* about mathematics also expressed stronger *Values* about mathematics.

Table 11.2b FIMSA-13-year-olds - Inner Model Betas

Variable	Beta	Correlation	Tolerance	R^2
Yearlevel				.01
Homeback	.11	.11	.00	
Classize				.01
Homeback	.11	.11	.00	
Views				.02
Gender	.09	.09	.0002	
Classize	-.09	-.09	.0002	
Values				.04
Gender	-.15	-.14	.01	
Classize	.09	.08	.01	
Views	.12	.10	.02	
Motivation				.12
Homeback	.09	.12	.02	
Gender	.16	.14	.02	
Yearlevel	.16	.18	.02	
Classize	.10	.15	.03	
Values	.22	.19	.03	
Timlearn				.24
Yearlevel	-.17	-.08	.03	
Motivation	.49	.47	.03	
Aspiration				.17
Homeback	.30	.32	.03	
Gender	-.08	-.09	.05	
Yearlevel	.11	-.05	.05	
Classize	.08	.14	.04	
Values	.10	.15	.08	
Motivation	.17	.21	.12	
Futmath				.18
Gender	-.12	-.12	.06	
Values	.12	.21	.07	
Motivation	.24	.29	.10	
Aspiration	.23	.31	.07	
Attitude				.12
Values	.24	.28	.05	
Futmath	.20	-.25	.05	
Mathachi				.40
Yearlevel	.49	.51	.05	
Classize	.17	.26	.04	
Motivation	.10	.27	.09	
Aspiration	.25	.28	.08	
Attitude	.09	.11	.03	
Mean R^2				.13

Table 11.2c indicated that *Homeback* (0.01) influenced this LV indirectly. This indirect influence showed that students from higher socioeconomic status backgrounds expressed stronger values about mathematics indirectly through the mediating variable *Classize* than students from lower socioeconomic backgrounds. However, it is important to point out that the effect for this LV was very small and the variance explained (R^2=0.04) of the construct was small (see Tables 11.2b and 11.2c). Therefore, from these results it would seem reasonable to conclude that 13-year-old students' values about mathematics were influenced by their *Gender*, *Classize* and *Views* about mathematics, and only indirectly by *Homeback*.

Table 11.2c FIMSA - 13-year-old students - Inner Model Statistics

Variable	Direct	Indirect	Total	Correlation	Fit	R^2
Yearlevel						.01
Homeback	.11	-	.11	.11	-	
Classize						.01
Homeback	.11	-	.11	.11	-	
Views						.02
Homeback	-	-.01	-.01	.01	.02	
Gender	.09	-	.09	.09	-	
Classize	-.09	-	-.09	-.09	-	
Values						.04
Homeback	-	.01	.01	-.002	-.02	
Gender	-.15	.01	-.14	-.14	-	
Classize	.09	-.01	.08	.08	-	
Views	.12	-	.12	.10	-	
Motivation						.12
Homeback	.09	.03	.12	.12	-	
Gender	.16	-.03	.13	.14	-	
Yearlevel	.16	-	.16	.18	-	
Classize	.10	.02	.12	.15	-	
Views	-	.03	.03	.07	.05	
Values	.22	-	.22	.19	-	
Timlearn						.24
Homeback	-	.04	.04	.09	.04	
Gender	-	.07	.07	.02	-.05	
Yearlevel	-.17	.08	-.09	-.08	-	
Classize	-	.06	.06	.07	.01	
Views	-	.01	.01	.05	.04	
Values	-	.11	.11	.08	-.02	
Motivation	.49	-	.49	.47	-	
Aspiration						.17
Homeback	.30	.02	.32	.32	-	
Gender	-.08	.01	-.07	-.09	-	
Yearlevel	.11	-.03	.08	-.05	-	
Classize	.08	.03	.11	.14	-	
Views	-	.02	.02	.003	.009	
Values	.10	.04	.13	.15	-	
Motivation	.17	-	.17	.21	-	
Futmath						.18
Homeback	-	.11	.11	.14	.034	
Gender	-.12	-.002	-.12	-.12	-	
Yearlevel	-	.02	.02	.04	.01	
Classize	-	.06	.06	.11	.04	
Views	-	.02	.02	.02	.02	
Values	.12	.08	.20	.21	-	
Motivation	.24	.04	.27	.29	-	
Aspiration	.23	-	.23	.31	-	
Attitude						.12
Homeback	-	.02	.02	.01	-.02	
Gender	-	-.06	-.06	-.09	-.03	
Yearlevel	-	-.004	-.004	-.06	-.06	
Classize	-	.03	.03	-.003	-.04	
Views	-	.03	.03	.05	.03	
Values	.24	.04	.28	.28	-	
Motivation	-	.05	.05	.11	-.001	
Aspiration	-	.05	.05	.15	.05	
Futmath	.20	-	.20	-.25	-	
Mathachi						.40
Homeback	-	.17	.17	.20	.03	
Gender	-	-.01	-.01	-.02	-.0004	
Yearlevel	.49	-.005	.49	.51	-	
Classize	.17	.04	.21	.26	-	
Views	-	.01	.01	-.04	-.03	
Values	-	.08	.08	.06	-.04	
Motivation	.10	.05	.15	.27	-	
Aspiration	.25	.004	.25	.28	-	
Futmath	-	.02	.02	.22	.06	
Attitude	.09	-	.09	.11	-	
Mean R^2						.13

Motivation towards Mathematics (Motivation)

Six LVs were hypothesised as factors that influenced students' *Motivation* towards mathematics. In Table 11.2c all the hypothesised factors are shown to influence this construct either directly, indirectly, or in both ways. *Homeback*, *Gender* and *Classize* influenced *Motivation* both directly and indirectly (see Table 11.2c and Figure 11.2), and the total effects of these variables were 0.12, 0.13 and 0.12 respectively. While *Yearlevel* (0.16) and *Values* (0.22) influenced *Motivation* only directly, *Values* (0.22) showed a greater direct effect on this construct than did the other LVs. However, *Views* influenced this construct only indirectly. The variance (R^2=0.12) explained for this construct was medium. In general, the findings can be summarised as follows:

(a) students from higher socioeconomic status backgrounds expressed stronger *Motivation* towards mathematics than students from lower socioeconomic backgrounds;

(b) girls expressed stronger *Motivation* towards mathematics than boys;

(c) students in higher grades expressed stronger *Motivation* towards mathematics than did students in lower grades;

(d) students from larger class groups expressed stronger *Motivation* towards mathematics than did students from smaller class groups; and

(e) students who expressed stronger values about mathematics also expressed stronger *Motivation* towards mathematics.

Therefore, student factors that influenced students' *Motivation* towards mathematics directly were their *Home Background, Gender, Class Size, Year Level, Values* about mathematics and *Views* only indirectly.

Time in Learning (Timlearn)

Seven LVs were hypothesised to influence this construct. However, the results of the analysis revealed that only *Yearlevel* (direct effect = -0.17, indirect effect =0.08, total effect =-0.09) showed direct and indirect effects while *Motivation* (0.49) showed a direct effect on this construct (see Tables 11.2c and Figure 11.2). Meanwhile, the remaining LVs showed indirect effects acting through the mediating variable *Motivation* (see Tables 11.2c). The R^2 (0.24) value for this construct was medium. Therefore, these results revealed that the time taken by students in learning mathematics was influenced directly by *Yearlevel* and *Motivation* towards mathematics and only indirectly by *Homeback, Gender, Classize, Views and Values.*

Aspiration towards Mathematics (Aspiration)

Homeback, Gender, Yearlevel, Classize, Views, Values, Motivation and *Timlearn* were hypothesised to influence *Aspiration* towards mathematics. Like *Motivation* and *Timlearn*, all the hypothesised factors showed either direct, indirect or both direct and indirect effects on this construct *Aspiration*. With the exception of *Motivation* (direct effect = 0.17) and *Views* (indirect effects = 0.02) all the other factors had both direct and indirect effects on *Aspiration* (see Tables 11.2c and Figure 11.2). *Homeback* (0.30) showed the highest direct effect and *Views* (0.02) the lowest indirect effect on *Aspiration*.

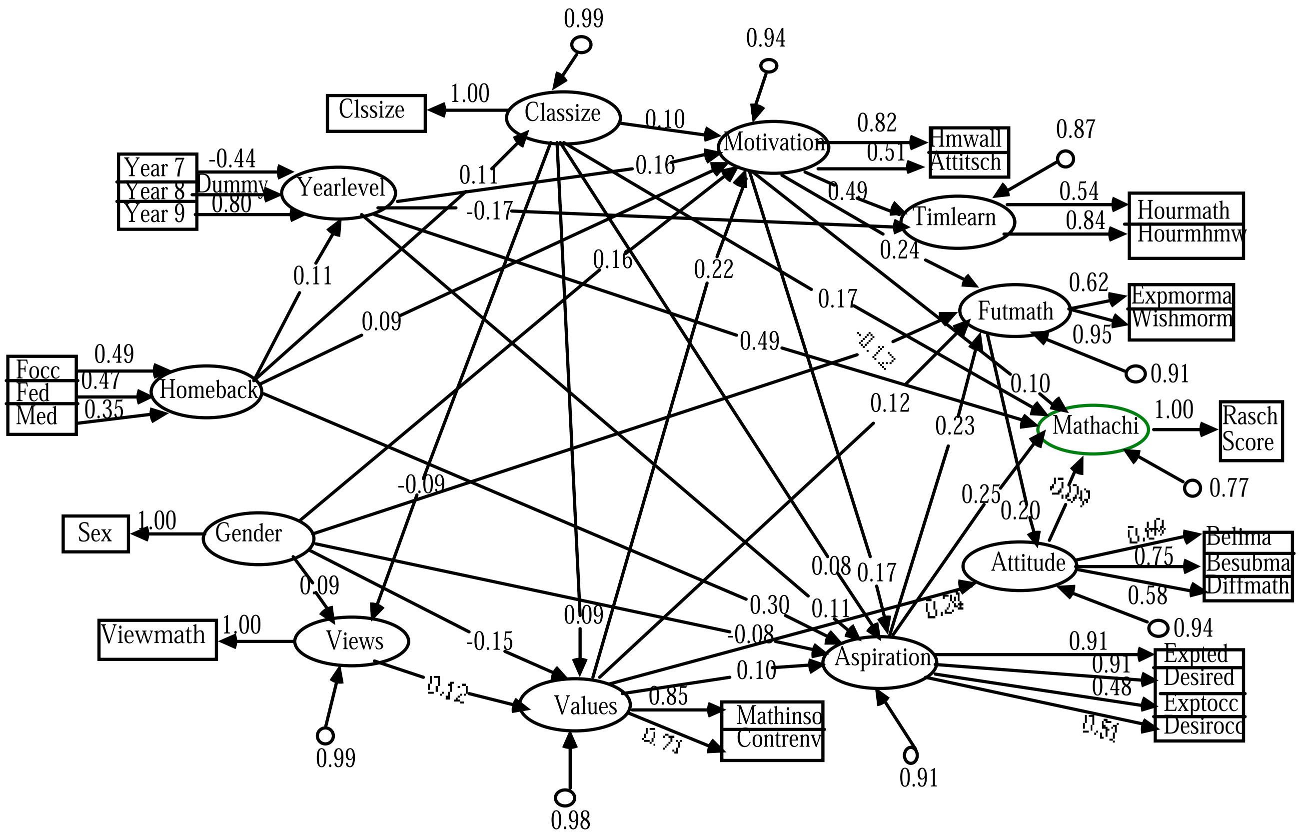


Figure 11.2 FIMSA - Student level factors influencing Mathematics Achievement

These direct and indirect effects on *Aspiration* revealed the following important points:

(a) students from higher socioeconomic status backgrounds showed higher *Aspiration* than students from other socioeconomic backgrounds;

(b) male 13-year-old students had stronger aspirations than female students;

(c) older students showed greater aspiration than younger students;

(d) students from larger class groups showed higher levels of aspiration than students from smaller class groups;

(e) students who showed stronger values also showed stronger aspiration; and

(f) more highly motivated students also showed higher levels of aspiration than other students.

Generally, *Homeback, Gender, Yearlevel, Classize, Values*, and *Motivation* were student level factors that influenced *Aspiration,* and *Views* had an effect only indirectly.

Future Mathematics (Futmath)

Nine LVs were hypothesised to influence this construct. The result of the PLSPATH analysis demonstrated that eight of the nine hypothesised factors directly or indirectly influenced this LV. The only LV that did not show either direct or indirect effects on this construct was *Timlearn* (Time in Learning). *Gender*, *Values* and *Motivation* showed both direct and indirect effects, while *Aspiration* showed only a direct effect. Meanwhile *Homeback*, *Yearlevel*, *Classize* and *Views* showed indirect effects through the mediating variable *Aspiration.* The largest total effect for this LV was from *Motivation* (0.27), while the smallest total effect was from *Yearlevel* (0.02) and *Views* (0.02). The explained variance for this LV was 0.18. Hence, the summary of the findings are:

(a) boys were more interested in taking more mathematics courses than girls ;

(b) students who expressed stronger *Values* about mathematics were more interested to take more mathematics courses than those students who expressed weaker *Values*;

(c) students who expressed stronger *Motivation* towards mathematics were also more interested to take more mathematics courses than students who expressed less *Motivation*; and

(d) students who showed higher *Aspiration* were also more interested to take more mathematics courses than students who expressed lower *Aspiration*.

Thus, *Gender, Values, Motivation* and *Aspiratio*n were student level factors that influenced *Future Mathematics*, while *Homeback, Yearlevel, Classize,* and *Views* influenced it only indirectly.

Attitudes towards Mathematics (Attitude)

Ten LVs were hypothesised to influence students' *Attitudes towards Mathematics*. The result of the PLSPATH analysis demonstrated that except for *Timlearn* all the remaining hypothesised LVs showed direct or indirect effects on this construct. It is interesting to observe that seven out of nine factors showed their effects only indirectly, and did not have any direct effect on this construct. The only factors that had a direct effect on *Attitudes towards Mathematics* were *Futmath* (0.20), and *Values,* which showed both direct (0.24) and indirect (0.04) effects. The total effect of *Values* was 0.28 (see Tables 11.2c and Figure 11.2). Thus, students who expressed stronger values about mathematics showed more positive attitudes towards mathematics than students who expressed weaker values and students who wished and showed a desire to take more mathematics courses showed more positive attitudes towards mathematics than students who showed less desire to take more mathematics courses. The interesting point here was that *Gender* did not have a direct effect on *Attitudes towards Mathematics*. However, the indirect effects indicated that there was an indirect *Gender* effect (-0.06) on this LV. Thus, boys held more favourable attitudes towards mathematics than girls. Previous research findings have revealed that boys expressed more positive attitudes towards mathematics than did girls (Keeves, 1972; Fraser, 1980; Schofield, 1981; Ballenden et al., 1985).

Mathematics Achievement (Mathachi)

Students' mathematics achievement level was hypothesised to be influenced by 11 LVs. The result of the PLS analysis revealed that four of the 11 factors influenced *Mathematics Achievement* both directly and indirectly, while five other factors influenced *Mathematics Achievement* indirectly (see Tables 11.2b and 11.2c and Figure 11.2) and one factor showed only a direct effect. The only factor that did not show any effect on the outcome variable was *Time in Learning*.

Direct Effects

The five factors that had a direct influence on *Mathematics Achievement* were *Yearlevel, Classize, Motivation, Aspiration and Attitude*.

Year Level (Yearlevel)

This LV influenced the mathematics achievement level of 13-year-old students directly (0.49) and the indirect influence was small. The total effect of the variable on *Mathematics Achievement* was also 0.49. *Year level* was the strongest of the factors that influenced this construct. This result indicated that students in a higher grade were likely to achieve at a higher level in mathematics than students in a lower grade within this 13-year-old age sample.

Class Size (Classize)

This was a LV that influenced *Mathematics Achievement* of students directly (0.17) and indirectly (0.04). The total effect was 0.21. This variable indicated that students from larger class groups achieved at a higher level in mathematics than students from smaller class groups. The question that must be asked here is how large is large? It seems important to undertake further study to recommend the maximum and the minimum number of students in mathematics classes for effective teaching and learning processes to take place.

Motivation towards Mathematics (Motivation)

Motivation was another factor showing direct (0.10) and indirect (0.05) effects on *Mathematics Achievement.* It was the fourth strongest variable to have an effect on this LV. The total effect of *Motivation* on this outcome variable was 0.15. The results indicated that highly motivated students towards mathematics were likely to achieve at a higher level in mathematics than less motivated students (see Table 11.2c and Figure 11.2).

Aspiration towards Mathematics (Aspiration)

Aspiration also had direct (0.25) and indirect (0.004) effects on *Mathematics Achievement.* It was the second strongest variable that had a direct effect on this outcome variable, the total effect was 0.25. The variable indicated that those students who expressed greater *Aspiration* towards mathematics were also higher achievers in the same subject (see Table 11.2c and Figure 11.2) than students who expressed less *Aspiration*.

Attitudes towards Mathematics (Attitude)

This LV influenced Mathematics Achievement directly. It was the least strong (ß = 0.09) variable to have a direct effect on *Mathematics Achievement* (see Tables 11.2b and 11.2c and Figure 11.2). The evidence showed that students who expressed more positive attitudes towards mathematics were higher achievers in mathematics than students who expressed less positive attitudes. This observation was consistent with previous research findings (Keeves, 1972; Schofield, 1981; Milne, 1992).

Indirect Effects

Homeback (0.17), *Gender* (-0.01), *Views* (0.01),*Values* (0.08) and *Futmath* (0.02) had indirect effects on mathematics achievement (see Table 11.2c). Among the factors that showed strongest indirect effects on this outcome measure were, *Homeback* (0.17), which was even higher than the direct effects of *Motivation* (0.10) and *Attitudes towards Mathematics* (0.09). This factor was mediated largely by *Aspiration*. Previous research findings on the effect of socioeconomic status on mathematics achievement have indicated that students from higher socioeconomic status family backgrounds were likely to achieve at a higher level than their class mates from lower socioeconomic backgrounds (Keeves, 1968; Rosier, 1980; Ainley et al., 1990). However, the results of this analysis indicated that this effect operated only indirectly through *Aspiration* and not directly as other analyses with less carefully specified models have contended.

Gender

Another factor that had an indirect effect on the outcome variable was *Gender*. Like *Homeback*, this factor was mediated by *Aspiration,* however, the effect was very small (-0.01). This small indirect effect revealed that boys were likely to achieve only slightly higher in mathematics than girls when other factors were taken into account. Previous research findings in Australia showed that sex difference in mathematics started to emerge in the lower secondary school. Researchers such as Fitzpatrick (1978), Keeves and Mason, (1980), Moss (1982), Carss (1980), Leder (1989, 1990), Willis (1989), Leder and Forgasz (1991) have argued that sex-related differences did not actually show up before the junior secondary school years and when they emerged, they mostly favoured boys. Therefore, this small indirect effect of gender for mathematics achievements would seem to indicate the beginning of a gender difference in achievement at the junior secondary school level. However, there is little sign of significant direct effect that Keeves (1968) reported from the analyses of the same data.

Values

Values was the only other factor that had recognisable indirect effects on this outcome measure. The total contribution of this indirect effect was 0.08 operating largely through *Aspiration.* The R^2 value for this outcome variable was 0.40 and the mean R^2, that indicated the model fit, was 0.13. The relatively high proportion of variance explained of the criterion variable *Mathematics Achievement* in this analysis was a consequence of the fact that an age-based sample was under survey, with substantial variation in the mathematics curricula across grade levels.

Conclusion

Twenty five MVs and 11 LV were hypothesised to influence the mathematics achievement level of 13-year-old students who participated in the First International Mathematics Study in Australia. Among the 25 MVs one was removed from further analysis, because the loading of *Views*ch was below the critical value of 0.30.

The results of the analysis revealed that among the 11 hypothesised LVs only five of them were identified as student level factors that directly influenced the mathematics achievements of 13-year-old students in Australia in 1964. These student level factors were *Yearlevel*, *Classize, Motivation*, *Aspiration* and *Attitude towards Mathematics*. While *Homeback, Gender, Views, Values* and *Futmath* influenced the criterion variable only indirectly. The mean of the R^2 values of the endogenous variables of the inner model was 0.13, which showed that the model was not a strong model. However, 40 per cent of the variance of Mathematics Achievement, the criterion variable, was explained by the predictor variables in the model.

Results of the FIMSB Data Set

In the previous section the results of the PLS analysis from the FIMSA (13-year-old students) data set are discussed, while in this section the results of the analysis from the FIMSB (Year 8 students) data set are addressed. Tables 11.3a and 11.3b show the outer model and the inner model results for the FIMSB data set which included all Year 8 students. Twelve LVs and 24 MVs were included in the path model. Thus the results of the PLSPATH analyses are discussed in the following two parts. The first part presents the results for the outer model and the second part discusses the inner model.

Outer Model Results

The weights and the factor loadings, the communality, redundancy and the tolerance values of each MV associated with a construct, are discussed with respect to the LV to which it contributes.

Home Background (Homeback)

Table 11.3a indicated that *Homeback* was formed in an inward mode and built by three MVs, namely *Focc, Fed* and *Med*. The weights for these three MVs forming this antecedent construct, were 0.44, 0.54 and 0.31 respectively.

Gender

This LV comprised a single MV called *Sex*.

Student Age (Studage)

The age of the student was considered initially for *Studage* in this causal model. However, this MV and the corresponding LV were removed from the analysis by default, because the LV formed by this MV did not contribute directly or indirectly to the path model. Commonly the younger students did better because they were brighter but there were shared relationships with other variables in the model. These effects generally cancelled out in a grade level sample. It should be noted that the sample under consideration was a grade sample with all students coming from the eighth grade or year level. Under these circumstances it was perhaps to be expected that the effects of a variable involving student age would be slight or non-existent.

Table 11.3a FIMSB (Year 8 Students) - Outer Model Results

Variable		Weight/Loading	Communality	Redundancy	Tolerance
Homeback[i]	*Focc*	.44	.59	.00	.24
	Fed	.54	.77	.00	.33
	Med	.31	.39	.00	.17
Gender[u]	*Sex*	1.00	1.00	.00	.00
Studage[u]	*Age*	Deleted by default			
Classize[u]	*Clssize*	1.00	1.00	.00	.00
Views[o]	*Viewmath*	1.00	1.00	.00	.00
	Viewsch	Deleted			
Values[o]	*Mathinso*	.84	.71	.02	.05
	Contrenv	.70	.50	.02	.05
Motivation[i]	*Hmwall*	.45	.89	.12	.69
	Attitsch	.60	.94	.13	.69
Timlearn[o]	*Hourmath*	.98	.96	.07	.87
	Hourmhmw	.99	.97	.07	.87
Aspiration[o]	*Exptedu*	.60	.36	.06	.47
	Desiredu	.84	.70	.11	.02
	Exptocc	Deleted			
	Desirocc	.59	.35	.06	.47
Futmath[o]	*Expmorma*	.72	.52	.11	.16
	Wishmorm	.92	.85	.18	.16
Attitude[o]	*Belima*	.70	.49	.06	.17
	Besuma	.75	.57	.07	.16
	Diffmath	.59	.35	.04	.01
Mathachi	*Rasch Score*	1.00	1.00	.22	.00
Mean Communality			0.71		

i = inward mode; o = outward mode; u = unity mode

Class Size (Classize)

This LV comprised a single MV called *Clssize*.

Views about Mathematics (Views)

Two MVs, namely, *Viewmath* and *Viewsch* reflected this outward mode construct. However, the latter was removed from further analysis, because its loading was below the critical value of 0.30. This MV was also removed from further analysis in the previous analyses, namely EMS and FIMSA. After the deletion of *Views*ch, *Viewmath* became the only MV to reflect this construct. As this latent variable consisted of only a single manifest variable it was estimated using unity mode.

Values about Mathematics (Values)

Two MVs, *Mathinso* and *Contrenv* combined in the reflective or outward mode to reflect this LV. The factor loadings indicate that both MVs had high loadings on the LV *Values*. However, *Mathinso* (0.84) dominated this construct (see Table 11.3a). The communality values were 0.71 and 0.50 respectively.

Motivation towards Mathematics (Motivation)

The two MVs, namely, *Attitsch* and *Hmwall* were selected to form this inward mode LV. Both MVs contributed to the formation of this LV. In the hypothetical model developed in Chapter 6, it was assumed to be in an outward mode, however, in the analyses, it was changed to an inward mode, in order to get better estimations for the MVs. The weights for the two MVs were 0.60 and 0.45 respectively. The tolerance value was 0.69 that indicates a clear sign of multicollinearity, but without deleterious effects being evident.

Time in Learning (Timlearn)

The MVs *Hourmath* and *Hourmhmw* reflected this LV in the outward mode. In the hypothetical model developed in Chapter 6, it was assumed to be in an inward mode, however, in the analyses, it was changed to an outward mode, in order to get better estimations for the MVs. Both MVs strongly contributed to reflect this construct..

Aspiration towards Mathematics (Aspiration)

Exptedu, Desiredu, Exptocc and *Desirocc* were selected to reflect this outward mode LV. *Exptoc*c was deleted because the loading was below the critical value of 0.30 (Campbell, 1996). As with the FIMSA data set the loadings revealed that both *Exptedu* and *Desiredu* were the strongest factors in this LV, while the least effective variate was *Desirocc* (0.59).

Future Mathematics (Futmath)

Expmorma (0.72) and *Wishmorm* (0.92) reflected this outward mode construct *Futmath*. The former variable involved students' expectations to take more mathematics courses, while the latter indicated their wishes to take more mathematics courses. Both MVs reflected the LV. However, *Wishmorm* was the stronger contributor, when compared with *Expmorma*.

Attitudes towards Mathematics (Attitude)

Three MVs namely, *Belima, Besuma*, and *Diffmath* were combined to reflect the LV *Attitude* in the outward mode. In *Belima* students indicated whether or not mathematics was their best liked subject, while in *Besuma*, they expressed whether or not their mathematics test results were for their best subject. Meanwhile, in *Diffmath* they reflected their perceived ease of learning mathematics. The loadings showed that the three MVs combined well to reflect this construct. The MV *Besuma* (0.75) was the strongest contributor compared to the other two, while the least reflector was *Diffmath* (0.59).

Mathematics Achievement (Mathachi)

This LV comprised a single MV, namely *Rasch score*. The loadings and the communality values were 1.00.

In summary, for the outer model, among the 24 hypothesised MVs that constituted the 12 constructs, only three MVs subsequently were removed from further analysis, because of their failure to contribute to the model while, the remaining 21 MVs contributed in forming the 12 LVs. The outer model was a good model because the average communalities value of the MVs which is considered as a measure of a good outer model was 0.71. The next section presents the results for the inner model.

Inner Model Results

Table 11.3b shows the beta (ß), correlation and tolerance values and Table 11.3c presents the direct, indirect and total effects, correlations, fit and R^2 values for the results of the analyses. There are initially 12 LVs in the inner model, and the results obtained from the analyses for these LVs are presented in Figure 11.3.

Among the 12 LVs, *Homeback*, *Gender* and *Studage* were exogenous LVs. Thus, the discussion in this section considers only those nine endogenous LVs that were assumed to be influenced by another LV in the hypothesised model.

Table 11.3b FIMSB YEAR 8 Students - Inner Model Betas

Variable	Beta	Correlation	Tolerance	R^2
Classize				.01
Homeback	.09	.09	.00	
Views				.02
Gender	.09	.07	.002	
Classize	-.09	-.09	.002	
Values				.03
Gender	-.15	-.14	.01	
Classize	.08	.06	.01	
Views	.10	.08	.02	
Motivation				.14
Homeback	.33	.34	.01	
Classize	.08	.12	.01	
Values	.12	.14	.01	
Timlearn				.08
Homeback	.09	.17	.11	
Motivation	.23	.26	.11	
Aspiration				.16
Gender	.18	.15	.03	
Views	.08	.13	.02	
Values	.19	.21	.04	
Motivation	.27	.29	.02	
Futmath				.21
Gender	-.08	-.08	.06	
Values	.12	.22	.08	
Motivation	.27	.36	.10	
Aspiration	.24	.34	.15	
Attitude				.13
Values	.23	.28	.05	
Futmath	.23	.28	.05	
Mathachi				.22
Homeback	.08	.21	.12	
Classize	.19	.24	.02	
Motivation	.29	.39	.22	
Futmath	.10	.27	.19	
Attitude	.09	.18	.09	
Mean R^2				.11

Class Size (Classize)

Three factors namely, *Homeback*, *Gender* and *Studage* were hypothesised to influence this construct. However, only *Homeback* (0.09) had a recognisable effect. This indicates that those students from a higher socioeconomic status background were in larger classes for mathematics than students from a lower socioeconomic status background.

However, the variance explained (R^2=0.01) for this construct was very small.

Table 11.3c FIMSB YEAR 8 Students - Inner Model Effects (On)

Variable	Direct	Indirect	Total	Correlation	Fit	R^2
Classize						.01
Homeback	.09	-	.09	.09	-	
Views						.02
Homeback	-	-.01	-.01	.004	.01	
Gender	.09	-	.09	.09	-	
Classize	-.09	-	-.09	-.09	-	
Values						.03
Homeback	-	.01	.01	.04	04	
Gender	-.15	.01	-.14	-.14	-	
Classize	.08	-.01	.07	.06	-	
Views	.10	-	.10	.08	-	
Motivation						.14
Homeback	.33	.01	.34	.34	-	
Gender	-	-.02	-.02	-.05	-.03	
Classize	.08	.01	.09	.12	-	
Views	-	.01	.01	.06	.06	
Values	.12	-	.12	.14	-	
Timlearn						.08
Homeback	.09	.08	.17	.17	-	
Gender	-	-.004	-.004	.01	.01	
Classize	-	.02	.02	-.01	-.04	
Views	-	.003	.003	-.03	-.03	
Values	-	.03	.03	.03	.002	
Motivation	.23	-	.23	.26	-	
Aspiration						.16
Homeback	-	.09	.09	.12	.03	
Gender	.18	-.02	.16	.15	-	
Classize	-	.03	.03	.08	.04	
Views	.08	.02	.10	.13	-	
Values	.19	.03	.23	.21	-	
Motivation	.27	-	.27	.29	-	
Futmath						.21
Homeback	-	.11	.11	.18	.07	
Gender	-.08	.02	-.06	.08	-	
Classize	-	.04	.04	.11	.06	
Views	-	.04	.04	.07	.04	
Values	.12	.09	.21	.22	-	
Motivation	.27	.07	.33	.36	-	
Aspiration	.24	-	.24	.34	-	
Attitude						.13
Homeback	-	.03	.03	.04	.01	
Gender	-	-.05	-.05	-.10	-.05	
Classize	-	.02	.02	.01	-	
Views	-	.03	.03	.07	.05	
Values	.23	.05	.28	.28	-	
Motivation	-	.08	.08	.18	.07	
Aspiration	-	.06	.06	.17	.05	
Futmath	.23	-	.23	.28	-	
Mathachi						.22
Homeback	.08	.13	.20	.21	-	
Gender	-	-.02	-.02	-.04	-.02	
Classize	.19	.03	.22	.24	-	
Views	-	.01	.01	-.05	-.03	
Values	-	.08	.08	.10	-.005	
Motivation	.29	.04	.33	.39	-	
Aspiration	-	.03	.03	.14	-.01	
Futmath	.10	.02	.13	.27	-	
Attitude	.09	-	.09	.18	-	
Mean R^2						.11

Views about Mathematics (Views)

Four LVs, namely *Homeback*, *Gender*, *Studage* and *Classize* were hypothesised to influence this LV. However, only three factors showed an effect on this criterion variable. *Gender* (0.09) and *Classize* (-0.09) influenced the LV directly, while

Homeback (-0.01) influenced this LV only indirectly (see Tables 11.3b and 11.3c and Figure 11.3). The variance explained (R^2=0.02) for this variable was very small. The analyses indicated that girls showed stronger *Views* than boys, and students from smaller classes indicated stronger *Views* than students from larger classes (see Table 11.3c and Figure 11.3). Moreover, the very weak indirect effect of *home background* revealed that students from lower socioeconomic status backgrounds indirectly exhibited stronger *Views* than students from higher socioeconomic status backgrounds (-0.01). Therefore, the *Views* of Year 8 Australian students about mathematics were influenced by *Gender* and *Class Size*, and only indirectly and slightly by their *Home Background*.

Values about Mathematics (Values)

Five LVs, namely *Homeback*, *Gender*, *Studage, Classize* and *Views* were hypothesised to influence this LV. The results of the PLSPATH analyses revealed that *Gender* and *Classize* showed both direct and indirect effects while *Views* had a direct effect. The only factor that showed an indirect effect on this criterion variable was *Homeback* (see Tables 11.3b and 11.3c and Figure 11.3). However, *Studage* did not show either a direct or an indirect effect on *Values*. The findings showed that at the Year 8 level: (a) boys expressed stronger values than girls; (b) students in larger class groups expressed stronger values than students from smaller class groups; and (c) students who expressed stronger views also expressed stronger values. Table 11.3c indicates that *Homeback* (0.01) influenced indirectly *Values* about mathematics, this means that students from higher socioeconomic status backgrounds expressed stronger values which were indirectly influenced through the mediating variable *Classize*. However, it is important to point out that the effect (0.01) on this criterion variable was very small. The variance explained for the construct was 0.03 (see Table 11.3c). Therefore, from this result it would seem reasonable to conclude that Year 8 students' *Values* about mathematics were influenced by their *Gender, Classize,* and *Views* about mathematics and only indirectly and to a slight extent by *Home Background*.

Motivation towards Mathematics (Motivation)

Six LVs, namely *Homeback*, *Gender*, *Studage, Classize, Views* and *Values* were hypothesised to influence this LV. It is important to note that except *Studage*, all the hypothesised factors namely *Homeback*, *Gender*, *Classize*, *Views* and *Values* influenced this LV either directly, indirectly or both directly and indirectly. *Homeback* and *Classize* had both direct and indirect influences (see Tables 11.3b and 11.3c and Figure 11.3), and their direct effects were 0.33 and 0.08 respectively. While *Values* (0.12) influenced *Motivation* directly, *Views* (0.01) and *Gender* (-0.02) influenced *Motivation* indirectly.

Homeback had a greater effect on this construct than the other LVs. The variance explained for this construct of *Motivation* was 0.14. In general, the findings revealed that, at the Year 8 level:

(a) students from higher socioeconomic status backgrounds showed stronger motivation towards mathematics than students from lower socioeconomic back grounds;

(b) students from larger class groups indicated stronger motivation than students from smaller class groups; and

(c) students who expressed stronger values also indicated stronger motivation.

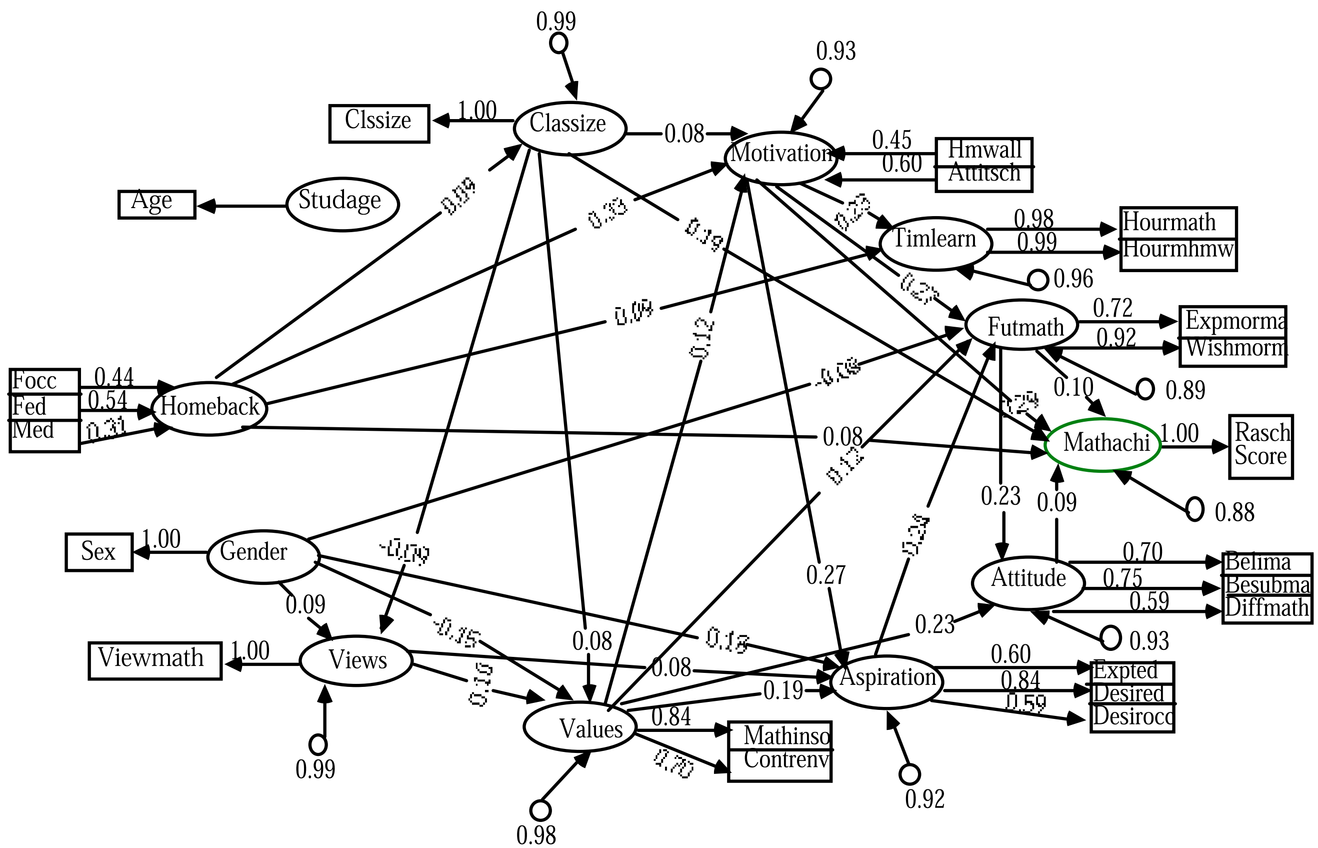


Figure 11.3 FIMSB - Student level factors influencing Mathematics Achievement

Meanwhile, the indirect effects on this criterion variable showed that:

(a) boys expressed stronger motivation towards mathematics than girls; and

(b) students who indicated stronger views also expressed stronger motivation.

Time in Learning (Timlearn)

Seven LVs, namely *Homeback*, *Gender*, *Studage, Classize*, *Views, Values* and *Motivation* were hypothesised to influence this LV. Like *Motivation* all the hypothesised LVs except *Studage* showed either direct, indirect or both direct and indirect effects on *Time in Learning*. The results of the analysis revealed that *Homeback* showed direct and indirect effects on this construct (see Tables 11.3b and 11.3c and Figure 11.3), and its total effect was 0.17. Moreover, *Motivation* indicated only a direct effect of 0.23. The variance explained (R^2=0.08) for this criterion variable was small. The results showed that at the Year 8 level in Australian schools: (a) students from higher socioeconomic status backgrounds; and (b) highly motivated students spent more time in learning mathematics than students from lower socioeconomic status backgrounds and less motivated students towards mathematics. Meanwhile, *Gender* (-0.004), *Classize* (0.02), *Views* (0.003) and *Values* (0.03) influenced this variable only indirectly through the mediating variable *Motivation* (see Table 11.3c).

Aspiration towards Mathematics (Aspiration)

Homeback, *Gender*, *Studage*, *Classize*, *Views*, *Values*, *Motivation* and *Timlearn* were hypothesised to influence *Aspiration* towards mathematics. All the hypothesised LVs except *Studage* and *Timlearn* showed either direct, indirect or both a direct and an indirect effect on this construct. *Gender* (total effect = 0.16), *Values* (total effect = 0.23) and *Views* (total effect = 0.10) showed both direct and indirect effects on *Aspiration. Motivation* (0.27) showed only a direct effect on the construct. Meanwhile, *Homeback* (0.09) and *Classize* (0.03) showed indirect effects on *Aspiration* (see Tables 11.3b and 11.3c and Figure 11.3). The R^2 (0.16) value for this construct was medium. Hence the results discussed above are summarised as follows:

(a) girls showed more *Aspiration* towards mathematics than boys;

(b) students who held stronger *Views* also expressed stronger *Aspiration*;

(c) students who held stronger *Values* also expressed stronger *Aspiration.*; and

(d) students who indicated stronger *Motivation* also showed stronger *Aspiration*.

Therefore, student factors that influenced *Aspiration* towards mathematics were *Gender, Views, Values* and *Motivation,* while *Home Background* and *Class Size* influenced this LV only indirectly.

Future Mathematics (Futmath)

Nine MVs, namely *Homeback*, *Gender*, *Studage, Classize, Views, Values*, *Motivation, Timlearn* and *Aspiration* were hypothesised to influence this LV. The result of the PLSPATH analysis demonstrated that seven of the nine hypothesised factors showed a direct, an indirect or both direct and indirect effects on this LV. Among these variables *Gender, Values* and *Motivation* showed both direct and indirect effects, while *Aspiration* showed only a direct effect. Meanwhile *Homeback*, *Classize* and *Views* showed only indirect effects (see Tables 11.3b and 11.3c and Figure 11.3). The highest total effect on this LV was from *Motivation* (0.33), while the lowest was from *Classize* and *Views* 0.04 each.

The variance explained (R^2=0.21) for this construct was moderate. A summary of the findings of factors influencing *Future* of studying Mathematics at the Year 8 level:

(a) boys were more interested in taking of more mathematics courses in the future than girls;

(b) students who expressed stronger values were more interested in taking mathematics courses than students who expressed weaker values about mathematics;

(c) students who were more motivated towards mathematics were also interested in taking mathematics courses in the future; and

(d) students who showed stronger *Aspiration* were also more interested to take further mathematics courses.

It is also important to point out that *Homeback, Classize*, and *Views* influenced this variable indirectly.

Attitudes towards Mathematics (Attitude)

Ten MVs, namely *Homeback*, *Gender*, *Studage, Classize, Views, Values*, *Motivation, Timlearn, Aspiration* and *Futmath* were hypothesised to influence this LV. The results of the PLSPATH analysis demonstrated that except *Studage* and *Timlearn* all the remaining hypothesised factors influenced either directly, indirectly or both directly and indirectly this LV. It is interesting to observe that only *Values* (total effect=0.28) influenced both directly and indirectly this criterion variable while *Future Mathematics* (0.23) influenced *Attitude* only directly (see Tables 11.3b and 11.3c and Figure 11.3). The R^2 (0.13) value for this criterion variable was medium. Thus, at the Year 8 level students who expressed stronger values also showed more positive attitudes towards mathematics, and students who wished and showed a desire to take more mathematics courses also expressed more positive attitudes towards mathematics.

Therefore, student factors that influenced attitudes towards mathematics were *Values,* and *Future Mathematics,* meanwhile *Home Background, Gender, Class Size, Views*, *Motivation* and *Aspiration* influenced this construct only indirectly.

Mathematics Achievement (Mathachi)

Eleven LVs, namely *Homeback*, *Gender*, *Studage, Classize, Views, Values*, *Motivation, Timlearn, Aspiration, Futmath* and *Attitude* were hypothesised to influence students' level of achievement in mathematics. The results of the PLSPATH analyses revealed that four of the 11 factors influenced *Mathematics Achievement* directly and indirectly, while one factor influenced the criterion variable only directly. However, four other factors influenced *Mathematics Achievement* indirectly (see Tables 11.3b and 11.3c and Figure 11.3). The factors that did not have any effect on the outcome measure were *Student Age* and *Time in Learning.*

Direct Effects

The factors that had direct influence on *Mathachi* are discussed in greater detailed below.

Home Background (Homeback)

Home Background was an antecedent variable that both directly and indirectly influenced mathematics achievement. The total effect was 0.20. This variable showed that students from higher socioeconomic backgrounds were more likely to achieve at a higher level in mathematics than students from lower socioeconomic status backgrounds. This finding was consistent with previous Australian research findings of the effect of socioeconomic status on mathematics achievement. These findings indicated that students from higher socioeconomic status family backgrounds were likely to achieve at a higher level than their classmates from lower socioeconomic backgrounds (Keeves, 1968; Rosier, 1980; Ainley et al., 1990).

Class Size (Classize)

Class size is the other LV that influenced Mathematics Achievement of students both directly and indirectly. The total effect was 0.22. This variable demonstrated that students from larger class groups achieved at a higher level in mathematics than students from smaller class groups. This is a somewhat surprising result, that would seem to require further careful examination.

Motivation towards Mathematics (Motivation)

This variable was another factor that showed a direct and an indirect effect on *Mathematics Achievement*. It was the strongest factor that indicated an effect on this outcome measure. The total effect of *Motivation* towards this outcome measure was 0.33. The variable indicated that more highly motivated students were likely to achieve at a higher level in mathematics than students with less *Motivation* towards mathematics (see Table 11.3c and Figure 11.3).

Future Mathematics (Futmath)

This LV also indicated a direct and an indirect effect on *Mathematics Achievement*. It was the fourth strongest variable that was found to have an effect on this outcome measure. Its total effect was 0.13. The variable indicated that those students who wished and showed a desire to take more mathematics courses were likely to achieve higher in mathematics (see Tables 11.3b and 11.3c and Figure 11.3) than students who showed less desire to take more mathematics courses.

Attitude towards Mathematics (Attitude)

This LV influenced *Mathematics Achievement* directly (see Tables 11.3b and 11.3c and Figure 11.3). The effect on the outcome measure was 0.09. The evidence showed that students who expressed more positive attitudes towards mathematics were likely to achieve at a higher level in mathematics than their classmates who expressed less positive attitudes towards mathematics. This observation was consistent with previous Australian research findings (Keeves, 1972; Schofield, 1981; Milne, 1992).

Indirect Effects

An interesting feature of the analysis was that Gender did not influence Mathematics Achievement directly. It influenced *Mathachi* indirectly through the mediating variable Future Mathematics, however, the effect was very small (-0.02). This small indirect effect revealed that boys were likely to achieve higher in mathematics than girls after other factors had been taken into account. Previous research findings in Australia have suggested that sex differences in mathematics achievement started to emerge at the lower secondary school level. Researchers such as Fitzpatrick (1978), Keeves and Mason, (1980), Moss (1982), Carss (1980), Leder (1989), Willis (1989), Leder (1990), and Forgasz (1991) have argued that sex-related differences did not actually show up before the junior secondary school years and when they emerged, they mostly favoured boys. Consequently the findings of this study indicated that sex differences were possibly starting to emerge in 1964 at the junior secondary school level in favour of the boys, but operating through expectation to participate in Mathematics courses in the future.

Values, Views and *Aspiration* were the other factors that showed indirect effects on this outcome measure of *Mathematics Achievement*. These indirect effects were 0.08, 0.01 and 0.03 respectively. The variance explained for this outcome measure was 0.22, which is a rather smaller proportion of the total variance explained than was recorded for the previous analysis with the 13-year-old sample. The effects of grade based curricula are seen in this reduced proportion of variance explained which involved a grade level sample rather than an age sample.

Conclusions

Twenty-four MVs and 11 LVs were hypothesised to influence the Mathematics Achievement of Year 8 students in FIMSB. From the 24 MVs only two were deleted from further analysis. *Viewsch* and *Exptocc* were deleted, because these MVs failed to contribute to or reflect their respective LVs. Age was removed from the analysis by default, because the LV formed by this MV did not contribute directly or indirectly to the path model. The results of the analyses showed that among the 11 hypothesised predictors only five were identified as student level factors that influenced directly the Mathematics Achievement of students at the Year 8 level. The factors identified were *Homeback, Classize, Motivation, Futmath* and *Attitude*. While *Gender, Views, Values* and *Aspiration* influenced the *Mathematics Achievement* level of Year 8 students only indirectly. The average R^2 of the variance explained for the endogenous variables, which was used to measure the strength of the inner model, was 0.11, and 22 per cent of the variance of *Mathematics Achievement*, the criterion variable, was explained by the predictor variables in the model.

Results of SIMS Data Set

Tables 11.4a and 11.4b show the outer and inner model results for the SIMS data set. Eleven LVs and 26 MVs were included in the path model.

Outer Model Results

In the following discussion the weights and the factor loadings of each MV within a construct are discussed with respect to the LV to which it contributed.

Home Background (Homeback)

Table 11.4a indicated that *Homeback* was formed in an inward mode and was built from five MVs. One MV, namely *Med* was deleted from the analysis, since the weight of the variable was below the critical value of 0.07. This indicated that *Med* failed to contribute in a meaningful way to the construct. However, the other four MVs contributed meaningfully to the construct *Homeback.* These four MVs that contributed meaningfully to the construct were *Focc, Fed, Homebook* and *Siblings.* The weights for *Focc, Fed, Homebook* and *Siblings* were 0.51, 0.12, 0.63 and -0.22 respectively.

Hence, apart from *Med* all other MVs were well suited to form the home background or socioeconomic status latent variable. The communality, redundancy and the tolerance values recorded in Table 11.4a show that all four of the remaining MVs contributed effectively to this construct LV *Homeback*.

Gender

The sex of the student was considered to indicate *Gender*. Thus, this LV comprised a single MV.

Ethnicity (Ethnicity)

Five MVs, namely *Fcntry* (Father's country), *Mcntry* (Mother's country), *Cntry* (country of birth) *Yrscntry* (Years in the country) and *Enghome* (English spoken at home) were hypothesised to contribute to the formation of this inward mode LV. However, three out of five MVs were removed from further analysis, because of their failure to contribute to the formation of the construct. The two MVs that contributed to the formation of the latent construct were *Yrscntry* and *Enghome*. It can be seen from Table 11.4a that the MV *Enghome* was the highest contributor (0.92) to the formation of the LV *Ethnicity*. The communality values for *Yrscntry* and *Enghome* were 0.18 and 0.94 respectively, while the redundancy values were 0.00 each and the tolerance values were 0.04 each. Hence, the weights, the communality, redundancy and tolerance values showed that only two MVs contributed effectively to the formation of this criterion variable.

Table 11.4a SIMS Outer Model Results

Variable		Weight/Loading	Communality	Redundancy	Tolerance
Homeback[i]	*Focc*	.51	.55	.00	.17
	Fed	.12	.22	.00	.15
	Med	Deleted			
	Homebook	.63	.66	.00	.11
	Siblings	-.22	.07	.00	.01
Gender[u]	*Sex*	1.00	1.00	.00	.00
Ethnicity[i]	*Fcntry*	Deleted			
	Mcntry	Deleted			
	Cntry	Deleted			
	Yrscntry	.25	.18	.00	.04
	Enghome	.92	.94	.00	.04
Yearlevel[i]	*Year7*	-.84	.84	.02	.04
	Year8	Dummy variable			
	Year9	.40	.31	.01	.04
Classize[u]	*Clssize*	1.00	1.00	.01	.004
Views	*Viewmath*	1.00	1.00	.01	.004
Values[o]	*Mathinso*	.91	.83	.03	.04
	Contrnev	.58	.34	.01	.04
Motivation[i]	*Hmwall*	.77	.73	.08	.02
	Attitsch	.53	.42	.05	.02
Timlearn[i]	*Hourmath*	.34	.11	.02	.001
	Hourmhmw	.95	.91	.13	.001
Attitude[o]	*Likemath*	.81	.65	.05	.23
	Mathmark	.86	.74	.06	.21
	Diffmath	.40	.16	.01	.04
Mathachi[u]	*Rasch score*	1.00	1.00	.36	.00
Mean Communality					0.60

i = inward mode; o = outward mode; u = unity mode

Year Level (Yearlevel)

Three variables namely *Year7, Year8,* and *Year9* formed this antecedent inward mode construct. It can be observed from Table 11.4a that the MV *Year7* had a high negative weight (-0.84) on the LV *Yearlevel*. This shows that *Year 7* was the strongest predictor of the LV *Yearlevel.* MV *Year8* was treated as a dummy variable, while *Year9* had a weight of 0.40. The weights, communality, redundancy and tolerance values in Table 11.4a indicated that the variables contributed well to the formation of this construct.

Class Size (Classize)

This LV comprised a single MV called *Clssize*. Since it was in unity mode the loading and the communality were each unity.

Views about Mathematics (Views)

A single MV, namely, *Viewmath* formed this construct. As this latent variable consisted of only a single manifest variable it was estimated using unity mode.

Values about Mathematics (Values)

It can be seen in Table 11.4a that two MVs, *Mathinso* and *Contrenv* combined in the reflective mode to contribute to this LV. The factor loadings indicated that *Mathinso* (0.91) dominated this LV while *Contrenv* had a loading of 0.58 (see Table 11.4a).

Motivation towards Mathematics (Motivation)

The two MVs, namely, *Attitsch* and *Hmwall* were selected to form this inward mode construct of *Motivation* towards mathematics. In the hypothetical model developed in Chapter 6, it was assumed to be in an outward mode, however, in the analyses, in order to increase the estimations of the MVs, it was changed to an inward mode. Preliminary exploratory PLSPATH analysis suggested that both MVs contributed to the formation of this LV. The weights for the two MVs were 0.53 and 0.77 respectively. The communality, redundancy and tolerance values in Table 11.4a revealed that these two MVs were strong contributors to the formation of the LV *Motivation.*

Time in Learning (Timlearn)

The MVs *Hourmath* and *Hourmhm*w formed this LV in the inward mode. The loading for *Hourmhmw* (0.95) was stronger than the MV *Hourmath* whose loading was 0.34. The communality values were 0.91 and 0.11 respectively.

Attitude towards Mathematics (Attitude)

Three MVs namely, *Likemath*, *Mathmark*, and *Diffmath* were combined to form the LV *Attitude* in the outward mode. The loadings showed that the three MVs combined well to form *Attitude*. The MV *Mathmark* (0.86) was the strongest contributor compared to the other two, while the least contributor was *Diffmath* (0.40).

Mathematics Achievement (Mathachi)

This LV consists of a single MV, namely *Rasch score*. Because it was in the unity mode the loading and the communality were both one.

In summary, for the outer model, among the 26 MVs that were hypothesised to contribute to form the 11 constructs, four MVs were removed from further analysis because of their failure to contribute to the formation of the respective LVs in the hypothesised path model and 22 MVs remained to form the 11 LVs. The mean communalities value (0.60) showed that the outer model was a good model. The next section presents the results for the inner model.

Inner Model Results

The results for the outer model were discussed in the previous section, while in this section the results for the inner model are addressed. Table 11.4b shows the beta (ß), weights, correlations and tolerance values and Table 11.4c indicates the direct, indirect and total effects, correlations, fit and R^2 values. There are 11 LVs in the inner model, the results of the analyses of these LVs are presented in Figure 11.4.

Among the 11 LVs, *Homeback*, *Gender* and *Ethnicity* were exogenous antecedent LVs. Thus, the discussion in this section considers only those eight LVs that were assumed to be influenced by an other LV in the hypothesised model.

Table 11.4b SIMS Inner Model Betas

Variable	Beta	Correlation	Tolerance	R^2
Yearlevel				.03
Homeback	.12	.13	.02	
Ethnicity	.09	.11	.02	
Classize				.01
Homeback	.12	.12	.00	
Views				.01
Yearlevel	-.12	-.12	.00	
Values				.03
Views	.18	.18	.00	
Motivation				.11
Homeback	.20	.20	.06	
Gender	.12	.10	.01	
Ethnicity	-.11	-.07	.03	
Yearlevel	.10	.11	.03	
Classize	.08	.10	.01	
Values	.18	.18	.01	
Timlearn				.14
Gender	-.09	-.05	.01	
Motivation	.37	.36	.01	
Attitude				.08
Homeback	.08	-.13	.04	
Gender	-.10	.08	.01	
Motivation	.25	-.25	.05	
Mathachi				.36
Homeback	.27	.38	.06	
Ethnicity	.07	.14	.03	
Yearlevel	.29	.38	.03	
Classize	.21	.27	.02	
Attitude	.28	.33	.02	
Mean R^2				.10

Year Level (Yearlevel)

Homeback, Ethnicity and Gender were the hypothesised factors to influence *Yearlevel*. The result in Tables 11.4b and 11.4c show that *Homeback* (0.12) and *Ethnicity* (0.09) influenced *Yearlevel* directly. This means that students from higher socioeconomic status backgrounds were in a higher year level than students from lower socioeconomic status backgrounds and Australian students were in higher year levels than non Australian students.

Class Size (Classize)

Four factors namely, *Homeback*, *Gender*, *Ethnicity* and *Yearlevel* were hypothesised to influence this construct. However, only *Homeback* (0.12) was found to have an effect. Tables 11.4b and 11.4c and Figure 11.4 show that the direct effect of *Homeback* was 0.12 and there was no other factor which had an indirect effect on this LV. The explained variance (R^2=0.01) for the LV *Classize* was very small. This indicated that students from higher socioeconomic status backgrounds were in larger class groups for mathematics compared with students from lower socioeconomic backgrounds. Thus, *Class size* was influenced only by the home background of students in this hypothesised model.

Table 11.4c SIMS Inner Model Effects (On)

Variable	Direct	Indirect	Total	Correlation	Fit	R^2
Yearlevel						.03
Homeback	.12	-	.12	.13	-	
Ethnicity	.09	-	.09	.11	-	
Classize						.01
Homeback	.12	-	.12	.12	-	
Views						.01
Homeback	-	-.01	-.01	-.05	-.0308	
Ethnicity	-	-.01	-.01	-.03	-.0215	
Yearlevel	-.12	-	-.12	-.12	-	
Values						.03
Homeback	-	-.002	-.002	.01	.01	
Ethnicity	-	-.002	-.002	-.04	-.04	
Yearlevel	-	-.02	-.02	-.06	-.04	
Views	.18	-	.18	.18	-	
Motivation						.11
Homeback	.20	.02	.22	.20	-	
Gender	.12	-	.12	.10	-	
Ethnicity	-.11	.01	-.10	-.07	-	
Yearlevel	.10	-.004	.09	-.11	-	
Classize	.08	-	.08	.10	-	
Views	-	.03	.03	.04	.02	
Values	.18	-	.18	.18	-	
Timlearn						.14
Homeback	-	.08	.08	.06	-.01	
Gender	-.09	.04	-.05	-.05	-	
Ethnicity	-	-.04	-.04	-.05	-.02	
Yearlevel	-	.03	.03	.01	-.03	
Classize	-	.03	.03	-.002	-.04	
Views	-	.01	.01	.03	.02	
Values	-	.07	.07	.04	-.03	
Motivation	.37	-	.37	.36	-	
Attitude						.08
Homeback	.08	.05	.13	.13	-	
Gender	-.10	.03	-.07	.08	-	
Ethnicity	-	-.03	-.03	.04	.03	
Yearlevel	-	.02	.02	.01	.03	
Classize	-	.02	.02	.06	.03	
Views	-	.01	.01	.11	.10	
Values	-	.05	.05	.28	.24	
Motivatio	.25	-	.25	.25	-	
Mathachi						.36
Homeback	.27	.09	.37	.38	-	
Gender	-	-.02	-.02	-.02	.004	
Ethnicity	.07	.02	.09	.14	-	
Yearlevel	.29	.01	.30	-.35	-	
Classize	.21	.01	.22	.27	-	
Views	-	.002	.002	-.09	-.05	
Values	-	.01	.01	.01	.002	
Motivation	-	.07	.07	.21	.04	
Attitude	.28	-	.28	-.33	-	
Mean R^2						.10

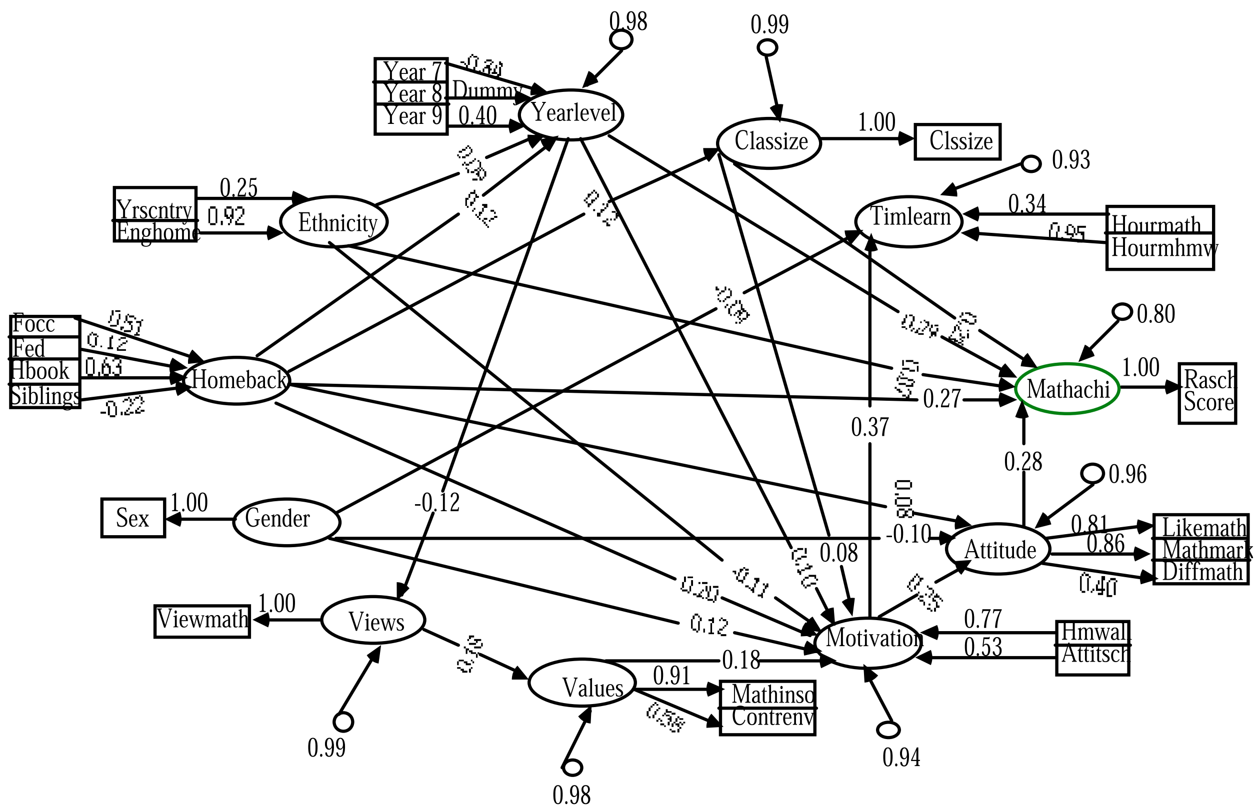


Figure 11.4 SIMS - Student level factors influencing Mathematics Achievement

Views about Mathematics (Views)

The *Views* of students about mathematics were hypothesised to be influenced by five LVs, namely *Homeback*, *Gender*, *Ethnicity*, *Yearlevel* and *Classize*. Among these factors only one LV, namely *Yearlevel* influenced this construct (see Tables 11.4b and 11.4c and Figure 11.4) directly. This observation indicated that students in the lower year level expressed stronger *Views* about mathematics than did students in the higher year levels (see Tables 11.4b and 11.4c and Figure 11.4). Table 11.4c indicated that *Home Background* and *Ethnicity* indirectly influenced *Views*. These two LVs influenced *Views* through the mediating variable *Yearlevel*. The indirect effects indicated that students from lower socioeconomic status backgrounds showed stronger *Views* than students from higher socioeconomic status backgrounds and non-Australian students showed stronger *Views* than Australian students. The R^2 (0.01) value of the construct was very small. Therefore, from this result it was possible to conclude that students' *Views* about mathematics were influenced directly by *Year level.* In addition, *Homeback* and *Ethnicity* influenced the *Views* of students indirectly through the mediating variable, *Yearlevel*.

Values about Mathematics (Values)

The *Values* of students about mathematics were hypothesised to be influenced by six LVs, namely *Homeback*, *Gender*, *Ethnicity*, *Yearlevel*, *Classize* and *Views*. Among these LVs only a single factor namely, *Views* influenced this construct directly (see Tables 11.4b and 11.4c and Figure 11.4). This indicated that those students who expressed stronger *Views* also expressed stronger *Values* about mathematics than those students who expressed weaker views about mathematics. Table 11.4c indicates that *Homeback* (-0.002), *Ethnicity* (-0.002) and *Yearlevel* (-0.02) indirectly influenced this particular LV. The indirect effects showed that those students from lower socioeconomic status backgrounds, non Australian students and younger students indirectly expressed stronger *Values* about mathematics than students from higher socioeconomic status backgrounds, Australian and older students. However, it is important to point out here that the effects for these LVs were very small and operated through the students' *Views*. The variance explained (R^2=0.03) for the construct was very small (see Table 11.4c). Therefore, from these results it was possible to conclude that students' *Values* about mathematics were influenced directly by their *Views* about mathematics, and only slightly influenced by *Homeback*, *Yearlevel* and *Ethnicity*.

Motivation towards Mathematics (Motivation)

The *Motivation* level of students towards mathematics was hypothesised to be influenced by seven LVs, namely *Homeback*, *Gender*, *Ethnicity*, *Yearlevel*, *Classize*, *Views* and *Values*. The interesting point to be observed here is that all the hypothesised factors show either direct, indirect or both effects. Three LVs show both direct and indirect effects, while three other LVs show direct effects and one variable shows only an indirect effect (see Tables 11.4b and 11.4c and Figure 11.4). The factors that had a direct and an indirect influence on *Motivation* were *Home Background* (students from higher socioeconomic backgrounds were more motivated towards mathematics than students from lower socioeconomic backgrounds), *Ethnicity* (non-Australian students were more motivated than Australian students) and *Yearlevel* (older students were more motivated than younger students). While, the other factors that showed only a direct effect on the criterion variable were *Gender* (girls were more motivated than boys), *Classize* (students in large class groups were more motivated than students from smaller class groups) and *Values* (students who expressed stronger values were more motivated than students who showed weaker values). It is interesting to note that only *Views* of the students about mathematics did not influence their *Motivation* directly. Furthermore, *Home Background* and *Values* had a greater influence on this construct than the other LVs. The variance explained (R^2=0.11) for the construct was medium. Therefore, this result revealed that students' *Motivation* for mathematics was influenced by their *Home Background*, *Gender, Ethnicity*, *Yearlevel, Classize* and *Values* about mathematics, and only indirectly by their *Views* about mathematics.

Time in Learning (Timlearn)

Eight LVs, namely *Homeback*, *Gender*, *Ethnicity, Year level, Classize*, *Views*, Values and *Motivation* were the factors which were expected to influence *Time in Learning*. However, the results of the analyses revealed that only one LV influenced this construct both directly and indirectly, while one other LV influenced this construct directly (see Tables 11.4b and 11.4c and Figure 11.4). The factor that had both a direct and an indirect influence on *Time in Learning* was *Gender*, direct (-0.09) and total (-0.05), while *Motivation* (highly motivated students indicated that they spent more time in learning mathematics than less motivated students) showed only a direct effect (0.37) on the construct. All the other six LVs influenced *Time in Learning* indirectly through the mediating variable *Motivation* (see Table 11.4c). The variance explained for this construct was 0.14. Therefore, this result revealed that the time taken by students in learning mathematics was influenced by their *Gender* and by their level of *Motivation* towards mathematics, while *Home Background, Ethnicity, Year level, Class Size, Views* and *Values* influenced only indirectly time spent in learning mathematics.

Attitudes towards Mathematics (Attitude)

Home background, *Gender*, *Ethnicity, Year Level, Class Size*, *Views*, *Values*, *Motivation* and *Time in Learning* were hypothesised to influence *Attitude towards Mathematics.* The result of the PLSPATH analysis demonstrated that all the hypothesised factors except *Timlearn* influenced this LV either directly, indirectly or both directly and indirectly (see Table 11.4c). Two of the factors showed both direct and indirect effects on this construct, while one factor influenced this criterion variable only directly. Meanwhile, five factors influenced *Attitudes towards mathematics* indirectly (see Tables 11.4b and 11.4c and Figure 11.4). The factors that had a direct and an indirect influence on *Attitudes towards Mathematics* were *Home Background* (total effect = 0.13) (those students from higher socioeconomic status backgrounds showed more positive attitudes towards mathematics than students from lower socioeconomic backgrounds) and, *Gender* (total effect = -0.07), (male students indicated more positive *attitudes* towards mathematics than female students). Furthermore, *Motivation* also influenced this LV directly (0.25). Students who indicated stronger *Motivation* towards mathematics also showed more positive attitudes towards mathematics than students who expressed less *Motivation* towards mathematics. It was surprising to observe that *Views* and *Values* about mathematics did not influence *Attitude* directly. It was hypothesised that students who expressed strong *Views* and *Values* would also exhibit positive attitudes towards mathematics. However, this hypothesis was supported only indirectly with *Values* acting through *Motivation*. *Ethnicity, Yearlevel*, and *Classize*, also influenced attitudes towards mathematics indirectly (see Table 11.4c).

The variance explained (R^2=0.08) for this construct was small. Therefore, this analysis showed that students' attitudes towards mathematics were influenced by their *Home Background*, *Gender* and their *Motivation* towards mathematics and only indirectly influenced by *Year Level, Ethnicity, Class Size, Views* and *Values*.

Mathematics Achievement (Mathachi)

It was hypothesised that the mathematics achievement levels of SIMS students were influenced by ten LVs, namely *Homeback*, *Gender*, *Ethnicity, Yearlevel, Classize*, *Views*, *Values*, *Motivation*, *Timlearn* and *Attitude.* The results of the PLSPATH analyses revealed that four of the nine factors influenced *Mathematics Achievement* both directly and indirectly, while one factor influenced it only directly. The remaining four factors influenced the criterion variable indirectly (see Tables 11.4b and 11.4c and Figure 11.4). The only factor that did not have any effect on the outcome measure was *Time in Learning*.

Direct Effects

The factors that had direct influence on *Mathachi* are discussed as follows.

Home Background (Homeback)

This factor influenced *Mathachi* directly (0.27) and indirectly (0.09) through *Attitude*. It was noted that the total effect of this LV on *Mathematics Achievement* was 0.37, furthermore, this LV showed the highest effect on *Mathematics Achievement* when compared with the other LVs. This analysis revealed that students from higher socioeconomic status backgrounds showed a higher achievement level in mathematics than students from lower socioeconomic backgrounds (see Tables 11.4b and 11.4c and Figure 11.4). Therefore, *Home Background* was found as one of the stronger student level factors influencing *Mathematics Achievement*.

Ethnicity

The LV *Ethnicity* influenced *Mathematics Achievement* directly (0.07) and indirectly (0.02). The total effect was 0.09. This finding implied that Australian students (Australian born students or those students who lived longer in Australia and who spoke English at home) achieved at a higher level in mathematics than non Australian students.

Year Level

The other LV that influenced *Mathematics Achievement* directly (0.29) and indirectly (0.01) was *Year Level*. The total effect was 0.30, which was the second strongest factor to influence *Mathematics Achievement* next to *Home Background*. This indicates that students in higher grade levels achieved at a higher level in mathematics than students at lower grade levels, even after controlling for the effects of *Home Background.*

Class Size

Class size was another LV that influenced the *Mathematics Achievement* of students both directly (0.21) and indirectly (0.01). The total effect was 0.22. This variable demonstrated that students from larger class groups achieved at a higher level in mathematics than students from smaller class groups. It would seem important to undertake further studies to examine in greater detail the manner in which this factor operated for effective teaching and learning to be undertaken.

Attitudes towards Mathematics (*Attitude)*

This LV influenced *Mathematics Achievement* directly (0.28). It was the third strongest variable influencing *Mathematics Achievement*. The variable indicated that students who expressed more positive attitudes towards mathematics were likely to achieve at a higher level in mathematics than students who expressed less positive attitudes towards mathematics (see Tables 11.4b and 11.4c and Figure 11.4).

Indirect Effects

Gender, *Views*, *Values* and *Motivation* were the remaining factors that influenced the *Mathematics Achievement* of students indirectly (see Table 11.4c). The indirect effect of *Gender* was -0.02, the negative sign indicated that boys achieved at a slightly higher level in mathematics than girls. However, the value was very small. This indirect influence was mediated through *Attitude*. Another factor that exhibited an indirect effect on this outcome measure was *Views*. Its effect was only 0.002 which was too small to be considered. *Values* (0.01) and *Motivation* (0.07) also influenced the mathematics achievement of students indirectly.

The variance explained (R^2=0.36) for this outcome measure was relatively large since an age sample not a grade sample was under investigation. The results of the analyses discussed above revealed that the student level factors that influenced mathematics achievement were *Home Background, Ethnicity, Year Level, Class Size*, and *Attitudes* towards mathematics. In addition, *Gender, Views, Values* and *Motivation* had effects on *Mathematics Achievement* only indirectly.

Conclusion

Twenty six MVs and ten LVs were hypothesised to influence the mathematics achievement levels of 13-year-old students in SIMS. Four MVs were deleted from the analyses, because these variables did not contribute to the formation of their respective LVs.

The results of the analysis revealed that among the ten hypothesised LVs to influence the *Mathematics Achievement* of SIMS students five factors namely, *Homeback*, *Yearlevel*, *Ethnicity, Classize* and *Attitude* were shown to be student level factors that influenced *Mathematics Achievement* at the 13-year-old level in Australia in 1978.

Result of TIMS Data Set

Tables 11.5a and 11.5b show the outer model and the inner model results for the TIMS Year 8 students data set. Ten LVs and 25 MVs were included in the model, and the results of the PLSPATH analyses are discussed in the following two parts. The first part addresses the results of the outer model and the second part considers the inner model.

Outer Model Results

Table 11.5a shows the weights and the factor loadings, the communalities, redundancies and the tolerances values of each MV within a construct and are discussed with respect to the LV to which it contributes.

Home Background (Homeback)

Table 11.5a shows that *Homeback* is reflected by six MVs, namely *Focc, Mocc, Fed*, *Med, Homebook* and *Siblings*. In the hypothetical model developed in Chapter 6, it was assumed to be in an inward mode. However, in the analyses, it was changed to an outward mode. Since this LV had as many as six observed or MVs the outward mode was chosen to avoid problems of multicollinearity. The loadings for five of the six MVs reflecting this antecedent construct, were 0.69, 0.59, 0.77, 0.74 and 0.58 respectively. *Siblings* were deleted from the model, because the factor loading was -0.06 and this value was below the critical value of 0.30. The important point here is that, in EMS and SIMS data sets the MV *Siblings* contributed to reflect *Homeback*. However, it did not reflect the same LV in TIMS. This suggested that the drop in the number of children per family and the greater uniformity in family size in 1994, when compared with 1978 in Australia has led to a decline in the importance of this factor. It is of interest to observe that both Fed and *Med* contribute more to the formation of the LV *Homeback* than do the other three variables. However, all were highly significant for the development of this construct..

Table 11.5a TIMS-Year 8 students-Outer Model Results

Variable		Weight/Loading	Communality	Redundancy	Tolerance
Homeback[o]	*Focc*	.69	.48	.00	.25
	Mocc	.59	.35	.00	.21
	Fed	.77	.59	.00	.48
	Med	.74	.54	.00	.47
	Homebook	.58	.33	.00	.09
	Siblings	Deleted			
Gender[u]	*Sex*	Deleted by default			
Studageu	*Age*	Deleted by default			
Ethnicityi	*Cntry*	Deleted by default			
	Fcntry	Deleted by default			
	Mcntry	Deleted by default			
	Enghome	Deleted by default			
Classize[u]	*Clssize*	1.00	1.00	.02	.00
Views[u]	*Studpart*	1.00	1.00	.007	.00
Motivation[o]	*Motiv1*	.75	.57	.02	.35
	Motiv2	.44	.19	.007	.11
	Motiv3	.83	.70	.02	.37
	Motiv4	.79	.63	.02	.24
Timlearn[i]	*Homworkf*	.82	.72	.03	.02
	Hourmhw	.27	.19	.009	.03
	Hourmath	.42	.19	.009	.01
Attitude[o]	*Mathmrk*	.83	.69	.18	.33
	Diffmath	.64	.40	.11	.21
	Likmath	.85	.73	.19	.26
Mathachi[u]	*Rasch score*	1.00	1.00	.34	.00
Mean Communality			0.57		

i = inward mode; o = outward mode; u = unity mode

Gender

The sex of the student was used to indicate *Gender*. Thus, this LV involved a single MV. However, this MV was deleted by default, since the LV *Gender* formed by this MV did not contribute to the inner model. It is important to observe that in the 1964, and the 1978 Australian data sets, this LV showed effects on some of the endogenous LVs and was considered as an important student level factor. This suggests that because of the policies advanced by the Australian Government to reduce the differences between boys and girls' attitudes' towards mathematics and schooling, gender is no longer a significant factor in the learning of mathematics at the lower secondary school level.

Ethnicity

This LV was formed from four MVs, namely *Cntry, Fcntry, Mcntry* and *Enghome*. However, this LV was deleted by default, since the LV *Ethnicity* formed by these MVs did not contribute to the inner model. In the 1978 data set, this LV showed a direct effect on mathematics achievement, however, after 16 years in 1994, it did not show any effect and was deleted by default. This indicates that students coming from the non-English speaking background are no longer suffering from serious handicaps in the learning of mathematics.

Class Size (Classize)

The number of students in a mathematics class was taken to indicate *Classize*. Hence, this LV comprised just a single MV called *Clssize*.

Views about Mathematics (Views)

A single MV, namely, *Studpart* reflected this unity mode construct. Studpart involves students' participation in pairs or small groups in undertaking different kinds of activities in mathematics, such as working together in pairs or in groups on a problem or project. As this latent variable consisted of only a single manifest variable it was estimated using unity mode with its loading being assigned as 1.0..

Motivation towards Mathematics (Motivation)

Four MVs, namely, *Motiv1, Motiv2, Motiv3* and *Motiv4* were selected to reflect this outward mode LV. *Motiv1* involves the students need to do well in mathematics to get the job they want, while *Motiv2* shows the students need to do well in mathematics to please their parents. The students need to do well in mathematics to get into a university or post-school course of their choice was indicated by *Motiv3,* and *Motiv4* involved the students' need to do well in mathematics to please themselves. Preliminary exploratory PLSPATH analysis suggested that all MVs contributed to the formation of this LV. The factor loadings for the four MVs were 0.75, 0.44, 0.83 and 0.79 respectively. This indicated that *Motiv3* contributed relatively more to the formation of this construct than the remaining variables, but that all four factors could be considered to reflect the LV *Motivation.*

Time in Learning (Timlearn)

The MVs *Homworkf, Hourmhw* and *Hourmath* formed this LV in an inward mode. *Homworkf* involved the frequency of mathematics homework being given to students in a week, while *Hourmhw,* was the time taken by the students to do their mathematics homework in a week. The time allowed for mathematics classes in a week was indicated by *Hourmath.* It can be seen in Table 11.5a that *Homworkf,* (0.82) was a noticeably higher contributor in the formation of this construct than the remaining variables. This indicates that the number of homework sessions given by the mathematics teachers in a week contributed strongly to the formation of this LV.

Attitudes towards Mathematics (Attitude)

Three MVs namely, *Mathmrk, Diffmath* and *Likmath* were combined to reflect this outward mode LV *Attitude.* The loadings showed that the three MVs combined well to reflect this construct. The analysis showed that MV *Likmath* (0.85) was the relatively strongest contributor in reflecting the construct compared to the other two observed variables, while the least contributor was *Diffmath* (0.64).

Mathematics Achievement (Mathachi)

This LV consists of a single MV, namely *Rasch score.*

In summary, for the outer model, among the 25 hypothesised MVs that contributed to the ten constructs, seven MVs were removed from further analysis, because they did not contribute to the model and one of the MVs, *Siblings,* was removed from the analysis, because its loading was below the critical value of 0.30. *Gender, Studage* and *Ethnicity* were hypothesised to influence the other predictors and the outcome measure in the inner model. However, these LVs failed to have any effect on any of the endogenous LVs. Hence, they were deleted from the analyses by default. Thus, the MVs namely *Sex, Age, Cntry, Fcntry, Mcntry* and *Enghome* which were hypothesised to form *Gender, Studage* and *Ethnicity respectively*, were also deleted from the analyses by default. Consequently, the remaining 18 MVs contributed to the construction of the remaining LVs. The average of the communalities of the MVs was 0.57 which indicated that the model was a sound model. The next section presents the results of the analysis of the inner model.

Inner Model Results

The results of the outer model are discussed in the previous section, while in this section the results of the inner model are presented. Table 11.5b shows the beta (ß), correlation and tolerance coefficients and R^2 and Table 11.5c indicates the direct, indirect and total effects, correlations, fit and R^2. There are ten LVs initially in the inner model, and the results of the analyses of these LVs are presented in Figure 11.5.

Among the ten LVs, *Homeback*, *Gender*, *Studage* and *Ethnicity* were exogenous LVs, which meant that they were not influenced by any other LV. Among the exogenous LVs which were hypothesised to influence the endogenous LVs *Gender*, *Studage* and *Ethnicity* were deleted from the analyses by default because the LVs did not have an effect on any of the endogenous variables. Thus, the discussion in this section considers only those six endogenous LVs that were assumed to be influenced by another LV in the hypothesised model.

Class Size (Classize)

Four factors namely, *Homeback*, *Gender*, *Studage* and *Ethnicity* were hypothesised to influence this construct. However, only *Homeback* (0.16) influenced the LV *Classize* (see Tables 11.5b and 11.5c and Figure 11.5). This meant that those students from higher socioeconomic status backgrounds were in larger class groups for mathematics. The explained variance (R^2=0.03) for this construct was very small.

Views about Mathematics (Views)

Five LVs were hypothesised to influence this LV. Among these factors only one LV namely *Classize* (-0.08) influenced this LV directly, while *Homeback* (-0.01) acted indirectly (see Tables 11.5b and 11.5c and Figure 11.52). The variance explained (R^2=0.01) for this construct was very small. The value for *Classize* was negative which indicated that students from small class groups expressed stronger *Views* about mathematics than students from large class groups (see Tables 11.5b and 11.5c and Figure 11.5). Furthermore, the indirect effect of *Homeback* (-0.01) revealed that students from lower socioeconomic status backgrounds indirectly expressed stronger

Views about mathematics than students from higher socioeconomic backgrounds. Therefore, from the analysis it would seem possible to conclude that students' *Views* about mathematics were influenced by *Classize*. It is also important to point out that *Home Background* influenced the students' *Views* about mathematics indirectly through the mediating variable, *Classize*.

Table 11.5b TIMS-Year-8 Students- Inner Model Betas

Variable	Beta	Correlation	Tolerance	R^2
Classize				.03
Homeback	.16	.16	.00	
Views				.01
Classize	-.08	-.08	.00	
Motivation				.04
Homeback	.11	.11	.0005	
Views	.15	.15	.0005	
Timlearn				.05
Homeback	.17	.18	.01	
Views	-.09	-.08	.02	
Motivation	.09	.10	.03	
Attitude				.26
Homeback	.09	.14	.01	
Views	.09	.16	.02	
Motivation	.48	.50	.03	
Mathachi				.34
Homeback	.23	.35	.06	
Classize	.20	.32	.07	
Views	-.12	-.11	.04	
Timlearn	.21	.34	.08	
Attitude	.30	.36	.06	
Mean R^2				0.12

Motivation towards Mathematics (Motivation)

Six LVs were hypothesised as factors that influenced students' *Motivation* towards mathematics. In Table 11.5c only three of the hypothesised factors are shown to influence this construct either directly, indirectly, or in both ways. Only *Homeback*, *Classize* and *Views* influenced *Motivation* (see Table 11.5c and Figure 11.5), and the total effects of these variables were 0.11, -0.01 and 0.15 for *Homeback, Classize* and *Views* respectively. *Homeback* influenced *Motivation* both directly (0.11) and indirectly (-0.002), while, *Views* (0.15) showed only a direct effect on this construct. However, *Classize* (-0.01) influenced this construct only indirectly and to a slight extent. The variance explained for this construct was 0.04. In general, the findings can be summarised as follows:

(a) students from higher socioeconomic status backgrounds expressed stronger *Motivation* towards mathematics than students from lower socioeconomic backgrounds; and

(b) students who expressed stronger *Views* about mathematics also expressed stronger *Motivation* towards mathematics.

Therefore, student factors that influenced students' *Motivation* towards mathematics were their *Home Background,* and their *Views* about mathematics while *Classize* had only a slight indirect influence.

Time in Learning (Timlearn)

Seven LVs were hypothesised to influence this construct. However, the result of the analysis revealed that only *Homeback* (direct effect = 0.17, indirect effect = 0.01, total effect = 0.18) and *Views* (direct effect = -0.09, indirect effect =0.01, total effect = -

0.07) showed direct and indirect effects while, *Motivation* (0.09) showed only a direct effect on this construct (see Table 11.5c and Figure 11.5). *Homeback* indicated that students from higher socioeconomic status backgrounds were likely to spend more time in learning mathematics than students from lower socioeconomic status backgrounds. Furthermore, *Motivation* also influenced this LV directly. Students who indicated stronger *Motivation* towards mathematics were likely to spend more time in learning mathematics than students who expressed less *Motivation* towards mathematics. The effect of *Views* on *Timlearn* is negative, this indicates that students who showed less participation in pair or group work in mathematics were likely to spend more time in learning mathematics than those students who participated more on group work. Meanwhile, *Classize* showed very small indirect effect (0.006) acting through the mediating variables *Views* and *Motivation* (see Table 11.5c). The remaining LVs did not show any influence on this LV. The R^2 (0.05) value for this construct was small. Therefore, these results revealed that the time taken by students in learning mathematics was influenced by *Homeback, Views* and *Motivation* towards mathematics and only indirectly by *Classize.*

Table 11.5c TIMS - Year 8 students - Inner Model Statistics

Variable	Direct	Indirect	Total	Correlation	Fit	R^2
Classize						.03
Homeback	.16	-	.16	.16	-	
Views						.01
Homeback	-	-.01	-.01	-.02	-.01	
Classize	-.08	-	-.08	-.08		
Motivation						.04
Homeback	.11	-.002	.11	.11	-	
Classize	-	-.01	-.01	.06	.05	
Views	.15	-	.15	.15	-	
Timlearn						.05
Homeback	.17	.01	.18	.18	-	
Classize	-	.006	.006	.21	.18	
Views	-.09	.01	-.07	-.08	-	
Motivation	.09	-	.09	.10	-	
Attitude						.26
Homeback	.09	.05	.14	.14	-	
Classize	-	-.01	-.01	.09	.08	
Views	.09	.07	.16	.16	-	
Motivation	.48	-	.48	.50	-	
Mathachi						.34
Homeback	.23	.11	.35	.35	-	
Classize	.20	.007	.21	.32	-	
Views	-.12	.03	-.08	-.11	-	
Motivation	-	.16	.16	.15	-.04	
Timlearn	.21	-	.21	.34	-	
Attitude	.30	-	.30	.36	-	
Mean R^2						.12

Attitudes towards Mathematics (Attitude)

Eight LVs were hypothesised to influence students *Attitudes towards Mathematics*. The result of the PLSPATH analysis demonstrated that four of the hypothesised LVs showed direct and/or indirect effects on this construct. *Home Background* and *Views about Mathematics* showed both direct and indirect effects on *Attitudes towards Mathematics*. The total effects of *Homeback* and *Views* were 0.14 and 0.16 respectively (see Table 11.5c). *Classize* showed only an indirect effect of -0.01. However, *Motivation* towards mathematics showed a direct effect (0.48) on *Attitudes towards Mathematics* and the effect of this LV was much stronger than the other two variables which showed both direct and indirect effects.

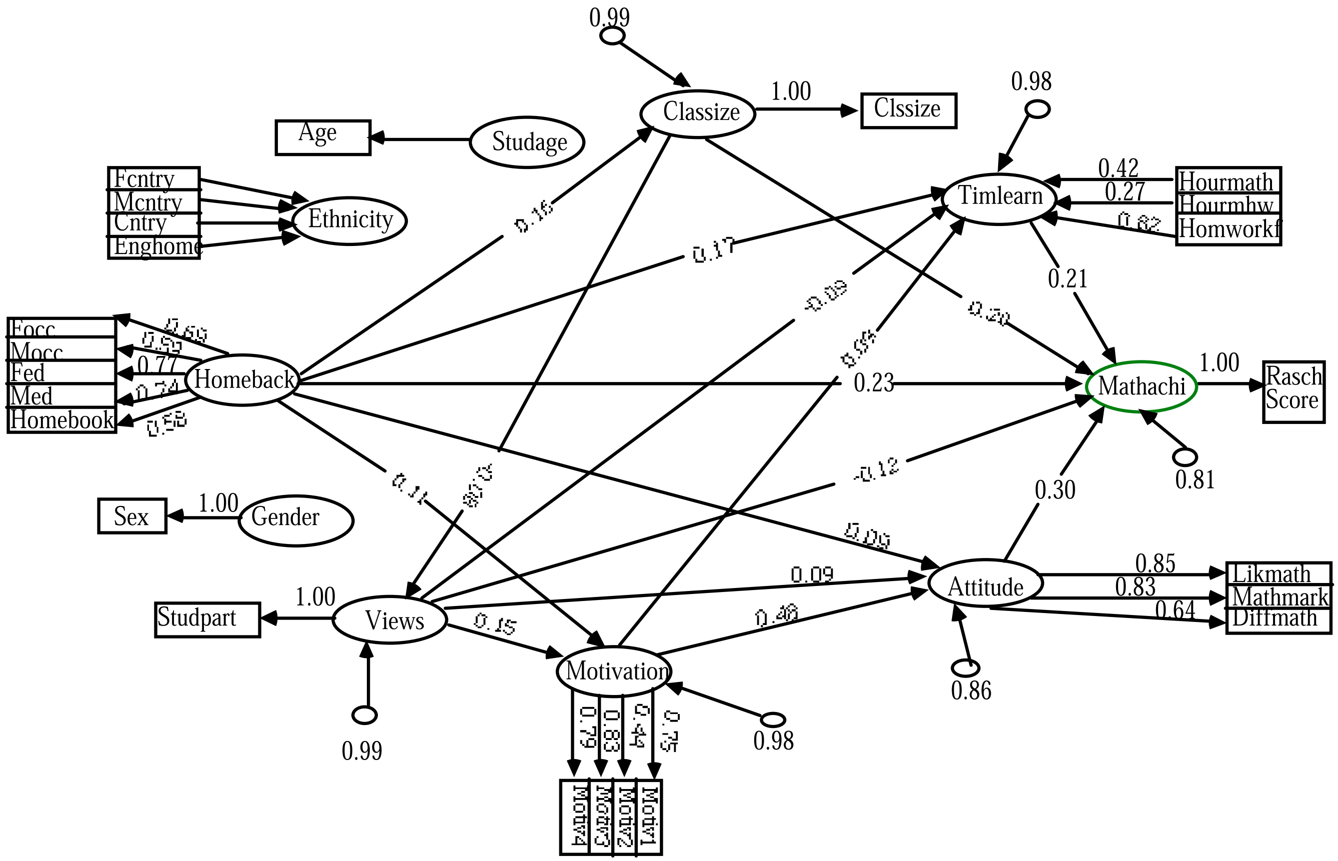


Figure 11.5 TIMS - Student level factors influencing Mathematics Achievement

Thus, students who expressed stronger motivation towards mathematics, students from higher socioeconomic status backgrounds and those students who indicated more positive views about mathematics expressed more positive attitudes towards mathematics than students who expressed weaker motivation, students from lower socioeconomic backgrounds and students who expressed weaker views about mathematics. The important point here was that *Gender* did not have any effect on *Attitudes towards Mathematics*. Previous Australian research findings have revealed that boys had more positive attitudes towards mathematics than did girls (Keeves, 1972; Fraser, 1980; Schofield, 1981; Ballenden et al., 1985). However, the findings here did not support these previous findings. In the previous three analyses, namely 1964, 13-year-old students, 1964 Year 8 students and 1978 13-year-old students data analyses showed that both 13-year-old and Year 8 students in 1964 indicated that boys expressed more positive attitudes towards mathematics than girls. However, in 1994, there was no gender effect either direct or indirectly on *Attitudes towards Mathematics.* Thus the effects of *Gender*, would appear to have changed over time with respect to the learning of mathematics.

Mathematics Achievement (Mathachi)

Students' level of mathematics achievement was hypothesised to be influenced by nine LVs. The result of the PLS analysis revealed that three of the nine factors influenced *Mathematics Achievement* both directly and indirectly, while one other factor influenced *Mathematics Achievement indirectly* (see Tables 11.5b and 11.5c and Figure 11.5) and two factors showed only a direct effect. The factors that did not show any effect on the outcome variable were *Gender, Studage* and *Ethnicity* and since these variables did not enter the model in any way, they do not appear in the final analyses presented. The five factors that had a direct influence on *Mathematics Achievement* were *Homeback, Classize, Views, Timlearn and Attitude*.

Direct Effects

The five factors that had a direct influence on Mathematics Achievement are discussed in greater detail as follows.

Home Background (Homeback)

This LV influenced the mathematics achievement level of Year 8 students directly (0.23) and also with a sizeable indirect effect (0.11). The total effect of the variable on *Mathematics Achievement* was 0.35. *Homeback* was the strongest of the factors that influenced this criterion variable. This result indicated that students from higher socioeconomic status backgrounds were likely to achieve at a higher level in mathematics than students from lower socioeconomic status backgrounds within this TIMS Year 8 sample. Previous Australian research findings into the effects of socioeconomic status on mathematics achievement have indicated that students from higher socioeconomic status family backgrounds were likely to achieve higher than their classmates from lower socioeconomic backgrounds (Keeves, 1968; Rosier, 1980; Ainley et al., 1990). A weaker effect was observed in the FIMSB sample which involved Year 8 students as did the TIMS sample, but the FIMSB sample was restricted to students from government schools only, whereas the TIMS sample included all types of schools.

Class Size (Classize)

This was a LV that influenced *Mathematics Achievement* of students directly (0.20) and indirectly (0.007). The total effect was 0.21. This variable indicates that students from larger class groups achieve at a higher level in mathematics than students from smaller class groups. This finding was consistent with the findings in FIMSA, FIMSB and SIMS data sets.

Views about Mathematics (Views)

Views also had direct (-0.12) and indirect (0.03) effects on *Mathematics Achievement.* This variable influenced *Mathematics Achievement* negatively. Thus, the variable indicated that those students who expressed less participation in working together in pairs or in small groups on different kinds of mathematical activities were likely to achieve higher in mathematics than those who indicated more participation in working together in pairs or in small groups in the same subject (see Table 11.5c and Figure 11.5).

Time in Learning Mathematics (Timlearn)

This LV influenced *Mathematics Achievement* directly. It was the third strongest (ß = 0.21) variable that had a direct effect on *Mathematics Achievement* (see Tables 11.5b and 11.5c and Figure 11.5). The evidence showed that students who spent more time in learning mathematics were higher achievers in mathematics than students who spent less time in learning mathematics. While this latent variable included time spent in mathematics classes, the most important variate forming this variable was the number of times mathematics homework was assigned in a week. In part the change in composition of this variable could account for its increased contribution in TIMS compared to FIMS and SIMS. However, its greater contribution could arise from greater variability in the 1990s to time given to learning in mathematics classes.

Attitudes towards Mathematics (Attitude)

This LV influenced *Mathematics Achievement* directly. It was the strongest (ß = 0.30) variable that had a direct effect on *Mathematics Achievement* (see Tables 11.5b and 11.5c and Figure 11.5). The evidence showed that students who expressed more positive attitudes towards mathematics were higher achievers in mathematics than students who expressed less positive attitudes. This observation was consistent with previous Australian research findings (Keeves, 1972; Schofield, 1981; Milne, 1992).

Indirect Effects

Motivation towards Mathematics (Motivation)

Motivation was the only factor that showed an indirect (0.16) effect on *Mathematics Achievement.* The results indicated that highly motivated students towards mathematics were likely to achieve at a higher level in mathematics than less motivated students (see Table 11.5c and Figure 11.5). The important point here was that gender did not exhibit any effect on *Mathematics Achievement.*

An interesting point was that the age of the student did not show any influence on the outcome variable *Mathematics Achievement,* even though a grade sample was under survey. The ethnic background of the student also did not show a direct or indirect effect on the outcome variable.

Conclusion

Twenty-five MVs and nine LV were hypothesised to influence the *Mathematics Achievement* level of Year 8 students who participated in the Third International Mathematics Study in Australia. Among the 25 MVs seven were removed from further analysis, because the loading of *Siblings* was below the critical value of 0.30, and the LVs which were formed by the remaining six deleted MVs did not contribute to the inner model and were deleted from the analysis by default.

The results of the analysis revealed that among the nine hypothesised LVs only five of them were identified as student level factors that influenced the mathematics achievements of Year 8 students in Australia in 1994. These students level factors were *Homeback, Classize, Views*, *Timlearn* and *Attitude.* While *Motivation* influenced the criterion variable only indirectly. The mean of the R^2 values of the endogenous variables of the inner model was 0.12, which showed that the model was not a strong model. However, 34 per cent of the variance of the criterion variable of *Mathematics Achievement* was explained by the latent variables in the model.

Comparisons between Different Occasions

Table 11.6 presents the direct and indirect effects of LVs identified as student level factors that influenced mathematics achievement on the different occasions. The first column shows the variables, while the remaining columns show the direct, indirect and total effects of each variable on Mathematics Achievement on each occasion. The direct effects of each variable on the outcome measure *Mathematics Achievement* were considered to indicate the relative strengths of the factors that influenced Mathematics Achievement on the different occasions.

Home Background

This construct showed direct and indirect influence on the outcome variable on all occasions except in FIMSA (see Table 11.6). In FIMSA, it showed only an indirect influence. When FIMSA and SIMS were compared it would appear that the impact of home background had increased markedly over time. In 1964, in FIMSA, the effect was indirect while 14 years later the effect was both direct and indirect. It is important to remember that in both groups were only 13-year-old students. Furthermore, when Year 8 students in FIMSB (0.08) and TIMS (0.23) were compared it would appear that the effect of home background had increased markedly over the last three decades. However, this is almost certainly a consequence not of greater inequity but of a difference in the sample design employed in 1978, and in 1994 when students from nongovernment schools were included in the investigation compared with 1964 when they were not included. Most of the students in non-government schools would be from higher socioeconomic status backgrounds, therefore, there was greater variability in this predictor LV in 1978 and 1994 compared with 1964, and hence stronger effects were detected on the two later occasions compared to the former occasion.

Gender

In all Australian groups, namely FIMSA, FIMSB and SIMS *Gender* had only an indirect effect on Mathematics Achievement, furthermore, it showed neither a direct not an indirect effect in TIMS. The findings suggested that the difference between boys and girls in mathematics achievement in Australia had been reduced over time. In 1964 and 1978 the effects of sex on *Mathematics Achievement* were indirect effects, but in 1994 there was not even an indirect effect. However, sex showed a direct effect in EMS. The important point here is that both the direct effect in EMS and the indirect effects in the three Australian data sets were in favour of boys. This findings leads to a research question. Does *Gender* difference emerge at different educational levels in different educational systems? Ethiopian research findings showed that the effect of *Gender* on Mathematics achievement emerged in elementary school in favour of boys (Tesfaye, 1987; Anbessu and Junge, 1988; Atsede and Kebede, 1988; Derese et al., 1990; Assefa, 1991; Gennet, 1991b; Behutiye and Wagner, 1993; Sewnet, 1995).

However, Australian research findings have indicated that it begins to emerge at the lower secondary school stage (Fitzpatrick, 1978; Carss, 1980; Keeves and Mason, 1980; Moss, 1982; Leder, 1989, 1990; Willis, 1989; Forgasz, 1991). Therefore, this issue needs further investigation with consideration given to the emergence of the *Gender* difference between boys and girls in mathematics achievement in different parts of the world and at different times over the past 30 or more years.

Table 11.6 Comparisons of Student factors that influence Mathematics Achievements in Different Occasions

	EMS			FIMSA			FIMSB			SIMS			TIMS		
Variable	D	I	T	D	I	T	D	I	T	D	I	T	D	I	T
Homeback	0.14	0.02	0.16	NE	0.17	0.17	0.08	0.13	0.20	0.27	0.09	0.37	0.23	0.11	0.35
Gender	-0.08	0.003	-0.08	NE	-0.01	-0.01	NE	-0.02	-0.02	NE	-0.02	-0.02	NE	NI	
Studage	-0.12	NI	-0.12	NC			NE	NI	-	NC			NE	NI	
Classize	-0.18	0.04	-0.15	0.17	0.04	0.21	0.19	0.03	0.22	0.21	0.01	0.22	0.20	0.01	0.21
Motivationa	NE	0.05	0.05	0.10	0.05	0.15	0.29	0.04	0.33	NE	0.07	0.07	NE	0.16	0.16
Timlearnb	0.11	NI	0.11	NE	NI		NE	NI	-	NE	NI		0.21	NI	0.21
Attitude	0.16	NI	0.16	0.09	NI	0.09	0.09	NI	0.09	0.28	NI	0.28	0.30	NI	0.30
Aspiration	NC			0.25	0.004	0.25	NE	0.03	0.03	NC			NC		
Futmath	NC			NE	0.02	0.02	0.10	0.02	0.13	NC			NC		
Yearlevel	NC			0.49	-0.005	0.49	NC			0.29	0.01	0.30	NC		
Viewsa	NE	0.04	0.04	NE	0.01	0.01	NE	0.01	0.01	NE	0.002	0.002	-0.12	0.03	-0.08
Values	NE	0.05	0.05	NE	0.08	0.08	NE	0.08	0.08	NE	0.01	0.01	NC		

D = Direct effect; I = Indirect effect; T = Total effect; NC = Not Considered; NE = No direct effect; NI = No indirect effect; *a*=The 1994 MVs that formed the LV were different from the MVs that formed the LV in 1964, 1978 and 1996; *b*= In 1994 there was additional MV that formed the LV additional to the MVs that formed the LV in 1964, 1978 and 1996

Student Age

All students in FIMSA and SIMS were 13-year-olds therefore, *Student Age* was not considered as a factor in these analyses. However, in EMS, FIMSB and TIMS student age was a potential factor, since these groups were all Year 8 students, who were not of the same age group. Therefore, student age was considered as a factor that could influence *Mathematics Achievement*. While student age was identified as a factor that influenced *Mathematics Achievement* in Ethiopia, there was no direct or indirect effect on achievement in both the Australian data sets. Thus, student age was not found to be a factor that influenced *Mathematics Achievement* in Australia. The reason might be related to small age differences among Australian students and larger differences among Ethiopian students. In FIMSB the age of students ranged from 11 to 16 years and for TIMS ranged from 12 to 16.3 years, while the range in EMS students was between 10 and 25 years.

Class Size

The other factor which was considered in all the five groups of students was *Classize*. In all groups it showed both direct and indirect effects on the outcome variable. However, the effect was in different directions for Ethiopian and Australian students. In the Ethiopian situation students in smaller class groups were likely to achieve at a higher level than students in larger class groups. For Australians the reverse was true. The findings seem contradictory, but they are not. It is important to remember that the average class size at this Year level in the Addis Ababa region of Ethiopia was more than 70 students (Ministry of Education, 1996). However, in Australian schools the class size is much smaller, less than 20 students, (Australian Bureau of Statistics, 1997a). Therefore, these findings call for further research to investigate the optimal size of a mathematics class, as well as why a positive relationship is being detected in Australia, in spite of widespread myths to the contrary.

The class size effect on *Mathematics Achievement* between 1964 (0.19) and 1994 (0.20) in Australia was of median strength. The interesting point is that students from larger groups were likely to achieve at a higher level in mathematics in 1964 and the same and perhaps an even greater effect was found after a 30-year period. Thus the finding suggests that to achieve better results, mathematics students must be in larger class groups. This finding was consistent with that of Pidgeon's (1967). In the analysis of the English FIMS data set, he outlined that "there is evidence particularly with pupils up to 'O' level that higher mathematics performance is associated with larger classes" (Pidgeon, 1967, p. 140). However, the findings from the EMS data set suggest that to achieve better results mathematics students should be in smaller class groups, hence, further investigation is needed to decide the upper and lower limits of class size in mathematics in both countries.

Motivation towards Mathematics

Motivation showed both direct and indirect effects on the outcome variable for FIMSA, FIMSB and only an indirect effect for EMS, SIMS and TIMS data sets. The direct and indirect effect for FIMSA, FIMSB and the indirect effect for SIMS and TIMS suggest that the effect of *Motivation* on *Mathematics Achievement* declined over time. It has been reduced from both direct and an indirect effects in 1964 to only an indirect effect in 1978 for 13-year-olds and 1994 for Year 8 students. However, the MVs forming this LV were not identical. In 1964 the variables were *Hmwall,* which involved the number of hours in a week given by students to all homework and *Attitsch*, a nine item scale measuring students' attitude towards school and school learning. Whereas in 1994 the MVs involved students' need to do well in mathematics. Therefore, the results might not be comparable, but the MVs were similar for 1964 and 1978 13-year-old students.

Time in learning Mathematics

Time in learning mathematics was hypothesised as a factor that would influence Mathematics Achievement for all the five groups of students. However, it was found that there was only a direct effect on the outcome variable in EMS and TIMS data sets. In both Australia (1994) and Ethiopia students who spent more time in learning mathematics were likely to achieve at a higher level than those who spent less time in learning the same subject. The interesting point is that in Australia both in 1964 and 1978 *Timlearn* did not show either direct or an indirect effects on *Mathematics Achievement.* However, in 1994 it showed a direct effect (0.21) on *Mathematics Achievement.* This suggests that the effects of *Time in learning mathematics* increased between occasions and it became one of student level factors that had a significant influence in *Mathematics Achievement*. The MVs which formed this LV in all samples were the time allocated for students to learn mathematics in a week and the time taken by students to do their mathematics homework in a week. These two variables formed the LV for the EMS, FIMSA, FIMSB, and SIMS data sets, but for the TIMS data set the frequency of homework given by the teacher in a week was an additional MV that helped to form *Timlearn*. The variable was also a strong contributor (0.82) in the formation of this construct. Therefore, the increased effect in TIMS might be a consequence of the availability of this additional observed variable. However, the changed effect might be due to greater variability among student groups in the time given to learning mathematics on the later occasion than on the earlier occasions, as well as in Ethiopian schools. There is also a possible conclusion that the frequency of homework given in a week is more important than the time spent in doing homework to achieve a higher level of performance in mathematics.

Attitudes towards Mathematics

This LV was found to influence *Mathematics Achievement* and it showed a direct effect for all groups of students. In Australia over time, that is between 1964, 1978 and 1994, *Attitudes towards Mathematics* greatly increased in its effects on *Mathematics Achievement.* In 1964 the effect was 0.09 while this direct effect after 30-year period had increased to 0.30. However, in 1964 *Aspiration* and *Futmath* were included as variables in the analysis, but in 1978 and in 1994 these two variables could not be introduced into the model. The total effects of *Attitude* (0.09), *Aspiration* (0.25) and *Futmath* (-0.02) for FIMSA data set was 0.32, while for FIMSB was 0.25 (see Table 11.6). If these two variables were excluded, from the analyses the effect of *Attitude* would probably have been increased. Hence, the increase in the effect of attitude on the outcome variable in 1978 and 1994 might have been due to the influence of these surrogate variables for *Attitude*. Moreover, in all samples those students who had positive attitudes towards mathematics were likely to be higher achievers in mathematics than other students. Clearly it is important to find ways and means to improve the attitudes of students towards mathematics.

Year Level

Students in FIMSA and in SIMS were from Years 7, 8 and 9. Hence, Year level was considered to be a factor that influenced *Mathematics Achievement* for these two groups of students. In both groups which were age samples *Year Level* showed direct and indirect effects on the outcome variable. This indicated that *Year Level* influenced *Mathematics Achievement,* that is to say, the higher a student's year level the higher his or her level of mathematics achievement. Year level continued to be a factor from 1964 to 1978. However, it is important to observe that the total effect was markedly stronger for FIMSA than SIMS. This may be consequence of the inclusion of primary school Year 7 students in FIMSA samples for New South Wales, the largest state. By the time the SIMS study was conducted in 1978 New South Wales had reorganised its educational system and Grade 7 was at the secondary school level.

Views about Mathematics

Views was considered to be a factor that would influence *Mathematics Achievement* for the EMS, FIMSA, FIMSB, SIMS and TIMS data sets. However, *Views* showed only a direct effect for the TIMS data set and very weak indirect effects for the other three Australian and for the EMS data sets. The effect of *Views* on the outcome variable for the TIMS data set showed that *Views* was a student level factor that influenced *Mathematics Achievement*. However, it is important to observe that the influence of *Views* was very small, specially for the SIMS study. Therefore, the effects of *Views* on the outcome measure were weak and did not change greatly between 1964 and 1978 in Australia. Furthermore, this factor showed change from an indirect effect in 1964 to a direct effect in 1994. However, the MV that formed the LV in 1964 involved students' views about the methods employed by their mathematics teachers during the teaching-learning process, while in 1994 the MV involved the students' participation in working together in pairs or in small groups on different kinds of mathematical activities. These findings would seem to suggest that such student participatory activities could have a detrimental effect on student learning as measured by the mathematics tests employed in these studies.

Values about Mathematics

Like *Views*, *Values* were considered to be a factor that would influence *Mathematics Achievement* for EMS, FIMSA, FIMSB and SIMS data sets. The variable was not considered in the TIMS data set. The variable was formed from two MVs namely, *Mathinso* which measured the attitude of students towards the place of mathematics in society and *Contrenv* which measured students' attitudes towards the relationship of man to his environment. This LV showed only indirect effects for the three Australian data sets and for the EMS data set. It is important to observe that the influence was very small, specially for SIMS where it was only 0.01. Consequently, it was possible to conclude that the effect of *Values* on the outcome measure did not change greatly over time in Australia.

The indicator used to select similarities and the differences between factors that influenced *Mathematics Achievement* over time and across nations was the direct effects of the same variable for all groups on the outcome variable. Consequently, the first comparison was across nations that is between Australia (FIMSB and TIMS) and Ethiopia, while the second comparison was over time in Australia.

Similarities and Differences between Australian and Ethiopian student factors

Among the nine LVs which were hypothesised to influence *Mathematics Achievement*, four factors influenced students *Mathematics Achievement* in both countries. These variables were *Home Background*, *Class Size*, *Time in Learning* and *Attitudes towards Mathematics*. However, *Class Size* influenced mathematics achievement in different directions. In EMS students from smaller class groups, while in Australia in FIMS, SIMS and TIMS, students from larger class groups were higher achievers in mathematics.

The other similarities between the Ethiopian and 1964 Australian students were that *Views* and *Values* did not show direct effects on the outcome variables. However, the variables showed small indirect influences for the 1964 Australian and EMS data sets.

The differences between Australian and Ethiopian students were in *Gender*, *Student Age*, and *Motivation.* In Australia, *Gender* and *Student Age* did not show any direct effect on the outcome variable, moreover *Gender* had only an indirect effect for 1964 students. However, in Ethiopia both factors *Gender* and *Student Age* showed direct effects on the outcome variable. Therefore, *Gender* and *Student Age* were clearly student level factors in Ethiopian schools that influenced *Mathematics Achievement.* While *Motivation* showed a direct effect and was considered an important factor in Australia in 1964 and 1978 (direct effect in FIMSA and FIMSB, there was only an indirect effect in SIMS and TIMS). However, it showed only an indirect effect in Ethiopia.

Hence, the similar student level factors that influenced mathematics achievement both in Australia and Ethiopia were *Homeback, Classize*, *Timlearn* and *Attitude*. While the factors that had different effects were: *Motivation* (which was a factor in Australia but not in Ethiopia, showed only an indirect effect), *Gender* and *Student Age* (which were factors that influenced *Mathematics Achievement* in Ethiopia but not in Australia).

Similarities and Differences between FIMSA and SIMS

Class Size, Attitudes towards Mathematics and *Year Level* continued to be factors that influenced *Mathematics Achievements* between FIMSA and SIMS, that is between 1964 and 1978. While *Gender*, *Views* and *Values* continued to influence *Mathematics Achievements* from 1964 to 1978 indirectly (see Table 11.6). Furthermore, Time in Learning did not show any effect either in 1964 or in 1978.

There were differences between 1964 and 1978 with respect to two variables, *Home Background* and *Motivation*. *Home background* was not directly a factor influencing the outcome variable for FIMSA, but it was found to be a strong factor in 1978. On the other hand *Motivation* was a direct factor in FIMSA and was not a direct factor in SIMS, since it showed only an indirect effect. This might be as a consequence of a difference in the sample design employed in 1978, in which students from nongovernment schools were included. Most of the students in non-government schools would be from higher socioeconomic status backgrounds, therefore, there would have been greater variability in this predictor variable and the greater likelihood of an effect being detected.

Therefore, *Class Size, Attitudes towards Mathematics* and *Year Level were* student level factors that had a stable influence on mathematics achievement over time. *Home Background* increased from an indirect effect in 1964 to a direct effect in 1978, while *Motivation* declined from a direct effect in 1964 to an indirect effect in 1978.

Similarities and Differences between FIMSB and TIMS

Among the nine LV which were hypothesised to influence *Mathematics Achievement*, three factors influenced students' *Mathematics Achievement* on both occasions in 1964 and 1994 at the Year 8 level. These variables were *Home Background*, *Class Size* and *Attitudes towards Mathematics* (see Table 11.6). All three LVs had effects on the outcome measure which had increased in 1994 when compared to the 1964 data set. These three variables are student level factors influencing mathematics achievement for the last 30 years. Consequently, the findings of this investigation indicate that students from higher socioeconomic status backgrounds, students from larger class groups and students who expressed more positive attitude towards mathematics are likely to achieve at a higher level in mathematics.

The other similarities between the two Australian groups was that *Student Age* did not show any effect on the outcome variable.

The differences between the FIMSB and TIMS studies were in *Gender*, *Views, Motivation* and *Time in Learning.* In FIMSB, *Gender* showed an indirect effect on the outcome variable, while *Gender* did not show any effect for the TIMS data set. *Motivation* showed a direct effect and was considered an important factor in FIMSB, however, its contribution was reduced to an indirect effect for the TIMS data set. Furthermore *Views* showed a direct effect in TIMS, but only an indirect effect for the FIMSB data set. The other major difference was for *Time in Learning* mathematics, this variable did not have any effect in 1964, but it showed a direct effect of 0.21 in 1994. However, it is important to recognise that the MVs which formed *Motivation* and *Views* in 1964 and 1994 were different in nature and an additional strong variable (frequency of mathematics homework) was included to form *Time in Learning* mathematics in 1994.

Hence, the student level factors that influenced mathematics achievement both in 1964 and 1994 were *Homeback, Classize* and *Attitude*. While the factor that showed direct effects in 1964 but not in 1994 was *Motivation.* Furthermore *Time in Learning* and *Views* about mathematics showed direct effects in 1994 but not in 1964.

Conclusions

In this Chapter student level factors that influenced mathematics achievement in Australia in 1964, 1978 and 1994 and in Ethiopia in 1996 are identified. The similarities and differences between 1964 and 1978, between 1964 and 1994 student factors and Ethiopian factors are also identified. A summary of the findings is presented below.

The 1964, Australia data were analysed as two separate data sets namely FIMSA and FIMSB. FIMSA included all 13-year-old students who participated in the study and FIMSB involved all Year 8 students. FIMSA was comparable with SIMS and FIMSB was comparable with EMS and TIMS data sets. Thus for FIMSA data set *Year Level, Class Size*, *Motivation*, *Aspiration and Attitude* were identified as student level factors that influenced *Mathematics Achievement* of 13-year-old 1964 Australian students. Furthermore, for the SIMS data set *Home Background*, *Ethnicity*, *Year Level, Class Size* and *Attitude* were identified as student factors that influenced *Mathematics Achievement* of 13-year-old Australian students in 1978.

For FIMSB, *Home Background, Class Size*, *Motivation*, *Future Mathematics* and *Attitude* were found to be student level factors that influenced *Mathematics Achievement* of Year 8 Australian students in 1964. While, in TIMS *Home Background, Class Size*, *Views, Timlearn* and *Attitude* were found to be student level factors influencing mathematics achievement in 1994. Furthermore, in EMS, *Home background, Gender, Student Age, Class Size, Time in Learning* and *Attitude* were found to be student level factors that influenced *Mathematics Achievement* of Year 8 students.

Similarities and differences between the two (FIMSB and TIMS) Australian and Ethiopian data sets were also identified. *Home Background, Class Size* [although its effect was in different directions] and *Attitude* were students level factors that influenced *Mathematics Achievement* of Year 8 students in both countries and over time in Australia. While, *Time in Learning* was identified as a student level factor for TIMS and EMS data sets. *Gender* and *Student Age* were student level factors of importance in Ethiopia, but not in Australia. *Motivation* was an important student level factor in Australia in 1964 but was not an important factor in Ethiopia or in Australia in 1994.

Similarities and differences between the 1964 and 1978 Australian data sets were also identified. *Class Size*, *Year Level* and *Attitude* were student factors that influenced Mathematics Achievement both in 1964 and 1978. While *Gender*, *Views* and *Values* maintained indirect effects from 1964 to 1978.

Meanwhile, in FIMSA *Home Background* influenced mathematics achievement only indirectly, but in SIMS it was identified as a significant student factor, because it showed both direct and indirect effects on the outcome variable. Therefore, *Home Background* would appear to have increased from an indirect effect (1964) to a direct effect (1978). On the other hand *Motivation* showed a direct effect on the outcome variable in 1964, but declined to an indirect effects in 1978. Both these effects are probably consequences of the changed design of the samples between 1964 and 1978 to include students from non-government schools in the 1978 sample, although they were not in the 1964 sample.

In view of the effects of home background, class size and attitude towards mathematics, in order to improve the achievement level of students in mathematics teachers, teachers' unions and the governments might need to provide appropriate assistance to students from lower socioeconomic backgrounds, and those students who expressed less positive attitudes towards mathematics.

In this study *Class Size* was found consistently to be an important student level factor that influenced the mathematics achievement of Australian student at the lower secondary school level over the last 30 years. The findings revealed that students in larger class groups were likely to achieve at a higher level in mathematics than students in smaller class groups. However, the finding in the EMS data set revealed that students in a smaller class groups were likely to achieve higher than students in larger class groups in the same subject. The difference would appear to be in the optimal number of students in a mathematics class. Therefore, further investigation would seem important to recommend the lower and upper limits of students in mathematics classes for both countries, as well as why a positive relationship is being detected in Australia.

Some of the variables employed here for comparison between 1964 and 1994 were not identical on the different occasions such as *Views*. Moreover, the composition of the samples were not identical, the 1964 samples were from government schools, while the 1978 and 1994 samples included students from nongovernment schools. Therefore, further investigation might be important using identical variables and samples to identify the changing effects of student level factors over time.

In Australia at the lower secondary school level *Gender* is no longer a student level factor that influences mathematics achievement. This is possibly due to the implementation of policies by Australian government authorities to reduce the differences between boys and girls in their achievements and attitudes towards mathematics and schooling. However, it is still one of the factors that influences the mathematics achievement of students in Ethiopia. Hence, further investigation would seem important in order to recommend ways and means of reducing sex differences in attitudes towards mathematics and schooling in Ethiopian schools.

12

Summary and Conclusions

The major purposes of the present study are to:

(a) develop a general theoretical model which considers the multivariate structure of the mathematics achievement tests items at the lower secondary school level in Australia and Ethiopia;

(b) examine the changes of the mathematics achievement level of lower secondary school Australian students over time;

(c) compare the mathematics achievement level of lower secondary school students between Australia a developed country and Ethiopia which is a developing country;

(d) investigate the views and attitudes of Australian and Ethiopian students towards mathematics and schooling and to compare the views and attitudes of students over time and across countries; and

(e) develop a theoretical model of student level factors influencing the mathematics achievement of lower secondary school students in Australia and Ethiopia, to examine the hypothesised interrelationships between variables, and to estimate the magnitudes of the causal paths in the models that were tested.

In order to achieve these purposes the following procedures were undertaken. A theoretical framework was proposed from general school learning models and previous research findings. From this framework, specific models were developed in which factors such as the home background of students or the time taken to learn mathematics, were hypothesised to influence mathematics achievement as an indicator of school learning. The models were advanced after taking into consideration the multivariate structure of the available data sets. From these considerations, general and specific research questions relating to the structure underlying mathematics learning and student level factors influencing mathematics achievement were formulated.

The confirmatory factor analysis procedure was employed to examine the different hypothesised models using the mathematics achievement data for EMS, FIMS and SIMS which were available in order to determine the model which indicated the best relative fit across the three occasions, and to infer the underlying dimensions (Marsh and Hocevar, 1983; Hattie, 1985) in the mathematics test. Meanwhile, partial least squares path analysis was employed to identify the student level factors that influenced the mathematics achievement of the 13-year-old and Year 8 students in EMS, FIMS, SIMS and TIMS.

Furthermore, the Rasch model was selected for use to equate student performance in mathematics on a common scale between the EMS, FIMS, SIMS and TIMS data sets. The Rasch model allowed the item parameters to be estimated independently of the students sampled, and the student parameters to be estimated independently of the sample of items employed. The samples employed in EMS, FIMS, SIMS and TIMS were large data sets and Lord (1974), and Kline (1993) have argued that with the Rasch model only large samples should be used if reliable, population-free scaling is to be established. Sontag (1984) also has shown the strength of the use of the Rasch model in the scaling of tests employed in IEA studies. Thus, the changes in the levels of mathematics achievement of students between 1964, 1978 and 1994 in Australia, as well as the difference in the levels of mathematics achievement between Australian and Ethiopian students were measured using the Rasch model. The Rasch scaled scores of mathematics achievement were also brought to a common mathematics scale.

In addition, the changes in views and attitudes of students towards mathematics and schooling between 1964 and 1978 in Australia and 1996 in Ethiopia were measured using the Rasch model. The Rasch scaled scores for views and attitudes were also brought to common view and attitude scales.

Results of the Study

The results of the study are summarised below by addressing the general research questions that are presented in Chapter 6. It seems important to provide a summary of the general research questions that could assist as a basis for discussion and future research.

Therefore in this section the general research questions are presented according to the categories: achievement in mathematics, views and attitudes towards mathematics and schooling, and student level factors influencing mathematics achievement.

Achievement in Mathematics

1. *Does mathematics achievement consist of separate skills or one underlying dimension?*

The results of the confirmatory factor analysis revealed that a nested model best fitted the data in which all the mathematics items were assigned to one general factor known as Mathematics, as well as to one of the several specific and correlated factors. Four different kinds of analyses, one by content area and three by cognitive processes were undertaken. However, all the four analyses supported the nested model which indicated the presence of a strong common factor as well as separate factors for the different fields of school mathematics (Tilahun, 1998).

Thus, the results of these analyses provided no evidence to prevent the calculation of a total score for the mathematics achievement tests. Therefore, a total score was considered as an appropriate measure of mathematics achievement of 13-year-old and Year 8 students on the different occasions in Australia and in Ethiopia. The result of the analyses also showed that most of the items fitted well the Rasch model which indicated the unidimensionality of the items.

2. *Have changes occurred in the level of achievement in mathematics of the 13-year-old students between 1964 and 1978 in Australia?*

There were 65 common items in the tests which were administered in both 1964 and 1978 to 13-year-old students in Australia. Hence, it was possible to measure mathematics achievement changes over the 14-year period using these common items. The Rasch model with concurrent equating and anchor item equating procedures was employed for measuring change over time.

The result of the Rasch analysis showed significant differences between the 1964 mean (460) and the 1978 mean (441) for 13-year-old students' mathematics achievement in Australia. The mean difference was 19 centilogits in favour of the 1964 students. Thus, in Australia the mathematics achievement level of the 13-year-old students declined over the 14-year period, between 1964 and 1978, to an extent that represented approximately two-thirds of a year of mathematics learning.

3. *Have changes occurred in the level of achievement in mathematics of the Year 8 students between 1964 and 1994 in Australia?*

There were nine common items in the tests which were administered to both the 1964 and 1994 Year 8 students in Australia. Therefore, it was possible to measure changes in mathematics achievement over the 30-year period using these common items. The Rasch model with concurrent equating, anchor item equating and common item difference equating methods were employed for measuring change over time.

The result of the Rasch analysis revealed differences between the 1964 mean (451) and the 1994 mean (426) for Year 8 students' achievement. The mean difference was in favour of the 1964 students. The difference represented almost a year of mathematics learning.

4. *Are there differences in the level of achievement in mathematics between Year 8 1994 Australian and 1996 Ethiopian students?*

Nine common items in the tests were employed for comparing achievement levels between the two groups of students, namely Year 8 1994 Australian (mean = 426) and 1996 Ethiopian (mean = 318) students. The Rasch model with concurrent, anchor item, and common item difference equating methods were used for measuring differences in mathematics achievement between Australian and Ethiopian students.

The result of the comparison between the two groups of students revealed differences in the levels of achievement in mathematics between Australian and Ethiopian students. The 1994 Australian students achieved at a markedly higher level than the Ethiopian students. These findings were consistent with the findings of previous research, not only in mathematics but also in science (Comber and Keeves, 1973; Thorndike, 1973; Postlethwaite and Wiley, 1991; Elley, 1992; Keeves, 1992; Beaton et al. 1996b) and in reading (Thorndike, 1973; Elley, 1992).

Even though, the finding of a difference between developed and developing countries is consistent with the expectations from previous studies, it would seem essential to identify the reasons that students in developing countries like Ethiopia, achieve at a lower level than their counterparts in the developed countries like Australia. Inkeles (1979) advanced four hypotheses that might account for the lower performance level of developing countries. These hypotheses are summarised as follows.

The first hypothesis is a lack of test-wiseness. Inkeles (1979) argued that the test when administered to developing country students might be new experiences for those students, which are more or less completely outside of their day to day test taking practices. Therefore the tests must be recognised as irrelevant for judging how much the students actually know the subject in which they are tested when compared with students who are experienced in taking this type of test in their own countries. Inkeles believed that this lack of test-wiseness of students from developing countries is one of the reasons for their lower achievement. In the present study there were 11 constructed response items (about 16% of the total test items) for EMS and FIMS students. Seventy-six, 59, 31, 40 and 72 per cent of FIMS students in Australia provided the correct responses for five out of 11 constructed response items, while the per cent for EMS students were 40, 3, 34, 1 and 22 respectively. For the remaining six items the percentage of students who provided correct responses ranged from 5 to 15 per cent for FIMS and from 0.2 to 2 per cent from EMS students. These results might show that the EMS students were also not familiar with responding to these constructed response items.

The second hypothesis advanced by Inkeles (1979) was poor curriculum. The presence or absence of a topic or a particular task in a given curriculum is an important factor in determining whether or not a student would provide a correct response on a test. Comber and Keeves (1973) have argued that the curriculum is important in making decisions about what is to be learned, and how much effort is needed from the student to learn a particular school subject. Thus, the explanation for the low achievement of students in developing countries is that the tests may have dealt with material which the students did not have an adequate opportunity to learn. Furthermore, Inkeles argued that whether or not the students had an opportunity to learn the material relevant to a given task could be determined by the judgement of their teachers as to whether or not any given test item was suitable for their students. Their judgements could be summarised using an ‘opportunity to learn’ rating for each test. He believed that students from developing countries were given less opportunity to learn the content and skills on which they were tested, and this deficiency would explain why their scores were markedly lower. In the present study, the Ethiopian mathematics teachers provided ratings of the opportunity to learn for the mathematics tests. More than 50 per cent of the teachers indicated that the Ethiopian students had few or no opportunities to learn similar types of problems for 17 per cent of the test items (12 out of 72 items). This shows that at least 17 per cent of the items were not taught to Ethiopian students.

Inkeles’ (1979) third hypothesis was concerned with poor school resources. Inkeles argued that the low achievement level of students from developing countries results from the effects of poverty, which is manifested by the limited resources available to students in their schools. These poor countries spend less money per student, and also student-teacher ratios are generally much less favourable in developing countries. Inkeles believed that their cumulative effects contribute to the lower achievement levels of students in developing countries. This argument would seem true in the Ethiopian context. In the 1995/96 academic year, the Ethiopian government provided less than A$100 per student (Ministry of Education, 1997), while the Australian governments provided over A$4000 per student (Australian Bureau of Statistics, 1997a, 1997b). The cost per student for both countries included both recurrent and capital budget.

In the same year the average number of students per teaching staff member in Addis Ababa region was 41. This figure was smaller when compared with the national norm. The national norm is 50 in both primary and secondary schooling (Ministry of Education, 1997, p.14). However, this number fell to 18 for primary schools and 13 for Australian secondary schools in 1996 (Australian Bureau of Statistics, 1997a, p.6).

The last hypothesis advanced by Inkeles (1979) is social deprivation. Inkeles argued that social deprivation is a separate factor from school effectiveness which is a residual effect after school effectiveness has been identified. Inkeles measures school effectiveness by the gain in a student's score realised during each year the student is exposed to the school's influence. He argued that on this measure schools in developing countries are often as effective as their more developed counterparts and that less successful overall outcomes should be attributed to social deprivation. In the present study there is insufficient data to present conclusions on school effectiveness and hence on the effects of social deprivation.

Overall, the evidence seems to provide *prima facie* support for Inkeles' hypotheses. However, the question of what factors cause differential achievement between schools in developed and developing countries is complex and requires further investigation.

5. *Are there gender differences in the achievement in mathematics for 1964, 1978 and 1994 Australian students?*

The estimated scores of boys and girls on each occasion were compared to examine whether or not gender differences existed in mathematics achievement in Australian lower secondary schools.

The comparisons between boys and girls in different groups of FIMS students, that is between 13-year-olds, Year 8 and the total students who participated in FIMS, revealed that boys achieved higher than girls. However, in all the comparisons the differences were not statistically significant. Hence, the sex difference in mathematics achievement in 1964 at the lower secondary school level in Australia was not significant and was very small in magnitude but generally in favour of boys.

These findings are slightly different from FIMS findings presented by Keeves (1968). In FIMS in all the comparisons the differences were in favour of boys, and were statistically significant. However, crude tests of significance were employed in the mid-1960s (Postlethwaite, 1967 p.36) that while recognising the problem, resulted in an erroneous finding. The mathematics achievement levels of boys and girls in government, nongovernment, restricted and total number of students who participated in SIMS in 1978 were compared. Out of the four comparisons, only in government schools, girls achieved at a slightly higher level than boys. However, in the remaining comparisons boys achieved at a marginally higher level than girls. None of the differences were significant.

The comparisons of mathematics achievement between boys and girls in government, restricted and total number of students who participated in TIMS revealed that girls achieved at a slightly higher level than boys. When the findings were compared with the findings in FIMS and SIMS, the achievement level of girls changed relative to those of the boys. In FIMS the differences were marginally in favour of boys, while in SIMS the difference was in favour of girls only in government schools. Furthermore, in TIMS the differences, except in nongovernment schools, were in favour of the girls. However, it is important to point out that all the differences were found not to be statistically significant.

In conclusion, on all the three occasions at the lower secondary school level sex differences in mathematics achievement were not significant. The findings seem to challenge the assertions of many Australian researchers who have argued that sex difference in mathematics achievement in Australian schools starts to emerge at the junior secondary school stage in favour of boys (Keeves, 1972; Carss, 1980; Leder, 1980, 1985, 1989, 1990; Moss, 1982; Willis, 1989; Ainley, Goldman and Reed, 1990; Leder and Forgasz, 1991). The evidence from this study indicates the possibility of emerging superior performance of girls at the lower secondary school level. However, these results are not statistically significant.

6. *Are there gender differences in achievement in mathematics for 1996 Ethiopian students?*

The mathematics achievement levels of boys and girls in government, nongovernment, and total number of students who participated in EMS were compared. The results of the comparisons showed that boys achieved at a higher level than girls. However, a significant difference was identified between boys and girls only in government schools. Thus, a significant sex difference in mathematics achievement was found only in Ethiopian government junior secondary schools that was in favour of boys, and no significant sex differences were found in nongovernment schools and for the total group of students who participated in EMS. The present study did not fully support the significant differences reported by Behutiye and Wagner (1993, 1995).

Views and Attitudes towards Mathematics and Schooling

7. *Are there changes in attitudes towards the facility of learning mathematics between 1964 and 1978 Australian students?*

Significant differences were found between 1964 and 1978 for 13-year-old Australian lower secondary school students' Attitudes towards Facility of Mathematics. The estimated mean score difference between the two groups of students was 17 centilogits in favour of the SIMS students. Consequently, it can be said that the attitudes of Australian students towards the facility of learning mathematics had improved over time. Thus students' attitudes towards facility of learning mathematics improved over the 14-year period.

8. *Are there differences in attitudes towards facility of learning mathematics between 1964 Australian and 1996 Ethiopian students?*

When the attitudes of Australian (mean = 630) and Ethiopian (mean = 581) students are compared, the Ethiopian students showed less favourable attitudes towards the facility of learning mathematics than the Australian students. This indicates that the Ethiopian students saw mathematics as a difficult subject which only the best students could learn. By contrast, the Australian students saw mathematics more as a subject that any one could learn.

9. *Are there gender differences in the attitudes towards facility of learning mathematics for 1964 and 1978 Australian students?*

In FIMS, boys expressed more positive attitudes towards facility of learning mathematics, but a statistically significant sex difference was found only for Year 8 students. Furthermore, when the mean score of boys was compared with that of girls in SIMS, the mean score of girls was higher than that of their male counterparts. However, the difference was not statistically significant. Thus, there was no significant sex difference in attitudes towards facility of learning mathematics for SIMS students. In summary, changes in sex differences were not found in Australian lower secondary school students in their attitudes towards the facility of learning mathematics over the 14-year period.

10. *Are there gender differences in the attitudes towards facility of learning mathematics for the 1996 Ethiopian students?*

Like their Australian counterparts, a significant sex difference was not observed between male and female Ethiopian students in attitudes towards facility of learning mathematics.

11. *How do the 1964 Australian students' attitudes towards schooling compare with those of the 1978 students and 1996 Ethiopian students?*

Attitudes towards school and school learning were taken as measures of students' attitudes towards schooling. The mean score of FIMS 13-year-old students (mean = 553) was higher than their SIMS peers (mean = 519). This indicates that the attitudes of Australian students towards school and school learning declined over the 14-year period.

Meanwhile, the Ethiopian students expressed more positive attitudes towards school and school learning than their 1964 Australian counter parts. This finding was consistent with findings reported by Walker (1976) and Kotte (1992) who reported that students from developing countries expressed more favourable attitudes towards school and school learning than students from the more developed countries.

12. *Are there gender differences in attitudes towards schooling between Australian and Ethiopian students?*

In 1964 girls in the Australian study expressed more favourable attitudes towards school and school learning than their male counterparts. Furthermore, there were significant differences observed between the two sexes in their attitudes towards schooling in Australian SIMS data set, except in nongovernment schools. This result would seem to indicate that the difference between boys and girls in their attitudes towards schooling was similar over the 14-year period. In Ethiopian junior secondary schools there was no significant difference between the two sexes in their attitudes towards school and school learning.

13. *How do the 1964 Australian students' views about the teaching of mathematics compare with those of the 1978 and the Ethiopian students?*

In the 14-year time period there were no significant changes between the FIMS (mean = 510) and SIMS (mean = 510) students in their views about mathematics teaching in Australia at the 13-year-olds level. However, when the 1964 Year 8 Australian students (mean = 511) were compared with their Ethiopian peers (mean = 573), the Ethiopian students expressed more positive views about mathematics teaching than their Australian counterparts.

14. *Do boys and girls share common views about the teaching of mathematics in both countries?*

Significant sex differences were not found in the two Australian data sets (1964 and 1978). This finding shows that boys and girls in Australia in 1964 and 1978 share common views about the teaching of mathematics in their schools. Boys and girls in Ethiopian junior secondary schools also share common views about the ways and the methods their mathematics teachers use to teach mathematics, except in nongovernment schools where girls showed more positive views than their male classmates.

15. *How do the 1964 Australian students' views about schooling compare with those of the 1996 Ethiopian students?*

The views about school and school learning scale was administered only to FIMS and EMS students, so the comparison was only between these two groups. When the Rasch estimate mean scores of EMS (mean = 503) and FIMS (mean = 503) were compared, there was no difference. This finding shows that Year 8 students in both countries share common views about their schools and school learnings.

16. *Are there gender differences between the 1964 Australian and the 1996 Ethiopian students' views about schooling?*

In both countries sex differences between students were not detected in their views about school and school learning. Both boys and girls expressed similar views about their schools and school learning.

Student level factors influencing achievement in mathematics

17. What student level factors influence student learning in mathematics achievement in Australia and Ethiopia?

Eleven student level factors were hypothesised to influence the mathematics achievement level of the 1964, 13-year-old Australian students. Among the 11 factors only *Year Level*, *Class Size, Motivation, Aspiration*, and *Attitude* were identified as student level factors that influenced mathematics achievement for this age group of students, whereas the student level factors for Year 8 students in 1964 were *Home Background, Class Size, Motivation, Future Maths* (students wish and desire to take more mathematics courses) and *Attitude.* It is important to observe that there were some differences in the factors identified for the two groups of students. *Aspiration* was a factor for 13-year-old students but not for Year 8 students, whereas, *Home Background* and *Future Maths* were identified as factors for Year 8 students but not for 13-year-old students. It is not clear why the significant factors were slightly different for the two groups of students. Hence, it seems important to have further investigation to identify the reason why the factors were different for the two groups of students. Perhaps the explanation is likely to be found in the differences in grade composition of the two groups of students.

Among the ten student level factors that were hypothesised to influence the mathematics achievement level of the SIMS students, only five factors namely, *Home Background, Year Level*, *Ethnicity, Class Size* and *Attitude* were found to be student level factors that influenced achievement in mathematics at the 13-year-old level in Australia in 1978.

Nine latent variables were hypothesised to influence mathematics achievement levels of Year 8 Australian students in TIMS. Among the nine latent variables, only five predictors directly influenced the outcome variable, *Mathematics Achievement*. The five factors that had a direct influence on *Mathematics Achievement* were *Home Background, Class Size*, *Views about Mathematics*, *Time in Learning Mathematics* and *Attitude*. In the Ethiopian study there were ten student level factors that were hypothesised to influence the mathematics achievement level of Ethiopian Year 8 students. However, only six factors namely *Home Background, Gender, Student Age, Class Size, Time in Learning* and *Attitude* proved to be student level factors that influenced the mathematics achievement of Year 8 Ethiopian students.

18. Are there differences between 1964, 1978 and 1994 in the student level factors that influenced the learning of mathematics in Australia and 1996 in Ethiopia?

Differences and similarities were identified between the 1964 and 1978 student level factors at the 13-year-old student level. *Year Level*, *Class Size,* and *Attitude* were common student level factors for both groups of students. Differences were found only in Australia for two factors namely *Home background*, which proved to be a factor in 1978 but not in 1964 and *Motivation,* which was identified as a factor in 1964 but not in 1978. The difference recorded for *Home Background* almost certainly resulted from the change in the composition of the samples to include students for nongovernment schools in 1978 but not in 1964.

Home Background, Class Size and *Attitude* were student level factors that influenced *Mathematics Achievement* both in 1964 and 1994 for Year 8 Australian students.

Differences and similarities were also found between the 1964 Australian and the 1996 Ethiopian Year 8 student level factors influencing mathematics achievement. The common student level factors were *Home Background, Class Size,* and *Attitude. Motivation* was a factor for 1964 Year 8 Australian students, but not for 1996 Ethiopian students, whereas *Gender, Student Age,* and *Time in Learning* were found to be factors for the Ethiopians but not for the Australians.

Home Background, *Class Size*, *Time in Learning Mathematics*, and *Attitude* influenced *Mathematics Achievement* levels of Year 8 1994 Australian and 1996 Ethiopian students, whereas, *Gender* and *Student Age* influenced Mathematics Achievement of Ethiopian students, but not the 1994 Australian students. It should be noted that there were differences and similarities in the student level factors influencing mathematics achievement between occasions and across-systems. Consequently, further investigation is needed to identify the student level factors that influence achievement across a wider range of school systems or countries.

Implications for Theory

The results of the confirmatory factor analyses revealed that none of the proposed factor structures appeared to provide a model that fitted the data well. Hence, further investigation is required to develop conceptual models which represent the dimensions of mathematics achievement more completely.

The main interest in developing a causal model in the present study was to identify the student level factors that influence mathematics achievement over time and across-nations. Hence, models of individual student level factors influencing mathematics achievement at the 13-year-old and at Year 8 levels were developed and investigated using the partial least squares (PLS) analysis procedure. The analysis demonstrated that the proposed variables did contribute to the explanation of differences in the mathematics achievement level of 13-year-old and Year 8 students in Australia and in Ethiopia. However, the variance explained by the models showed that only a limited amount of variance of mathematics achievement was accounted for. It is evident that there are other factors outside of the models which contribute to the difference in student's mathematics achievement, such as school factors which were not considered in this study. The inclusion of such variables in the hypothetical model might increase the variance explained on all the four occasions. Thus, the general theoretical framework presented in this study was a first step towards developing a more advanced theoretical framework which would include all the factors that influence mathematics achievement. Further work is required to expand the current attempt aimed at identifying the factors that influence mathematics achievement across the world.

Implications for Practice

The results of the path analyses for the four different data sets (EMS, FIMS, SIMS and TIMS) have revealed that the home background of students, number of students in class and attitudes of students towards mathematics are student level factors influencing achievement in mathematics over the last 30 years. Time in learning mathematics did not show any influence both in 1964 and 1978. However, it influenced the 1994 students' achievement directly, the main reason perhaps being the inclusion of a new and strong manifest variable, frequency of mathematics homework in a week. This variable also influenced the achievement level of Ethiopian students. These findings are consistent with Carroll's (1963) model of school learning. Carroll has argued that perseverance (motivation, attitude) and time for learning are some of the factors that influence school learning. These factors reported by Carroll continued to influence the Australian and Ethiopian students' mathematics achievement. Therefore, these factors are influencing both developed (Australia) and developing (Ethiopia) countries lower secondary school students' achievement in mathematics.

Thus, teachers, school administrators and curriculum designers need to consider the following points to improve the achievement level of students in mathematics.

1. Teachers should be aware of the importance of attitudes towards mathematics in the teaching learning process in mathematics, since students who expressed more positive attitudes towards mathematics, achieved higher than those students who expressed less positive attitudes towards the same subject. In view of the importance of mathematics for the technological development of nations, teachers should develop ways and means to improve the attitudes of students towards mathematics.
2. School administrators should also be ready to provide the necessary assistance for those students who are from a lower socioeconomic background in order to improve their achievement level in mathematics, since the evidence suggested that students from a higher socioeconomic background status families achieved better in mathematics than those students from lower socioeconomic status families.
3. Curriculum designers should provide enough time for students to learn mathematics and to provide regular homework during a school week, since the number of mathematics homework sessions in a week showed more influence on mathematics achievement than the time the students spent in doing their assignments.
4. School administrators should seek to reduce excessively large class groups. The evidence suggests the importance of reducing class sizes such as the Ethiopian average of more than 70 to more manageable levels. However, where manageable class sizes have been achieved, such as the Australian average of less than 20, further decreases in class sizes do not appear to have benefits. Rather, resources should be devoted to other priorities.

Implications for Future Research

The IEA-Mathematics studies administered in 1964, 1978 and 1994 for Australian lower secondary school students provided a unique opportunity to investigate changes over time. There were sufficient common mathematics test items and student background information questions to compare changes between 1964 and 1978. However, there were very few common mathematics test items and student background information questions between the two earlier studies and 1994. It is important to realise that, a larger number of common mathematics test items and student background information questions would be needed to improve the stability of the estimates, and the meaningfulness of comparisons over time. Future research studies of change in achievement over time should attempt to incorporate more common test items and common background information questions.

The findings of the comparisons in the mathematics achievement level of 13-year-old students between 1964 and 1978, and the Year 8 students between 1964 and 1994 revealed that the mathematics achievement levels of Australian students at the lower secondary school level have declined over the last three decades. However, the decline was not consistent in all Australian states. More recently, Tilahun and Keeves (1997b) examined changes in students' mathematics achievement over time in five

Australian states at the lower secondary school level. Two comparisons were made, the first comparison was for 13-year-old government school students in the five Australian states between 1964 and 1978. Only in one state did achievement improve over time (but the improvement was not statistically significant), and there was no difference between the two occasions in a further state. However, in the remaining three states, achievement over time declined, but the decline was significant only in one state and over the five state systems taken together.

The second comparison made by Tilahun and Keeves (1997b) was for the mathematics achievement level of Year 8 government school students in the five states between 1964 and 1994. The findings indicated that one state had improved in mathematics achievement over the last three decades (the improvement was not statistically significant), whereas in the remaining states the achievement level of Year 8 students level declined over the last 30 years. However, significant declines were recorded only in two states, and across the five state systems overall.

The question related to these findings would be why achievement in one state had improved over the 30-year period and had declined in the remaining four states and overall in Australia. The answer to this question might be related to the conditions of school learning in each school in Australia. Carroll (1963) has identified factors that facilitate or hinder learning in schools. The factors which were identified by Carroll are summarised here.

1. Ability: A student's ability to understand the nature of the subject he or she is going to learn.
2. Aptitude: A student's prior learning or specific level of knowledge about the subject he or she is going to learn.
3. Perseverance: the level of attitudes and motivation of a student to learn the subject.
4. Time in learning (opportunity to learn): the amount of time allowed for a student to learning the subject.
5. Quality of instruction: The level of the presentation, explanation and arrangements of the subject to be learned.

Among the five factors that influence school learning, attitude and time in learning were factors which were examined in the present study to determine whether or not they influenced mathematics achievement at the lower secondary school level in Australia. The result of the present study showed that attitude appeared as a significant variable on all the three occasions, while time in learning was a significant variable only in TIMS. The influence of the time spent in was a significant variable in the analysis for 1994 probably because of greater variability of the time allowed for students to learn mathematics across Australian schools. Therefore, further investigation would seem to be important to identify the reasons for the decline in mathematics achievement between 1964 and 1994 in four Australian states and overall in Australia and to suggest solutions for the problem. This decline in achievement in mathematics might not only be an Australian phenomenon. Thus it might be of interest to conduct a similar study on those countries that participated in the three international mathematics studies to investigate the trend on achievement in mathematics internationally over time.

Furthermore, the results of the present study show that analytical efforts should aim to employ methods that are appropriate for the data available. Thus, the results discussed in the present study have demonstrated that the use of advanced measurement and analytical procedures such as item response models and partial least squares path analysis can contribute greatly to understanding of educational phenomena. Thus, this study has indicated that the comparative study of education provides an understanding of phenomena involving educational outcomes as well as the factors that influence such outcomes.

Finally, the present study has indicated that the processes involved in mathematics achievement and factors influencing the achievement level of students at the 13-year-old and Year 8 level across countries of the world have more in common than is widely acknowledged.

13

References

Abera Mekonnen & Zewdu Desta (1992) Female participation in the Ethiopian educational system. Unpublished manuscript, Addis Ababa: Planning and External Relation Service, Ministry of Education.

Adams, R. J. & Gonzalez, E. J. (1996). The TIMSS test design. In M O Martin & D L Kelly (eds), *Third International Mathematics and Science Study: Technical Report vol. 1: Design and Development,* Boston: IEA, pp. 3-1 - 3-36.

Adams, R. J. & Khoo, S.T. (1993). *Quest- The Interactive Test Analysis System.* Hawthorn, Victoria: ACER.

Adams, R. J. (1984). *Sex Bias in ASAT?* Hawthorn, Victoria: ACER.

Ademe Mekonen & G. Meskel G. Eyesus. (1989). Does students' class activity indicate their academic achievement? *The Ethiopian Journal of Education, 11 (1),* 73-87.

Ademe, M & Gebre, B (1990). The impact of objective type tests on the learning process of high school mathematics. *The Ethiopian Journal of Education, 11 (2),* 28-49.

Ainley, J., Goldman, J. & Reed, R. (1990). *Primary Schooling in Victoria: A Study of Students' Attitudes and Achievements in Years 5 and 6 of Government Primary Schools*. Hawthorn, Victoria: ACER.

Almaz Eshetie (1989). Girls' participation and performance in education. Paper presented at the workshop on the psychology of female learners, Bahir Dar, Ethiopia.

Amir, N., Foa, E. B.& Coles, M. E. (1997). Factor structure of the Yale-Brown obsessive compulsive scale. *Psychological Assessment, 9 (3)*, 312-316.

Anbessu Biazen & Junge, B. (1988). *Problems in Primary School Participation and Performance in Bahir Dar Awraja.* Addis Ababa: Ministry of Education and UNICEF,

Anderson, J. (1989). Sex-related differences on objective tests among undergraduates. *Educational Studies in Mathematics, 20* (2), 165-177.

Anderson, L. W. (1994). Attitude measures. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2nd ed.), Oxford: Pergamon Press, pp. 380-390.

Anderson, L. W. (1997). Attitudes, measurement of. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook (*2nd ed.), Cambridge, UK: Pergamon Press, pp. 885-895.

Andrich, D. & Masters, G. N. (1988). Rating scale analysis. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 297-303.

Andrich, D. (1997). Rating scale analysis. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook (*2nd ed.), Cambridge, UK: Pergamon Press, pp. 874-880.

Assefa Beyene (1991). *Female Participants and Performance in Rural Primary Schools in Ethiopia.* Addis Ababa: Curriculum Evaluation and Educational Research Division, Institute for Curriculum Development and Research.

Atsede Wondimagegnehu & Kebede Teku (1988). Participation of Ethiopian women in education. Paper presented to the World Bank.

Atsede Wondimagegnehu (1991). Women in science and technology in Ethiopia. In Tsehai Berhane-Selassie (ed.), *Gender Issues in Ethiopia,* Addis Ababa Institute of Ethiopian Studies, Addis Ababa University, pp. 101-119.

Atweh, W. F.(1980) *Sexist Attitudes in Mathematics Education.* Mathematics education monograph, Kelvin Grove : College of Advanced Education.

Australian Association for Mathematics Teachers. (1989). School mathematics in New South Wales: Recent and current activities. *The Australian Mathematics Teacher, 45* (2), 5.

Australian Association for Mathematics Teachers. (1990a). Discipline review of teacher education in mathematics and science: A response from AAMT. *The Australian Mathematics Teacher 46 (1)*, 2-5.

Australian Association for Mathematics Teachers. (1990b). Girls and mathematics: Is there a problem? *The Australian Mathematics Teacher, 46 (3)*, 10-11.

Australian Bureau of Statistics (ABS). (1997a). *1996 Schools Australia.* Canberra: Commonwealth of Australia.

Australian Bureau of Statistics (ABS). (1997b). *1995-96 Expenditure on Education.* Canberra. Commonwealth of Australia.

Australian Council for Educational Research (ACER). (1964). *Primary School Mathematics: Report of a Conference of Curriculum Officers of State Education Departments, held in Melbourne 16th-20th March 1964*. Hawthorn, Victoria: ACER.

Australian Council for Educational Research (ACER). (1965). *Background in Mathematics. A Guidebook to Elementary Mathematics for Teachers in Primary Schools*. Melbourne: The education Department of Victoria and ACER.

Australian Education Council and Curriculum Corporation (1991). *A National Statement on Mathematics for Australian Schools.* Carlton, Victoria: Australian Education Council and Curriculum Corporation.

Australian Education Council. (1989). *Mapping the Australian curriculum: Preliminary Document. vol. I. the General Curriculum Executive Summary*. Canberra: Australian Education Council.

Australian Education Council. (1990a). *A Community Guide to a National Statement on Mathematics for Australian Schools.* A joint project of the States, Territories & the Commonwealth of Australia initiated by the Australian Education Council. Melbourne: Australian Education Council.

Australian Education Council. (1990b). *A National Statement on Mathematics for Australian Schools.* Canberra: Australian Education Council.

Bachelor, P., Michael, W. B. & Kim, S. (1994). First-order and higher-order semantic and figural factors in structure of intellect divergent production measures. *Educational and Psychological Measurement, 54 (3)*, 608-619.

Baker, D. P. & Jones, D. P. (1993). Creating gender equality: Cross-national gender stratification and mathematical performance. *Sociology of Education, 66*(2), 91-103.

Baker, D. P. (1997). Good news, bad news, and international comparisons: Comment on Bracey. *Educational Researcher, 26 (3)*, 16-17.

Baker, F. B. & Al-karni, A. (1991). A comparison of two procedures for computing IRM equating coefficients. *Journal of Educational Measurement, 28* (2), 147-162.

Baker, F. B. (1977). Advances in item analysis. *Review of Educational Research, 47*, 151-178.

Ballenden, C., Davidson, M. & Newell, F. (1985). *Better Chances for Girls: A Handbook of Equal Opportunity Strategies for use in Schools.* Melbourne: Victorian Institute of Secondary Education.

Baltes, P. B. & Nesselroade, J. R. (1979). History and rationale of longitudinal research. In Paul B. Baltes and John R. Nesselroade (eds), *Longitudinal Research in the Study of Behavior and Development*, New York: Academic Press, pp. 1-40.

Baxter, J. & Brinkworth, P. (1989). On questioning a national mathematics curriculum framework. *The Australian Mathematics Teacher, 45* (4), 7-9.

Beard, J. G. & Pettie, A. L. (1979). A comparison of linear and Rasch equating results for basic skills assessment tests. Florida state university: ERIC.

Beaton, A. E. (1996). Foreword. In M O Martin & D L Kelly (eds), *Third International Mathematics and Science Study: Technical Report vol 1: Design and Development,* Boston : IEA, p. ix.

Beaton, A. E., Martin, M. O, Mulls, I. V. S., Gonzalez, E. J., Smith, T. A. & Kelly, D. L. (1996b). *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Boston: IEA.

Beaton, A. E., Mulls, I. V. S., Martin, M. O, Gonzalez, E. J., Kelly, D. L. and Smith, T. A. (1996a). *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study.* Boston: IEA.

Behutiye, N. & Wagner, J. (1993). *Some Aspects of Primary School Pupils Cognitive Development and Their Relationship to Scholastic Achievement.* Addis Ababa: Curriculum Evaluation and Educational Research Division, Institute for Curriculum Development and Research.

Behutiye, N. & Wagner, J.(1995). *Some Aspects of Primary School Pupils Cognitive Development and Their Relationship to Scholastic Achievement: Comprehensive Report.* Addis Ababa: Curriculum Evaluation and Educational Research Division, Institute for Curriculum Development and Research.

Benson, J. (1987). Detecting item bias in affective scales. *Educational and Psychological Measurement, 47 (1)*, 55-67.

Berhanu Beyene (1986). An investigation of the effect of family background on academic success of students in Menelik II secondary school and its psychological implications. Unpublished BA thesis, Faculty of Education, Addis Ababa University.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F M Lord & M R Novick (eds), *Statistical Theories of Mental Test Scores*, Reading , MA: Addison-Wesley, pp. 395-479.

Bishop, F. & Clements, M. A. (1994). Predicting performance differences between years 5 and 6 boys and girls on pencil-and-paper mathematics items. Paper presented at the annual conference of the Australian Association for Research in Education, the University of Newcastle, NSW.

Blakers, A. L. (1976). Mathematics education in Australia: Past and present. *The Australian Mathematics Teacher, 32* (4), 145-155.

Blakers, A. L. (1977). Mathematics education in Australia: Past and present. *The Australian Mathematics Teacher, 33* (1), 17-22.

Blakers, A. L. (1978). Change in mathematics education since the late 1950's- Ideas and realisation Australia. *Educational Studies in Mathematics, 9*, 147-158.

Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York: Mc Graw-Hill.

Bonda, K. & Schwarz, V. (1985). *Education for Girls.* Melbourne: Australian Educational Council.

Bourke, S. F. (1984). *The Teaching and Learning of Mathematics: National Report of the Second Phase of the IEA Classroom Environment Study*. Hawthorn, Victoria: Australian Council for Educational Research.

Bourke, S. F. (1986). Achievement and attitudes in mathematics: Some relationships between classroom contexts, teaching practices and student outcomes. *Research in Mathematics Education in Australia, 6* (2), 2-9.

Bourke, S. F., Mills, J. M., Stanyon, J. & Holzer, F. (1981). *Performance in Literacy and Numeracy: 1980.* Canberra: Australian Education Council.

Brick, J. M., Broene, P., James, P. & Severynse, J. (1996). *A User's Guide to WesVarPC.* (Version 2.0). Boulevard: Westat, Inc.

Brick, J. M., Broene, P., James, P. & Severynse, J. (1997). *A User's Guide to WesVarPC.* (Version 2.11). Boulevard: Westat, Inc.

Brinkworth, P. (1985). Numeracy: What, why, how? *The Australian Mathematics Teacher, 41* (4). 8-10.

Brown, H. C. (1917). *Intelligence and Mathematics*. Henly Holt.

Brueckner, L. J. (1942). *Arithmetic We Use.* Chicago: J.C. Winston.

Bukowski, W. M., Hoza, B. & Boivin, M. (1993). Popularity, friendship, and emotional adjustment during early adolescence. *New Directions for Child Development, No 60*, 23-37.

Cameron, L. C., Ittenbach, R. F., McGrew, K. S., Harrison, P. L., Taylor, L. R. & Hwang, Y. R. (1997). Confirmatory factor analysis of the K-ABC with gifted referals. *Educational and Psychological Measurement, 57 (5),* 823-840.

Campbell, J. R. (1996). *PLSPATH Primer* (2nd ed.). New York: St. John's Press.

Carroll, J. B. (1963). A model of school learning. *Teachers College Record, 64,* 723-733.

Carss, M. C. (1980). Girls, mathematics and language: Some observations from classrooms. In Adelaide College of the Arts and Education (ed.), *Improving Maths for Girls: Report of a Conference Held at Raywood Inservice Centre, 30 May-1 June, 1980*, pp. 16-18.

Carss, M. C. (1984). The Australian mathematics education program. In P Costello, S Ferguson, K Slinn, M Stephens, D Trembath & D Williams (eds), *Facets of Australian Mathematics Education,* Adelaide: Australian Association of Mathematics Teachers for the Fifth International Congress on Mathematics Education, pp. 27-44.

Carss, M. C. (1988). What would a nice girl like you want with a calculator? In S Taylor & M Henry (eds), *Battlers and Bluestockings Women's Place in Australian Education,* Deakin: Australian College of Education, pp. 95-100.

Cheung, K. C. & Keeves, J. P. (1990). The problems with current analytical procedures. *International Journal of Educational Research 14 (3)* 233-244.

Choppin, B. H. & Wolf, R. M. (1992). Correction for guessing. In J P Keeves (ed.), *The IEA Technical Handbook*, The Hague : IEA, pp. 171-176.

Choppin, B. H. (1974). *The Correction for Guessing on Objective Tests*. Stockholm, Sweden.: IEA.

Choppin, B. H. (1988). Correction for guessing. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 384-386.

Christiansen, N. D., Lovejoy, M. C., Szymanski, J. & Lang, A. (1996). Evaluating the structural validity of measures of hierarchical models: An illustrative example using the social problem-solving inventory. *Educational and Psychological Measurement, 56(4)*, 600-625.

Clarkson, P. (1979). RAMP-A resource for school-based curriculum development. *Research in Mathematics Education in Australia 1979 vol. I.*, Kelvin Grove: MERGA.

Clarkson, P. C. (1991). *Bilingualism and Mathematics Learning.* Geelong, Victoria: Deakin University Press.

Clarkson, P. C. (1992). Language and mathematics: A comparison of bilingual and monolingual students of mathematics. *Educational Studies in Mathematics, 23* (4), 417-429.

Clarkson, P. C. (1993). Students who talk in two languages. In J Mousley & M Rice (eds), *Mathematics: Of Primary Importance,* Victoria: The Mathematics Association of Victoria, pp. 357-366.

Clements, M. A. (1989). *Mathematics for the Minority: Some Historical Perspectives of School Mathematics in Victoria*. Geelong: Victoria: Deakin University.

Cohen, A. S. & Kim, S. H. (1993). A Comparison of Lord's chi square and Raju's area measures in detection of DIF. A*pplied Psychological Measurement; 17 (4*), 39-52.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155-159.

Cole, N. S. & Moss, P. A. (1989). Bias in test use. In R L Linn (ed.)*, Educational Measurement,* (3rd ed.), London: Collier Macmillan Publishers, pp. 201-219.

Comber, L. C. & Keeves, J. P. (1973). *Science Education in Nineteen Countries.* Stockholm: Almqvist and Wiksell.

Cook, L. L. & Eignor, D. R. (1991). IRM equating methods. *Educational Measurement: Issues and Practice, 10* (3), 37-45.

Costello, P.(1978.) Make it with maths. *Vinculum 15* (2), 47-51.

Cowban, R. H. (1987). Historical perspectives (The first interstate seminar for teaching mathematics in secondary schools, Sydney, 1962). *Vinculum 24* (3), 27-29.

Cox, P. (1993). Sex and CATs: Gender related CAT bias in VCE mathematics. In J Mousley & M Rice (eds), *Mathematics: Of Primary Importance,* Melbourne: Mathematics Association of Victoria, pp. 139-144.

Cox, P. (1994). Sex and Mathematics CATs (gender differences in 1993 VCE mathematics assessment). In C Beesey & D Rasmussen (eds), *Mathematics Without Limits*, Melbourne: Mathematics Association of Victoria, pp. 20-28.

Crawford, K. P. (1988). Equality of opportunity and equality of conditions in mathematics learning for girls. *Unicorn, 14* (3), 155-160.

Damtew Gebru (1972). Certain social factors affecting scholastic achievement of fifty Grade 5 Pupils from Menelik II school. Unpublished BA thesis, Faculty of Education, Addis Ababa University.

Daniel Desta (1995). School attendance and achievement among boys and girls: The case of selected grade six students in Addis Ababa. *Proceedings of the National Workshop on Strengthening Educational Research,* Addis Ababa: Institute of Educational Research, Addis Ababa University, pp. 253-264.

Dawe, L. (1983). Bilingualism and mathematical reasoning in English as a second language. *Educational Studies in Mathematics 14,* 325-353.

Dekkers, j., De Laeter, J. R. & Malone, J. A. (1986). *Upper Secondary School Science and Mathematics Enrolment Patterns in Australia, 1970 - 1985.* Bentley, Western Australia: Science and Mathematics Education Centre, Western Australia Institute of Technology.

Dekkers, j., De Laeter, J. R. & Malone, J. A. (1991). *Upper Secondary School Science and Mathematics Enrolment Patterns in Australia, 1970-1989.* Perth: Curtin University.

Derese Mengistu, Wagner, J. & Alemayehu Minas (1990). *Factors Affecting Achievement of Lower Primary School Pupils.* Addis Ababa.: Institute of Curriculum Development and Educational Research.

Diamond, J. J. (1992). A graphic procedure for studying differential item functioning. *Journal of Experimental Education, 60 (4)*, 351-357.

Dorans, N. J. (1990). Equating methods and sampling designs. *Applied Measurement in Education, 3* (1), 3-17.

Doron, R. (1986). The "average system"-A new technique for equating scores on different tests constructed from item bank. *Studies in Educational Evaluation, 12,* 169-175.

Dungan, J. F. & Thurlow, G. R. (1989). Students' attitudes to mathematics: A review of the literature. *The Australian Mathematics Teacher, 45* (3), 8-11.

Durell, C. V. (1930). *New Algebra for Schools. Parts I, and II.* London: Bell.

Durell, C. V. (1931). *New Algebra for Schools. Parts III.* London: G. Bell.

Durell, C. V. (1939). *New Geometry for Schools: Stage B Parts I, II and III*. London: Bell.

Durell, C. V. (1950). *General Arithmetic for Schools. Parts I, II and III*. London: Bell.

Eckstein, M. A. (1982). Comparative school achievement. In H E Mitzel (ed.), *Encyclopedia of Educational Research, vol*. 1, New York, NY: The Free Press, pp. 323-329.

Edwards, A. L. & Kilpatrick, F. P. (1948). Scale analysis and the measurement of social attitudes. *Psychometrika, 13 (2)*, 99-114.

Edwards, A. L. (1957). *Techniques of Attitude Scale Construction.* New York: Appleton-Century-Crofts, Inc.

Elley, W. B. (1992). *How in the World do Students Read?* The Hague: IEA.

Elley, W. B. (1994). *The IEA Study of Reading Literacy: Achievement and Instruction in Thirty-Two School Systems.* Oxford: Pergamon Press.

Ellis, B. B, Becker, P. & Kimmel, H. D. (1993). An item response theory evaluation of an English version of the Trier Personality Inventory (TPI). *Journal of Cross-Cultural Psychology, 24 (2)*, 133-48.

Endalkachew W. Mariam (1990). Factors that affect students' academic performance in some selected senior secondary schools in Addis Ababa. Unpublished BA thesis, Faculty of Education, Addis Ababa University.

Eshetu Alemu (1988). Comparative study of family back-ground of high achievers and low achievers at Entoto technical, vocational and academic school. Unpublished BA thesis, Faculty of education, Addis Ababa University.

Ethington, C. A. (1990). Gender differences in mathematics: An international perspective. *Journal for Research in Mathematics Education, 21 (1)*, 74-80.

Ethiopian School Leaving Certificate Examinations Office. (1985). *The Ethiopian School Leaving Certificate Examinations Handbook.* Addis Ababa: Ethiopian School Leaving Certificate Examinations Office, Addis Ababa University.

Evaluative Research of the General Education System in Ethiopia. (1986). *A Quality Study: Summary Report Presented to the Executive Committee of ERGESE*. Addis Ababa: Evaluative Research of the General Education System in Ethiopia Secretariat.

Falk, F. R. & Miller, N. B. (1992). *A Primer for Soft Modelling*. Akron, Ohio: the University of Akron Press.

Falk, F. R. (1987). *A Primer for Soft Modelling.* Berkeley: Institute of Human Development, University of California.

Feldt, L. S. & Brennan, R. L. (1982). Reliability. In P W Holland & D B Rubin (eds), *Test Equating,* New York: Academic Press, pp. 105-145.

Feldt, L. S. & Brennan, R. L. (1989). Reliability. In R L Linn (ed.), *Educational Measurement,* (3rd ed.), London: Collier Macmillan Publishers, pp. 105-146.

Fields, D. L & Herold, D. M. (1997). Using the leadership practices inventory to measure transformational and transactional leadership. *Educational and Psychological Measurement, 57 (4),* 569-579.

Finn, J. D. (1997). Hypothesis testing. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (2nd ed.), Cambridge, UK: Pergamon Press, pp. 556-561.

Firth, D. (1981). School mathematics: What a mess! In A Rogerson (ed.), *Maths, Myths and Realities*, Mathematics Association of Victoria 18th Annual Conference, pp. 36-42.

Fitzpatrick, J. & Brown, S. (1983). Sex based enrolment patterns in secondary school subjects. *Curriculum Perspectives, 3* (2), 47-52.

Flanagan, J. C. (1982). Discussion of "some issues in test equating." In P W Holland & D B Rubin (eds), *Test Equating,* New York: Academic Press, pp. 243-246.

Foy, R., Rust, K. & Schleicher, A. (1996). Sample design. In M O Martin & D L Kelly (eds), *Third International Mathematics and Science Study: Technical Report Vol 1: Design and Development,* Boston: IEA, pp. 4-1 to 4-17.

Francis, E. C. (1951). Highway Arithmetic. London, Longmans.

Fraser, B. J. (1979). Evaluation of science-based curriculum. In H J Walberg, (ed.), *Educational Environments and Effects: Evaluation, Policy, and Productivity*, Berkeley: McCutchan, pp. 218-234.

Fraser, B. J. (1980). Grade level and sex differences in attitude to several school subjects. *The Australian Journal of Education, 24* (2), 128-136.

Freudenthal, H. (1975). Pupil's achievements internationally compared. *Educational Studies in Mathematics, 6,* 127-186.

Fuson, K., Stigler, J. & Bartsch, K. (1988). Grade placement of addition and subtraction topics in Japan, China, the USSR, Taiwan, and the U.S. *Journal for Research in Mathematics Education 19*, 449-456.

Garden , R. A. & Orpwood, G. (1996). Development of the TIMSS Achievement tests. In M 0. Martin & D L Kelly (eds), *Third International Mathematics and Science Study: Technical Report, vol. 1: Design* and Development*,* Boston: IEA, pp. 2-1 to 2-19.

Garden, R. A. (1987). The second IEA mathematics study. *Comparative Education Review, 31* (1), 47-68.

Gavin, W. D. A. & Herry, Y. (1996). The French self-perception profile for children score validity and reliability. *Educational and Psychological Measurement, 56 (4)*, 678-700.

Gennet Zewdie (1991a). *Women in Education: A Study of Academic Performance and Participation of Female Students in the High Schools of Addis Ababa*. Addis Ababa: Department of Business Education, Faculty of Education, Addis Ababa University,

Gennet Zewdie (1991b). Women in primary and secondary education. In Tsehai Berhane-Selassie (ed.), *Gender Issues in Ethiopia.* Addis Ababa.: Institute of Ethiopian Studies, Addis Ababa University, Addis Ababa, pp. 88-98.

Gieral, M. J. & Rogers, W. T. (1996). A confirmatory factor analysis of the test anxiety inventory using Canadian high school students. *Educational and Psychological Measurement, 56(2)*, 315-324.

Goldman, M. S., Greenbaum, P. E. & Darkes, J. (1997). A confirmatory test of hierarchical expectancy structure of predictive power: Discriminant validation of the alcohol expectancy questionnaire. *Psychological Assessment 9 (2),* 145-157.

Grace, N. & Carss, M. (1989). A national curriculum or the national curriculum. *The Australian Mathematics Journal, 45* (4), 12-13.

Grace, N. (1989). Years 1 to 10 mathematics curriculum development in Queensland. *The Australian Mathematics Teacher, 45* (2), 4-5.

Green, D. R. (1975). What does it mean to say a test is biased? *Educational and Urban Society, 8,* 33-52.

Gruijter, D. N. M. & Kamp, L. J. T. (1991). Generalizability theory. In R K Hambleton and J N Zaal (eds), *Advances in Educational and Psychological Testing*, Boston: Kluwer Academic Publishers, pp. 45-69.

Guttman, L. (1947). On Festinger's evaluation of scale analysis. *Psychological Bulletin, 44*, 451-465.

Hair, J. F., Anderson, R. E., Tatham, R. L. & Black, W. C. (1995). *Multivariate Data Analysis with Readings,* (4th ed.), Englewood, New Jersey: Prentice-Hall International, Inc.

Hambleton, R. K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Application*. Boston MA: Kluwer Academic Publishers.

Hambleton, R. K. (1989). Principles and selected applications of item response theory. In R Linn (ed.), *Educational Measurement,* (3rd ed.), New York: Macmillan. pp. 147-200.

Hambleton, R. K. (1993). Principles and selected applications of item response theory. In R L Linn (eds.), *Educational Measurement* (3rd ed.), Phoenix, AZ: United States of America, Oryx Press, pp. 147-200.

Hambleton, R. K.& Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement, 14 (2)*, 75-96.

Hambleton, R. K., Zaal, J. N. & Pieters, J. P. M. (1991). Computerized adaptive testing: theory, applications, and standards. In R K Hambleton & J N Zaal (eds), *Advances in Educational and Psychological Testing*, Boston: Kluwer Academic Publishers, pp. 341-366.

Hanna, G. & Kündiger, E. (1986). Differences in mathematical achievement levels and in attitudes for girls and boys in twenty countries. Paper presented at the Annual Conference of the American Educational Research Association, San Francisco.

Hanna, G. (1989). Mathematics achievement of girls and boys in grade eight: Results from twenty countries. *Educational Studies in Mathematics, 20* (2), 225-232.

Harman, H. H. (1976). *Modern Factor Analysis*. Chicago, IL: University of Chicago Press.

Harmon, M. G., Morse, D. T. & Morse, L. W. (1996). Confirmatory factor analysis of the Gibb experimental test of testwiseness. *Educational and Psychological Measurement, 56 (2)*, 276-286.

Harnisch, D. L., Walberg, H. J., Shiow-Ling, T., Takahiro, S. & Fyans, L. J.Ir. (1985). Mathematics productivity in Japan and Illinois. *Evaluation in Education: An International Review Series, 9 (3)*, 277-284.

Harnischfeger, A. & Willey, D. E. (1978). Conceptual issues in models of school learning. *Curriculum Studies, 10* (3), 215-231.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9 (2)*, 139-164.

Havercamp, S. M. & Reiss, S. (1997). The reisss screen for maladaptive behavior: Confirmatory factor analysis. *Behav. Res. Ther. 35 (10)*, 967-971.

Helfers, M. (1986). Children's attitudes to mathematics. *Rhombus, 14,* 17-23.

Hirut Immru (1986). Research priorities and support needs on women and development in Ethiopia: Educational training and employment needs in Ethiopia. Paper presented at OSSREA workshop on women and development, Nazareth.

Hochwarter, W. A., Harrison, A. W. & Amason, A. C. (1996). Testing a second-order multidimensional model of negative affectivity: A cross-validation study. *Educational and Psychological Measurement, 56 (5),* 791-808.

Hoijtink, H. & Molenaar, I. W. (1992). Testing for DIF in a model with single peaked item characteristic curves: The PARELLA Model. *Psychometrika, 57 (3)*, 383-397.

Holland, P. W. & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H Wainer & H I Braun (eds), *Test Validity,* Hillside, NJ: Lawrence Erlbaum, pp. 129-145.

Holland, P. W. Rubin, D, B. (1982). Introduction: Research on test equating sponsored by Educational Testing Service, 1978-1980. In P W Holland & D B Rubin (eds), *Test Equating,* New York: Academic Press, pp. 1-6.

Hopkins, C. D. & Antes, R. L. (1990). *Classroom Measurement and Evaluation*, (3rd ed.), F. E. Itasca, Illinois: Peacock Publishers.

Horne, M & Stephens, M. (1989). Post-compulsory mathematics in Victorian schools. *The Australian Mathematics Teacher, 45* (2), 6.

Hungi, H. (1997). Measuring basic skills across elementary school grades. Unpublished MA thesis, School of Education, the Flinders University of South Australia.

Husén, T. (1996). Lessons from the IEA studies. *International Journal of Educational Research, 25 (3)*, 207-218.

Husén, T. (ed.), (1967). *International Study of Achievement in Mathematics (vols 1 & 2)*. Stockholm: Almquist & Wiksell.

Imperial Ethiopian Government (1956). *Five Year Development Plan 1957-1961.* Addis Ababa, Imperial Ethiopian Government.

Inkeles, A. (1977). A Review of " international studies in education". In *Proceedings of the National Academy of Education, 4,* 139-200.

Inkeles, A. (1979). National differences in scholastic performance. *Comparative Education Review, 23 (3),* 386-407.

Jacobs, J. E. (1991). Influence of gender stereotypes on parent child mathematics attitudes. *Journal of Educational Psychology, 83 (4),* 518-527.

Jamieson, R. & Karmelita, W. (1989). Change and development in the mathematics syllabus in Western Australia. *The Australian Mathematics Teacher, 45* (2), 8-9.

Jeffery, P. (ed.) (1975). *Primary School Mathematics in Australia: Review and Forecast.* Hawthorn, Victoria, Australia: Australian Council for Educational Research.

Jones, E. P. (1973). *Comparative Education: Purpose and Method.* Brisbane: University of Queensland Press.

Jones, G. A. (1979). Change in mathematics education in Queensland since the late 1950s. Part II. *Teaching Mathematics, 4* (1), 20-25.

Jöreskog, K. L. & Sörbom, D. (1993a). *LISREL8: Structural Equation Modelling with the SIMPLIS™ Command Language,* Hillsdale, N. J.: Lawrence Erlbaum Associates

Jöreskog, K. L. & Sörbom, D. (1993b). *New Features in PRELIS™2*. Scientific Software. Chicago.

Keeves J. P. & Bourke, S. F. (1976). *Australian Studies in School Performance, Literacy and Numeracy in Australian schools: A First Report*. Canberra: Australian Government publishing service.

Keeves, J. P. & Adams, D. (1997). Comparative methodology in education. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (2nd ed.), Cambridge, UK: Pergamon Press, pp. 31-41.

Keeves, J. P. & Kotte, D. (1992). Disparities between the sexes in science education: 1970-84. In J P Keeves (ed.), *The IEA Study of Science III : Changes in Science Education and Achievement: 1970 to 1984*, Oxford: Pergamon Press, pp. 263-290.

Keeves, J. P. & Kotte, D. (1994). Sex differences and educational outcomes. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2n ed.), Oxford: Pergamon Press, pp. 5423-5433.

Keeves, J. P. & Kotte, D. (1996). The Measurement and reporting of key competencies. In *Teaching and Learning the Key Competencies in the Vocational Education and Training sector,* Adelaide: Flinders Institute for the Study of Teaching, pp. 139-168.

Keeves, J. P. & Mason, J. (1980). Sex differences in attitudes towards, achievement in and participation in mathematics in school. In Adelaide College of the Arts and Education (ed.), *Improving Maths for Girls: Report of a Conference Held at Raywood Inservice Centre, 30 May-1 June, 1980*, pp. 31-38.

Keeves, J. P. & Radford, W. C. (1969). *Some Aspects of Performance in Mathematics in Australian schools*. Hawthorn, Victoria: Australian Council for Educational Research.

Keeves, J. P. & Schleicher, A. (1992). Changes in Science Achievement: 1970-84. In J P Keeves (ed.), *The IEA Study of Science III : Changes in Science Education and Achievement: 1970 to 1984*, Oxford: Pergamon Press, pp. 141-151.

Keeves, J. P. (1966). Students' attitudes concerning mathematics. Unpublished MEd thesis, University of Melbourne.

Keeves, J. P. (1966a). Evaluation of Achievement in Mathematics. Memorandum No 4. Melbourne: ACER.

Keeves, J. P. (1968). V*ariation in Mathematics Education in Australia*: Some Interstate *Differences in the Organization, Courses of Instruction, Provision for and Outcomes of Mathematics Education in Australia.* Hawthorn, Victoria: ACER.

Keeves, J. P. (1972). *Educational Environment and Student Achievement: A Multivariate Study of the Contribution of the Home, the School and the Peer Group to Change in Mathematics and Science Performance During the First Year at Secondary School.* Stockholm: Almqvist & Wiksell.

Keeves, J. P. (1975). The home, the school and achievement in mathematics and science. *Science Education, 59 (4)*, 207-218.

Keeves, J. P. (1976). *Curricular Factors Influencing School Learning: Time and Opportunity to Learn. [IEA (Australia) Report 1976:2]*. Hawthorn, Victoria: Australian Council for Educational Research.

Keeves, J. P. (1988a). Models and model building. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 559-566.

Keeves, J. P. (1988b). Path analysis. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 723-731.

Keeves, J. P. (1991). *Changes in Science Education and Achievement*: 1970-1984. Pergamon Press, Oxford.

Keeves, J. P. (1992a). The design and conduct of the second science study. In J P Keeves (ed.), *The IEA Study of Science III: Changes in Science Education and Achievement: 1970 to 1984,* Oxford, Pergamon Press, pp. 42-67.

Keeves, J. P. (1992b). Scaling achievement test scores. In J P Keeves (ed.), *The IEA Technical Handbook*, The Hague : IEA, pp. 107-125.

Keeves, J. P. (1992c). Missing data and non-response. In J P Keeves (ed.), *The IEA Technical Handbook*, The Hague: IEA, pp. 261-270.

Keeves, J. P. (1994). Models and model building. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2nd ed.), New York: Pergamon Press, pp. 3865-3872.

Keeves, J. P. (1995). *The World of School Learning: Selected Key Findings from 35 Years of IEA Research.* The Hague, The Netherlands: The International Association for the Evaluation of Education.

Keeves, J. P. (1997a). Models and model building. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*, (2nd ed.), Cambridge, UK: Pergamon Press, pp. 386-394.

Keeves, J. P. (1997b). Longitudinal research methods. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook*, (2nd ed.), Cambridge, UK: Pergamon Press, pp. 138-149.

Keith, T. Z. (1993). Latent variable structural equation models: LISREL in special education research. *Remedial and Special Education, 14(6)*, 36-46.

Kennedy, J. P. (1981). Current trends -future practice? In A Rogerson (ed.), *Maths, Myths and Realities*, Maths Association of Victoria 18th Annual Conference, pp. 496-498.

Kings, E. J. (1975). Some recent changes in the curriculum, content and organisation of primary mathematics and their effects in Australian schools. In P Jeffery (ed.), *Primary School Mathematics in Australia: Review and Forecast.* Hawthorn, Victoria, Australia: Australian Council for Educational Research, pp. 136-162.

Kiryluk, S. (1980). What the pupils think. *Mathematics Teacher, 91*, 42-44.

Klieme, E. & Stumpf, H. (1991). DIF: A computer program for the analysis of differential item performance. *Educational and Psychological Measurement, 51(3)*, 669-671.

Kline, P. (1993). *Rasch Scaling and Other Scales. The Handbook of Psychological Testing*. London: Routledge.

Kolen, M. (1994). Equating of tests. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2nd ed.), New York: Pergamon Press, pp. 6349-6355.

Kolen, M. (1997). Equating of tests. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,(* 2nd ed.), Cambridge, UK: Pergamon Press, pp. 730-737.

Kolen, M. J. & Whitney, D. R. (1981). Comparison of four procedures for equating the tests of general educational development. Paper presented at the annual meeting of the American Educational Research Association. Los Angeles, California.

Kolen, M. J. (1988). Defining score scales in relation to measurement error. *Journal of Educational Measurement, 25,* 97-110.

Koretz, D. *et al*. (1993). *Omitted and Not-Reached Items in Mathematics in the 1990 National Assessment of Educational Progress*. National Center for Research on Evaluation, standards, and Student Testing, Los Angeles, CA. National Center for Education Statistics (ED), Washington, DC.

Kotte, D. (1992). *Gender Differences in Science Achievement in 10 Countries*. Frankfurt am Main: Peterlang.

Lautenschlager, G. J., Flaherty, V. L. & Park, D. (1994). IRT differential item functioning: An examination of ability scale purifications. *Educational and Psychological Measurement, 54(1)*, 21-31.

Leach, D. J. & Tunnecliffe, M. R. (1984). The relative influence of time variables on primary mathematics achievement. *The Australian Journal of Education, 28* (2), 126- 131.

Leder, G. & Forgasz, H. (1992). Gender: A critical variable in mathematics education. In B Atweh & J Watson (eds), *Research in Mathematics Education in Australasia 1988-1991,* pp. 67-95.

Leder, G. & Forgasz, H. J. (1991). Learning mathematics: Some student perspectives. In J O'Reilly & S Wettenhall (eds), *Mathematics Ideas.* Parkville: Mathematical Association of Victoria, pp. 3-9.

Leder, G. (1979). Interest session: Attitudes to mathematics. In G Booker (ed.), *Research in Mathematics Education in Australia,* vol. 2, Brisbane: Mathematics Education Research Group of Australia, pp. 150-158.

Leder, G. (1980). Bright girls, mathematics and fear of success. *Educational Studies in Mathematics, 11,* 411-423.

Leder, G. (1985). Sex-related differences in mathematics : An overview. *Educational Studies in Mathematics, 16,* 304-309.

Leder, G. C. (1989). Girls and mathematics: A question of equity. In B Doing (ed.), *Everyone Counts,* Parkville: Mathematical Association of Victoria, pp. 18-23.

Leder, G. C. (1990). Gender differences in mathematics: An overview. In E Fennema & G C Leder (eds). *Mathematics and Gender.* New York: Teachers College Columbia University, pp. 10-26.

Levine, M. S. (1977). *Canonical Analysis and Factor Comparison.* Beverly Hills: Sage.

Lietz, P. (1992). Factors influencing reading Achievement at the 14-year-old Level in 15 Educational Systems. Unpublished Master's thesis, The Flinders University of South Australia, Adelaide, Australia.

Lietz, P. (1996). *Changes in Reading Comprehension across Culture and Over Time*. German: Waxman, Minister.

Linn, R. L. & Gronlund, N. E. (1995). *Measurement and Assessment in Teaching*. Englewood Cliffs, N. J.: Merrill.

Lokan, J., Ford, P. & Greenwood, L. (1996). *Maths and Science on the Line: Australian Junior Secondary Students' Performance in the Third International Mathematics and Science Study*. Camberwell: ACER.

Long, J. S. (1983a). *Confirmatory Factor Analysis: A Preface to LISREL*. Beverly Hills: Sage Publications.

Long, J. S. (1983b). *Covariance Structure Models: Introduction to LISREL*. Beverly Hills: Sage Publications.

Longford, N. T. (1994). Models for scoring missing responses to multiple-choice items. *Program Statistics Research Technical Report No. 94-1*. Educational Testing Service, Princeton, N. J.

Lord, F. M. & Stocking, M. L. (1988). Item response theory. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 269-272.

Lord, F. M. (1974). Quick estimates of the relative efficiency of two tests as a function of ability level. J*ournal of Educational Measurement , 11,* 247-254.

Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, new Jersey: Lawrence Elbuam, Associates.

Lovitt, C., Clarke, D. & Stephens, M.(1985). Mathematics teachers and curriculum renewal- A process of change and growth. *The Australian Mathematics Teacher, 41* (3), 12-13.

Lowe, I. (1979). A profile of the RIME project. *Vinculum, 16* (5), 147-150.

Ludlow, L. H. Omitted and not-reached responses. Draft paper produced for the Technical advisory Committee of the International Association for the Evaluation of Educational Achievement. The Hague, The Netherlands.

MacCann, R. G. (1989). A derivation of Levine's formulae (for equating unequally reliable tests using random groups) without the assumption of parallelism. *Educational and Psychological Measurement, 49* (1), 53-58.

MacDonald, T. H. (1977). Resolution in mathematics education: Suggestions and prediction. *The Australian Mathematics Teacher, 33* (2), 54-58.

Mallinson, V. (1975). *An Introduction to the Study of Comparative Education*. London: Heineman.

Manolis, C., Keep, W. W., Joyce, M. L. & Lambert, D. V. (1994). Testing the underlying structure of a storage image scale. *Educational and Psychological Measurement, 54 (3)*, 628-645.

Manolis, C., Levin, A. & Dahlstrom, R. (1997). A generation, x scale: Creation and validation. Educational and Psychological Measurement, 57 (4), 666-684.

Marsh, H. W. & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement 20 (3),* 231-249.

Martin, M. O. (1996). Third international mathematics and science study: An overview. In M O Martin & D L Kelly (eds), *Third International Mathematics and Science Study: Technical Report vol 1: Design and Development*, Boston: IEA, pp. 1.1-1.19.

Maxwell, B. (1996).Translation and cultural adaptation of the survey instruments. In M O Martin & D L Kelly (eds), *Third International Mathematics and Science Study Technical Report vol 1: Design and Development*, Boston: IEA, pp. 8.1-8.11.

McCollam, K. M., Embretson, S. E., Mitchell, D. W. & Horowitz, F. D. (1997). Using confirmatory factor analysis to identify newborn behavior structure with the NBAS. *Infant Behavior and Development 20 (2)*, 123-131.

McQualter, J. W. (1980). Mathematics education in Australia 1945-1979. In D Williams & H Crawford (eds), *Mathematics Theory into Practice, 1980,* Armidale: MERGA, pp. 52-60.

Mehrens, W. A. & Lehmann, I. J. (1984). *Measurement and Evaluation in Education and Psychology,* (3rd ed.), New York: Holt, Rinehart & Winston.

Mellenbergh, G. J. (1994). A unidimensional latent trait model for continuous item responses. *Multivariate Behavioral Research, 29 (3)*, 223-236.

Meredith, W. & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika, 57 (2)*, 289-311.

Meredith, W. & Milsap, R. E. (1985). On component analyses. *Psychometrika, 50,* 495-507.

Milne, L. (1992). Bridging mathematics students: Attitudes, autonomous learning behaviours, and problem solving. In B Southwell, B Perry & K Owens.(eds), *Proceedings of the Fifteenth Annual Conference Mathematics Education Research Group of Australasia (MERGA)* UNWS Conference Centre , Richmond: The Mathematics Education Group of Australasia, pp. 382-389.

Ministry of Education & Fine Arts (1951/52). *Secondary School Curriculum*. Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education (1975). *Ethiopian Education Direction: Draft.* Provisional military government of Ethiopia: Addis Ababa: Ministry of Education.

Ministry of Education (1982/83). *The Importance, Objectives and Contents of the New Mathematics Education*.: Addis Ababa: Provisional Military Government of Ethiopia: Ministry of Education.

Ministry of Education and Fine Arts (1947/48). *Elementary School Curriculum Years I-VIII*. Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1950/51). *Educational Leaders Annual Conference Proceedings 1940-1943 EC.* Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1955). *A Ten-Year Plan for the Controlled Expansion of Ethiopian Education.* Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1958/59a). *Elementary Community School Curriculum Years I-VI: Experimental*. Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1958/59b). *Secondary School Curriculum (provisional)*. (2nd ed.). Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1962/63). *Elementary School Curriculum Years I- VI*. Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1963/64). *Secondary School Curriculum Book 1*. Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1967). *Programme of Work Submitted by the Division of Curriculum and Teaching Material. Approved by the Major Policy Committee.* Addis Ababa: Imperial Ethiopian Government Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1967/68a). *Elementary School Curriculum Years I-VI.* (2nd ed.), Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1967/68b). S*econdary School Curriculum, Book 1*. (2nd ed..), Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1970/71a). *Elementary School Curriculum Years I-VI*. (3rd ed.), Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1970/71b). *The General Secondary School Curriculum Book 1.* (3rd ed.), Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1972/73a). *Elementary School Curriculum Years I-VI.* (4th ed.), Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts (1972/73b). *The General Secondary School Curriculum Book 1.* (4th ed.), Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts, and HSIU Faculty of Education (1967). *Secondary School Curriculum Development and E. S. L. C. Examination Seminar. Final Report.* Addis Ababa: Ministry of Education.

Ministry of Education and Fine Arts. (1968a). Report on education development in 1967 - 1968. Presented at the 31st session of the conference on public education. Addis Ababa: Ministry of Education and Fine Arts.

Ministry of Education and Fine Arts. (1968b), *Brief Report on Educational Developments in Ethiopia During 1968-1969.* Addis Ababa, Ethiopia: Ministry of Education.

Ministry of Education. (1980). *New Educational Objectives and Directions for Ethiopia General Introduction to and Summary of the Proposals for Educational Reform.* Addis Ababa: Ministry of Education.

Ministry of Education. (1981). *Hints Concerning Content, Methods and Organization of Mathematics Instruction for Grades 9-12.* Addis Ababa.: Curriculum Department, Ministry of Education.

Ministry of Education. (1996). *The Federal Republic of Ethiopia Educational Statistics Annual Abstracts 1987 E. C. 1994/95.* Addis Ababa: Educational Management Information Systems EMIS, Ministry of Education.

Ministry of Education. (1997). *The Federal Republic of Ethiopia Educational Statistics Annual Abstracts 1988 E. C. 1995/96.* Addis Ababa: Educational Management Information Systems EMIS, Ministry of Education.

Mislevy, R. J. (1987). Recent developments in IRT. In E Z Rothkoph (ed.), *Review of Research in Education, 14,* 239-275.

Mitchell, F. W. (1938). *The Nature of Mathematical Thinking*. Melbourne: Melbourne University Press.

Miura, I. T. & Okamoto, Y. (1989). Comparisons of U.S. and Japanese first grader's cognitive representation of number and understanding of place value. *Journal of Educational Psychology, 81 (1)*, 109-113.

Mohandas, R. (1996). Test equating, problems and solutions : Equating English test forms for the Indonesian junior secondary school final examination administered in 1994. Unpublished Masters thesis, The Flinders University of South Australia.

Molenaar, I. W. (1995). Some background for item response theory and the Rasch Model. In G H Fischer & I W Molenaar (eds), *Rasch Models Foundations, Recent Developments, & Applications*. New York: Springer-Verlag, pp. 3-14.

Morgan, D. (1986). *Girls Education and Career Choice What the Research Says.* Sydney: NSW Joint Non-Government Schools Participation and Equity Program Committee.

Morgenstern, C. & Keeves, J. P. (1994). Attitudes toward schooling, comparative studies in. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2n ed.), Oxford: Pergamon Press, pp. 390-396.

Morris, R. W. (1978). The role of language in mathematics. *Prospects: Quarterly Review of Education, 8* (1), 73-81.

Morris, S. B., McDaniel, M. A., Worst, G. J.& Timm, H. (1995). Vanity-motivated overspending: personnel screening for positions of trust. *Educational and Psychological Measurement, 55(1)*, 95-104.

Morrison, C. A. & Fitzpatrick, S. J. (1992). *Direct and Indirect Equating: A Comparison of Four Methods Using the Rasch Model.* Austin, Texas: Measurement and Evaluation Center, the University of Texas.

Moss, J. D. (1982). *Towards Equality: Progress by Girls in Mathematics in Australian Secondary Schools.* Hawthorn, Victoria: ACER.

Mueser, K. T., Curran, P. J. & McHugo, G. J. (1997). Factor structure of the brief psychiatric rating scale in schizophrenia. *Psychological Assessment, 9 (3),* 196-204.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S. & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105 (3)*, 430-445.

Mullan, E., Markland, D. & Ingledew, D. K. (1997). A graded conceptualisation of selfdetermination in the regulation of exercise behaviour: Development of a measure using confirmatory factor analytic procedures. *Person. individ. Diff. 23 (5)*, 745-752.

Nandakumar, R. (1995). Application of SIBTEST in dealing with issues of DIF in the context of multidimensional data. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA, April 19-21.

National Council of Teachers of Mathematics. (1980). *An Agenda for Action: Recommendation for School Mathematics of the 1980s*. Virginia: National Council of Teachers of Mathematics.

Nisbet, S. (1978). New mathematics in Australia: What happened, and why. The *Australian Mathematics Teacher, 34* (4), 117-128.

Nisbet, S. (1978). New mathematics in Australia: What happened, and why. *The Australian Mathematics Teacher, 34* (5), 117-128. Continued from 34, 4.

Noonan, R. D. & Wold H. (1988). Partial least squares. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 710-716.

Norusis, M. J. (1993). *SPSS for Windows: Base System User's Guide: Release 6.0*. Chicago: SPSS Inc.

Osterlind, S. J. (1983). *Test Item bias.* Sage University Paper Series on Quantitative Application in Social Sciences, 07-001, Beverly Hills: Sage Publications.

Owen, J. M., Dowsey, J. & Hurworth, R. (1983). *A Review of Recent Research in Science and Mathematics Education with Particular Reference to Australia.* Melbourne College of Advanced Education: Centre for Program Evaluation.

Pajares, F. & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology, 86 (2)*, 195-203.

Parker, L. & Offer, J. (1987). 'Girls, boys and lower secondary school achievement: The shifting scene 1972-1986'. *Unicorn 13* (3), 148-154.

Pattison, P. (1992). Gender and mathematics. In P Grimshaw, R Fincher, & M Campbell, (eds), *Studies in Gender, Essays in Honour of Norma Grieve*, Melbourne, Parkville, Victoria, Australia: The University of Melbourne, pp. 105-116.

Peaker, G. F. (1969). How should national part scores be weighted? *International Review of Education, 15*, 229-237.

Peaker, G. F. (1975). *An Empirical Study of Education in Twenty-One Countries: A Technical Report.* Stockholm, Sweden: Almqvist & Wiksell International.

Pedhazur, E. J. (1982). *Multiple Regression in Behavioural Research: Explanation and Prediction.* New York: Holt, Rinehart, and Winston.

Petersen, N. S., Cook, L. L. & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics, 8* (2), 137-156.

Petersen, N. S., Kolen, M. J. & Hoover, H. D. (1989). Scaling, norming, and equating. In R L Linn (ed.), *Educational Measurement,(*3rd ed.), New York: American Council on Education.

Pidgeon, D. A. (1967). *Achievement in Mathematics a National Study in Secondary Schools.* The Mere: London: National Foundation for Educational Research in England and Wales.

Pizzini, E. L. & Shepardson, D. P. (1992). A Comparison of the classroom dynamics of a problem-solving and traditional laboratory model of instruction using path analysis. *Journal of Research in Science Teaching, 29 (3)* 243-258.

Plucker, J. A., Taylor v, J. W., Callahan, C. M & Tomchin, E. M. (1997). Mirror, mirror, on the wall: Reliability and validity evidence for self-description questionnaire-ii with gifted students. *Educational and Psychological Measurement, 57 (4)*, 704-713.

Postlethwaite, T. N. (1967). *School Organization and Student Achievement a Study Based on Achievement in Mathematics in Twelve Countries.* Stockholm: Almqvist & Wiksell.

Postlethwaite, T. N. (1971). International association for the evaluation of educational achievement (IEA)-the mathematics study: *Journal for Research in Mathematics Education*, *2*, 69-103.

Postlethwaite, T. N. (1992). Data processing and data analysis. In J P Keeves (ed.), *The IEA Technical Handbook*, The Hague : IEA, pp. 91-103.

Postlethwaite, T. N.& Wiley, D. E. (1991). *The IEA Study of Science II: Science Achievement in Twenty-Three Countries.* Oxford: Pergamon.

Potenza, M. T. & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. *Applied Psychological Measurement, 19 (1)*, 23-37.

Raju, N. S., Drasgow, F. & Slinde, J. A. (1993). An empirical comparison of the AREA methods, Lord's chi-square test, and the Mantel-Haenzel technique for assessing differential item functioning. *Educational and Psychological Measurement, 53(2)*, 301-314.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research.

Rasch, G. (1961). On the general laws and the meaning of measurement in psychology. In J Neyman (ed.), *Proceedings of the Fourth Symposium on Mathematical Statistics and Probability, 4,* 321-333.

Rasch, G. (1966). An individualistic approach to item analysis. In P F Lasarsfeld & N W Henry (eds), *Reading in Mathematical Social Science,* Chicago: Science Research Associates.

Rasch, G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests,* Chicago: University of Chicago Press.

Reeves, H. (1989). Action on all fronts in Tasmania. *The Australian Mathematics Teacher, 45* (2), 9-10.

Rentz, R. R. & Bashaw, W. L. (1975). *Equating Reading Tests with the Rasch Model, Vol. I Final Report*. Athens, Georgia: University of Georgia: Educational Research Laboratory, College of Education.

Reyonlds, A. J. & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology, 83 (1),* 97-107.

Reyonlds, A. J. & Walberg, H. J. (1992). A structural model of high school mathematics outcomes. *Journal of Educational Research, 85 (3),* 150-158.

Robitaille, D. F. & Donn, J. S. (1992). The third international mathematics and science study (TIMSS): A brief introduction. *Educational Studies, 23* (2), 203-210.

Robitaille, D. F. & Travers, K. J. (1992.) International studies of achievement in mathematics. In D A Grouws, (ed.), *Hand book of Research on Mathematics Teaching and Learning,* New York: Macmillan, pp. 687-709.

Robitaille, D. F. (1990). Achievement comparisons between the first and second IEA studies of mathematics. *Educational Studies in Mathematics, 21* (5), 395-414.

Robitaille, D. F. (1994). *Curriculum Frameworks for Mathematics and Science.* Vancouver, Canada: Pacific Educational Press.

Rogers, A. L. (1918). *Experimental Tests of Mathematical Ability and Their Prognostic Value.* New York.

Rogers, H. J. & Swaminathan, H. (1993). A comparison of the logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement, 17 (2)*, 105-16.

Rogosa, D. (1979). Causal models in longitudinal research: Rationale, formulations and interpretation. In P B Baltes & J R Nesselroade (eds), *Longitudinal Research in the Study of Behavior and Development*, New York: Academic Press, pp. 263-302.

Rosier, M. J. & Ross, K. N. (1992). Sampling and administration. In Keeves J P (ed.) *The IEA Technical Handbook,* The Hague, The Netherlands: IEA, pp. 51-90.

Rosier, M. J. & Ross, K. N. (1997). Sampling in survey research. In Keeves J P (ed.) *Educational Research, Methodology, and Measurement: An International Handbook,* (2nd ed.), Oxford: Pergamon, pp. 427-438.

Rosier, M. J. (1980). *Changes in Secondary School Mathematics in Australia.* Hawthorn: The Australian Council for Research.

Sampson, S. (1982). Trends in research on the education of girls. In *Educational Research in the 1980s Collected Papers, vol 2*, Brisbane: Australian Association for Research in Education 1982 Annual Conference Brisbane 10-14 November, AARE.

Scheuneman, J. & Bleistein (1994). Item bias. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2n ed.), Oxford: Pergamon Press, pp. 3043-3051.

Scheuneman, J. & Bleistein (1997). Item bias. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (2n ed.), Cambridge, UK: Pergamon Press, pp. 742-749.

Scheuneman, J. (1979). A method of assessing bias in test items. *Journal of Educational Measurement, 16 (3)*, 143-152.

Schmidt, W. H. & Cogan, L. S. (1996). Development of the TIMSS context questionnaires. In M O Martin & D L Kelly (eds), *Third International Mathematics and Science Study: Technical Report Vol 1: Design and Development,* Boston : IEA, pp. 5-1 to 5-22.

Schofield, H. L. (1981). Sex, grade level and the relationship between mathematics attitude and achievement in children. In M J Lawson, & R Linke (eds), *Inquiry and Action in Education: Papers Presented at the 1981 Annual Conference, vol.* 1., Adelaide: Australian Association for Research in Education, pp. 174-181.

Seleshi Zeleke (1995). Gender differences in mathematics achievement as a function of attitudes in grades 8 through 11 (in Northern Shewa Region). Unpublished Masters thesis, School of Graduate Studies Addis Ababa University.

Sellin, N. & Keeves, J. P. (1994). Path analysis with latent variables. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2nd ed.), Pergamon, pp. 4339-4352.

Sellin, N. & Keeves, J. P. (1997). Path analysis with latent variables. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.), Cambridge, UK: Pergamon Press, pp. 633-640.

Sellin, N. (1989). *PLSPATH Version 3.01 Application Manual.* Hamburg, West Germany.

Sellin, N. (1990). *PLSPATH Version 3.01 Program Manual.* Hamburg, West Germany.

Sellin, N. (1992). Partial least squares path analysis. In J P Keeves (ed.), *The IEA Technical Handbook,* The Hague, The Netherlands: IEA, pp. 397-412.

Sewnet Mamo (1995). Some factors affecting scholastic achievement of elementary school pupils. Unpublished MA thesis, School of Graduate Studies, Addis Ababa University.

Seyoum Teferra (1986). The education of women in Ethiopia: A missing piece in the development puzzle. *The Ethiopian Journal of Education*, *10 (1)*, 5-19.

Shaffer, B., Percy, P. M. & Tepper, B. J. (1997). Further assessment of the structure of Hinkin and Schriesheim's measures of interpersonal power. *Educational and Psychological Measurement, 57 (3)*, 505-514.

Shealy, R. & Stout, W. (1993). A model-based standardization approach that separates true bias/DIF from group ability differences and detects test bias/DTF as well as item bias/DIF. *Psychometrika, 58 (2)*, 159-94.

SIDA (1993). *Country Gender Analysis for Ethiopia.* Addis Ababa: SIDA.

Skaggs, G. & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research, 56* (4), 495-529.

Slinde, J. A. & Linn, R. L. (1978). An exploration of the adequacy of the Rasch model for the problem of vertical equating. *Journal of Educational Measurement, 15* (1), 23-35.

Smith, R. (1994). Detecting item bias in the Rasch rating scale model. *Educational and Psychological Measurement, 54 (4)*, 886-896.

Smith, R. M. & Kramer, G. A. (1992). A comparison of two methods of test equating in the Rasch model. *Educational and Psychological Measurement, 52 (4)*, 835-846.

Sobolewski, S. J & Doran R. J. (1996). Replication of a path analysis model of secondary physics enrolments: 20 years later. *Journal of Research in Science Teaching, 33 (5),* 501-512.

Song, M., & Ginsburg, H. P. (1987). The development of informal and formal mathematics thinking in Korea and U.S. children. *Child Development, 58*, 1286-1296.

Sontag, L. M. (1984). Vertical equating methods: A comparative study of their efficacy. *DAI, 45-03B*, p. 1000.

South Australian Education Department. (1975). *Primary School Mathematics Interim Revision, 1975*. Adelaide: South Australian Education Department.

South Australian Education Department. (1986). *Mathematics Curriculum Materials Evaluation.* Adelaide: Directorate of Studies, South Australian Education Department.

Stephens, M. (1989a). How well does mathematics count? *Prime Number, 4* (3), 3-7.

Stephens, M. (1989b). Where is the primary mathematics curriculum heading? *Prime Number, 4* (4), 4-11.

Stevenson, H. W., Lee, S. Y. & Stigler, W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 231*, 693-699.

Stigler, J. W. & Baranes, R. (1988). Culture and mathematics learning. In E Z Rothkopf (ed.), *Review of Research in Education,* vol. 15, Washington , DC: American Educational Research Association, pp. 253-306.

Stigler, J. W. & Perry, M. (1988). Cross-cultural studies of mathematics teaching and learning : Recent findings and new directions. In D A Grouws, T J Cooney, & D Jones (eds), *Effective Mathematics Teaching,* Reston, VA: National Council of Teachers of Mathematics, pp. 194-223.

Stigler, J. W., Lee, S., Licker, G. W. and Stevenson, H. W. (1982). Curriculum and achievement in mathematics: A study of elementary school children in Japan, Taiwan, and the United States. *Journal of Educational Psychology, 74 (3)*, 315-322.

Stocking, M. L. (1997). Item response theory. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* (2nd ed.), Cambridge, UK: Pergamon Press, pp. 836-840.

Tadesse Getaneh (1993). In-school factors affecting the attitudes of students towards learning in selected senior secondary schools in eastern Shoa. Unpublished BA thesis, Faculty of Education, Addis Ababa University.

Teese, R. (1988). Australian education in cross-national perspective: A comparative analysis with France. *Comparative Education, 24* (3), 305-316.

Tesfaye Fekadu (1987). Sex differences in language and mathematics ability. Unpublished BA thesis, Faculty of Education, Addis Ababa University.

Teshome Abebe (1993). A comparative study of academic achievement of students brought up under SOS children's village and under parental care at Higher 23 Secondary School. Unpublished BA thesis. Faculty of Education, Addis Ababa University.

Theisen, G. L., Achola, P. W. & Boakari, F. M. (1983). The underachievement of cross-national studies of achievement. *Comparative Education Review, 27* (1), 46-68.

Thissen, D., Steinberg, L. & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H Wainer & H I Braun (eds), *Test Validity,* Hillside, NJ: Lawrence Erlbaum, pp. 147-168.

Thorndike, R. L. (1967a). Mathematics test and attitude inventory scores. In T Husén (ed.), *International Study of Achievement in Mathematics: A Comparison of Twelve Countries,* vol 1, Stockholm: Almqvist & Wiksell, pp. 90-108.

Thorndike, R. L. (1967b). Mathematics test and attitude inventory scores. In T Husén (ed.), *International Study of Achievement in Mathematics: A Comparison of Twelve Countries,* vol 2, Stockholm: Almqvist & Wiksell, pp. 21-48.

Thorndike, R. L. (1973). *Reading Comprehension Education in Fifteen Countries.* Stockholm: Almqvist and Wiksell.

Thorndike, R. L. (1982). *Applied Psychometrics.* Boston, MA.: Houghton-Mifflin.

Tilahun Mengesha Afrassa. (1994*). National examinations in Ethiopia and South Australia: A Comparative Study*. Unpublished Master 's dissertation, The Flinders University of South Australia, Adelaide, Australia.

Tilahun Mengesha Afrassa (1995). National examinations in Ethiopia and South Australia: A Comparative Study. *Studies in Educational Evaluation, 21 (3),* 281-299.

Tilahun Mengesha Afrassa (1996). Students attitudes towards mathematics and school over time: A Rasch analysis. Paper Presented at the Joint Conference of Educational Research Association, Singapore and Australian Association for Research in Education, 26-30 November, Singapore, Polytechnic, Singapore.

Tilahun Mengesha Afrassa & Keeves, J. P. (1997a). Student level factors that influence the mathematics achievement of Ethiopian students: A path analysis. In K Fukui, E Kurimoto, & M Shigeta (eds), *Ethiopia in Broader Perspective: Papers of the 13th International Conference of Ethiopian Studies, Vol I-III,*, Kyoto , Japan: Shokado Book Sellers.

Tilahun Mengesha Afrassa & Keeves, J.P. (1997b). Changes in students mathematics achievement in Australian lower secondary schools over time: A Rasch analysis. Paper Presented at the Australian Association for Research in Education annual conference, Brisbane, 30 November - 4 December.

Tilahun Mengesha Afrassa & Keeves, J.P. (2001). Change in differences between the sexes in mathematics achievement at the lower secondary school level in Australia: Over time. *International Education Journal*, 2(2), 96-108.

Tsigie Haile (1991) *An Assessment of the Academic Performance of Female Students in Higher Education Institutions in Ethiopia.* Addis Ababa, Ethiopia.

Tsion Dessie (1990). *A Critical Look into Women's Education*. Addis Ababa.

Tuijnman, A. C. & Keeves, J. P. (1994). Path analysis and linear structural relations analysis. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2nd ed.), Pergamon, pp. 4339-4352.

Tuijnman, A. C. & Keeves, J. P. (1997). Path analysis and linear structural relations analysis. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.), Cambridge, UK: Pergamon Press, pp. 621-633.

Tupa, D. J., Wright, M.O'D. & Fristad, M. a. (1997). Confirmatory factor analysis of the WISC-III with child psychiatric inpatients. *Psychological Assessment, 9 (3)*, 302-306.

Turner, R. L. (1980). Language factors in mathematics of students with English speaking backgrounds compared with those with a non-English speaking background. Paper presented to the fourth annual conference of the Mathematics Education Research Group of Australia. Hobart,: MERGA.

Vale, C. (1993). Sex difference in achievement in VCE mathematics. In B Atweh, C Kanes, M Carss, & G Booker (eds), *Contexts in Mathematics Education,* Proceedings of the sixteenth annual conference of the Mathematics Education Research Group of Australasia (MERGA), Brisbane: MERGA.

Vijver, F. R. & Poortinga, Y. H. (1991). Testing across cultures. In R K Hambleton & J N Zaal (eds), *Advances in Educational and Psychological Testing,* Boston, MA: Kluwer Academic Publishers, pp. 277-308.

Voelkl, K. E. (1996). Measuring students' identification with school. *Educational and Psychological Measurement, 56 (5)*, 760-770.

Wainer, H. (1995). Precision and differential item functioning on a testlet- based test: The 1991 law school admissions test as an example. *Applied Measurement in Education*; *8 (2)*,157-186.

Walberg, H. J. (1981). A Psychological theory of educational productivity. In F H Farley & N Gordon (eds), *Psychology and Education.* Chicago: National Society for the Study of Education.

Walker, D. A. (1976). *The IEA Six Subject Survey: An Empirical Study of Education in Twenty-One Countries*. Stockholm, Sweden: Almqvist & Wiksell, IEA.

Weiss, D. J. & Yoes, M. E. (1991). Item response theory. In R K Hambleton & J N Zaal (eds), *Advances in Educational and Psychological Testing*, Boston: Kluwer Academic Publishers, pp. 69-95.

Westers, P. & Kelderman, H. (1991). Examining differential item functioning due to item difficulty and alternative attractiveness. *Psychometrika, 57 (1)*, 107-118.

Widdup, D. (1980). Review of research on sex differences in mathematics. In Adelaide College of the Arts and Education (ed.), *Improving Maths for Girls: Report of a Conference held at Raywood Inservice Centre, 30 May-1 June.*

Willett, J. B. (1997). Change, measurement of. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.), Oxford, UK: Pergamon Press, pp. 327-334.

Williams, D. (1991). A reaction to the national statement on mathematics for Australian schools. *Prime Number, 6* (3), 15-16.

Willis, S. (1989). *'Real Girls don't do Mathematics', Gender and the Construction of Privilege*. Geelong, Victoria: Deakin University.

Wingersky, M. S., Barton, M. A. & Lord, F. M. (1982). *LOGIST User's Guide*. Princeton, NJ: Educational Testing Service.

Wolf, R. M. (1967). Construction of descriptive and attitude scales. In T Husén (ed.), *International Study of Achievement in Mathematics: A Comparison of Twelve Countries,* vol 1, Almqvist & Wiksell, Stockholm, pp. 109-122.

Wolf, R. M. (1994). Rating scales. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* (2n ed.), Oxford: Pergamon, Press, pp. 4923-4930.

Wolf, R. M. (1997). Rating scales. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook* (2nd ed.),Cambridge, UK: Pergamon Press, pp. 958-965.

Wood, R. (1985). Item analysis. In T Husén & T N Postlethwaite (eds), *The International Encyclopedia of Education,* Oxford: Pergamon Press, pp. 376-384.

Wood, R. (1988). Item analysis. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 376-384.

Wright, B. D. (1968) Sample free test calibration and person measurement. *Proceedings of the 1967 Invitational Conference on Testing Problems*, Princeton, New Jersey: Educational Testing Service, pp. 85-101.

Wright, B. D. (1977). The Rasch model. *Journal of Educational Measurement 14,* 97-116.

Wright, B. D. (1988). Rasch measurement models. In J P Keeves (ed.), *Educational Research, Methodology, and Measurement: An International Handbook,* Oxford: Pergamon Press, pp. 286-292.

Wright, B. D. (1995). 3PL or Rasch? *Rasch Measurement Transactions, 9* (1), 408-409.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement.* Chicago: Mesa Press.

Wright, B. D., Mead, R. J. & Drada, R. E. (1976). Detecting and correcting test item bias with a logistic response model. *Research Memorandum No. 22*. Statistical laboratory, department of Education, University of Chicago.

Xin Ma. (1995). Gender differences in mathematics achievement between Canadian and Asian education system. *The Journal of Educational Research, 89 (2)*, 118-127.

Yelfign Worku, Zewdu Desta, Alemnesh H/Mariam & Anbessu Biazen (1995). Study on primary school female participation and performance in Cheha district. Unpublished manuscript, Addis Ababa: Ministry of Education.

Young, D. J. & Fraser, B. J. (1994). Gender differences in science achievement: Do school effects make a difference? *Journal of Research in Science Teaching, 31 (8),* 857-871.

Young, F. (1981). Quantitative analysis of qualitative data. *Psychometrika, 46,* 357-388.

Zwick, R. & Others. (1994). A simulation study of methods for assessing differential item functioning in computerized adaptive tests. *Applied Psychological Measurement, 18 (2),* 121-40.

Appendix A: Sampling Procedure Employed in Ethiopia

The sample required for data collection for the Ethiopian Mathematics Study was 1200 Grade 8 students from 40 government and nongovernment schools in the Addis Ababa Region. In addition, it was planned that 30 students should be selected from each school. The schools and the students were selected randomly using a random sampling procedure with probability proportional to the size of the school (pps sampling). The sampling details involved in selecting the schools and the students are presented in Table A1. For administrative purposes the Addis Ababa Region is divided into six zones. The number of schools and students in each zone and the number of selected schools and students are listed in Table A1.

Table A1 Number of schools, Grade 8 students, selected number of schools and students in the Addis Ababa Region

	Number of Schools and students in each Zone		**Selected Schools and students in each Zone**		**13-year-olds in Gov Schools**		**13-year-olds in Non-Gov Schools**		**13-year-olds in Total**		**Students Not 13-year-olds in selected Schools**		
Zone	Sch	Stu	Sch	Stu	M	F	M	F	M	F	M	F	Total
1	18	8705	6	180	6	11	-	3	6	14	70	88	178
2	36	11898	8	240	4	10	13	17	17	27	78	117	239
3	26	8823	6	180	8	5	7	9	15	14	82	67	178
4	52	16366	12	360	14	13	11	11	25	24	148	159	356
5	30	9244	6	180	3	9	4	21	7	30	60	80	177
6	7	2594	2	60	3	4	-	-	3	4	23	30	60
Sub					38	52	35	61	73	113	461	541	
Total	169	57630	40	1200	90		96		186		1014		1188[a]

Gov= Government; Non-Gov= Non-government; Sch = schools; Stu = students; Sub = sub-total; M= Male; F= Female; a = 12 students did not write the age, therefore are not included in the table

In the 1995/96 academic year there were 169 schools in the region that had Grade 8 level classes. The number of schools varied from one zone to another, for example, Zone 4 had the highest number of schools, namely 52, while Zone 6 had only seven schools (see Table A1). The total number of students at Grade 8 level in the region was 57,630. The number of students also varied from one zone to another, for example, Zone 6 had 2594 students, while, Zone 4 had 16,366 students. Therefore, it was necessary to select 40 schools out of 169 and 1200 students from 57,630 students.

In order to obtain a representative sample of 57,630 students in 169 schools in the region, random sampling procedures with probability proportional to size (pps) were employed. This sampling procedure employed in the Ethiopian Mathematics Study is presented here.

A list of all schools with the number of their Grade 8 students was collected from the Addis Ababa Regional Education Office. The information collected from the Regional Education Office was used to list the schools from 1 to 169. The schools were stratified by zone and the first school listed was from Zone 1 and the 169th school was from Zone 6 (see Table 5.4) within each zone the schools were listed alphabetically. The next step was to divide the total number of Grade 8 students in each school by 30 and to write the result in the next column giving this variable the name TICKETS (see Table 5.4). The total number of Grade 8 students in each school was divided by 30, since a sample of 30 students was to be drawn from each school.

The third step was to calculate the cumulative frequency of the number of tickets for the 169 schools in ascending order and to write the cumulative frequency in the third column of the table as shown in Table A2.

Table A2 Procedures and results of sample selection process in Ethiopian Mathematics Study

Code of Schools	No of tickets	Cumulative frequencies of tickets	Selected random number	Code of the randomly selected school
1	53	53	30	1
2	45	98		
3	22	120	128	3
.				
.				
.				
28	2	416	422	28
.				
.				
99	23	1304	1304	99
.				
.				
167	20	1942	1941	167
168	3	1945		
169	2	1947		

The fourth step was to select the schools. The following formula was applied in selecting the schools.

$$\frac{\text{Total number of tickets}}{\text{Total sample schools}} = \text{Interval(I)}$$

This gives $\frac{1955}{40} = 48.9 = 49$

The interval between the first selected ticket and the next was 49. In order complete the fourth column of Table A2, that is to start selecting the schools, the formula I - R = X was used, where I is the interval and R is the first random number selected by the investigator. R the first random number, which was randomly selected by the investigator was 19 and the interval I was 49, so, 49 - 19= 30. Therefore, the first school was the one whose ticket cumulative frequency was 30, which was the school with the code of 1 (see Table A2). Hence, school code 1 was chosen as the first sampled school. The remaining schools were selected using the X + I procedure where X is the selected ticket and I is the interval between two selected schools. Hence the first selected school was 30 and the second school was 30 + 49 = 79 which was the second school. The last school was 1892 +49 = 1941 (see Table A2).

The next step after the selection of the 40 schools from 169 schools in the Addis Ababa Region was to select 30 students from each of the chosen schools. The investigator visited the 40 schools to obtain the list of Grade 8 students in each school and then selected randomly the 30 students from each school. The method which was used for selecting students was:

$$\frac{\text{Total number of Grade 8 students in the selected school}}{\text{Total number of students to be selected in the school}} = \text{Interval}$$

The first student was selected using the formula: I - R = X, where I is the interval and R is the first random number selected by the investigator.

The remaining 29 students were selected using the X + I procedure where X is the selected student and I is the interval between two selected students. This method was used to select the 1200 students from the 40 schools.

During the sampling process the differences between the government and nongovernment schools were not taken into consideration, because the numbers of government and nongovernment schools in the Addis Ababa region were approximately equal. From the whole sample, 50 per cent of the schools were government schools and the other 50 per cent were nongovernement schools (see Table A3).

Table A3 School Visits in Addis Ababa Region

Number of visits in each school	Number of schools	Total visits
2	19	38
3	13	39
4[a]	7 [a]	28 [a]
5	1	5
7	1	7
Total	41	117

a = a school was replaced by an other school after four visits

After selecting the schools and the students using the above procedures, the investigator visited the 40 schools and administered the tests and questionnaires to the selected 30 students in each school. Because of lack of cooperation from the principal and staff at one school, one of the selected schools was replaced after four visits. The replacement school was selected by applying the formula, X $\pm$ 1 = N, where X is the replaced school, and N is the replacing school.

Since the schools were listed in the sampling frame stratified by zones and in a random order within zone, the replacement schools would be from the same zone. This formula gives a choice to use, either X+1 or X-1. Therefore, a decision had to be made in advance, before starting the replacement work. Hence, the investigator, after the selection of the schools, and before a first visit to the schools chosen randomly to use X+1, if there was a need to replace schools. The decision was made before knowing the withdrawal of this particular school. Table 5.5 shows the number of visits to the schools.

After selecting the schools and the students and arranging the data collection period in each school, the investigator started the testing program by mid-February, and the program was completed by the end of April.

Shannon Research Press
Adelaide, South Australia
ISBN: 0-9580704-4-X