DOCUMENT RESUME

ED 286 70C SE 047 894

AUTHOR	Garden, Robert A.
TITLE	Second IEA Mathematics Study. Sampling Report.
INSTITUTION	International Association for the Evaluation of Educational Achievement, Wellington (New Zealand).; New Zealand Dept. of Education, Wellington.
SPONS AGENCY	Center for Education Statistics (OERI/ED), Washington, DC.
REPORT NO	CS-87-333
PUB DATE	Mar 87
CONTRACT	300-83-0212
NOTE	229p.
PUB TYPE	Reports - Research/Technical (143) -- Statistical Data (110)
EDRS PRICE	MF01/PC10 Plus Postage.
DESCRIPTORS	Data Analysis; Educational Assessment; *Guidelines; *International Studies; *Mathematics Achievement; Mathematics Curriculum; Mathematics Instruction; *Research Methodology; *Sampling; Secondary Education; *Secondary School Mathematics; Surveys; Testing
IDENTIFIERS	*Second International Mathematics Study

ABSTRACT

This report provides users of data from the Second International Mathematics Study with a summary of the survey procedures used by participating countries. The introductory chapter includes definitions, cross-sectional and longitudinal components, and recommended sampling procedures. Chapter 2 presents national population definitions and sampling procedures for the 20 countries surveying Population A (grade 8), while chapter 3 presents corresponding information for the 15 countries surveying Population B (grade 12). Response rates for each population are considered in chapters 4 and 5 and representativeness of the sample is the focus in chapters 6 and 7. The distribution of rotated forms is described in chapter 8, weighting in chapter 9, sampling errors in chapter 10, and non-sampling errors in chapter 11. A brief conclusion is included as chapter 12. Appendices present achieved sampling fractions and the text of the sampling manual. (MNS)

Second IEA Mathematics Study

Sampling Report

Robert A. Garden
New Zealand Department of Education

International Association for the Evaluation of
Educational Achievement (IEA)

Larry E. Suter, Project Officer
Center for Education Statistics

Prepared in part for the Center for Education Statistics under contract OE 300-83-0212. Opinions, conclusions or recommendations contained herein are those of the author, and not necessarily those of the U.S. Department of Education.

March 1987

CS 87-333

Foreword

The purpose of this report is to provide users of the data derived from this study with a summary of the survey procedures used by the countries participating in this study. The information about sampling procedures, population definitions, and response rates were prepared by each of the national centers which participated in the Second IEA International Mathematics Study. Each of the research centers submitted statements of the sampling procedures to the International Coordinator, Mr. Robert Garden at the New Zealand Department of Education, who prepared this report at the request of the U.S. Department of Education's Center for Education Statistics. The research center in each country was responsible for the proper implementation of the sampling procedures described in the report attached as Appendix II.

The U.S. sample was designed and implemented by a designated U.S. national center located at the University of Illinois. Participation of school districts and schools in this study was strongly affected by the length of the survey instrument which demanded several hours of student and teacher participation. The Center for Education Statistics wishes to thank each national center for its cooperation and participation in the study.

CONTENTS

TABLES

SECOND IEA MATHEMATICS STUDY

SAMPLING REPORT

1 INTRODUCTION

1.1 Purpose of the Report

In this comparative study of secondary school mathematics education, data was collected for variables at system, school, teacher, classroom and student levels. It is essential that the statistics obtained from measures used to quantify these variables be able to be evaluated for the degree of accuracy with which they estimate within country parameters and for the extent to which they are comparable between countries. This report summarizes the known characteristics of the samples in participating countries and is thus concerned with sample comparability. In making cross-national comparisons between statistics for some Study variables it should be remembered that structural features of education systems, curricular differences and cultural differences must also be considered.

1.2 International Population Definitions

Two populations were specified by the IEA International Mathematics Committee. These were selected because of intrinsic interest in mathematics education at these levels and also in order to allow comparisons to be made with results of the First IEA Mathematics Survey (Husen, 1967). Population A, the younger population, is at an age when all students are still in school in most of the participating countries and Population B is the group of students studying the highest level of mathematics taught in the school system of each country. The formal definitions are as follows:

Population A: All students in the grade (year level) where the majority has attained the age of 13.00 to 13.11 years by the middle of the school year.

Note: National Centers were advised that in the event of the 13-year old population being split equally over two grades in any country, then the grade for which the cognitive mathematics tests were most appropriate to the curriculum should be chosen

Population B: All students who are in the normally accepted terminal grade of the secondary education system and who are studying mathematics as a substantial part (approximately five hours per week) of their academic program.

Note: In the event students in the target population in most countries study mathematics for somewhat less than 5 hours per week.

Some National Centers found it necessary or desirable to depart from the intention of these definitions in defining the populations at national level. For Population A, Nigeria and Swaziland students studying at an appropriate curriculum level have a mean age considerably greater than 13.00 to 13.11 years. On the other hand, students in Hong Kong and Ontario are, on average, about one year younger.

At Population B level, Ontario and Scotland have two grade levels which can be regarded as "the normally accepted terminal grade." Ontario designated one of these (grade 13) as containing the target population but Scotland's Population B sample contains students from S5 and S6 (grades 11 and 12). The Hungary sample contains a substantial proportion of students who, although studying mathematics for "approximately 5 hours per week", are taking courses which are not pre-university type mathematics. These discrepancies will be noted under the separate country sections of the report.

1.3 Cross-sectional and Longitudinal Components of the Study

The full mathematics Study at Population A level was envisaged as a longitudinal study with pre-testing early in the school year and post-testing late in the same school year. The focus of interest was on the teaching and learning of mathematics at the classroom level.

The recommended sampling design was thus:

i) Stratification based, where possible, on groupings seen by each National Center as having some significance for education in their country.

ii) Random selection of schools with probability proportional to size of the target group within each school.

iii) Random selection of two classes within each school at the target grade level.

The alternative strategies used by various countries are described below under the separate country sections of the report.

Some National Centers judged that the full study would make more demands on teachers and resources than could be easily justified in their countries and others had as their main interest either a comparison with First IEA Mathematics Survey results or an assessment of the extent to which mathematics objectives were currently being met. These countries chose to administer a cross-sectional study based on the post-test and background instruments.

Countries/systems which took part in the two components of the study are:

Longitudinal Study	Cross-sectional Study
Belgium (Flemish)	Belgium (French)
British Columbia	England and Wales
France	Finland
Japan	Hungary
New Zealand	Hong Kong
Ontario	Israel
Thailand	Luxembourg
USA	The Netherlands
	Nigeria
	Scotland
	Swaziland
	Sweden

At Population B level a longitudinal study was not seen as feasible for most countries and was designated a national option. Countries participating at this level were:

Belgium (Flemish)
Belgium (French)
British Columbia
England and Wales
Finland
Hungary
Hong Kong
Israel
Japan
New Zealand
Ontario
Scotland
Sweden
Thailand
USA

In addition USA and Ontario undertook longitudinal studies.

Note: i) School questionnaires for both components were identical.

Teacher questionnaires for the cross-sectional component were a subset of those used for the longitudinal component.

Student questionnaires for both components were identical.

Student cognitive mathematics tests contained 157 items common to both components. Comparisons between countries are based on subtests drawn from these common items. Results for all 20 countries are thus included in the report of the cross-sectional study.

ii) In Swaziland a longitudinal study based on a reduced pre-test was carried out. Cross-sectional results only have been included in the international reports.

1.4 The International Sampling Committee

The Sampling Committee for the Second IEA Mathematics Study had the following members:

Dr Malcolm Rosier, Australian Council for Educational Research, (Chairman)

Dr John Keeves, Australian Council for Educational Research

Mr Ian Livingstone, New Zealand Council for Educational Research

Mr Ken Ross, Australian Council for Educational Research

Dr Rosier was appointed Sampling Referee for the Study.

The Sampling Committee met at the Australian Council for Educational Research in Melbourne in February 1979 and prepared a sampling manual (IEA (MATHS-NZ)/A/122) which was based on the authors' experience in previous IEA studies. In addition, considerable weight was given to the published reports of Gilbert Peaker, who was sampling consultant for earlier IEA studies (Husen, 1967, Volume 1: Chapter 9 and Peaker, 1975) and to a monograph by Ross (1979). The 68-page manual contained six sections:

A. an introduction in which populations were defined and the aims of the study related to sampling designs;

B. basic sampling theory with sampling decisions tables and examples in their use;

C. factors to be considered in preparing a sampling design for the cross-sectional study and detailed procedures for each of several possible designs;

D. additional considerations and procedures needed for the longitudinal study;

E. an action schedule related to sampling indicating steps which National Centers needed to take with an appropriate time scale; and

F. questionnaires to be completed at National Centers which sought details about their population definitions, sample designs, marker variables, estimated sampling errors and schedules.

1.5 Further Guidance for National Centers

National Centers forwarded details of their proposed sampling procedures to the Sampling Referee. Dr Rosier either approved the sampling plans or, in the case of many National Centers, sought further information or recommended modifications that were to be made before his approval could be given.

During the phase of the Study when sampling was a major concern for National Centers, or when issues relating to samples arose, Dr Rosier issued sampling memoranda to all National Centers.

These had as subjects:

October 1980	Surv/80.18	The necessity for full sampling information from countries with an explanation of the purposes for which each element of information is needed.
		General comments on sampling designs.
		Summary of the current status of national center sampling plans.
November 1980	Surv/80.35	Achieved samples and weighting procedures.
May 1981	Surv/81.23	Problems associated with sampling areas and intact classes.
February 1983	Surv/83.16	Comments on SIMS Sampling and Weighting.

National Research Coordinators were also able to discuss their sampling plans and any problems they were encountering in person with Dr Rosier at international meetings in Osnabruk and Bielefeld in January 1980 and with Mr G Pollock (Scottish Council for Research in Education) acting on behalf of the Sampling Committee at an international meeting held at Urbana in December 1980.

1.6 Recommended Sampling Procedures

The Sampling Manual (IEA (Maths-NZ)/A/122) detailed a variety of procedures which could be followed at each stage of sampling. The most common pattern followed by National Centers was:

i) Stratification by geographical region, school type or some other variable(s) of interest in a particular country.

ii) Systematic ordering of schools within strata followed by pseudo-random selection of schools by the random start—constant interval method.

iii) Random selection of one or two intact classes within selected schools.

iv) Replacement of refusing schools either from a parallel sample or by selecting the next on the list.

Intended sample size was determined by a priori calculation of the sample size required to meet specific confidence limits for statistics. The calculations were based on values of intraclass correlations from previous national studies, where these were known.

In general, sampling and data collection were well executed by participating countries. Deviations from the above procedures are outlined in the separate country sampling descriptions in sections 2 and 3 of this report and where samples are such that there is reason to be cautious in interpreting statistics derived from them this is indicated. A conservative approach has been taken and, even for those countries in which less than very good samples and response rates have been obtained, enough is known about the achieved samples for informed interpretations within country, and comparison between countries, to be made.

References

Husen, Torstin (ed) — International Study of Achievement in Mathematics; John Wiley and Sons; New York; 1967.

Peaker, Gilbert F. — An Empirical Study of Education in Twenty-One Countries : A Technical Report; John Wiley and Sons; New York; 1975.

Ross, Ken — Searching for Uncertainty, A.C.E.R., Melbourne, 1979.

2 NATIONAL POPULATION DEFINITIONS AND SAMPLING PROCEDURES - POPULATION A

2.1 Belgium (Flemish)

2.1.1 Population Definition

All students in the second year of the general secondary education, technical secondary education, and vocational secondary education programs in both Type I and Type II forms of school organization.

Note: Type I refers to schools in which a modernization of the organization and curriculum had occurred; Type II refers to schools still operating in a traditional mode.

2.1.2 Excluded Population

Students in special schools for the handicapped. Students in Provincial "General and Technical" and "General" schools (0.6% of the population).

2.1.3 Stratification

Stratification variables were initially:

Stratum Number	Description
1	Organizing authority: Catholic General and technical (comprehensive) school, Type I
2	Organizing authority: Catholic General school, Type II
3	Organizing authority: Catholic Technical school, Type II
4	Organizing authority: Catholic Vocational schools, Type I and II
5	Organizing authority: State General and Technical (comprehensive) school, Type I
6	Organizing authority: State General school, Type II No schools in this stratum

Stratum Number	Description
7	Organizing authority: State Technical school, Type II No schools in this stratum
8	Organizing authority: State Vocational schools, Type I
9	Organizing authority: Provincial General and technical, Type I No sample schools
10	Organizing authority: Provincial General, Type II No sample schools
11	Organizing authority: Provincial Technical, Type II
12	Organizing authority: Provincial Vocational schools, Types I and II
13	Organizing authority: Communal General and technical, Type I
14	Organizing authority: Communal General, Type II
15	Organizing authority: Communal Technical, Type II
16	Organizing authority: Communal Vocational, Type I and Type II

These sixteen strata were collapsed to six at the International Center for two reasons. First, the National Center advised that during the course of the study the process of "modernization" which was occurring within the school system meant that the balance between Type I and Type II schools changed rapidly and, second, some strata contained too few schools to allow reliable weighting.

The new strata formed were as follows:

Stratum 1 : 1 + 2 above

Stratum 2 : 3 + 4 above

Stratum 3 : 13 + 14 above

Stratum 4 : 11 + 12 + 15 + 16 above

Stratum 5 : 5 above

Stratum 6 : 8 above

Thus the strata for weighting consist of:

Stratum Number	Percent of Population	Description
1	36.4	Catholic "General and Technical" and "General" schools
2	34.5	Catholic "Technical" and "Vocational" schools
3	2.9	Communal "General and Technical" and "General" schools
4	5.2	Provincial and Communal "Technical" and "Vocational" schools
5	15.5	State "General and Technical" schools
6	5.6	State "Vocational" schools

2.1.4 Selection of Sample

Schools were ordered by (National Center) strata and by geographical criteria within strata.

The random start—constant interval method was used to select schools with probability proportional to size of target grade.

One class was then randomly selected within school.

2.2 Belgium (French)

2.2.1 Population Definition

All students in the second year of the "general, technical and vocational" program in both Type I and Type II forms of (school) organization.

Note: Type I and Type II as for Belgium (Flemish)

2.2.2 Excluded Population

Students in special schools for the handicapped.

2.2.3 Stratification

Stratification variables were initially:

Statum Number	Description
1	Organizing authority: Catholic Comprehensive academic school (general education) - non traditional
2	Organizing authority: Catholic Comprehensive technical and vocational school - non traditional
3	Organizing authority: Catholic Traditional academic school
4	Organizing authority: Catholic Traditional technical and vocational education
5	Organizing authority: Local authorities or boards Comprehensive academic school - non traditional
6	Organizing authority: Local boards Comprehensive technical and vocational education - non traditional
7	Organizing authority: Local boards Traditional academic school
8	Organizing authority: Local boards Traditional technical and vocational education
9	Organizing authority: State Comprehensive academic - non traditional
10	Organizing authority: State Comprehensive technical and vocational - non traditional

These ten strata were collapsed to six at the International Center on the advice of the National Center because of the rapid change in the distribution of students between Type I and Type II schools during the course of the study.

The new strata formed were as follows:

Stratum 1 : 1 + 3 above

Stratum 2 : 2 + 4 above

Stratum 3 : 5 + 7 above

Stratum 4 : 6 + 8 above

Stratum 5 : 9 above

Stratum 6 : 10 above

Thus the strata for weighting consist of

Stratum Number	Percent of Population	Description
1	40.0	Catholic general education (academic) schools
2	8.8	Catholic technical and vocational schools
3	13.0	Local board general academic schools
4	10.2	Local board technical and vocational schools
5	21.7	State general academic schools
6	6.4	State technical and vocational schools

2.2.4 Selection of Sample

Schools were ordered by (National Center) strata and by geographical criteria within strata. The random start—constant interval method was used to select schools with probability proportional to size of the target grade.

One class was then randomly selected within school.

2.3 British Columbia

2.3.1 Population Definition

All students enrolled in regular grade 8 classes in September, 1980 in the British Columbia public school system.

2.3.2 Excluded Population

i) Slower students requiring extensively modified programs to suit their needs (approximately 5% of age cohort).

ii) Students enrolled in private schools (approximately 5% of age cohort).

The total excluded population is thus of the order of 10% of the age cohort.

2.3.3 Stratification

Stratification by geographical zone.

Stratum Number	Percent of Population	Description
1	14.7	Zone 1
2	38.5	Zone 2
3	10.5	Zone 3
4	18.0	Zone 4
5	6.7	Zone 5
6	11.5	Zone 6

2.3.4 Selection of Sample

Samples were drawn independently from each stratum. For sample selection an additional stratification variable, school size, was used.

In effect schools and classes were simultaneously selected with probability proportional to number of grade 8 classes. In all but a few schools the procedure resulted in one class per school being selected.

Note: Schools agreeing to cooperate were informed that the desired procedure was to use the randomly selected classes but that if this was not feasible it would be left to the schools' judgment as to which classes were included. The number of schools that made their own selection of a class cannot be ascertained.

2.4 England and Wales

2.4.1 Population Definition

All pupils in the third year of normal secondary schools (or their equivalent where a middle school operated) who were born between 1 September 1966 and 31 August 1967.

2.4.2 Excluded Population

Pupils in special schools for the educationally subnormal or severely maladjusted, or in special units for similar pupils in normal schools.

2.4.3 Stratification

Four stratification variables were initially used:

School type	a)	Comprehensive to age 16
	b)	Comprehensive to age 18
	c)	Other maintained
	d)	Independent
Region	a)	North
	b)	Midlands
	c)	South
	d)	Wales
Location	a)	Metropolitan
	b)	Non-metropolitan
School size by size of target group	a)	up to 80 pupils
	b)	81 - 160 pupils
	c)	161 - 240 pupils
	d)	more than 240 pupils

This gave 128 possible strata. Many cells were found to be empty or to include very few schools and for this and other reasons the strata were collapsed to 16.

Stratum Number	Percent of Population	Description (Region x Size of Target group x School Type
1	3.1	North, 1-160, Comprehensive to 16
2	2.2	North, 1-160, Comprehensive to 18
3	6.4	North, 161+, Comprehensive to 16
4	16.4	North, 161+, Comprehensive to 18
5	2.3	North, all, Other maintained
6	3.1	Midlands, 1-160, Comprehensive to 16
7	1.6	Midlands, 1-160, Comprehensive to 18
8	15.3	Midlands, 161+, All comprehensive
9	1.8	Midlands, all, Other maintained
10	2.1	South, 1-160, Comprehensive to 16
11	4.6	South, 1-160, Comprehensive to 18
12	7.0	South, 161+, Comprehensive to 16
13	19.8	South, 161+, Comprehensive to 18
14	5.9	South, all, Other maintained
15	5.9	Wales, all, All maintained
16	2.3	All, all, Independent

2.4.4 Sampling Procedures

A random sample of schools was drawn for each stratum and then a random sample of students from the selected schools. The proportion of students sampled from each school was made inversely proportional to the size of the target population in the school by selecting only those students born during a particular range of days in each month.

Note: Classes were not the sampling unit in England and Wales.

2.5 Finland

2.5.1 Population Definition

Pupils receiving standard mathematics instruction in the normal comprehensive school or corresponding schools at a grade-level where the majority of pupils are 13 years old (in the mddle of the school year). In Finland this age cohort is concentrated in grade 7 of the comprehensive school.

2.5.2 Excluded Population

Schools in the province of Ahvinanmaa.

Schools for the aurally, visually or motor handicapped.

Schools in which the language of instruction is other than Swedish or Finnish. These schools represent approximately 1% of the population.

2.5.3 Stratification

The Finnish National Center stratified first by language of instuction (Finnish, Swedish). Finnish speaking schools were stratified by geographical region, 11 provinces, while Swedish speaking schools constituted one stratum. The third stratification variable was school location (urban, rural). Thus there were 24 (national) strata.

A complication due to the sampling procedure (q.v.) necessitated post hoc stratification by course type (long course, Short course and Heterogeneous course) at the International Center. This gave rise to a total of 53 strata.

Stratum (National) Center	Stratum International Center (Weighting)	Percent of Population	Description	
01	01	3.2	Uusimaa, Urban,	Short course
	25	11.0		Long course
	48	2.0		Heterogeneous course
02	02	0.7	Uusimaa, Rural,	Short course
	26	2.6		Long course
03	03	2.1	Turku & Pori, Urban,	Short course
	27	6.4		Long course
04	04	0.5	Turku & Pori, Rural,	Short course
	28	2.1		Long course
	49	2.5		Heterogeneous course
05	05	1.3	Häme, Urban,	Short course
	29	7.1		Long course
06	06	1.1	Häme, Rural,	Short course
	30	3.9		Long course

Stratum (National) Center	Stratum International Center (Weighting)	Percent of Population	Description	
07	07	1.3	Kymi, Urban,	Short course
	31	3.2		Long course
08	08	0.5	Kymi, Rural	Short course
	32	2.1		Long course
09	09	0.2	Mikkeli, Urban,	Short course
	33	0.5		Long course
	50	1.3		Heterogeneous course
10	10	0.6	Mikkeli, Rural,	Short course
	34	0.2		Long course
11	11	0.3	Vaasa, Urban,	Short course
	35	0.3		Long course
	51	1.7		Heterogeneous course
12	12	0.7	Vaasa, Rural,	Short course
	36	3.5		Long course
13	13	0.2	Keski-Suomi, Urban,	Short course
	37	1.7		Long course
14	14	0.6	Keski-Suomi, Rural,	Short course
	38	2.6		Long course
15	15	0.3	Kuopi, Urban,	Short course
	39	2.5		Long course
16	16	0.2	Kuopi, Rural,	Short course
	40	0.9		Long course
	52	1.7		Heterogeneous course
17	17	0.7	Pohjois-Karjala, Urban,	Short course
	41	1.0		Long course
18	18	0.3	Phjois-Karjala, Rural,	Short course
	42	1.6		Long course
19	19	0.9	Oulu, Urban,	Short course
	43	3.0		Long course
20	20	1.0	Oulu, Rural,	Short course
	44	5.1		Long course
21	21	2.1	Lappi, Urban,	Heterogeneous course
22	22	0.4	Lappi, Rural,	Short course
	45	2.2		Long course

Stratum (National) Center	Stratum International Center (Weighting)	Percent of Population	Description
23	23	0.4	Swedish Speaking, Urban, Short Course
	46	2.6	Long Course
	53	0.2	Heterogeneous Cours
24	24	0.5	Swedish Speaking, Rural, Short course
	47	1.6	Long course

2.5.4 Sampling Procedures

Schools were randomly selected with probability proportional to size of target grade using random start-constant interval.

Two classes per school were randomly selected, one from the Short Course and one from the Long Course. From schools where no sets existed two (or sometimes more) heterogeneous classes were randomly selected.

This procedure resulted in Short Course (low ability) classes being very much over-represented. The International Center introduced a further stratifying variable (Course Type) resulting in 53 strata.

2.6 France

2.6.1 Population Definition

All students in class de 4e (grade 8) of colleges, private and public education in metropolitan France.

2.6.2 Excluded Population

Students in eighth grade classes of public and private colleges in overseas territories and departments of France (4%). Students in Technical Education (1%).

2.6.3 Stratification

The stratification variables are State/Private education and school location.

Stratum Number	Percent of Population	Descripition
1	4.6	State education, rural outside industrial and urban regions.
2	3.3	State education, rural within industrial and urban regions
3	48.3	State education, urban
4	5.3	State education, Paris conurbation

Stratum Number	Percent of Population	Description
5	2.2	Private education, rural outside industrial and urban regions
6	0.9	Private education, rural within industrial and urban regions
7	17.3	Private education, urban
8	4.3	Private education, Paris conurbation

2.6.4 Selection of Sample

Systematic drawing of 6 acadamies (university regions) out of the 26 acadamies in metropolitan France. For this acadamies were arranged in decreasing order according to percent of private education students. Regions selected were: Levres, Dijon, Lyon, Toulouse, Versailles, Reims. Information supplied by National Center indicates SES distribution for the sample matches distribution for the population very closely.

Schools were selected with probability proportional to size of eighth grade.

Two classes were randomly selected within each school.

Note: Pseudoschools were created by combining two small schools where only one eighth grade class existed in a selected school.

2.7 Hong Kong

2.7.1 Population Definition

All students in Form 1/Middle 1 with mathematics offered as part of the school curriculum.

Note: This corresponds to the grade level in which the majority of students reach the age of 13 years by the middle of the school year.

Form 1 - schools with English as the medium of instruction.

Middle 1 - schools with Cantonese as the medium of instruction.

2.7.2 Excluded Population

None stated:

2.7.3 Stratification

Stratification variables were School Types (Public/Private), Language of Instruction (English/Cantonese) and Gender of School Population (male, female, coeducational).

Stratum Number	Percent of Population	Description
1	8.6	Public, Boys, English
2	1.0	Public, Boys, Cantonese
3	6.4	Public, Girls, English
4	2.0	Public, Girls, Cantonese
5	21.7	Public, Coeducational, English
6	5.5	Public, Coeducational, Cantonese
7	0.6	Private, Boys, English
*8	- -	Private, Boys, Cantonese
9	5.0	Private, Girls, English
*10	- -	Private, Girls, Cantonese
11	44.1	Private, Coeducational, English
12	5.2	Private, Coeducational, Cantonese

2.7.4 Selection of Sample

Class was used as the sampling unit. All classes were listed within each stratum and selected using random start and constant interval.

Classes were thus chosen with probability proportional to size.

2.8 Hungary

2.8.1 Population Definition

All pupils in the 8th grades of elementary schools where classes contain 8th grade pupils only. (This excludes a small number of ungraded village schools).

2.8.2 Excluded Population

Ungraded village schools. Schools for the handicapped. (Note: The excluded population is less than 5% of the total population.)

2.8.3 Stratification

Stratification was by a combination of community size and cultural/administrative weight categorization.

Stratum Number	Percent of Population	Description
1	14.5	Capital (Budapest)
2	7.8	Large towns
3	26.2	Smaller towns
4	7.4	More significant villages (better cultural facilities)
5	44.1	Less significant villages (poorer cultural facilties)

2.8.4 Selection of Sample

Classrooms were listed within stratum and then selected by random start—constant interval. They were selected with probability proportional to number of classes in a stratum.

2.9 Israel

2.9.1 Population Definition

All students in grade 8 classes of schools in which Hebrew is the language of instruction.

2.9.2 Excluded Population

Students in schools in which Arabic is the language of instruction.

2.9.3 Stratification

Stratification variables in the sampling plan approved by the sampling referee were:

1 Size of school (schools having one or two parallel grade 8 classes/schools having more than two parallel grade 8 classes).

2 Type of school (Old system (elementary) having grades 1-8/Reformed system (secondary) having grades 7-9).

3 Organizing authority (State/Religious)

4 Percentage of culturally disadvantaged learners in the school (0-20%/21-40%/41-60%/61-80%/81-100%).

The sampling plan was revised at the time of data collection to have only two stratification variables, Type of School and Percent of Culturally Disadvantaged Learners.

Stratum Number	Percent of Population	Description
1	18.5	Elementary school, 0 - 20% disadvantaged
2	16.9	Elementary school, 21-40% disadvantaged
3	10.4	Elementary school, 41-60% disadvantaged
4	6.8	Elementary school, 61-80% disadvantaged
5	4.7	Elementary school, 81-100% disadvantaged
6	3.1	Secondary school, 0-20% disadvantaged
7	7.0	Secondary school, 21-40% disadvantaged
8	5.1	Secondary school, 41-60% disadvantaged
9	3.2	Secondary school, 61-80% disadvantaged
10	5.4	Secondary school, 81-100% disadvantaged

Stratum Number	Percent of Population	Description
11	3.4	Elementary school, no information about disadvantaged
12	15.4	Secondary school, no information about disadvantaged

2.9.4 Selection of Sample

Schools were clustered in cells of the original sampling frame (four stratification variables) and listed by size of school within cells.

Schools were then selected by the random start, constant interval method. Different intervals were used in small schools than in large schools (more than 2 grade 8 classes) because in small schools all grade 8 students were tested while in large schools only 2 grade 8 classes were tested. Intervals were determined by average class size in school types so the procedure gives an approximate probability proportional to size method.

Classes within large schools were randomly selected.

2.10 Japan

2.10.1 Population Definition

Students in grade 1 Lower Secondary School (U.S. grade 7 equivalent).

2.10.2 Excluded Population

Students of private schools and schools for the handicapped.

Note: Statistics from "Educational Statistics Japan", 1976 edition, Ministry of Education, Science and Culture indicate that approximately 3% Lower Secondary students attend private schools and approximately 1% of students are in special classes.

2.10.3 Stratification

Stratification variables were Community Size and School Size.

Stratum Number	Percent of Population	Description
		Town/village, population <50,000
11	2.6	School size <150
12	14.4	School size 150-499
13	12.3	School size 500-999
14	2.5	School size 1000-1499
		Small city, population <200,000
21	0.4	School size <150
22	3.5	School size 150-499
23	12.9	School size 500-999
24	6.6	School size 1000-1499
25	0.7	School size >1500
		Large city, population <1,000,000
31	0.2	School size <150
32	2.3	School size 150-499
33	10.3	School size 500-999
34	10.5	School size 1000-1499
35	2.3	School size >1500
		Metropolis, population >1,000,000
42	1.3	School size 150-499
43	9.6	School size 500-999
44	5.8	School size 1000-1499
45	0.8	School size >1500
56	0.8	National Schools

Note: National schools select high ability students for enrollment.

2.10.4 Selection of Sample

Schools were ordered by stratum and selected with probability proportional to size.

One class per school was then randomly selected.

2.11 Luxembourg

2.11.1 Population Definition

Population A comprises all students in normal classes at year 8 level across all school types in the whole country.

2.11.2 Excluded Population

All students of "classes spéciales" and "classes de fin d'études". Students of the "European School" of Luxembourg. Excluded population estimated at 7%.

2.11.3 Stratification

Classes selected directly, one class in every two chosen. The sample is thus approximately half of the population and all school types are represented in this ratio.

Post hoc stratification was by two variables, School Type and Streaming/Non-streaming.

Stratum Number	Percent of Population	Description
10	21.0	Only classes of Lycée, no streaming
20	23.0	Only classes of Lycée secondaire technique, no streaming
21	11.8	Only classes of Lycée secondaire technique, streaming
30	10.4	Only "complementaire" classes, no streaming
40	10.6	Classes of Lycée and one other type, either "Lycée secondaire technique" or "complementaire", no streaming
41	2.7	Classes of Lycée and one other type, either "Lycée secondaire technique" or "complementaire", streaming
50	3.2	Classes of Lycée secondaire technique and of complementaire, no streaming
51	6.5	Classes of Lycée secondaire technique and of complementaire, streaming in at least some classes

Stratum Number	Percent of Population	Description
60	5.5	Classes of Lycée, Lycée secondaire technique and complementaire in the school, no streaming
61	5.3	Classes of Lycée, Lycée secondaire technique and complementaire in the school, streaming in at least some classes

2.11.4 Selection of Sample

Approximately 50% of classes in the population selected by random start—constant interval. Selection is thus with probability proportional to size of class.

2.12 The Netherlands

2.12.1 Population Definition

All students in the second year of VWO/Havo, Mavo, LTO and LHNO (School types).

Note: i) The year level in The Netherlands is AE8.

ii) The school system is very complex and this definition includes approximately 80% of students at the year 8 level.

2.12.2 Excluded Population

Students in some lines of vocational education

LAO	(agricultural)
LEAO	(commercial)
LAVO	(general)
LMO	(tradesman)
LNO	(nautical)
ITO	(individual technical)
IHNO	(individual domestic science)
IAO	(individual agricultural)

This is approximately 20% of students at the year 8 level.

2.12.3 Stratification

The only stratification variable was course type.

Stratum Number	Percent of Population	Description
1	31.9	VWO/Havo
2	42.0	Mavo
3	14.4	LTO
4	11.7	LHNO

2.12.4 Selection of Sample

Within strata, schools were selected with probability proportional to size using the random start—constant interval technique.

Within school, one class was selected by the interval method with the number of students the size factor.

Note: Strata 3 and 4 were oversampled to allow adequate between strata comparisons.

2.13 New Zealand

2.13.1 Population Definition

"All students who are in normal classes in Form 3". This is the year level where the majority has attained the age 13.00 to 13.11 years by the middle of the school year.

2.13.2 Excluded Population

Students enrolled with the Correspondence School and those in special schools for the handicapped.

The excluded population is 0.6% of the target population.

2.13.3 Stratification

Stratification Variables were School Type (Private and *Integrated/State) and Sex of Students (Boys/Girls/Coeducational).

* Integrated schools are schools which were formerly private (mostly Roman Catholic) schools which have now been integrated into the state system. At the time of the study these schools had integrated comparatively recently and it was judged that their characteristics would resemble those of private schools on a number of study variables.

Stratum Number	Percent of Population	Description
1	5.8	Private and Integrated, Boys
2	5.7	Private and Integrated, Girls
3	1.6	Private and Integrated, Coeducational
4	9.8	State, Boys
5	9.0	State, Girls
6	68.1	State, Coeducational

2.13.4 Selection of Sample

Schools were ordered by geographical criteria within strata and selected, with probability proportional to number of students in the target grade, by the random start—constant interval method. The random start—constant interval method used to select schools also identified the first class. The second class in each school was randomly selected. Intact classes were sampled.

2.14 Nigeria

2.14.1 Population Definition

All students who were

i) in Form 3 in state-owned Secondary Grammar Schools which prepare students for the West African School Certificate Examination.

ii) attending regular classes in the year of data collection.

iii) in the 8 (of 10) Southern states defining the strata.

Note: The target population was originally intended to include students from all states. Logistic and financial constraints caused the National Center to reduce this to the 10 Southern States (which included 89.6% of school enrolments). Of these 10 states no data was received from one and only 1 school (22 students) returned data from another. These strata were discarded.

2.14.2 Excluded Population

Students in Trade Schools, Technical and other Vocational and Pre-Vocational institutions.

Students in schools which have been established for less than 5 years or in schools for the handicapped. (Percent of population not known).

2.14.3 Stratification

The sample was stratified by state.

Stratum Number	Percent of Population	Description
1	16.8	Anambra
3	19.9	Bendel
11	6.6	Kwara
12	15.3	Lagos
14	7.0	Ogun
15	10.3	Ondo
16	16.0	Oyo
18	8.1	Rivers

2.14.4 Selection of Sample

Schools were selected in each state with probability proportional to the number of schools in each state. One class per school was randomly selected and at the final stage 30 students were randomly selected in each class.

2.15 Ontario

2.15.1 Population Definition

Students enrolled in normal grade 8 classrooms in Ontario.

2.15.2 Excluded Population

Special schools (military, hospital, reformatory, handicapped, etc).

Very small schools (fewer than 10 students in grade 8).

The total excluded population is estimated by the Ontario National Center to be less than 2%.

2.15.3 Stratification

Stratification variables were:

Size of School - Big (50 or more grade 8 students)
- Small (fewer than 50 grade 8 students)

School Type - Public (English language)
Separate (English language)
Private (English language)
French language

Location
R1 City of Toronto
R2 Etobicoke and York Metropolitan Toronto Boroughs
R3 East and North York Metropolitan Toronto Boroughs
R4 Scarborough Metropolitan Toronto Borough
R5 Toronto Suburbs (Mississuaga, Brampton, Oshawa)
R6 Ottawa
R7 Windsow
R8 London
R9 Waterloo, Kitchener, Cambridge
R10 Hamilton
R11 Northern Ontario Cities (Thunder Bay, Sault Ste Marie, Sadbury)
R12 Smaller Southern Ontario Cities (Sarnia, Brantford, St Catharines, Burlington, Oakville, Barrie Kingston, Peterborough)
R13 Rural Eastern Ontario (Ottawa Valley)
R14 Rural Northwest Ontario (Thunder Bay area)
R15 Rural North Centre Ontario (Sudbury area)
R16 Rural Northeast Ontario (North Bay area)
R17 Rural Southwest Ontario (Windsor Area)
R18 Rural Central Southwest Ontario (Kitchener area)
R19 Rural Niagara area
R20 Rural Central Ontario (Barrie area)
R21 Rural East Central Ontario (Lindsay area)
R22 Rural Southeastern Ontario (Kingston area)

Stratum Number	Percent of Population	Description
1	4.7	Small Public R1-R12
2	2.5	Small Public R13-R22
3	2.0	Small Public R14, R15, R16
4	3.3	Small Public R17, R18
5	3.0	Small Public R19, R21
6	2.5	Small Separate R1-R5
7	3.8	Small Separate R6-R12
8	4.3	Small Separate R13-R22
9	2.4	Small French
10	1.9	Private
11	3.2	Big Public R1
12	2.8	Big Public R2
13	4.3	Big Public R3
14	3.3	Big Public R4
15	4.7	Big Public R5
16	4.7	Big Public R6, R8, R9
17	3.3	Big Public R7, R10, R11
18	4.2	Big Public R12
19	4.8	Big Public R13, R22
20	4.0	Big Public R14-R16, R20
21	5.7	Big Public R17, R18
22	6.5	Big Public R19, R21
23	5.7	Big Separate R1-R5
24	4.8	Big Separate R6-R12
25	4.3	Big Separate R13-R22
26	2.8	Big French

2.15.4 Selection of Sample

Small schools (on the stratum list) are those with less than 50 grade 8 students (median 25).

Schools were chosen with equal probability for strata 1-9 and with probability proportional to size (of grade 8) within stratum for strata 10-26. For strata 1-9 all students were selected, in stratum 10 one class was randomly selected and in strata 11-26 two classes were randomly selected.

Five schools (with replacements) were drawn for each stratum. Numbers of schools and classes were chosen to give correct representation to small schools and large schools.

Note: Not all schools declining to participate were able to be replaced and there are minor deviations from the above plan.

Mean cluster sizes vary considerably between strata.

2.16 Scotland

Note: Scotland did not draw a fresh sample but followed up a national sample of students drawn when the students were in their final year of primary school in 1978.

2.16.1 Population Definition

Students at state schools in the second year of secondary schooling (S2) who were in the final year of Scottish primary schools in 1978.

2.16.2 Excluded Population

Students in independent schools (approx 1.7%)
Students in special schools for the handicapped etc (Approximately 1.9%)
Immigrants to Scotland since 1978 (a very small number)

2.16.3 Stratification

For the sample drawn in 1978 the stratification variables were:

Local authority (including grant-aided);
Size of school in 1974.

Samples were confirmed in 1978 as being representative of primary schools at that date.

2.16.4 Selection of sample

For the 1978 sample 24 students were chosen from each school by date of birth, or where the number of students at the P7 grade level was less than 24, all students were included in the sample. Only students in P7 in 1978 were selected. These students were therefore in S2, the IEA target grade, in 1980 since grade repeating is almost non-existent in Scottish schools.

2.17 Swaziland

2.17.1 Population Definition

Students in Form 2, ie. the grade level in which 13 year old students should be found according to the school system.

Note: In Swaziland 13 year old students are distributed across all 10 grades of schooling with more than 90% not having reached Form 2. Form 2 is the grade level where 13 year olds would be found if they entered grade 1 at 5 years of age and did not repeat grades. More significantly, it is the grade level at which the curriculum was judged by the National Committee to be most appropriate for the IEA cognitive tests.

The actual age distribution of the sample was:

Age	12	13	14	15	16	17	18	19	20+
Percent	1.8	10.3	20.6	22.5	18.1	17.2	4.7	2.7	2.8

2.17.2 Excluded Population

In terms of the defined population the excluded population is nil. It should be noted that in Swaziland in 1980 19.9% of 12-17 year olds were in school. (World Bank Education Sector Policy Paper 1980)

2.17.3 Stratification

No stratification used.

2.17.4 Selection of sample

The approved sampling plan was for random selection of 25 schools with probability proportional to size.

In the event, only 35 of the 82 Swaziland secondary schools responded to a circular asking whether they were willing to participate. Of these 27 responded positively and 8 negatively. Two of the schools responding positively were excluded (no information on the method of exclusion is available) and the remaining 25 were formally invited to participate. All agreed to do so and hence comprise the sample. One class from each school was selected at random by the National Research Coordinator.

2.18 Sweden

2.18.1 Population definition

Students in grade 7 of the compulsory school. These students study either a general course in mathematics or an advanced course.

2.18.2 Excluded population

Not stated

2.18.3 Stratification

Sweden is divided into 24 administrative provinces which consist of some 270 municipalities. The National Center created 14 strata consisting of municipalities stratified by 4 variables:

Number of inhabitants;
Percentage of socialist seats in local government;
Percentage employed in the local administration;
Percentage of immigrant students.

A fifth stratifying variable, type of course, was introduced for weighting purposes because the selection procedure resulted in a disproportionate sampling of advanced course and general course classes.

Stratum Number	% of Population	Population	% Socialist eats in govt.	% in local admin	% immigrant students	Course
1	2.7	25,000	50%	25%	8%	General
2	2.1	25,000	50%	25%	8%	General
3	1.1	25,000	50%	25%	8%	General
4	1.3	25,000	50%	25%	8%	General
5	2.3	25,000	50%	25%	8%	General
6	4.8	Information not supplied				General
7	0.9	"	"	"		General
8	0.6	"	"	"		General
9	1.4	"	"	"		General
10	1.2	"	"	"		General
11	1.4	"	"	"		General
12	0.6	"	"	"		General
13	3.2	"	"	"		General
14	2.7	"	"	"		General
15	7.1	25,000	50%	25%	8%	Special (Advanced)
16	6.1	25,000	50%	25%	8%	Special
17	3.2	25,000	50%	25%	8%	Special
18	3.2	25,000	50%	25%	8%	Special
19	7.7	25,000	50%	25%	8%	Special
20	14.2	Information not supplied				Special
21	2.2	"	"	"		Special

Stratum Number	% of Population	Population	% Socialist seats in govt	% in local admin	% immigrant students	Course
22	1.7	Information not supplied				Special
23	2.9	" " "				Special
24	3.1	" " "				Special
25	4.1	" " "				Special
26	1.6	" " "				Special
27	7.9	" " "				Special
28	8.6	" " "				Special

2.18.4 Selection of sample

Schools were randomly selected with probability proportional to size of target grade within each of the 14 national center strata (ie. Strata, 1, 15; Strata 2, 16, etc).

Two classes per school were selected, one class taking the advanced course. Classes were selected by drawing a student at random from each of the two course lists provided by the school and letting the classes those two students belong to be represented in the sample.

2.19 Thailand

2.19.1 Population definition

All students in normal classes in grade 8 in all 71 provinces.

2.19.2 Excluded population

None stated but note that approximately 85% of the age cohort was enrolled in grade 8 at the time of the Study.

2.19.3 Stratification

Stratification is by geographical region. Approved sampling plans indicated 12 regions, but in the executed sample Bangkok was included as a separate region to give 13 strata.

Stratum Number	Percent of Population	Description		
1	6.9	Description not supplied		
2	2.2	"	"	"
3	11.8	"	"	
4	2.7	"	"	"
5	5.7	"	"	"
6	8.7	"	"	"
7	6.4	"	"	"
8	7.9	"	"	"
9	7.1	"	"	"
10	8.1	"	"	"
11	7.8	"	"	"
12	6.1	"	"	"
13	18.5	Bangkok		

2.19.4 Selection of sample

Schools were randomly selected with probability proportional to size of target grade.

One class per school was then randomly selected by the National Center.

2.20 United States of America

2.20.1 Population Definition

All students in the eighth grade of mainstream public and non-public schools.

2.20.2 Excluded Population

Students with disabilities (mental, physical, emotional or learning) (sufficiently severe to require their placement in special education classes rather than in mainstream classes).

2.20.3 Stratification

Stratification variables were:

School Type (Public/Private);
Regional Standard Metropolitan Statistical Area (SMSA) Location (East-Central/South-West);
Metropolitan Status Grade (City/Suburb/other or district outside SMSA);

Stratum Number	Percent of Popultation	Description
1	10.4	East-Central/SMSA City
2	20.4	East-Central/SMSA Suburb
3	11.5	East-Central/Non-SMSA
4	10.7	South-West/SMSA City
5	20.3	South-West/SMSA Suburb
6	15.6	South-West/Non-SMSA
7	11.1	Private

2.20.4 Selection of Sample

Separ[illegible]te national probability samples were drawn for [illegible]lic and private schools.

The [illegible]onal probability sample of public schools was [illegible] two stages: (administrative) district and school within district. In the first stage districts were selected with probability proportional to size of grade eight enrolment. In the second stage public schools were selected without replacement, two per grade eight level, with probability proportional to the estimated number of 8th grade students in district schools.

The national probability sample of private schools was selected with probability proportional to size of total school enrolment. From both school types two intact classes per school were selected with equal probability from content - ability substrata.

Sampling plans called for the total number of school districts selected to be dependent on the co-operation rate among school districts, i.e. for a co-operation rate of 50%; 140 school districts were to be sampled to achieve the designed sample size of 70 school districts. The co-operation rate did prove to be of this order.

3 NATIONAL POPULATION DEFINITIONS AND SAMPLING PROCEDURES - POPULATION B

3.1 Belgium (Flemish)

3.1.1 Population Definition

All students who are in the normally accepted terminal grade of secondary education and who are studying a minimum of 5 hours of mathematics per week.

3.1.2 Excluded Population

Defined by National Center as those students in the normally accepted terminal grade of secondary education who are studying mathematics for less than 5 hours per week.

Note: National Center estimated 25-30% of students in the terminal grade constitutes Population B.

Approximate size of age cohort = 90,000
Number in population B = 12,900
i.e. Population B is of the order of 14% of the age cohort (International Center estimate).

3.1.3 Stratification

Education Authority: State,
Catholic,
Local Board ("Provincial" and "Communal")

by

Curriculum: Academic type 1 - Renewed - comprehensive
Technical type 1 - Renewed - comprehensive
Academic type 2 - Traditional - selective
Technical type 2 - Traditional - selective

Stratum Number	Percent of Population	Description
1	3.7	Catholic, academic type 1
2	0.3	Catholic, technical type 1
3	70.4	Catholic, academic type 2

Stratum Number	Percent of Population	Description
4	2.6	Catholic, technical type 2
5	1.9	Local Board, Academic type 1
6	0.2	Local Board, technical type 1
7	0.7	Local Board academic type 2
8	0.1	Local Board technical type 2
9	11.1	State, academic type 1
10	2.1	State, technical type 1
11	6.5	State, academic type 2
12	0.3	State, technical type 2

3.1.4 Selection of Sample

Schools were ordered by geographical criteria within strata.

"Tickets" were allocated, one for each school with 40 or less students, two for each school with more than 40 students and then schools selected by the random start—constant interval method. Where a selected school had 40 or less students all students were tested. Where a selected school had more than 40 students half of the students were included in the sample. These students may be drawn from several classes.

3.2 Belgium (French)

3.2.1 Population Definition

All students in the sixth year of the secondary school system who are studying mathematics for a minimum of 5 hours a week.

3.2.2 Excluded Population

All students studying mathematics for less than 5 hours a week. Population B is approximately 14% of the age cohort.

3.2.3 Stratification

Initially stratification was School type (Catholic, Local Board, State) by Curriculum type (General, Traditional) by Course Type (General, Technical) giving 12 strata.

By the time data collection was carried out the proportion of Traditional Curriculum type versus Renewed type had changed considerably so a reduced stratification frame was used at the suggestion of the Belgium (French) National Center.

This was School type (Catholic, Local Board, State) by Course type (General, Technical) giving 6 strata.

Stratum Number	Percent of Population	Description
1	47.5	Catholic, general
2	1.5	Catholic, technical
3	8.6	Local board, general
4	2.2	Local board, technical
5	38.8	State, general
6	1.3	State, technical

3.2.4 Selection of Sample

Identical to that for Belgium (Flemish). See 3.1.4.

3.3 British Columbia

3.3.1 Population Definition

All students in the British Columbia public schools who are enrolled in the course Algebra 12 as of September, 1980.

3.3.2 Excluded Population

Students enrolled in private schools at grade 12 level. (Less than 3% excluded.)

3.3.3 Stratification

Stratification was by geographical zone.

Stratum Number	Percent of Population	Description
1	13.0	Zone 1
2	48.2	Zone 2
3	6.8	Zone 3
4	18.1	Zone 4
5	5.8	Zone 5
6	8.1	Zone 6

3.3.4 Selection of Sample

Samples were drawn independently from each zone. Within zone the total number of classes was determined and classes selected with probability proportional to size of Population B enrolment. In most schools only one class was selected but in a few with large Population B enrolments 2 or 3 classes were drawn.

3.4 England and Wales

3.4.1 Population Definition

Final year Sixth form pupils in the second year of study for A or S level qualifications in mathematics including pupils in sixth form colleges and independent schools.

3.4.2 Excluded Population

A very small number of students taking similar courses at polytechnics and other further education institutions.

Note: Appr[illegible]ately 16% of the age cohort is in scho[illegible] . this level. Of these approximatel[illegible] [illegible]0% study (Population B) mathematics. Population B is thus approximately 6% of the age cohort.

3.4.3 Stratification

Stratification variables were Region, Location, Size of Target Grade, School Type.

Stratum Number	Percent of Population	Description
1	3.2	North, Metropolitan, target grade 1-35, Comprehensive to 18
2	1.9	North, Non-Metropolitan, target grade 1-35, Comprehensive to 18
3	3.6	North, Metropolitan, target grade 36-60, Comprehensive to 18
4	2.4	North, Non-metropolitan, target grade 36-60, Comprehensive to 18
5	4.8	North, Metropolitan, 61+ Comprehensive to 18
6	3.3	North, Non-metropolitan, 61+ Comprehensive to 18
7	2.5	North, All, All, Other Maintained
8	5.9	North, All, All, 6th form colleges
9	1.4	Midlands, Metropolitan, 1-35, Comprehensive to 18
10	2.8	Midlands, Non-metropolitan, 1-35 Comprehensive to 18
11	1.4	Midlands, Metropolitan, 35-60, Comprhensive to 18
12	3.4	Midlands, Non-metropolitan, 35-60, Comprehensive to 18
13	4.5	Midlands, All, 61+ Comprehensive to 18
14	2.4	Midlands, All, All, Other maintained
15	3.3	Midlands, All, All, 6th form colleges
16	3.7	South, Metropolitan 1-35, Comprehensive to 18

Stratum Number	Percent of Population	Description
17	4.5	South, Non-metropolitan, 1-35 Comprehensive to 18
18	4.1	South, Metropolitan, 35-60 Comprehensive to 18
19	5.7	South, Non-metropolitan, 35-60, Comprhensive to 18
20	3.3	South, Metropolitan, 61+, Comprehensive to 18
21	7.2	South, Non-metropolitan 61+, Comprehensive to 18
22	7.2	South, All, All, Other maintained
23	7.7	South, All, All, Sixth form colleges
24	3.2	North, All, All, Independent
25	1.5	Midlands, All, All, Independent
26	4.2	South, All, All, Independent
27	0.2	Wales, All, All, Independent
28	0.8	Wales, All, All, Other maintained

3.4.4 Selection of Sample

A two stage stratified sample was drawn. Schools were stratified as above and a random sample of schools drawn from each stratum combination. In the second stage a random sample of students was drawn from the selected schools. The sampling proportion of students in a school was inversely proportional to school size.

3.5 Finland

3.5.1 Population Definition

Students studying the long course in mathematics (four 45 minute periods per week) in grade 3 of Finnish speaking upper secondary schools.

3.5.2 Excluded Population

Swedish speaking upper secondary schools
Evening classes of upper secondary schools

Province of Uusimaa:	Alppila upper secondary school Helsinki French-Finnish school Finnish-Russian school Rudolph Steiner school
Province of Vaasa:	upper secondary school of music Kaustinen

Note: Disregarding evening classes, the excluded sample is probably of the order of 5% of the target population (International Center estimate). Exact statistics not available.

Population B is 12.4% of the age cohort.

3.5.3 Stratification

Stratification variables were Province and Location (Urban/Rural)

Stratum Number	Percent of Population	Description
01	19.3	Uusimaa, towns
02	2.1	Uusimaa, rural
03	10.3	Turku and Pori, towns
04	4.9	Turku and Pori, rural
05	9.7	Häme, towns
06	4.3	Häme, rural
07	6.7	Kymi, towns
08	-	Kymi, rural
09	3.1	Mikkeli, towns
10	1.9	Mikkeli, rural
11	3.5	Vaasa, towns
12	3.9	Vaasa, rural
13	2.4	Keski - Suomi, towns
14	3.1	Keski - Suomi, rural
15	4.1	Kuopio, towns
16	2.7	Kuopio, rural
17	1.9	Pohjiois - Karjala, towns
18	1.8	Pohjiois - Karjala, rural
19	5.0	Oulu, towns
20	5.0	Oulu, rural
21	2.5	Lappi, towns
22	1.9	Lappi, rural

Note: Stratum 08 was represented by only 1 school in the designed sample and data was not received for this school. The stratum was thus eliminated and N adjusted accordingly.

3.5.4 Selection of Sample

Schools were selected with probability proportional to size of target population by the random start—constant interval method.

One class per school was randomly selected.

3.6 Hong Kong

3.6.1 Population Definition

Population B is made up of two sub-populations:

Population B1. All students in Lower Six or Middle Six who are studying mathematics as a substantial part (approximately 5 hours or more per week) of their academic program.

Population B2. All students in Upper Six or Form 7 studying mathematics as a substantial part (approximately 5 hours or more per week) of their academic program.

Note: The situation in Hong Kong is complex as there are two grade levels which are pre-university years. The ages of Lower Six and Middle Six students correspond to those of students in their terminal year in most countries. Upper Six and Form 7 students are one year older. The four groups are collectively referred to as Form 6 or matriculation classes.

For the purposes of international analyses the two sub-populations are treated as one combined population, which can be described as:

All students in matriculation classes who are studying mathematics as a substantial part (approximately 5 hours or more per week) of their academic program.

3.6.2 Excluded Population

Nil

Note: The target population is a highly selected group within the Hong Kong school system (approximately 6% of the age cohort).

3.6.3 Stratification

Stratification variables are School Type (Public/Private) by Sex of Students (Boys/Girls/Coeducational) by Language of Instruction (English/Cantonese)

Stratum Number	Percent of Population	Description
1	14.6	Public, Boys, English
2	0.8	Public, Boys, Cantonese
3	7.8	Public, Girls, English
4	1.6	Public, Girls, Cantonese
5	3.2	Public, Coeducational, English
6	6.6	Public, Coeducational, Cantonese
7	0.9	Private, Boys, English
8	-	Private, Boys, Cantonese
9	-	Private, Girls, English
10	-	Private, Girls, Cantonese
11	55.5	Private, Coeducational, English
12	9.1	Private, Coeducational, Cantonese

Note: Strata 8 and 10 contain no schools. Stratum 9 contains 6 schools but was not included in the sample.

3.6.4 Selection of Sample

Classes were listed within strata and selected by the random start—constant interval method, ie. with probability proportional to size of class.

3.7 Hungary

3.7.1 Population Definition

The set of all pupils in the 4th grades of Hungarian grammar schools, specialised vocational secondary schools and technical schools.

Note: (International Center). Although they study mathematics for approximately 5 hours per week a substantial proportion of students at specialised vocational secondary schools and technical schools are undertaking courses at a lower level than would be considered pre-university courses. Population B as defined above is approximately 50% of the age cohort.

3.7.2 Excluded Population

The 4th grades of Workers' Schools are excluded. Terminal grades of institutions for skilled workers, schools of shorthand and typing, secondary schools of health care and special education classes.

Note: (International Center). A negligible number of the above would fall within the population B definition and thus the excluded population is nil.

3.7.3 Stratification

The original sampling plan (approved by the sampling referee) had three stratification variables; type of school (Grammar School/ Specialised Vocational Secondary Schools/ Technical Schools); Type of Settlement (Large Town/Small Town/Village); Type of Curriculum (7 categories, 3 present in Grammar Schools and 4 in SVSS).

For international purposes the Type of Settlement variable was not used. It should also be noted that Technical Schools are almost "extinct" and none were drawn in the sample.

Stratum Number	Percent of Population	Description
1	41.1	Grammar Schools, Curriculum type CG1
2	3.1	Grammar Schools, Curriculum type CG2

Stratum Number	Percent of Population	Description
3	0.2	Grammar Schools, Curriculum type CG3
14	45.1	SVSS, Curriculum type CS1
15	6.6	SVSS, Curriculum type CS2
16	3.6	SVSS, Curriculum type CS3
17	0.3	SVSS, Curriculum type CS4

3.7.4 Selection of Sample

Classrooms were listed by region within strata and selected with probability proportional to number of classes in stratum column by random start—constant interval. Some cells with very few classrooms were oversampled.

3.8 Israel

3.8.1 Population Definition

Students in Hebrew speaking schools offering extended mathematics programs in the terminal year of schooling.

Note: Not all schools offer such courses and the number of schools containing target population students is much smaller than the number of all secondary schools in the country.

3.8.2 Excluded Population

Students in Arabic speaking schools. Students of 6 schools deleted from list of qualifying schools through lack of information. Students of schools (approximately 4) from strata from which no data was collected.

3.8.3 Stratification

The approved sampling plan was based on two stratification variables:

- Type of School (Academic, Vocational, Continuation and Agricultural)
- Extent of Mathematics Programmes (schools with 4 point (360 periods) programmes, schools with 4 or 5 point (450 periods) programmes).

Vocational and agricultural schools do not offer 5 point programmes and there were thus 6 strata.

This plan was altered before data collection to Type of School (as above) x (Recognised, Not Recognised) ie. 8 strata. The terms "recognised" and "Not recognised" were not defined.

Information relating to the first and second frames could only be reconciled by constructing a frame based on School type only. Thus for weighting purposes there are four strata:

Academic
Vocational
Continuation
Agricultural

Stratum Number	Percent of Population	Description
1	79.4	Academic
2	8.9	Vocational
3	3.6	Continuation
4	8.0	Agricultural

3.8.4 Selection of Sample

Schools were classified by Type of School, Extent of Mathematics Programmes and Number of Parallel Classes in the Terminal Grade. Schools were listed according to the resulting clusters and 5 schools out of each consecutive 7 were selected. (The third and seventh were discarded).

The designed sample was 96 out of 133 schools.

All students in Population B mathematics classes in the selected schools were tested.

3.9 Japan

3.9.1 Population Definition

All students who are in the normally accepted terminal grade (grade 12) of the upper secondary school and who are studying mathematics as a substantial part (more than 5 hours per week) of their academic programme.

Note: This is 29% of all students in the terminal secondary level (National Center). About half the age cohort is in Upper Secondary Schools at this level (structure and diagram, Educational Statistics Japan, 1976 edition, Ministry of Education, Science and Culture). Population B is thus approximately 14-15% of the age cohort.

3.9.2 Excluded Population

All students of technical colleges, vocational courses of Upper Secondary and Special schools. The proportion of these students taking "substantial" mathematics courses cannot be determined from available information, but is probably very small. Only 0.6% of the age group is in technical and non-technical colleges.

3.9.3 Stratification

Stratification variables were School Type (Public/Private/National) and Percent of Students in the Target School who entered University in the Year prior to Testing (i.e. in 1979).

Stratum Number	Percent of Population	Description
11	26.6	Public School, 0 - 34% entered University in 1979
12	49.7	Public School, 35 - 64% entered University in 1979
13	9.2	Public School, 64 - 100% entered University in 1979
21	3.4	Private School, 0 - 34% entered University in 1979
22	7.1	Private school, 35 - 64% entered University in 1979
23	3.3	Private school, 65 - 100% entered University in 179
33	0.7	National school

3.9.4 Selection of Sample

Schools were selected with probability proportional to size followed by random selection of one class in each school. In some schools an additional class was randomly selected.

3.10 New Zealand

3.10.1 All students who are in Form 7 and who are studying Pure Mathematics as a substantial part (approximately 5 hours per week) of their academic program.

Form 7 is the terminal year of secondary education in New Zealand. Those studying mathematics comprise 11% of the age cohort.

3.10.2 Excluded Population

Those students enrolled with the Correspondence School and those in special schools for the handicapped. The excluded population is 0.4% of the target population.

3.10.3 Stratification

Stratification variables were School Type (Private and Integrated/State) and Sex of Students (Boys/Girls/Coeducational).

Note: Integrated schools were formerly private schools but are now integrated into the state system. At the time of the study the process of integration was taking place and these schools were judged likely to be more comparable to Private than to state schools on study variables.

Stratum Number	Percent of Population	Description
1	12.4	Private and Integrated, Boys
2	6.8	Private and Integrated, Girls
3	1.8	Private and Integrated, Coeducational
4	16.2	State, Boys
5	9.1	State, Girls
6	53.7	State, Coeducational

3.10.4 Selection of Sample

Schools were ordered within strata by geographical criteria and selected by random start—constant interval with probability proportional to size of Population B grade enrolment. The same process identified the intact class to be tested.

3.11 Ontario

3.11.1 Population Definition

Students in grade 13 who are taking two or more of the courses "Relations", "Calculus", "Algebra".

3.11.2 Excluded Population

Students in schools specialising in foreign students or schools with no fixed timetable.

3.11.3 Stratification

Stratification variables are Geographical Region or Category, Size of Community and Ratio of Grade 13 to Grade 12 students.

Stratum Number	Percent of Population	Description	
1	5.4	Toronto, Small, Low	
2	5.1	" " High	
3	5.4	" Large, Low	
4	6.0	" " High	
5	8.2	Cities outside Toronto except North, Small, Low	
6	8.0	" " " " " " High	
7	7.2	" " " " " Large, Low	
8	7.4	" " " " " " High	
9	3.2	Rural North and Northern Cities, Rural Ottawa, Small, Low	
10	3.8	" " " " " " " Large, Low	
11	5.7	Rural West	Small
12	5.8	" "	Large
13	5.2	Rural Central and East	Small
14	5.4	" " " "	Large
15	5.6	Private English	Small
16	5.8	" "	Large
17	6.7	French, (Public and Private)	

3.11.4 Selection of Sample

From each stratum five schools were drawn with probability proportional to size (of students in grade 13).

The sample of students from a school was determined upon investigation of the actual number of students by course, semester and the like school by school.

For the international sample it appears one class from each of the courses "Relations", "Calculus" and "Algebra" was selected. Students within those classes taking two or more of the courses comprise the population B sample.

3.12 Scotland

3.12.1 Population Definition

All pupils in the 5th and 6th year of secondary schooling who are studying for either

i) SCE Higher Mathematics

ii) ACE Advanced Level Mathematics

iii) Scottish Certificate of Sixth Year Studies in Mathematics

in either Local Authority or Grant-aided Schools.

3.12.2 Excluded Population

Those pupils in independent schools (not in the state system) are excluded. (Approximately 3.3% of the IEA Population B).

3.12.3 Stratification

Local authority schools were stratified by "sizeband" where "sizeband" is determined by the number of presentations in Higher and Scottish Certificate of Sixth Year Studies in 1978.

Grand-aided schools form a separate stratum.

Stratum Number	Percent of Population	Description
1	17.8	Local authority x (average) 19 presentations per school
2	37.6	Local authority x (average) 56 presentations per school
4	22.8	Local authority x (average) 100 presentations per school
6	12.0	Local authority x (average) 150 presentations per school
9	9.8	Grand aided

Note: Limits of size bands for Local Authority Schools not available. Averages included to give indication of ranges.

3.12.4 Selection of Sample

The sampling frame was stratified by presentation size factor and school roll (1 - 800, 800 - 1400, 1400 and over).

i) Local Authority Schools

Each school was allocated a size factor of 1, 2, 4 or 6. Schools were then ordered by Local Authority Region and by size factor within each region. Within each major region a systematic 1:12 sample was drawn from a random start giving schools of size 6 six chances in the draw, schools of size 4 four chances and so on.

ii) Grant-aided schools

The list was divided into Boys', Girls' and Mixed schools. Since schools were of similar size within these divisions a simple random selection was made to give the correct pro-rata split of the 6 schools required (out of 20).

Pupils within schools sampled with probability inversely proportional to size factor.

3.13 Sweden

3.13.1 Population Definition

Students in grade 3 of the natural sciences line and the technical line. The mathematics course is the same for these students.

3.13.2 Excluded Population

Not stated.

3.13.3 Stratification

The sampling plan approved by the Sampling Referee had 14 strata consisting of municipalities stratified by 4 variables:

A Population
B Percentage of Socialist Seats in the Local Government
C Percentage Employed in Public Administration
D Percentage of Immigrant Students.

Note: Sweden is divided into 24 administrative provinces which consist of some 270 municipalities.

Stratum Number	Percent of Population	Description			
		A	B	C	D
1	9.9	> 25000	> 50%	> 25%	> 8%
2	9.9	> 25000	> 50%	> 25%	< 8%
3	4.6	> 25000	> 50%	< 25%	> 8%
4	4.6	> 25000	> 50%	< 25%	< 8%
5	12.8	> 25000	< 50%	> 25%	> 8%
6	25.2				
7	2.4				
8	0.9				
9	1.2	(Information not supplied. 1-5 given as example)			
10	1.4				
11	3.5				
12	0.3				
13	0.9				
14	21.9				

Note: This sampling plan gave disproportionate representation to the two course types available. A fifth stratifying variable, Type of Course, was introduced at the International Center for weighting purposes. Each of the existing strata was divided on the basis of the Long and Short courses, giving 28 strata.

3.13.4 Selection of Sample

Schools were randomly selected with probability proportional to size of target grade within each of the national center strata.

One class per school was randomly selected.

3.14 Thailand

3.14.1 Population Definition

All students in normal classes at the terminal grade of the secondary education system (grade 12) who were studying mathematics six periods per week (1 period = 50 Minutes).

3.14.2 Excluded Population

Two strata (educational regions) were not included in the designed sample. Five percent of potential Population B students were thus excluded.

3.14.3 Stratification

Stratification of data sent to the International Center was by educational region. There are 13 educational regions but the two smallest of these (in terms of number of schools) were not included in the designed sample.

Stratum Number	Percent of Population	Description
1	5.1	None supplied
3	9.6	"
5	5.0	"
6	6.4	"
7	7.2	"
8	9.4	"
9	8.0	"
10	11.9	"
11	9.5	"
12	5.1	"
13	22.8	Bangkok

3.14.4 Selection of Sample

The NRC report describes the sampling method as selection of 64 schools with probability proportional to size and random selection of intact classes within schools.

This oversimplifies the procedures.

The selection of schools was based on stratification by number of classrooms per school and the number of classes per school chosen ranged from 1 to 4 depending on school size.

Designed samples based on this stratification variable or on the regional stratification variable do not indicate strict probability proportional to size sampling. The two stratification variables appear to have been used independently.

However, from information supplied by the NRC and by combining the sampling frames very good national estimates of statistics can be obtained. In effect the random selection was of classes with probability proportional to number of classes.

3.15 United States of America

3.15.1 Population Definition

All students in mainstream public and non-public schools in (typically terminal) fourth year advanced mathematics courses that require as prerequisites three years of secondary level mathematics (typically two years of algebra and one of geometry).

3.15.2 Excluded Population

Students in the normally accepted terminal grade

i) who are in classes typically consisting almost of students from lower grade levels (eg. a geometry class made up mostly of grade 10 students)

ii) whose mathematics work consists primarily of remedial mathematics, business, shop or other vocational mathematics as opposed to a terminal year academic program in mathematics.

3.15.3 Stratification

Stratification variables were:

School Type (Public/Private);
Regional Standard Metropolitan Statistical Area (SMSA) Location (East-Central/South-West);
Metropolitan Status Code (City/Suburb/other or district outside SMSA)

Stratum Number	Percent of Population	Description
1	10.7	East-Central/SMSA, City
2	21.5	East-Central/SMSA, Suburb
3	11.8	East-Central/Non-SMSA
4	11.0	South-West/SMSA, City
5	20.6	South-West/Non-SMSA
6	15.8	South-West/Non-SMSA
7	8.5	Private

3.15.4 Selection of Sample

Separate national probability samples were drawn for public and private schools.

The national probability sample of public schools was in two stages: (administrative) district and school within district. In the first stage districts were selected with probability proportional to size of grade 12 enrolment. In the second stage public schools were selected without replacement, two per grade 12 level, with probability proportional to the estimated number of 12th grade students in district schools. The national sample of private schools was selected with probability proportional to size of total school enrolment. From both school types two intact classes per school were selected with equal probability from content ability substrata. Twice as many school districts as were needed to provide an adequate number of data points were invited to participate in the expectation of a 50% cooperation rate at this level. This expectation proved fairly accurate. Some replacement occurred at school level.

4 RESPONSE RATES - POPULATION A

National Centers submitted their sampling plans to the Sampling Referee, Dr Malcolm Rosier, ACER. Where these met the criteria for representativeness and precision they were approved immediately. In several cases approval was granted only after the National Center had agreed to modify their designs to improve their sample and had resubmitted their sampling plans.

In the interval between having their designed samples approved and executing the sample a few National Centers found it necessary to amend their designed samples. In some cases (e.g. Belgium Flemish and Belgium French) this was because the curriculum structure of the school system was changing rapidly. In others (e.g. The Netherlands) decisions were taken to over-sample in some strata to allow particular within - country analyses. There are thus differences between the designed sample and the executed sample for some systems with the size of the executed sample exceeding the size of the designed sample in some cases. Response rates are therefore calculated as a percent of the executed sample.

The achieved sample refers to the data used for analysis. Where data were received from a school or class but the number of cases was so small that the data could not be used in any analysis the school or class does not form part of the achieved sample. For Nigeria, the number of cases in 2 strata was judged too low and these 2 strata were eliminated and the national population redefined. In all other systems there were sufficient cases in all strata to allow viable parameter estimates using weighting, because where the achieved samples for strata were small, the populations for those strata were also small.

Sampling plans were constructed with the aim of confining sampling errors within acceptable limits (see Sampling Manual). Since systems designed their samples to varying limits within those advocated as the minimum acceptable there is no single response rate at national or stratum level which can be designated as the minimum acceptable for specific analyses, i.e. one cannot say that response rates of less than 70% (say) will necessarily give inadequate achieved samples. The adequacy of a sample can be judged against marker variables, where these are available, and against the calculated design effects (see section 9).

A further problem in calculating response rates at some levels lies in the fact that where a system calculated the number of schools (say) needed for the sample, the number of students at the target level in classes which would ultimately be selected had to be estimated. This resulted in some systems having a greater number of students in the achieved sample than were estimated in the designed sample. Similarly, for systems where two classes per school were to be chosen, it sometimes happened that in some selected schools there was only one class at the target level.

Response rates are therefore discussed below system by system with the most appropriate response rates for particular countries calculated. The levels at which these are quoted depend on the sampling units and the degree of accuracy with which statistics for the sampling frame at these levels were known when the frame was constructed.

Not all teachers and students in the achieved sample returned data on all instruments and through misadventures at two national centers (England and Wales and Belgium (Flemish)) some instruments for parts of the samples were lost to the study. The remaining data set in both cases is quite adequate for some research questions but is dubious for others. Response rates (as a percent of the achieved sample) are given by instrument.

The general level of response rates for schools (or classes) is:

Response rate	No. of systems
> 90%	12
80% - 89%	4
70% - 79%	2
60% - 69%	2

4.1 Belgium (Flemish)

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate %
Schools	200	Slightly under 200	158	> 80%
Classes	200		158	
Teachers	200	"	158	
Students			3103	

Achieved sampling fraction (schools) = 0.095

As can be seen in the table below a full set of student cognitive data is available.

Instrument	N	% of Achieved Sample
School Questionnaire	158	100
Teacher Background and Attitudes	154	97
Opportunity to Learn		
Form Core	137	87
Form A	138	87
Form B	138	87
Form C	138	87
Form D	136	87
Student Background and Attitudes*	1385	45
Cognitive Form Core	3073	99
Form A) 25% of total	767	99
Form B) sample to do	760	98
Form C) each form	759	98
Form D)	761	98

* National Center mishaps. The lost data was spread across all strata almost proportionately. Comparison between cognitive results for this 1385 students and total achieved sample reveals that little, if any, bias is likely to be introduced for most student background variables. However, use of data from this questionnaire in a causal model is dubious.

Comparison on Selected Cognitive Items between Students For Whom Students Questionnaire Data is Available and Total Sample.

Item		Reduced Sample p-value	Total Sample p-value
Core	7	73	73
	15	83	80
A	7	94	92
	15	64	64
B	7	83	82
	15	76	76
C	7	73	72
	15	77	76
D	7	59	56
	15	73	68

4.2 Belgium (French)

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate %
Schools	150	125	108	86
Classes	150		108	
Teachers	150		108	
Students			3103	

Achieved sampling fraction (schools) = 0.084

Instrument	N	% of Achieved Sample
School Questionnaire	108	100
Teacher Background and Attitudes	105	100
Teacher Opportunity to		
Learn Form Core	Not	-
Form A	administered	-
Form B	in	-
Form C	Belgium	-
Form D	(French)	-
Student Background and Attitudes	2054	99
Cognitive Form Core	2025	98
Form A	501	97
Form B	488	94
Form C	499	96
Form D	501	97

4.3 British Columbia

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate %
Schools	105	93	89	
Classes	105	93	89	96%
Teachers	105	93	89	
Students	2748		2228	

Instrument	N	% of Achieved Sample
School Questionnaire	89	100
Teacher Background and Attitudes	89	100
Teacher Opportunity to		
Learn Form Core	78	88
Form A	78	88
Form B	77	87
Form C	78	88
Form D	78	88
Student Background and Attitudes	2158	97
Student Cognitive Form Core	2168	97
Form A	519	93
Form B	535	96
Form C	528	95
Form D	522	94

4.4 England and Wales

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate %
Schools	133	114	94	82%
Students	4041	3206	2678	84 %

The sampling procedure selected schools and then students (not classes) in the target population within schools. Thus within schools students were typically drawn from several classes. In some schools all teachers with students in the sample completed questionnaires, in others only one or some completed questionnaires.

Instrument	N	% of Achieved Sample
School Questionnaire	94	100
Teacher Background and Attitudes	244	-
Teacher Opportunity to Learn Form Core	396	-
Form A	380	-
Form B	379	-
Form C	378	-
Form D	379	-
Student Background and Attitudes	2619	98
Student Cognitive Form Core	2612	98
Form A	652	97
Form B	642	96
Form C	644	96
Form D	643	96

Data was collected from 21 more schools than are included in the achieved sample. (See Section 2.4.5)

4.5 Finland

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	103	103	98	95
Classes	206	220	206	94
Teachers	206	220	206	
Students	5665	4914	4484	

The designed sample overestimated the number of students expected to be in sampled classes and experiments with heterogeneous classes being conducted in some schools led to more than 2 classes being selected in these schools.

Instrument	N	% of Achieved Sample
School Questionnaire	98	100
Teacher Background and Attitudes	206	100
Teacher Opportunity to Learn Form Core	198	96
Form A	199	97
Form B	199	97
Form C	200	97
Form D	199	97

Instrument	N	% of Achieved Sample
Student Background and Attitudes	4484	100
Student Cognitive Form Core	4382	98
Form A	1071	96
Form B	1095	98
Form C	1094	98
Form D	1082	97

4.6 France

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	194	188	187	99
Classes	388	367	365	99
Teachers	388	353 *	362	99
Students			8889	

* 14 teachers taught 2 sample classes. In the achieved sample such teachers are counted twice.

Instrument	N	% of Achieved Sample
School Questionnaire	187	100
Teacher Background and Attitudes	347	96
Teacher Opportunity to		
Learn Form Core	335	93
Form A	333	92
Form B	333	92
Form C	331	91
Form D	331	91
Student Background and Attitudes	8329	94
Student Cognitive Form Core	8317	94
Form A	2088	94
Form B	2102	95
Form C	2089	94
Form D	2080	94

4.7 Hong Kong

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools			125	
Classes	120-150		130	> 90
Teachers			130	
Students			5548	

Selection based on classes at target level.
Achieved sampling fraction (classes) = 0.055.

Instrument	N	% of Achieved Sample
School Questionnaire	125	100
Teacher Background and Attitudes	130	100
Teacher Opportunity to		
Learn Form Core	Not	-
Form A	Administered	-
Form B	to	-
Form C	Adequate	-
Form D	Sample	-
Student Background and Attitudes	5548	100
Student Cognitive Form Core	5495	99
Form A	1382	100
Form B	1367	99
Form C	1367	99
Form D	1373	99

4.8 Hungary

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	70	70	70	100
Classes	70	70	70	100
Teachers	70	70	70	100
Students		1843	1754	95

Instrument	N	% of Achieved Sample
School Questionnaire	70	100
Teacher Background and Attitudes	70	100
Teacher Opportunity to		
Learn Form Core	64	91
Form A	64	91
Form B	63	90
Form C	63	90
Form D	63	90
Student Background and Attitudes	1754	100
Student Cognitive Form Core	1754	100
Form A	441	100
Form B	439	100
Form C	442	100
Form D	432	99

4.9 Israel

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
School	101	99	81	82
Classes		150 *	140	
Teachers		150 *	140	
Students		4877	3819	78

* These are approximate. Selection of 1 or 2 classes depended on size of school and, in addition, home room classes commonly split into smaller classes for mathematics instruction.

Instrument	N	% of Achieved Sample
School Questionnaire	81	100
Teacher Background and Attitudes	140	100
Teacher Opportunity to		
Learn Form Core	140	100
Form A	136	97
Form B	137	98
Form C	133	95
Form D	135	95
Student Background and Attitudes	3587	94
Student Cognitive Form Core	3524	92
Form A	879	92
Form B	897	94
Form C	857	90
Form D	890	93

4.10 Japan

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	220	220	213	97
Classes	220	220	213	97
Teachers	220	220	213	97
Students	8200 *	8200 *	8091	

* Approximate.

Instrument	N	% of Achieved Sample
School Questionnaire	213	100
Teacher Background and Attitudes	212	100
Teacher Opportunity to		
Learn Form Core	209	98
Form A	211	99
Form B	211	99
Form C	209	98
Form D	209	98
Student Background and Attitudes	8091	100
Student Cognitive Forms Core	8091	100
Form A	2041	100
Form B	2030	100
Form C	2028	100
Form D	1992	98

4.11 Luxembourg

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	46	43	42	98
Classes	116	110	107	97
Teachers	116	110	107	97
Students	2390	2184	2106	96

Note: 1 school out of every 2 sampled.

Instrument	N	% of Achieved Sample
School Questionnaire	42	100
Teacher Background and Attitudes	107	100
Teacher Opportunity to Learn Form Core	85	92
Form A	84	91
Form B	84	91
Form C	84	91
Form D	82	89
Student Background and Attitudes	2106	100
Student Cognitive Form Core	2038	97
Form A	505	96
Form B	504	96
Form C	501	95
Form D	509	97

4.12 The Netherlands

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	215	236	236	100
Classes	215	236	236	100
Teachers	215	236	236	100
Students	5145		5500	

Instrument	N	% of Achieved Sample
School Questionnaire	236	100
Teacher Background and Attitudes	236	100
Teacher Opportunity to Learn Form Core	230	97
Form A	228	97
Form B	224	95
Form C	223	94
Form D	223	94
Student Background and Attitudes	5500	100
Student Cognitive Form Core	5418	99
Form A	1353	98
Form B	1337	97
Form C	1341	98
Form D	1365	99

4.13 New Zealand

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	100	100	100	100
Classes	200	199	199	100
Teachers	200	199	199	100
Students	5400 *		5218	

* Approximate

Instrument	N	% of Achieved Sample
School Questionnaire	100	100
Teacher Background and Attitudes	189	95
Teacher Opportunity to Learn Form Core	175	88
Form A	170	85
Form B	169	85
Form C	169	85
Form D	168	84
Student Background and Attitudes	5218	100
Student Cognitive Form Core	5176	99
Form A	1297	99
Form B	1319	100
Form C	1303	100
Form D	1294	99

4.14 Nigeria

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	67	67	48	72
Classes	67	67	48	72
Teachers	67	67	48	72
Students		2010	1456	72

Instrument	N	% of Achieved Sample
School Questionnaire	48	100
Teacher Background and Attitudes	45	95
Teacher Opportunity to Learn Form Core	30	62
Form A	31	65
Form B	30	62
Form C	30	62
Form D	31	65
Student Background and Attitudes	1456	100
Student Cognitive Form Core	1414	97
Form A	359	99
Form B	359	99
Form C	384	100
Form D	349	96

4.15 Ontario

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	130	130	112	86
Classes	210	210	183	87
Teachers	210	210	183	87
Students	5050		5013	

Instrument	N	% of Achieved Sample
School Questionnaire	112	100
Teacher Background and Attitudes	173	95
Teacher Opportunity to		
Learn Form Core	160	87
Form A	160	87
Form B	159	87
Form C	159	87
Form D	157	86
Student Background and Attitudes	4885	97
Student Cognitive Form Core	4666	93
Form A	1183	94
Form B	1179	94
Form C	1165	93
Form D	1174	94

4.16 Scotland

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools			76	
Classes *			4563	
Teachers			354	
Students	2021		1356	67

* Intact classes not sampled - follow-up sample

Instrument	N	% of Achieved Sample
School Questionnaire	76	100
Teacher Background and Attitudes	354	100
Teacher Opportunity to		
Learn Form Core	Instruments	
Form A	not	
Form B	administered	
Form C	in	
Form D	Scotland	
Student Background and Attitudes	1356	100
Student Cognitive Form Core	1320	97
Form A	344	100
Form B	339	100
Form C	336	99
Form D	337	99

4.17 Swaziland

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	25	25	25	100
Classes	25	25	25	100
Teachers	25	25	25	100
Students			904	

Instrument	N	% of Achieved Sample
School Questionnaire	25	100
Teacher Background and Attitudes	25	100
Teacher Opportunity to Learn Form Core	24	96
Form A	24	96
Form B	23	92
Form C	24	96
Form D	24	96
Student Background and Attitudes	904	100
Student Cognitive Form Core	817	89
Form A	412	91
Form B	405	90
Form C	399	88
Form D	409	90

Each student took 2 rotated forms so the expected sample for each rotated form is 452.

4.18 Sweden

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	100	100	96	96
Classes	200	200	188 *	94
Teachers	200	200	186	93
Students	4020	4067	3585	88

* Includes 2 pseudo classes.

Instrument	N	% of Achieved Sample
School Questionnaire	96	100
Teacher Background and Attitudes	186	100
Teacher Opportunity to Learn Form Core	180	97
Form A	174	94
Form B	177	95
Form C	177	95
Form D	176	95
Student Background and Attitudes	3585	100
Student Cognitive Form Core	3451	96
Form A *	1659	92
Form B *	1689	94
Form C *	1664	93
Form D *	1691	94

* 2 rotated forms per student administered, thus expected number for each form is 50% of 3585

4.19 Thailand

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Schools	100	100	99	99
Classes	100	100	99	99
Teachers	100	100	99	99
Students	4233	4233	4023	95

Instrument	N	% of Achieved Sample
School Questionnaire	99	100
Teacher Background and Attitudes	99	100
Teacher Opportunity to Learn Form Core	90	91
Form A	90	91
Form B	90	91
Form C	90	91
Form D	90	91
Student Background and Attitudes	3821	95
Student Cognitive Form Core	3824	95
Form A	937	93
Form B	939	93
Form C	965	96
Form D	971	97

4.20 USA

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate%
Districts	70	185	93	50.3
Schools	125	180	150	83.3
Classes	250	360	280	77.8
Teachers	250	360	280	77.8
Students	5,000	9,000	6,858	76.2

* At this level. See section 6.20

Instrument	N	% of Achieved Sample
School Questionnaire	157	100
Teacher Background and Attitudes	276	99
Teacher Opportunity to Learn Form Core	269	96
Form A	269	96
Form B	269	96
Form C	268	96
Form D	267	95
Student Background and Attitudes	6683	97
Student Cognitive Form Core	6648	97
Form A	1692	100
Form B	1653	99
Form C	1695	100
Form D	1649	99

RESPONSE RATES - POPULATION B

Almost all National Centers chose to sample one intact class per school. In most countries a relatively small proportion of the age cohort takes mathematics at the advanced level defined for Population B. Thus although the executed and achieved samples fell well short of the designed sample as approved by the Sampling Referee, the achieved sampling fractions are still high. Comments for Population A (Section 4) are also applicable for Population B.

The general level of response rates for schools/classes are:

Response Rate	No of Countries
≥ 90%	9
80% - 89%	3
70% - 79%	2
60% - 69%	

5.1 Belgium (Flemish)

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	150	150	131	87
Classes			197	
Teachers			197	
Students			2859	

Instrument	N	% of Achieved Sample
School Questionnaire	131	100
Teacher Background and Attitudes	180	91
Teacher Opportunity to		
Learn Form 1	193	98
Form 2	193	98
Form 3	193	98
Form 4	193	98
Form 5	193	98
Form 6	193	98
Form 7	193	98
Form 8	193	98
Student Background and Attitudes	2858	100
Student Cognitive Form 1	716	100
Form 2	714	100
Form 3	723	100
Form 4	702	98
Form 5	714	100
Form 6	713	100
Form 7	721	100
Form 8	706	99

5.2 Belgium (French)

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	152	113	87	77
Classes			153	
Teachers			151	
Students			2062	

Although the executed sample is considerably smaller than the designed sample it should be noted that the achieved sampling fraction for schools is 0.19.

Instrument	N	% of Achieved Sample
School Questionnaire	87	99
Teacher Background and Attitudes	151	100
Teacher Opportunity to		
Learn Form 1	Not	
Form 2	administered	
Form 3	in	
Form 4	Belgium	
Form 5	(French)	
Form 6		
Form 7		
Form 8		
Student Background and Attitudes	2018	98
Student Cognitive Form 1	508	99
Form 2	490	95
Form 3	502	97
Form 4	503	98
Form 5	505	98
Form 6	487	94
Form 7	505	98
Form 8	507	98

5.3 British Columbia

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools			78	
Classes	105	105	95	90
Teachers	105	105	95	
Students			1954	

Instrument	N	% of Achieved Sample
School Questionnaire	88	100
Teacher Background and Attitudes	95	100
Teacher Opportunity to		
Learn Form 1	93	98
Form 2	93	98
Form 3	93	98
Form 4	93	98
Form 5	92	97
Form 6	90	95
Form 7	92	97
Form 8	94	99
Student Background and Attitudes	1948	100
Student Cognitive * Form 1	241	99
Form 2	248	100
Form 3	236	97
Form 4	244	100
Form 5	247	100
Form 6	240	98
Form 7	239	98
Form 8	233	95

* Each student took 1 rotated form so the expected number of students per form is 244.

5.4 England and Wales

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	399	346	312	90
Classes			-	
Teachers			678	
Students	3996	3703	3578	

Instrument	N	$ of Achieved Sample
School Questionnaire	312	100
Teacher Background and Attitudes	613	90
Teacher Opportunity to		
Learn Form 1	507	75
Form 2	502	74
Form 3	500	74
Form 4	503	74
Form 5	495	73
Form 6	497	73
Form 7	496	73
Form 8	492	73

Instrument	N	% of Achieved Sample
Student Background and Attitudes	3436	96
Student Cognitive Form 1	842	98
Form 2	848	99
Form 3	868	100
Form 4	850	99
Form 5	849	99
Form 6	857	100
Form 7	847	99
Form 8	836	97

Sampling was of random selection of students within schools so several teachers per school received questionnaires. Thus although not all teachers completed the teacher Opportunity-to-Learn questionnaires, good Opportunity-to-Learn data is available for all but 3 schools.

5.5 Finland

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	88	88	81	92
Classes	88	88	81	92
Teachers	88	88	81	91
Students	1632	1759	1550	88

Instrument	N	% of Achieved Sample
School Questionnaire	81	100
Teacher Background and Attitudes	81	100
Teacher Opportunity to Learn Form 1	76	94
Form 2	76	94
Form 3	76	94
Form 4	76	94
Form 5	76	94
Form 6	76	94
Form 7	76	94
Form 8	76	94
Student Background and Attitudes	1550	100
Student Cognitive Form 1	379	98
Form 2	379	98
Form 3	381	98
Form 4	373	96
Form 5	378	98
Form 6	369	95
Form 7	371	96
Form 8	376	97

5.6 Hong Kong

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools			112	
Classes *	150 approx.	150	125	83%
Teachers			125	
Students			3294	

* Intact classes sampled directly.

Achieved sampling fraction (classes) = 0.18

Instrument	N	% of Achieved Sample
School Questionnaire	112	100
Teacher Background and Attitudes	125	100
Teacher Opportunity to		
Learn Form 1	No	-
Form 2	data	-
Form 3	returned	-
Form 4	from	-
Form 5	National	-
Form 6	Center	-
Form 7		-
Form 8		-
Student Background and Attitudes	3294	100
Student Cognitive Form 1	815	99
Form 2	814	99
Form 3	817	99
Form 4	816	99
Form 5	820	100
Form 6	799	97
Form 7	803	98
Form 8	791	96

5.7 Hungary

Level	Designed Sample	Executed Sample*	Achieved Sample	Response Rate
Schools	75	92	92	100
Classes	78	95	95	100
Teachers	78	95	94	100
Students	2009	2540	2455	97

* Some cells of sampling frame oversampled to enable between stratum comparisons.

Instrument	N	% of Achieved Sample
School Questionnaire	92	100
Teacher Background and Attitudes	94	100
Teacher Opportunity to		
Learn Form 1	90	96
Form 2	90	96
Form 3	90	96
Form 4	90	96
Form 5	90	96
Form 6	90	96
Form 7	90	96
Form 8	90	96
Student Background and Attitudes	2443	100
Student Cognitive Form 1	649	100
Form 2	589	96
Form 3	587	96
Form 4	599	98
Form 5	610	99
Form 6	689	100
Form 7	529	86
Form 8	612	100

5.8 Israel

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	96	92	64	70
Classes		*	108	
Teachers			108	
Students		2650	1905	72

* Number of classes per school chosen dependent in size of school. Exact number not known at International Center.

Instrument	N	% of Achieved Sample
School Questionnaire	64	100
Teacher Background and Attitudes	82	76
Teacher Opportunity to		
Learn Form 1	79	73
Form 2	79	73
Form 3	79	73
Form 4	78	72
Form 5	78	72
Form 6	76	70
Form 7	77	71
Form 8	77	71

Instrument		N	% of Achieved Sample
Student Background and Attitudes		1810	95
Student Cognitive	Form 1	420	88
	Form 2	411	86
	Form 3	424	89
	Form 4	421	88
	Form 5	433	91
	Form 6	415	87
	Form 7	416	87
	Form 8	410	86

5.9 Japan

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	220	207	192	93
Classes	220	207	207 *	100
Teachers	220	207	207	100
Students	8200	7982	7954	100

* Two classes chosen in some schools.

Instrument		N	% of Achieved Sample
School Questionnaire		192	100
Teacher Background and Attitudes		207	100
Teacher Opportunity to Learn	Form 1	200	97
	Form 2	201	97
	Form 3	201	97
	Form 4	201	97
	Form 5	200	97
	Form 6	200	97
	Form 7	201	97
	Form 8	199	96
Student Background and Attitudes		7954	100
Student Cognitive	Form 1	1986	100
	Form 2	1970	99
	Form 3	1995	100
	Form 4	1999	100
	Form 5	1994	100
	Form 6	1982	100
	Form 7	1994	100
	Form 8	1988	100

5.10 New Zealand

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	80	80	79	99
Classes	80	80	79	99
Teachers	80	80	79	99
Students	1200 (approx)	1214	1193	98

Instrument	N	% of Achieved Sample
School Questionnaire	79	100
Teacher Background and Attitudes	79	100
Teacher Opportunity to Learn Form 1	78	99
Form 2	78	99
Form 3	78	99
Form 4	78	99
Form 5	78	99
Form 6	78	99
Form 7	78	99
Form 8	78	99
Student Background and Attitudes	1186	99
Student Cognitive Form 1	304	100
Form 2	296	99
Form 3	279	94
Form 4	280	94
Form 5	288	97
Form 6	294	99
Form 7	304	100
Form 8	284	95

5.11 Ontario

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	85	85	79	93
Classes				86
Teachers	245	245	210	86
Students	3000 (approx)		3214	

Instrument	N	% of Achieved Sample
School Questionnaire	79	100
Teacher Background and Attitudes	187	89
Teacher Opportunity to		
Learn Form 1	194	92
Form 2	197	94
Form 3	192	91
Form 4	194	92
Form 5	196	93
Form 6	194	92
Form 7	195	93
Form 8	190	90
Student Background and Attitudes	3190	99
Student Cognitive Form 1	699	87
Form 2	716	89
Form 3	682	85
Form 4	692	86
Form 5	713	89
Form 6	694	86
Form 7	732	91
Form 8	715	89

5.12 Scotland

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	67	67	54	81
Classes *				
Teachers			272	
Students	1700 (approx)		1501	

* Sampling not be intact classes.

Instrument	N	% of Achieved Sample
School Questionnaire	54	100
Teacher Background and Attitudes	218	80
Teacher Opportunity to		
Learn Form 1	Instrument	
Form 2	not	
Form 3		
Form 4	administered	
Form 5		
Form 6		
Form 7		
Form 8		
Student Background and Attitudes	1501	
Student Cognitive Form 1	373	99
Form 2	367	98
Form 3	373	99
Form 4	368	98
Form 5	364	97
Form 6	379	100
Form 7	371	99
Form 8	371	99

5.13 Sweden

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	129	129	127	98
Classes	129	130	134 *	
Teachers	129	129	127	98
Students	2999	2929	2712	93

* Some classes split into pseudo-classes on the basis of course.

Instrument	N	% of Achieved Sample
School Questionnaire	127	100
Teacher Background and Attitudes	127	100
Teacher Opportunity to		
Learn Form 1	124	98
Form 2	123	97
Form 3	124	98
Form 4	124	98
Form 5	124	98
Form 6	124	98
Form 7	124	98
Form 8	124	98
Student Background and Attitudes	2712	100
Student Cognitive Form 1	622	92
Form 2	609	90
Form 3	609	90
Form 4	623	92
Form 5	619	91
Form 6	638	94
Form 7	612	90
Form 8	626	92

5.14 Thailand

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Schools	64	64	64	100
Classes	107	107	107	100
Teachers	107	107	107	100
Students	4150	4150	3747	90

Instrument	N	% of Achieved Sample
School Questionnaire	64	100
Teacher Background and Attitudes	107	100
Teacher Opportunity to		
Learn Form 1	100	93
Form 2	99	93
Form 3	98	92
Form 4	99	93
Form 5	99	93
Form 6	98	92
Form 7	98	92
Form 8	98	92
Student Background and Attitudes	3747	100
Student Cognitive Form 1	945	100
Form 2	935	100
Form 3	959	100
Form 4	930	99
Form 5	931	99
Form 6	916	98
Form 7	934	100
Form 8	920	98

5.15 USA

Level	Designed Sample	Executed Sample	Achieved Sample	Response Rate
Districts	70	194	93	47.9
Schools	125	216	150	69.4
Classes	250	303	252	83.2
Teachers	250	303	252	83.2
Students	5,000	6,060	4,671	77.1

Instrument	N	% of Achieved Sample
School Questionnaire	150	69
Teacher Background and Attitudes	250	83
Teacher Opportunity to		
Learn Form 1	250	99
Form 2	250	99
Form 3	250	99
Form 4	250	99
Form 5	250	99
Form 6	250	99
Form 7	249	99
Form 8	249	99
Student Background and Attitudes	4643	99
Student Cognitive Form 1	1129	97
Form 2	1138	98
Form 3	1138	98
Form 4	1148	99
Form 5	1157	100
Form 6	1141	98
Form 7	1116	96
Form 8	1143	98

* National Center estimates
School districts over sampled to allow for refusals. Cooperation rate at district level of the order of 50%

6 REPRESENTATIVENESS OF SAMPLES - POPULATION A

In this and the next sections certain characteristics of the samples are examined in order to assist in judging the representativeness of the samples. Cross-national studies pose particular problems in this respect. Variables defined for international purposes do not necessarily match comparable within country variables which are usually used as marker variables. An example of this is the variable Father's Occupation. For the purposes of the study instructions were issued as to how national centres should go about classifying these to form scales which might allow between country comparisons. Thus most national centers had to adapt existing national scales or, in some cases, create a coding system appropriate to the IEA scale. Comparison of the IEA occupational scale with results for particular countries, where often the occupational classification system is not intended as a SES scale, then becomes almost meaningless. It is also difficult to obtain statistics on some (proposed) marker variables from some countries.

Below, each system is considered in turn and what relevant information is available is presented. For certain systems where loss of data, lower response rates or sample attrition indicated a possible problem with representativeness special efforts to obtain marker variable data were made and extended reports are given for these. In general, the methods by which national centers carried out sampling and data collections, and good response rates, ensured that the samples were representative.

Some of the marker variables for which results are presented for Population A include:

i Gender Distribution - Students. For almost all systems virtually 100% of students are in school and form the (Population A) population at this level. The expected proportion for each gender is thus approximately 50% with the caveat that excluded populations which have a preponderance of students of one gender may cause a deviation from this.

ii Student Age. Early in the Study national centers supplied figures for the distribution of 13 year olds across grades. The purpose of this was to enable the Sampling Referee to ensure that the target grade chosen was in keeping with the international population definition. Data from the Study gave age distribution within grade. A reasonable comparison between distributions (making some strong assumptions), might have been possible if the statistics supplied by the national centers had been gathered at the same time of year as IEA data collection took place. This was not the case. Age comparisons are thus useful only in providing an assurance that the correct grade (in terms of the population definition) was tested.

iii Father's Occupation. For some countries it was possible to obtain the proportion of males in various classifications of occupations. These can be used to give comparisons of trends but congruence should not be expected for two major reasons. First, the distribution of occupations for all males is likely to be significantly different from the distribution of males that are fathers of 13 year old students. Second, classifications of occupations for individual countries only approximate those for the IEA study.

Most of the occupational group statistics are taken from the Yearbook of Labour Statistics 1983, International Labour Office, Geneva.

Occupational groups have been combined to give an approximation to the IEA classifications as follows:

IEA Classification		ILO Category
1 Professional and Managerial	(1	Professional, Technical and Related Workers
	(2	Administrative and Managerial Workers
2 Clerical and Sales	(3	Clerical and Related Workers
	(4	Sales Workers
	(5	Service Workers
3 Skilled Workers)	(6	Agriculture, Animal Husbandry and Forestry Workers, Fishermen and Hunters
4 Unskilled Workers)	(7	Production and Related Workers, Transport Equipment Operators and Labourers

iv Sundry Variables. For a few systems data on other variables which provided reasonable checks on the sample were able to be obtained and are included for these systems.

Most data supplied by national centers with sampling plans or as part of the National Case Study material came from annual collections of education statistics undertaken by ministries of education or other departments of government. These were referred to by national centers as Official Statistics etc and in many cases there is no reference to the title of the publication from which they are taken.

In addition to the information above, for each system the distribution of responses to two teacher questionnaire items from the Study are presented. The first of these items asked teachers to judge whether their target class was lower, about the same or higher in average ability than other comparable classes in the school. In a system in which streaming or setting is widely employed it could be expected that similar proportions of teachers would choose "lower" and "higher". In systems in which streaming is rare the same result could be expected. Where systems have a mixture of streaming practices - ie some schools streaming and some not, it can be expected that greater proportions of teachers will choose "lower" than "higher" since providing for special or remedial mathematics classes is more common than providing for accelerated classes. It is therefore suggested that for a system with a high proportion of teachers choosing "higher" relative to the proportion choosing "lower" there is possible bias.

The second item asked teachers to judge how many students in the target class would rate in the top one-third of students nationally, how many in the middle one-third, how many in the bottom one-third, and for how many students they were unable to judge. When the data are aggregated to national level, assuming perfect judgment on the part of teachers, equal numbers in the "top", "middle" and "bottom" thirds would be expected. In fact the proportion of students judged to be in the "middle one-third" was much greater than proportions in the other "one-third" categories, perhaps because of the pervasive influence of the normal curve. It was also most common across countries for higher proportions to be judged to be in the bottom one-third than the top one-third but although it can be assumed that there will be national differences in teacher response to this item the data can still be regarded as an indicator of sample representativeness. Where an unduly high proportion of students is judged to be in the "top one-third" in relation to students in the "bottom one-third" there is a suggestion of possible upward achievement bias in the sample.

6.1 Belgium (Flemish) A

6.1.1 Gender Distribution - Students

	IEA Sample	
Male	47.6	All students at this grade level
Female	52.4	take Population A mathematics.

6.1.2 Student Age

IEA Sample Mean 14.2 years at post-test. At the middle of the school year the modal age would thus lie between 13 years and 14 years.

6.1.3 Teacher Judgment of Ability of Class (Percent)

No Other Class	Lower	About the Same	Higher
9	20	54	16

Incidence of Streaming/Setting : 27% of schools

6.1.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
10	29	42	19

6.2 Belgium (French) A

6.2.1 Gender Distribution - Students

	IEA Sample
Male	53.4
Female	46.6

All students at this grade level take Population A mathematics.

6.2.2 Student Age

IEA Sample Mean 14.5 years at post-test. This is somewhat higher than the Belgium (Flemish) mean and in part results from slightly differing grade retention practices.

6.2.3 Teacher Judgment of Ability of Class (Percent)

No Other Class	Lower	About the Same	Higher
2	37	51	11

6.2.4 Teacher Judgment of Student Ability (Percent)

Item not included.

6.3 British Columbia

6.3.1 Gender Distribution - Students

	IEA Sample	Grade Population*
Male	49.7	51.1
Female	50.3	48.9

* National Enrolment Figures, Sept 1977, Ministry of Education.

6.3.2 Student Age

IEA Sample Mean	14.0 years	at testing (May)
Grade Population Mean	13.5 years	at official Ministry data collection.

Assuming official Ministry data collection early in the school year, while IEA testing was towards the end of the school year, these mean values are not inconsistent. Standard deviations for both age distributions were of the order of 6 months.

6.3.3 Occupational Groups (Percent)

IEA	1		2		3+4	
ILO 1981		1+2		3+4+5		6+7
	37	23	10	27	54	50

Note: The ILO figures are for all Canada.

6.3.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
30	0	5	65

Incidence of Streaming/Setting : 70% of schools.

6.3.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
6	21	42	31

6.3.6 Possible Bias of Sample

Where principals or department heads selected classes it is likely that they tended to choose average or higher ability classes.

Three cognitive items used in a British Columbia province-wide assessment in 1981 were very similar to those used in the Second IEA Mathematics Study (there was a difference in the alternatives) and two others were close enough to be comparable. The mean percent correct for these items was 71.8 in the province-wide assessment and 75.6 in the IEA study.

It is thus very probable that the British Columbia Population A sample was biased upwards.

6.4 England and Wales

6.4.1 Gender Distribution

	IEA Sample	13 year old Population*
Male	46.0	51.3
Female	54.0	48.7

* As at 31 August 1979. School Leavers and Examinations, DES, London, and Statistics of Education in Wales, No 5, 1980, Welsh Office, Cardiff.

Note: i Comparison group is of 13 year olds, not third form.

ii The lower than representative proportion of boys in the sample is probably due to higher refusal rate from boys' schools. One of the stratifying variables was school type so weighting would have adjusted for this.

6.4.2 Student Age

IEA Sample mean 14.1 years at testing. In the middle of the school year the modal age would thus have been between 13 years and 13 years 11 months, as required by the population definition. No comparative population statistics available at the International Center.

6.4.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
2	45	20	34

6.4.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
2	30	37	30

6.4.5 Possible Bias of Sample

i For 21 schools (622 students) no stratum number was supplied. Most of these schools had apparently changed stratum during the course of the study and the England and Wales National Center was unable to, or preferred not to, allocate a stratum number. These schools were deleted from the sample because they were unable to be included in the weighting calculations.

The mean of the 40 item core test for these 622 students is 51.0 compared with a mean of 49.3 for the accepted IEA sample. Differences in percent correct for individual items ranged from 6.8 in favor of the rejected group to 4.2 in favor of the IEA sample. In general differences were small. Thus the loss of students who could not be assigned strata may have given a small downward bias to the IEA sample.

ii The intended Population A sample was 133 schools. Of a total of 248 schools which had to be invited to participate in order to achieve this target, 64 did not reply and 47 refused. Refusals and non-reply occurred across strata and while there were some differences in per strata proportions of refusal/ non-reply, no strata were eliminated. However, the relative within strata characteristics of the schools which refused or did not reply is not known.

Since this sampling procedure might be expected to result in bias through schools less confident of their students performing well refusing to participate, a more detailed examination of marker variables is included as Appendix 1. The material included above and in Appendix 1 does not indicate likelihood of upward bias in achievement.

6.5 Finland

6.5.1 Gender Distribution - Students

	IEA Sample	Grade Population
Male	52.4	All students in Population A
Female	47.6	

6.5.2. Student Age

IEA Sample mean 13.8 years at post-test.

6.5.3 Regional Distribution of Sample (Percentages)

Province	Schools		Students	
	Grade Population	Sample	Grade Population	Sample
Ilusimaa	17.6	19.4	20.7	20.5
Turku and Pori	12.8	11.2	13.3	14.5
Häme	12.3	13.3	13.4	12.5
Kymi	4.5	6.1	3.7	5.4
Bohjois-Karjala	4.0	5.1	4.6	6.9
Mikkeli	4.9	7.1	7.2	5.0
Vaasa	7.8	5.1	5.0	5.2
Keski-Suomi	5.8	6.1	5.0	5.2
Kuopio	5.8	4.1	5.5	3.2
Oulu	9.9	10.2	9.6	10.3
Lappi	6.3	6.1	5.0	5.5
Swedish Speaking Schools	6.1	5.1	4.9	4.5

6.5.4 Occupational Groups

IEA ILO 1980	1	1+2	2	3+4+5	3+4	6+7
	8	25	14	39	78	59

Note: ILO figures for Finland include both sexes.

6.5.5 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
25	22	45	8

Incidence of Streaming/Setting : 92% of schools.

6.5.6 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
6	39	39	17

6.6 France

6.6.1 Gender Distribution

IEA Sample Population 1979-80

Male	43.5	46.2	At the end of grade 7 older boys are commonly switched to technical education while girls remain in general education.
Female	56.5	53.8	

6.6.2 Student Age

IEA Sample Mean 14.2 years at post-test. (May)

Grade Population* Mean 13.8 years at date of official statistics collectior

* France 1978-79 Official Statistics (Ministry) 1980. Age is at 1.1.79.

Students between 13 years and 13 years 11 months are fairly equally split between grades 4e and 5e at the middle of the school year. The higher of the two grade levels (4e) was chosen on the basis of curricular fit to the tests.

6.6.3 Teacher Gender

	IEA Sample	Grade Population Teachers*
Male	51.7	53.2
Female	48.3	46.8

*, France 1979-80 Official Statistics (Ministry) 1980.

6.6.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
2	21	50	27

Incidence of Streaming/Setting : 15% of schools.

6.6.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
16	26	43	15

6.6.6 Because of grade repeating in France prior to the testing year the target grade contains students who have made normal progress through the grades, students who have repeated a year and, in some cases, students who had repeated two years.

6.7 Hong Kong

6.7.1 Gender Distribution - Student

	IEA Sample	Grade Population*
Male	50.9	50.9
Female	49.1	49.1

* Figures supplied by Hong Kong Education Department statistics section.

6.7.2 Student Age

IEA Sample Mean 13.2 years at post-test.

13 year olds are spread across several grades in Hong Kong. The grade selected was that which had the greatest number of 13 year olds by the middle of the school year.

6.7.3 Occupational Groups

IEA ILO 1981	1	1+2	2	3+4+5	3+4	6+7
	12	9	12	38	76	53

Note: ILO figures for Hong Kong include both sexes.

6.7.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
0	24	64	13

Incidence of Streaming/Setting : 23% of schools.

6.7.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
12	38	37	13

6.8 Hungary

6.8.1 Gender Distribution - Student

	IEA Sample	
Male	48.2	100% of students in school and taking mathematics at this level.
Female	51.8	

6.8.2 Student Age

IEA Sample Mean 14.2 years at testing. Modal age at mid-year is less than 14 years.

6.8.3 Occupational Groups

IEA ILO 1980	1	1+2	2	3+4+5	3+4	6+7
	14	13	20	11	66	75

6.8.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
22	34	29	15

Incidence of Streaming/Setting : 0% of schools.

6.8.5 Teacher Judgment of Student Ability (Percent)

Item not administered in Hungary.

6.9 Israel

6.9.1 Gender Distribution

	IEA Sample	Grade Population*
Male	50.9	49.5
Female	49.1	50.5

* Official statistics, 1977.

6.9.2 Student Age

IEA Sample Mean 14.0 years at time of testing. Modal age in the middle of the school year would thus fall within the range quoted in the international population definition. No comparative population data is available at the International Center.

6.9.3 Occupational Groups (Percent)

IEA	1		2		3+4	
ILO 1981		1+2		3+4+5		6+7
	10	23	39	28	51	49

6.9.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
21	34	19	26

Incidence of Streaming/Setting : 71% of schools.

6.9.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
2	35	39	24

6.9.6 Possible Bias in the Sample

There is no indication of bias with respect to the defined population, but it must be recalled that Arabic-speaking schools were not included in the defined population so that with respect to the whole Israel school system the sample is likely to be biased.

6.10 Japan

6.10.1 Gender Distribution

	IEA Sample	Grade Population*
Male	51.5	51.1
Female	48.5	48.9

* Educational Statistics, Japan, 1976 edition; Ministry of Education, Science and Culture.

6.10.2 Student Age

At the time of the post-test mean student age was 13.5 years. 91.2% of the sample were aged between 13 and 14 years. This is consistent with there being no grade repeating in Japan.

6.10.3 Teacher Gender

	IEA Sample	Grade (Teacher) Population*
Male	77.4	70.1
Female	22.6	29.9

* Full-time teachers, grade 7. Educational Statistics, Japan, 1976 edition.

6.10.4 Class Size

IEA Sample		Educational Statistics, Japan 1976	
Interval	% of classes	Interval	% of classes
29-36	11.0	31-35	10.0
37-40	27.1	36-40	28.9
41-44	44.3	41-45	46.5

Note: Intervals are different.

6.10.5 Occupational Groups

Because of sensitivity about this type of item in Japan no response was received from 43% of the sample..

6.10.6 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
8	27	62	3

Incidence of Streaming/Setting : less than 2% of schools.

6.10.7 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	30	38	29

6.11 Luxembourg

6.11.1 Gender Distribution - Students

	IEA Sample	
Male	49.3	All students in this level
Female	50.7	in Population A.

6.11.2 Student Age

IEA Sample Mean 14.5 years at post-test.
At mid-year 13 year olds are divided fairly evenly between two grades. The higher grade was chosen on the basis of curricular fit of the IEA items.

6.11.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
10	24	54	11

Incidence of Streaming/Setting : 38% of schools.

6.11.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
8	35	43	13

6.12 The Netherlands

6.12.1 Gender Distribution - Students

	IEA Sample	
Male	50.9	All students in school types
Female	49.1	sampled take mathematics.

6.12.2 Student Age

IEA Sample Mean 14.4 years at testing.

At about the middle of the school year ages are distributed as follows in the grades AE7 and AE8*

	12 years	13 years	14 years	Other
AE7	52.3%	37.2%	8.5%	2.0%
AE8	0.2%	45.2%	39.0%	15.5%

AE8 was chosen on the basis of curricular fit of the IEA tests.

* Official Statistics 1978/79

6.12.3 Occupational Groups (Percent)

IEA ILO 1979	1	1+2	2	3+4+5	3+4	6+7
	21	21	25	40	55	39

Note: ILO figures for the Netherlands include both sexes.

6.12.4 Teacher Judgment of Class Ability (Percent)

Item not administered in the Netherlands.

6.12.5 Excluded Population

There is no indication of bias (that cannot be corrected by weighting) with respect to the defined population. With respect to the total AE8 population, however, there is an upward achievement bias. Students in the excluded population are, in general, of lower ability than those in the IEA population and the excluded population is approximately 20% of the age group.

6.13 New Zealand

6.13.1 Gender Distribution - Student

	IEA Sample	Grade Population*
Male	50.5	50.8
Female	49.5	49.2

* Educational Statistics, Department of Education, 1981.

6.13.2 Student Age

IEA Sample Mean 14.0 at time of post-test (Nov)
Population Mean 13.7 at 1 July.

6.13.3 Occupational Groups

IEA Elley-Irving SES Scale	1	1+2	2	3	3	4	4	5+6
	24	14	27	27	29	29	20	30

Note: The Elley-Irving SES Scale is New Zealand developed but figures are for all males in the work force.

It is of interest to compare the ILO/IEA ratings.

IEA	1		2		3+4	
ILO		1+2		3+4+5		6+7
	24	18	27	23	49	62

6.13.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
<1	30	45	25

Incidence of Streaming/Setting : 75% of schools

6.13.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	30	45	21

6.14 Ontario

6.14.1 Gender Distribution - Students

	IEA Sample
Male	50.2
Female	49.8

All students are in school at this level and are taking Population A mathematics.

6.14.2 Student Age

IEA Sample Mean 13.4 years at post-test.
Modal age would be between 13 years and 14 years at mid-year.

6.14.3 Occupational Group (Percent)

IEA	1		2		3+4	
ILO 1981		1+2		3+4+5		6+7
	17	23	21	27	63	50

Note: The ILO figures are for all Canada.

6.14.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
24	8	59	9

Incidence of Streaming/Setting : 23% of schools

6.14.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
6	28	46	20

6.15 Nigeria

6.15.1 Gender Distribution - Students

	IEA Sample
Male	72.8
Female	27.2

The enrolment rate is low in Nigeria and since mathematics is compulsory for all students in Nigerian secondary schools it is apparent that the enrolment rate is much higher for boys than for girls. In the states which participated in the Study enrolment rates ranged from 180.8 per 10 000 of state population to 391.2 (British Council, 1979, Education Profile : Nigeria, London: British Council).

6.15.2 Student Age

IEA Sample Mean 16.7 years at testing.

The ages of Form 3 students in Nigeria range from 12 years to over 20 years. The grade was chosen on the basis of curricular fit rather than by age definition.

6.15.3 Teacher Judgement of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
14	22	58	5

Incidence of Streaming/Setting : 26% of schools.

6.15.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	22	35	40

Note: The population for this Study was confined to eight southern states. All ten southern states were in the designed sample. Although only approximately 50% of the population of Nigeria lives in the south, approximately 90% of the enrolment of secondary grammar/commercial schools is in these states. The 8 states remaining in the study have some 80% of the enrolment. However, low response rates and some doubt by the national center about the accuracy of coding and punching makes the representativeness of the sample, even for the 8 states defining the population, open to question.

6.16 Scotland

6.16.1 Gender Distribution

	IEA Sample	
Male	53.8	All students at this level
Female	46.2	take Population A mathematics.

6.16.2 Student Age

IEA Sample Mean 14.0 years at testing. The modal age of students at mid-year would thus be between 13 years and 13 years 11 months.

6.16.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
<1	31	33	35

Note: Intact classes were not selected. These figures refer to classes within which students in the sample were treated.

6.16.4 Teacher Judgment of Student Ability (Percent)

Item not administered in Scotland.

6.16.5 Since the sample used was a "follow-up" one there is a necessity to find whether sample attrition had introduced bias. An account of the examination undertaken by Mr G Thorpe, Scottish Council for Research in Education, is included as Appendix 2. The results indicate that the IEA sample is representative of the population.

6.17 Swaziland

6.17.1 Gender Distribution

	IEA Sample	Grade Population*
Male	46.1	50.8
Female	53.9	49.2

* Official Statistics

6.17.2 Students Age

IEA Sample Mean 15.7 years at testing. The target grade in Swaziland contains a wide range of ages. The grade was selected on the basis of curricular fit.

6.17.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
12	0	56	32

Incidence of Streaming/Setting : 8% of schools.

6.17.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
0	23	48	38

6.17.5 Examination Rankings (National Center)

Schools were ranked on their pass rates in external examinations and grouped into three categories on the basis of the rankings. Schools in Population A were distributed: Top group 10 schools; Middle group 8 schools; Bottom group 7 schools.

If the schools are grouped into four groups on the examination success ranking, the distribution is:

Top ¼:	8 schools
Second ¼:	5 schools
Third ¼:	7 schools
Bottom ¼:	5 schools

6.17.6 Possible Bias of Sample

From the above sections upward bias in achievement with respect to the population is indicated.

6.18 Sweden

6.18.1 Gender Distribution - Students

	IEA Sample	
Male	52.4	100% of the age cohort of
Female	47.6	this grade in school.

6.18.2 Student Age

IEA Sample Mean 13.9 years at testing. At mid-year the modal age lies between 13 years and 14 years.

6.18.3 Occupational Groups

IEA	1		2		3+4	
ILO 1981		1+2		3+4+5		6+7
	20	26	30	18	50	56

6.18.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
8	27	53	12

Incidence of Streaming/Setting: 100% of schools.

6.18.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	32	40	24

6.19 Thailand

6.19.1 Gender Distribution - Student

	IEA Sample	
Male	52.0	Approximately 85% (National Center) of the age cohort in school at time of data collection.
Female	48.0	

6.19.2 Student Age

IEA Sample Mean 14.2 years at post-test.
Modal age mid-year is between 13 years and 14 years.

6.19.3 Occupational Groups

IEA	1		2		3+4	
ILO 1980		1+2		3+4+5		6+7
	15	5	27	11	58	85

Note: Approximately 15% of the age cohort are not in schooling at this level. Those not in school can be expected to have fathers at the lower end of the occupational scale.

6.19.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
5	24	50	20

Incidence of Streaming/Setting: 49% of schools.

6.19.5 Teacher judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
15	38	33	14

6.20 USA

6.20.1 Gender Distribution - Students

	IEA Sample	
Male	48.1	100% of students in
Female	51.9	school at this level.

6.20.2 Student Age

IEA Sample Mean 14.1 years at post-test. Modal age was between 13 years and 14 years at mid-year.

6.20.3 Occupational Groups (Percent)

IEA ILO 1981	1	1+2	2	3+4+5	3+4	6+7
	16	31	36	21	48	48

6.20.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
5	20	41	33

Incidence of Streaming/Setting: 77% of schools.

6.20.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
2	26	44	28

6.20.6 While there is little indication of bias in the above, relatively low response rates, particularly at school district level, in spite of some replacements being made, called for a more extensive investigation. This is included as Appendix 3. If anything, there is a possibility of upward achievement bias for population A but this would be slight.

7. REPRESENTATIVENESS OF SAMPLES - POPULATION B

For most education systems the best indication of sample representativeness is the care with which the approved sampling methods have been followed and the size of the response rate.

In all systems, except Hungary, the Population B mathematics group is a subset of the grade population. Official statistics for the grade population are available for most systems but usually it is not possible to make useful comparisons between these statistics and the Population B statistics. For example, gender distribution for terminal year students taking mathematics is usually very different from the distribution for all students in the grade because of a tendency for greater numbers of boys than girls to take advanced mathematics in most systems.

Comparison of SES distributions (Father's Occupation, say) for Population B with SES distributions for the total population is not fruitful. The grade population is biased with respect to the total population to an extent determined by the selectivity of the system and it is not uncommon for the distribution for the group taking advanced mathematics to be biased with respect to that for the grade population. Selectivity with respect to both schooling versus non-schooling and mathematics versus non-mathematics for 17 - 19 year olds varies markedly across countries.

In this section of the report comparisons on variables for which available statistics seemed likely to give a reasonable indication of the nature of the sample relative to the population are presented.

Population A teachers were asked to judge the ability of their target class relative to other classes in the school and to judge how many students in the target class would fall into the top, middle and bottom one-thirds of a national ability distribution. National estimates were obtained by aggregation. These judgments were more difficult for teachers of Population B classes because Population B was a subset of the grade population.

Teachers were intended to compare the ability of their mathematics class with the abilities of comparable mathematics classes in the school but cross-tabs of this variable against school size reveal that, especially in some systems, they made a general ability comparison with other subject classes and/or with classes taking less advanced mathematics courses (e.g. in schools with only one Population B class some teachers judged the ability of their target class to be higher than comparable classes in the school).

Similarly, in judging how many of their students fell into each one-third of the national ability distribution there appeared to be a tendency to use general ability for the grade as a criterion in some systems.

As stated above, judgments about sample representativeness depend on more than will be presented in this section, or indeed in this report. To a large extent they are built up over the period of the Study from discussion and correspondence with national research coordinators about step by step progress, and occasionally problems, related to sampling and data collection and to knowledge of the idiosyncracies of the systems being sampled.

In the following country by country summary the amount of relevant information about systems varies. Where there is real doubt about the representativeness of a sample, this is mentioned.

7.1 <u>Belgium (Flemish)</u>

7.1.1. Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
31	19	32	18

7.1.2. The item calling for teacher judgment of the number of students in the target class who would be in the top, middle and bottom one-thirds of a national ability distribution, was not included in the Belgium (Flemish) questionnaire. However, 20% of teachers judged the range of ability of students in their target class to be "very wide" and 61% judged the range to be "fairly wide".

7.1.3 The achieved sample is 22% of the population so given the sampling method and stratification variables utilised, weighting ensures representativeness.

7.2 <u>Belgium (French)</u>

7.2.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
38	15	23	24

7.2.2 The Teacher Judgment of Student Ability item was not administered in Belgium (French). 31% of teachers judged the range of ability of their target class to be "very wide" and 49% judged the range to be "fairly wide".

7.2.3 The achieved sample was 22% of the population. Sampling methods and stratification variables utilised make sampling bias in computed statistics very improbable.

7.3 British Columbia

7.3.1 Gender Distribution - Students

	IEA Sample %	Grade Population *
Male	59.7	60-70% of students taking courses from which Population B is drawn are male.
Female	40.3	

* Summary report of British Columbia Mathematics Assessment, 1981 : A Report to the Ministry of Education, Province of British Columbia.

7.3.2 Student Age

IEA Sample	Mean 17.9 years (at testing)
Grade Population*	Mean 17.5 years (at time of official Ministry data collection)

* National enrolment figures, Sept 30 1977, Form 1 (presumably Ministry of Education, Province of British Columbia).

7.3.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
13	11	43	34

7.3.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
1	22	44	33

7.3.5 The achieved sample is 14% of the population.

7.4 England and Wales

For comparisons with marker variable statistics see Appendix 1.

7.4.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
37	16	27	20

7.4.2 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	21	31	35

Note: Students were not sampled by intact class. These statistics describe teacher perceptions with respect to the classes in which IEA sample students are located.

7.4.3 Loss from the Executed Example

Twenty-four schools (301 students) changed stratum during the course of the study. The National Center was unable to supply stratum numbers for these schools so they could not be included in weighting calculations and hence were deleted from the sample. A comparison on cognitive form means indicates that there is a small downward achievement bias in the achieved sample.

	Means for Students not assigned to strata	Achieved Sample Mean
Form 1	11.68	11.17
Form 2	10.49	10.16
Form 3	9.10	8.70
Form 4	10.89	10.57
Form 5	10.44	9.67
Form 6	10.70	10.46
Form 7	10.62	9.80
Form 8	9.57	9.05

7.4.4 In order to achieve the intended sample of 384 schools, 712 had to be invited to participate. Of these, 156 did not reply and 162 refused to participate. The relative within-strata characteristics of schools which refused to take part or did not reply is not known. The direction of bias, if any, is not known.

7.5 Finland

7.5.1 Student Age

	IEA Sample Distribution at Testing	Grade Population* Distribution autumn term, 1978
16 years	0.1	0.02
17 years	10.1	3.1
18 years	75.3	68.0
19 years	13.2	23.7
20 years+	1.3	5.2

* Official Statistics.

7.5.2 Regional Distribution of Sample (Percentages)

Province	Schools		Students (Pop B)	
	Population	Sample	Population	Sample
Uusimaa	20.2	19.7	21.1	20.6
Turku and Pori	14.1	13.6	15.0	13.1
Häme	12.7	12.3	13.7	12.3
Kyme	7.1	4.9	8.2	4.9
Mikkeli	5.6	4.9	5.0	4.9
Vaasa	7.8	8.7	7.2	8.8
Keski-Suomi	6.3	4.9	5.4	4.7
Kuopio	6.1	7.4	6.6	8.2
Pohjois-Karjala	4.4	4.9	3.6	4.6
Ouli	9.7	12.3	9.9	12.5
Lappi	6.1	6.1	4.3	4.3

7.5.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
63	9	23	5

7.5.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
2	26	40	33

7.6 Hong Kong

7.6.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
50	11	18	21

7.6.2 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
3	28	43	27

7.7 Hungary

7.7.1 Gender Distribution - Students (Percentages)

	IEA Sample	Grade Population*
Male	37.7	41.9
Female	62.3	58.1

* Official statistics, Hungarian Ministry of Culture, 1980/81.

For Hungary the grade population is virtually identical with the national Population B.

7.7.2 Student Age

IEA Sample	Mean 18.1	(at testing)
Grade Population*	Mean 17.6	(beginning of school year)

* Official Statistics, 1980/81, Hungarian Ministry of Culture. The standard deviations for age for the sample and the grade population are both of the order of four months. Assuming that there was about six months between the official Ministry of Culture data collection and IEA testing the means and standard deviations indicate that with respect to age the sample is representative of the population.

7.7.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
1	37	43	19

7.7.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
1	50	40	9

This distribution appears to be a result of teachers in vocational schools judging none of their students to be in the top one-third and teachers in grammar schools being rather conservative in their estimates - probably through taking grammar school achievement as a criterion. 50% of the age cohort formed Population B in Hungary and vocational school students do not follow a pre-university course.

7.8 Israel

7.8.1 Gender Distribution - Student

At this grade level in Israel almost 70% of students are girls but in the Physical Track the proportion of girls is only 37.6%. It is assumed that the majority of students taking extended mathematics courses would be students from the Physical Track.

	IEA Sample	Physical Track*
Male	57.1	62.4
Female	42.9	38.6

* Statistics from National Center.

7.8.2 Student Age

IEA Sample Mean at Testing, 17.9 years.

7.8.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
60	6	16	17

7.8.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
2	22	45	32

7.8.5 Only 65 of the 96 schools in the executed sample returned data. In view of this, and of inconsistencies in the sampling information, it is not possible to be confident that the sample is representative. On the other hand, the achieved sampling fraction (students) was 0.63.

7.9 Japan

7.9.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
5	3	40	51

7.9.2 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	25	34	37

Note: Approximately 23% of the grade cohort takes mathematics so in comparison with all classes and all students at this grade level, given the probability that those students who take mathematics are more able, these judgments are likely to be reasonably sound.

7.10 New Zealand

7.10.1 Gender Distribution - Students

	IEA Sample	Population*
Male	64.0	60.5
Female	36.0	39.5

* Educational Statistics, Department of Education, Wellington, 1982.

7.10.2 Student Age

IEA Sample Mean 17.8 years at testing.
Grade Population* Mean 17.5 years at mid-year.

* Educational Statistics, Department of Education, Wellington, 1982.

7.10.3 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
41	1	20	27

Note: "Comparable classes" was taken to mean Form 7 classes generally, rather than Form 7 mathematics classes. Mathematics tends to be taken by higher ability students.

7.10.4 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
3	26	45	26

7.11 Ontario

Marker variable statistics are taken from Education Statistics Ontario, 1982, Ministry of Education Ontario, 1982.

7.11.1 Gender Distribution - Students (Percentages)

	IEA Sample	Population*
Male	61.4	60.6
Female	38.6	39.4

* Successful Grade 13-level candidates by sex and subject (pure mathematics), 1982.

7.11.2 Teacher Age (Years)

IEA Sample Median 40.0

Secondary Teachers* Median 39.8

* Full-time teachers by age, 1982. Estimate based on gender medians weighted.

7.11.3 Teacher Gender

	IEA Sample	All Secondary Teachers*
Male	79.4	70.2
Female	12.3	29.8

* Full-time Teachers by Age, 1982. It is likely that a greater proportion of male teachers than the all-grade statistics is teaching mathematics at grade 13 level.

7.11.4 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
27	9	56	9

7.11.5 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
4	21	41	35

7.12 Scotland

7.12.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
11	24	36	29

7.12.2 Teacher Judgment of Student Ability (Percent)

Item not administered in Scotland.

The Scottish sample is drawn from two grade cohorts so it is not easy to judge representativeness. Given that the sampling method was appropriate and that there was no stratum in which response rates were not adequate, it is probable that statistics without bias could be constructed for both (grade) sub-populations. For the purposes of this Study the sample has been regarded as being drawn from a single population. Bias due to over-representation of either S5 (grade 11) or S6 (grade 12) students is likely to be negligible.

7.13 Sweden

7.13.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
15	19	45	21

7.13.2 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
1	22	41	36

Given the sampling methods and stratification variables utilised bias is unlikely.

7.14 Thailand

7.14.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
17	35	34	15

7.14.2 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
11	48	31	10

7.14.3 The statistics in the above section imply a downward achievement bias but the sampling methods (which were faithfully executed) and high response rates point to the sample being representative. The fact that Thailand teachers at this level were less experienced (on average) than those of any other system may be relevant.

7.15 USA

For comparisons with marker variable statistics see Appendix 3.

7.15.1 Teacher Judgment of Class Ability (Percent)

No Other Class	Lower	About the Same	Higher
12	13	40	35

7.15.2 Teacher Judgment of Student Ability (Percent)

Unable to Judge	Bottom $\frac{1}{3}$	Middle $\frac{1}{3}$	Top $\frac{1}{3}$
2	16	40	42

The USA national definition for the target population (which is an appropriate match for the international population definition) includes a subset of mathematics classes at grade 12 level. This subset contains classes of higher ability students (notably calculus students) and hence the distributions above. The above statistics should thus not be taken as an indication that the sample is other than representative.

8. DISTRIBUTION OF ROTATED FORMS

The tables below show how national centers distributed rotated forms of the cognitive tests.

For population A there was a core test of 40 items administered to all students and four rotated forms, at least one of which was to be taken by each student.

Procedures which, if followed, ensured random assignment of rotated forms to students were detailed to national centers. Most national centers chose to administer the core test and one rotated form randomly assigned to students. Thus for most countries approximately 25% of the sample took each rotated form.

Table 1 shows the numbers of students taking each combination. C1 is the core test plus rotated form A,C2 the core test and rotated form B and so on.

In each country a small proportion of students took only one form and was absent for the test session where the other was administered.

In Swaziland and Sweden each student took the core test plus two rotated forms and in Nigeria a few students took more than one rotated form.

It can be seen from the table that in each system almost equal proportions of the sample took the appropriate number of test combinations. Furthermore, analysis of test distribution at classroom level (not included here) indicates that approximately equal numbers of rotated forms were assigned in each class/school in each country so that it seems probable that procedures for random assignment were correctly followed.

Table 1.--Number and percent of students in population A who were distributed core and rotation forms of the cognitive test, by country

Form	All 20 participating countries	Belgium (Flemish)	Belgium (French)	British Columbia	Ontario	England & Wales	Finland	France	Hong Kong	Hungary
Students in sample	79,055	3,454	2,086	2,228	5,013	2,678	4,484	8,889	5,548	1,754
Took 1 form only										
Core only	1,644	56	73	105	178	72	[illegible]	219	31	
Rotation form A	349	6	12	7	53	8	[illegible]	57	7	
Rotation form B	364	4	9	7	43	9	24	64	5	
Rotation form C	356	8	9	8	68	11	13	70	7	
Rotation form D	378	12	7	19	49	13	30	70	6	
Took Core and-										
Rotation form A	17,684	761	489	512	1,130	644	1,047	2,031	1,375	441
Rotation form B	17,636	756	479	528	1,136	633	1,071	2,038	1,362	439
Rotation form C	17,611	751	490	520	1,097	633	1,081	2,019	1,360	442
Rotation form D	17,557	749	494	503	1,125	630	1,052	2,010	1,367	432
Took 2 rotation forms										
Forms A and B	14									
Forms A and C	11									
Forms A and D	17									
Forms B and C	11									
Forms B and D	7									
Forms C and D	13									
Took Core and-										
Rotation forms A and B	663									
Rotation forms A and C	680									
Rotation forms A and D	663									
Rotation forms B and C	685									
Rotation forms B and D	697									
Rotation forms C and D	692									
Rotation forms A,B, and C	1									
No cognitive test	1,322	351	24	19	134	25	11	311	28	

Table 1.--Number and percent of students in population A who were distributed core and rotation forms of the cognitive test, by country--Continued

Form	Israel	Japan	Luxem-bourg	Nether lands	New Zea-land	Nigeria	Scot-land	Swazi-land	Sweden	Thai-land	U.S.A.
Students in sample	3,819	8,091	2,106	5,500	5,401	1,465	1,356	904	3,585	3,836	6,858
Took 1 form only											
Core only	217		43	67	127	107		10	91	22	95
Rotation form A	58		5	9	41	10	6		5	2	39
Rotation form B	50		7	11	46	11	14		12	6	42
Rotation form C	49		7	9	37	14	11		9		26
Rotation form D	57		5	16	40	10	5		8	2	29
Took Core and-											
Rotation form A	821	2,041	500	1,344	1,256	313	338	3	45	935	1,653
Rotation form B	846	2,030	497	1,326	1,273	309	325	1	43	933	1,611
Rotation form C	807	2,028	494	1,332	1,266	288	325	1	43	965	1,669
Rotation form D	833	1,992	504	1,349	1,254	302	332		40	969	1,620
Took 2 rotation forms											
Forms A and B								1	13		
Forms A and C						1		2	8		
Forms A and D						2		3	12		
Forms B and C	1					1		1	8		
Forms B and D								1	6		
Forms C and D									13		
Took Core and-											
Rotation forms A and B						3		133	527		
Rotation forms A and C						22		131	527		
Rotation forms A and D						2		139	522		
Rotation forms B and C						29		133	523		
Rotation forms B and D						5		135	557		
Rotation forms C and D						28		131	533		
Rotation forms A,B, and C						1					
No cognitive test	80		44	37	61	2		79	40	2	74

123

Table 1.--Number and percent of students in population A who were distributed core and rotation forms of the cognitive test, by country--Continued

Form	All 20 participating countries	Belgium (Flemish)	Belgium (French)	British Columbia	Ontario	England & Wales	Finland	France	Hong Kong	Hungary
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Took 1 form only										
Core only	2.1	1.6	3.5	4.7	3.6	2.7	2.9	2.5	.6	
Rotation form A	.4	.2	.6	.3	1.1	.3	.5	.6	.1	
Rotation form B	.5	.1	.4	.3	.9	.3	.5	.7	.1	
Rotation form C	.5	.2	.4	.4	1.4	.4	.3	.8	.1	
Rotation form D	.5	.3	.3	.9	1.0	.5	.7	.8	.1	
Took Core and-										
Rotation form A	22.4	22.0	23.4	23.0	22.5	24.0	23.3	22.8	24.8	25.1
Rotation form B	22.3	21.9	23.0	23.7	22.7	23.6	23.9	22.9	24.5	25.0
Rotation form C	22.3	21.7	23.5	23.3	21.9	23.6	24.1	22.7	24.5	25.2
Rotation form D	22.2	21.7	23.7	22.6	22.4	23.5	23.5	22.6	24.6	24.6
Took 2 rotation forms										
Forms A and B	.0									
Forms A and C	.0									
Forms A and D	.0									
Forms B and C	.0									
Forms B and D	.0									
Forms C and D	.0									
Took Core and-										
Rotation forms A and B	.8									
Rotation forms A and C	.9									
Rotation forms A and D	.8									
Rotation forms B and C	.9									
Rotation forms B and D	.9									
Rotation forms C and D	.9									
Rotation forms A,B, and C	.0									
No cognitive test	1.7	10.2	1.2	.9	2.7	.9	.2	3.5	.5	

Table 1.--Number and percent of students in population A who were distributed core and rotation forms of the cognitive test, by country--Continued

Form	Israel	Japan	Luxem-bourg	Nether lands	New Zea-land	Nigeria	Scot-land	Swazi-land	Sweden	Thai-land	U.S.A.
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Took 1 form only											
Core only	5.7		2.0	1.2	2.4	7.3		1.1	2.5	.6	1.4
Rotation form A	1.5		.2	.2	.8	.7	.4		.1	.1	.6
Rotation form B	1.3		.3	.2	.9	.8	1.0		.3	.2	.6
Rotation form C	1.3		.3	.2	.7	1.0	.8		.3		.4
Rotation form D	1.5		.2	.3	.7	.7	.4		.2	.1	.4
Took Core and-											
Rotation form A	21.5	25.2	23.7	24.4	23.3	21.7	24.9	.3	1.3	24.4	24.1
Rotation form B	22.2	25.1	23.6	24.1	23.6	21.1	24.0	.1	1.2	24.3	23.5
Rotation form C	21.1	25.1	23.5	24.2	23.4	19.7	24.0	.1	1.2	25.2	24.3
Rotation form D	21.8	24.6	23.9	24.5	23.2	20.6	24.5		1.1	25.3	23.6
Took 2 rotation forms											
Forms A and B								.1	.4		
Forms A and C						.1		.2	.2		
Forms A and D						.1		.3	.3		
Forms B and C	.0					.1		.1	.2		
Forms B and D								.1	.2		
Forms C and D									.4		
Took Core and-											
Rotation forms A and B						.2		14.7	14.7		
Rotation forms A and C						1.5		14.5	14.7		
Rotation forms A and D						.1		15.4	14.6		
Rotation forms B and C						2.0		14.7	14.6		
Rotation forms B and D						.3		14.9	15.5		
Rotation forms C and D						1.9		14.5	14.9		
Rotation forms A,B and C						.1					
No cognitive test	2.1		2.1	.7	1.1	.1		8.7	1.1	.1	1.1

Table 2 includes comparable statistics for Population B. For Population B there were 8 rotated forms to be randomly assigned to students at the recommended rate of at least 2 per student. The procedures called for all possible combinations (two at a time) to be administered. Thus each rotated form was to be allocated to (at least) one quarter of the sample.

Countries which deviated from this pattern were:

Belgium (Flemish) and Belgium (French) randomly allocated four pairs of rotated forms (1 and 5, 2 and 6, 3 and 7, 4 and 8). There is thus no (sample) link between most combinations.

England and Wales randomly allocated the combinations 1 and 2, 2 and 3, 3 and 4, 4 and 5, 5 and 6, 6 and 7, 7 and 8, and 8 and 1.

Neither of these deviation precludes any analyses (for the purposes of the study) except certain latent trait analyses.

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country

Forms	All 14 participating countries	Belgium (Flemish)	Belgium (French)	Ontario	England & Wales	Finland	Hong Kong	Hungary
Students in sample	40,486	2,852	1,985	2,549	3,307	1,456	3,212	2,417
Forms A and B	1,632			79	424	57	114	116
Forms A and C	1,212			90	1	53	117	107
Forms A and D	1,195			91		52	117	110
Forms A and E	2,380	711	500	115		47	114	99
Forms A and F	1,154	1	1	85		53	119	111
Forms A and G	1,170			86		51	117	103
Forms A and H	1,472	4	2	101	393	53	112	1
Forms B and C	1,605			89	400	51	118	95
Forms B and D	1,110			77		51	117	104
Forms B and E	1,165			94		54	117	88
Forms B and F	2,367	711	481	90		50	115	91
Forms B and G	1,038		1	103		53	113	2
Forms B and H	1,149	1		107	1	52	117	92
Forms C and D	1,631		1	95	436	52	122	96

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	Israel	Japan	New Zealand	Scotland	Sweden	Thailand	U.S.A.
Students in sample	1,622	7,954	1,136	1,478	2,307	3,731	4,480
Forms A and B	57	310	48	50	85	129	163
Forms A and C	61	293	40	49	79	147	175
Forms A and D	60	290	42	55	80	138	160
Forms A and E	57	270	42	53	92	125	155
Forms A and F	61	262	46	54	88	127	146
Forms A and G	57	288	43	56	78	141	150
Forms A and H	56	273	36	55	87	136	163
Forms B and C	59	313	39	51	86	140	164
Forms B and D	53	276	41	51	73	125	142
Forms B and E	63	301	33	50	79	139	147
Forms B and F	54	269	40	56	85	137	188
Forms B and G	60	251	39	52	81	136	147
Forms B and H	51	250	48	57	77	128	168
Forms C and D	54	309	36	56	81	138	155

Table 2.-- Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	All 14 participating countries	Belgium (Flemish)	Belgium (French)	Ontario	England & Wales	Finland	Hong Kong	Hungary
Forms C and E	1,150	1		97		53	118	113
Forms C and F	1,187		1	85	2	52	113	85
Forms C and G	2,235	719	496	87		50	113	1
Forms C and H	1,114	2	2	68		53	109	86
Forms D and E	1,446	1		77	387	54	119	3
Forms D and F	1,122	1		88		52	112	88
Forms D and G	1,162	1	1	101	1	50	112	96
Forms D and H	2,431	698	498	98		50	112	98
Forms E and F	1,661			92	442	52	114	96
Forms E and G	1,176			79		56	115	103
Forms E and H	1,179	1	1	79		52	114	104
Forms F and G	1,602			98	397	52	110	104
Forms F and H	1,078			98		49	107	109
Forms G and H	1,663			100	423	52	115	116

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	Israel	Japan	New Zealand	Scotland	Sweden	Thailand	U.S.A.
Forms C and E	65	247	39	52	70	140	155
Forms D and F	57	315	33	57	80	132	175
Forms C and G	59	228	49	53	89	132	159
Forms C and H	55	290	39	53	86	130	141
Forms D and E	60	270	39	55	88	129	164
Forms D and F	63	247	37	49	80	134	171
Forms D and G	60	271	41	53	85	131	159
Forms D and H	62	336	38	48	86	132	175
Forms E and F	58	333	47	54	87	126	160
Forms E and G	54	291	42	50	75	135	176
Forms E and H	59	282	41	48	84	137	177
Forms F and G	52	332	41	53	90	130	143
Forms F and H	50	224	40	55	79	129	138
Forms G and H	65	333	37	53	77	128	164

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	All 14 participating countries	Belgium (Flemish)	Belgium (French)	Ontario	England & Wales	Finland	Hong Kong	Hungary
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Forms A and B	4.0			3.1	12.8	3.9	3.5	4.8
Forms A and C	3.0			3.5	.0	3.6	3.6	4.4
Forms A and D	3.0			3.6		3.6	3.6	4.6
Forms A and E	5.9	24.9	25.2	4.5		3.2	3.5	4.1
Forms A and F	2.9	.0	.1	3.3		3.6	3.7	4.6
Forms A and G	2.9			3.4		3.5	3.6	4.3
Forms A and H	3.6	.1	.1	4.0	11.9	3.6	3.5	.0
Forms B and C	4.0			3.5	12.1	3.5	3.7	3.9
Forms B and D	2.7			3.0		3.5	3.6	4.3
Forms B and E	2.9			3.7		3.7	3.6	3.6
Forms B and F	5.8	24.9	24.2	3.5		3.4	3.6	3.8
Forms B and G	2.6		.1	4.0		3.6	3.5	.1
Forms B and H	2.8	.0		4.2	.0	3.6	3.6	3.8
Forms C and D	4.0		.1	3.7	13.2	3.6	3.8	4.0

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	Israel	Japan	New Zealand	Scotland	Sweden	Thailand	U.S.A.
Total percent	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Forms A and B	3.5	3.9	4.2	3.4	3.7	3.5	3.6
Forms A and C	3.8	3.7	3.5	3.3	3.4	3.9	3.9
Forms A and D	3.7	3.6	3.7	3.7	3.5	3.7	3.6
Forms A and E	3.5	3.4	3.7	3.6	4.0	3.4	3.5
Forms A and F	3.8	3.3	4.0	3.7	3.8	3.4	3.3
Forms A and G	3.5	3.6	3.8	3.8	3.4	3.8	3.3
Forms A and H	3.5	3.4	3.2	3.7	3.8	3.6	3.6
Forms B and C	3.6	3.9	3.4	3.5	3.7	3.8	3.7
Forms B and D	3.3	3.5	3.6	3.5	3.2	3.4	3.2
Forms B and E	3.9	3.8	2.9	3.4	3.4	3.7	3.3
Forms B and F	3.3	3.4	3.5	3.8	3.7	3.7	4.2
Forms B and G	3.7	3.2	3.4	3.5	3.5	3.6	3.3
Forms B and H	3.1	3.1	4.2	3.9	3.3	3.4	3.8
Forms C and D	3.3	3.9	3.2	3.8	3.5	3.7	3.5

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	All 14 participating countries	Belgium (Flemish)	Belgium (French)	Ontario	England & Wales	Finland	Hong Kong	Hungary
Forms C and E	2.8	.0		3.8		3.6	3.7	4.7
Forms C and F	2.9		.1	3.3	.1	3.6	3.5	3.5
Forms C and G	5.5	25.2	25.0	3.4		3.4	3.5	.0
Forms C and H	2.8	.1	.1	2.7		3.6	3.4	3.6
Forms D and E	3.6	.0		3.0	11.7	3.7	3.7	.1
Forms D and F	2.8	.0		3.5		3.6	3.5	3.6
Forms D and G	2.9	.0	.1	4.0	.0	3.4	3.5	4.0
Forms D and H	6.0	24.5	25.1	3.8		3.4	3.5	4.1
Forms E and F	4.1			3.6	13.4	3.6	3.5	4.0
Forms E and G	2.9			3.1		3.8	3.6	4.3
Forms E and H	2.9	.0	.1	3.1		3.6	3.5	4.3
Forms F and G	4.0			3.8	12.0	3.6	3.4	4.3
Forms F and H	2.7			3.8		3.4	3.3	4.5
Forms G and H	4.1			3.9	12.8	3.6	3.6	4.8

Table 2.--Number and percent of students in population B who were distributed rotation forms of the cognitive test, by country--Continued

Forms	Israel	Japan	New Zealand	Scotland	Sweden	Thailand	U.S.A.
Forms C and E	4.0	3.1	3.4	3.5	3.0	3.8	3.5
Forms C and F	3.5	4.0	2.9	3.9	3.5	3.5	3.9
Forms C and G	3.6	2.9	4.3	3.6	3.9	3.5	3.5
Forms C and H	3.4	3.6	3.4	3.6	3.7	3.5	3.1
Forms D and E	3.7	3.4	3.4	3.7	3.8	3.5	3.7
Forms D and F	3.9	3.1	3.3	3.3	3.5	3.6	3.8
Forms D and G	3.7	3.4	3.6	3.6	3.7	3.5	3.5
Forms D and H	3.8	4.2	3.3	3.2	3.7	3.5	3.9
Forms E and F	3.6	4.2	4.1	3.7	3.8	3.4	3.6
Forms E and G	3.3	3.7	3.7	3.4	3.3	3.6	3.9
Forms E and H	3.6	3.5	3.6	3.2	3.6	3.7	4.0
Forms F and G	3.2	4.2	3.6	3.6	3.9	3.5	3.2
Forms F and H	3.1	2.8	3.5	3.7	3.4	3.5	3.1
Forms G and H	4.0	4.2	3.3	3.6	3.3	3.4	3.7

9. WEIGHTING

Although the recommended sampling method was designed to give self-weighting samples, data from all systems, with the exception of Swaziland Population A and Scotland Population A, have had weights applied in the computation of cognitive statistics. For many systems this made little difference to subscores and p-values but other systems for which differential response rates across strata were obtained or in which some small strata were over-sampled weighting was clearly necessary.

Swaziland and Scotland Population A samples were not stratified.

Almost all countries sampled intact classes because a principal aim of the study was to detect teacher effects. For between-class analyses for this purpose weighting of cognitive data is of doubtful value.

Teacher Opportunity to Learn data was also weighted.

The effect of weighting on other teacher variables and on student background variables was found to be negligible.

9.1 Weights for Cognitive Data.

Weights calculated for estimates of national parameters of student cognitive sub-scores and p-values depended for each sample on the sampling unit, the amount of variation in cluster (school or class) sizes and various other factors.

9.1.1 Stratum Weights

These were calculated for all samples using the formula

$$w_i = \frac{n}{N} \cdot \frac{N_i}{n_i}$$

where w_i is the weight for stratum i

n is the total sample size

N is the total population size

n_i is the stratum i sample size

and N_i is the stratum i population size.

Stratum weights were used to weight England and Wales data. In England and Wales students (not classes) were sampled within school and this, coupled with the loss of data at the data preparation stage, gave a large variation in (school) cluster size.

Stratum weights gave p-values and sub-score means which were more stable than obtained using school weights.

9.1.2 School Weights

School weights were calculated where sampling was by schools and where the variance of class size within school was substantial. The formula used was:

$$w_{ij} = \frac{n}{N} \cdot \frac{N_i}{s_i N_{ij}}$$

where w_{ij} is the weight for school j in stratum i

s_i is the number of schools in the sample for stratum i

N_{ij} is the number of students in the sample in school j in stratum i.

n, N and N_i are as in 9.1.1

Systems for which school weights were applied are:

Belgium (Flemish) Populations A and B, Belgium (French) AB, British Columbia A, England and Wales B, France A, Israel A, Japan AB, New Zealand AB, Ontario AB, Scotland B, Thailand AB, U.S.A. AB.

Note: where only one class per school was chosen the terms school weight and class weight are synonymous.

9.1.3 Class Weights

Where sampling was by classes the weights were calculated by the formula in 9.1.2 but with s_i = number of classes in the stratum i sample and n_{ij} = number of students in the sample in class j of stratum i.

Samples for which class weights were calculated are:

Hong Kong AB, Hungary AB, Luxembourg A, British Columbia B, Finland AB, Israel B, Sweden AB.

Note: where only one class per school was chosen the terms school weight and class weight are synonymous.

9.1.4 Weighted p-values and Subscores.

i) At school or class level (depending on the sampling method) the number of students responding correctly to an item was counted (and school or class level p-values obtained).

ii) National estimates of p-values were computed using $\frac{\Sigma p_{ij} w_{ij}}{\Sigma w_{ij}}$ where p_{ij} and w_{ij} are the p-values and weights for school/class j in stratum i.

w_{ij} used in this way is an estimate for the weight which would be obtained if the number of schools/classes in the population and in each stratum were known. Σw_{ij} will bw approximately equal to the number of schools/classes in the sample.

iii) Weighted p-values were summed across sub-test items to give sub-test means.

It should be noted that for many countries there was little difference (1 or 2%) between unweighted and weighted p-values and sub-test means. In addition, use of school/class weights gave very similar results to the use of stratum weights.

Calculation of p-values using $\frac{\Sigma X_{ij} w_{ij}}{\Sigma n_{ij} w_{ij}}$ where X_{ij} is the sum of correct responses to an item and n_{ij} is the number of students in school/class j of stratum i also produced very similar results at subtest level, although non-systematic differences of several points were evident for some items a for a few samples. Differences can be expected where cluster sizes vary considerably and class response patterns are very different.

9.1.5 Weighting Teacher Opportunity-to-Learn.

The calculated stratum weights were used to weight teacher OTL.

$$w_{ij} = \frac{n}{N} \cdot \frac{N_i}{n_i}$$

where w_{ij} = weight for teacher j in stratum i.

n = total number of students in the sample.

N = total number of students in the population.

n_i = number of students in the stratum i sample

N_i = number of students in the stratum i population.

$$\frac{n}{N} \approx \frac{n_c}{N_c} \approx \frac{n_t}{N}$$

and $$\frac{N_i}{n_i} \approx \frac{N_{ci}}{n_{ci}} \approx \frac{N_{ti}}{N_{ti}}$$

where the "c" ratios are school/class ratios and the "t" ratios are teacher ratios.

10. SAMPLING ERRORS

Standard errors have been calculated for cognitive forms Core and A at population A level and forms 1 and 7 at population B level and these are displayed in the tables below. The standard errors are, in general, stable across forms for both populations and will be representative of the error levels for subscores.

Intraclass correlations, and consequently Design Effects, were considerably higher than was anticipated. In spite of this errors for almost all countries lie within acceptable limits.

The high intraclass correlation coefficients (Rho) result from several factors:

i) Intact mathematics classes were sampled;

ii) The widespread practice of streaming/setting mathematics classes results in a considerable reduction in within class heterogeneity;

iii) Sampling systems with differing school types, or wide course variations in curricula between school/course types leads to relatively greater degree of within school/class homogeneity.

iv) Learning in mathematics is probably more sensitive to curricular and instructional differences than is learning in most other school subjects.

Thus population A intraclass correlation coefficients are high in Belgium, Hong Kong, Luxembourg, The Netherlands (differing school types) in Finland, Sweden and the USA (differing course types) and in New Zealand (a high level of streaming).

In some countries a combination of these factors applies. Lowest intraclass correlations occurred in Japan where the school system is almost uniform and where streaming/setting of classes is not practised.

Low intraclass correlations also occur where the tests were too difficult for a large majority of the samples (Nigeria and Swaziland) so that between class variance is considerably depressed.

Standard errors for Scotland population A were calculated by a jack-knifing procedure since a relatively small sample was spread across a great number of schools. Sampling was not by selection of schools or classes so calculation of design effects is inappropriate.

For population B the intraclass correlation coefficient is affected by the factors mentioned above but, in addition, the retentivity of the school system has a marked effect. In school systems in which retention in grade 12 mathematics is low, between-class variance is likely to be low, as is within-class variance and the relative changes with respect to these are not easy to predict.

For rotated forms the clusters completing a given form have been treated as though they were complete "schools/classes" although they were, in effect, random selections of students within school/classes. The standard errors for rotated forms are therefore conservative. Furthermore, sampling fractions for some countries were sufficiently large to justify adjusting the variance by a factor $(1-\frac{a}{A})$ where 'a' clusters are selected from a population of 'A' clusters. The extreme case is Luxembourg where $\frac{a}{A} = \frac{1}{2}$. Thus for Luxembourg (for example) the sampling error for the mean will be considerably less than is shown in the tables.

SECOND IEA MATHEMATICS STUDY

DESIGN EFFECTS - STANDARD ERRORS

Population A

Country	Test Form	Rho	DEFF	Standard Error of mean as a proportion of s	Standard Error of Mean	S.E as a % of the Mean
Belgium (Flemish)	Core A	0.65 0.57	13.55 3.32	0.066s 0.066s	0.54 0.42	2 2
Belgium (French)	Core A	0.71 0.86	14.30 4.37	0.083s 0.093s	0.63 0.62	3 3
British Columbia	Core A	0.31 0.35	0.03 3.00	0.064s 0.076s	0.52 0.50	2 3
Ontario	Core A	0.25 0.25	8.98 2.53	0.042s 0.046s	0.34 0.29	2 2
England	Core A	0.38 0.38	10.27 3.02	0.062s 0.068s	0.58 0.49	3 3
Finland	Core A	0.47 0.50	10.87 3.25	0.049s 0.051s	0.38 0.37	2 2
France	Core A	0.28 0.27	7.38 2.32	0.029s 0.033s	0.19 0.20	1 1
Hong Kong	Core A	0.51 0.49	22.52 5.81	0.063s 0.065s	0.51 0.44	2 3
Hungary	Core A	0.32 0.28	8.94 2.52	0.071s 0.076s	0.58 0.52	2 3
Israel	Core A	0.37 0.37	9.40 2.82	0.050s 0.057s	0.42 0.39	2 2
Japan	Core A	0.07 0.08	3.69 1.75	0.021s 0.029s	0.16 0.20	1 1
Luxembourg	Core A	0.53 0.50	10.54 2.88	0.071s 0.075s	0.46 0.43	3 3
The Netherlands	Core A	0.69 0.65	16.80 4.25	0.055s 0.056s	0.47 0.39	2 2
New Zealand	Core A	0.55 0.50	16.00 4.01	0.056s 0.056s	0.46 0.36	2 2
Nigeria	Core A	0.27 0.22	9.59 2.60	0.081s 0.085s	0.48 0.38	3 3
Scotland	Core A					2 2
Swaziland	Core A	0.28 0.17	11.30 2.40	0.11s .076s	0.64 0.37	5 3

Design Effects - Standard Errors (cont'd)

Country	Test Form	Rho	DEFF	Standard Error of mean as a proportion of s	Standard Error of Mean	S.E as a % of the Mean
Sweden	Core A	0.52 0.42	10.83 4.74	0.055s 0.053s	0.37 0.33	2 2
Thailand	Core A	0.42 0.33	18.22 4.10	0.069s 0.066s	0.53 0.38	3 3
USA	Core A	0.57 0.57	15.48 4.19	0.048s 0.050s	0.44 0.33	2 2

Notes

i Mean scores on the core test ranged from 13.6 to 26.9 and rotated form A from 12.5 to 21.7.

ii All students in all participating countries took the 40 item Core Test. In all countries except Sweden rotated forms were randomly assigned to students with one form per student. Thus in these countries ¼ of the sample took each rotated form.

iii In Sweden 2 rotated forms were randomly assigned to each student. Thus ½ the sample took each rotated form. Rotated forms contain 34 items for the cross-sectional study and 35 for the longitudinal study.

iii $Rho = \frac{bSa^2 - S^2}{(b-1)S^2}$

Rho is the intraclass correlation.
b is the mean cluster size (½ of mean class size for Sweden, ¼ of mean class size for all others)
Sa^2 is the variance between clusters and S^2 is the variance between students.

iv $DEFF = 1 + (b-1)Rho$

v Standard error of the mean as a proportion of the student standard deviation
$= S/\sqrt{\frac{n}{DEFF}}$ where n is the sample size (for a given form).

$\frac{n}{DEFF}$ is the simple equivalent sample.

SECOND INTERNATIONAL MATHEMATICS STUDY

DESIGN EFFECTS - STANDARD ERRORS

Population B

Country	Rotated Form	Rho	DEFF	Standard Error of mean as proportion of s	Standard Error of Mean	S.E as a % of the Mean
Belgium (Flemish)	1 7	0.66 0.67	2.91 2.91	0.064s 0.064s	0.18 0.21	2 3
Belgium (French)	1 7	0.49 0.47	2.22 2.17	0.066s 0.065s	0.21 0.21	3 3
British Columbia (One rotated form per student)	1 7	0.77 0.71	4.75 4.42	0.14s 0.13s	0.35 0.35	6 7
Ontario	1 7	0.31 0.30	2.60 2.57	0.057s 0.057s	0.17 0.18	2 2
England	1 7	0.27 0.30	1.41 1.47	0.040s 0.041s	0.12 0.11	1 1
Finland	1 7	0.26 0.27	2.00 1.73	0.072s 0.067s	0.20 0.19	2 2
Hong Kong	1 7	0.63 0.59	4.69 4.43	0.074s 0.072s	0.23 0.25	2 2
Hungary	1 7	0.55 0.61	4.06 4.44	0.081s 0.085s	0.26 0.29	4 5
Israel	1 7	0.37 0.57	2.30 3.02	0.069s 0.080s	0.21 0.27	3 4
Japan	1 7	0.60 0.57	6.47 6.16	0.057s 0.056s	0.19 0.20	2 2
New Zealand	1 7	0.27 0.12	1.80 1.36	0.078s 0.068s	0.25 0.19	3 2
Scotland	1 7	0.05 0.03	1.20 1.14	0.057s 0.055s	0.14 0.14	2 2
Sweden	1 7	0.21 0.11	1.96 1.50	0.054s 0.047s	0.16 0.14	2 1

Country	Rotated Form	Rho	DEFF	Standard Error of mean as proportion of S	Standard Error of Mean	S.E as a % of the Mean
Thailand	1 7	0.46 0.50	5.48 5.90	0.076s 0.079s	0.22 0.26	4 5
USA	1 7	0.48 0.49	3.04 3.17	0.051s 0.052s	0.15 0.16	2 3

Notes:

i Forms 1 and 7 each contain 17 items. Country means range from Hong Kong to Hungary.

ii With the exception of British Columbia national centres randomly assigned 2 forms per student.

iii $\text{Rho} = \frac{bs_a^2 - s^2}{(b-1)s^2}$ Intraclass correlation where b is the mean cluster size,

bsa^2 is the variance between clusters and s^2 in the variance between students.

Note that mean cluster size is ¼ mean class/school size for all countries except British Columbia (1/8th).

iv DEFF = 1 + (b-1)Rho.

v Standard error of the mean as a proportion of the student standard deviation $= s\Big/\sqrt{\frac{n}{DEFF}}$ where n is the sample size. $\frac{n}{DEFF}$ is the simple equivalent sample.

11. NON-SAMPLING ERRORS

Some non-sampling errors and sources of bias have been discussed in previous individual country sections. These include errors due to loss of data at data collection and data processing phases. Where possible achieved samples in these cases have been examined for bias and the very few cases in which bias seemed either present or possible reported.

Throughout the course of the study the International Center provided extensive advice to National Centers on procedures which should be followed to ensure the highest possible response rates and achieved samples. This advice was disseminated by means of manuals encompassing sampling, data collection and preparation, memoranda and letters to individual National Research Coordinators where problems specific to a particular country were encountered.

At the International Center gargantuan efforts were made to ensure that loss of data at the cleaning and editing stage was kept to an absolute minimum. This necessitated many letters, cables and telephone calls to National Centers and, while the process resulted in delays, has paid off in terms of the magnitudes and qualities of the achieved samples. Other possible sources of non-sampling error are discussed below.

11.1 Non-coverage.

An intention of the study was to obtain measures of outcomes of mathematics education based on the attainments and attitudes of all students in normal classes at the grade level in which most 13 year olds are found. Excluded samples included students in special schools for the intellectually handicapped and the like. While almost all countries defined their national populations in the spirit of this intention there is variation in the proportions of 13 year olds in non-normal classes from country to country, ranging from less than 1% to about 5%. Errors in estimates of parameters due to these differences would be very slight. On the other hand, for the Netherlands where a substantial group of students was not included in the defined population, and for Nigeria where 12 of 20 states (albeit containing a smallish proportion of the school population) were excluded comparisons with measures from other systems can be made, but with caution.

On the other hand, national definitions for Hungary and for Scotland at population B encompassed a wider range of students than was envisaged by the international population definition and cognitive measures for these countries are somewhat lower than would have been the case if grammar school students (Hungary) and S6 students (Scotland) had contained the national populations.

11.2 Non-response

Errors resulting from mistakes made at National Centers in preparing tests and questionnaires were extremely few. All national test forms and questionnaires were checked in the form in which they were presented to respondents except those written in languages such as Hungarian and Hebrew where back translations were checked.

Some National Centers chose to delete (non-cognitive) items from questionnaires or not to administer opportunity-to-learn instruments. Cases in which a deletion rendered an important variable unusable for a country were very small in number.

Loss of data at the England and Wales and Belgium (Flemish) national centers does not appear to have introduced any important bias and the achieved sample for cognitive instruments is high. Estimates of subtest means and p-values are sound.

The possible effect of lower response rates has been discussed earlier. The method used by England and Wales to obtain schools in sufficient numbers for the designed sample and by the USA to obtain sufficient cooperating school districts, namely inviting about twice as many as were needed in the expectation of a 50% cooperation rate, might be expected to produce a bias in achievement scores but no evidence of this has been found.

11.3 Cultural Bias

Lengthy negotiations were conducted with National Centers with respect to methodology, instruments and items and an aim in this process was to eliminate cultural bias wherever possible from all levels of the study. A full account of the procedures adopted to validate the items is given in Bulletin 5 of the Second IEA Mathematics Study.

11.4 Systematic Variation on Class Size with Ability

The practice common in many countries of making low ability classes smaller than higher ability classes may have produced a bias in the calculation of national achievement parameters given the method of applying weights which assumes equal (or near equal) cluster sizes. However, comparison of parameter estimates from raw scores, and estimates using two different weighting systems failed to detect any systematic effect due to this cause.

12. CONCLUSION

Twenty educational systems provided population A data and fifteen population B data. Thirty five samples ranging in size from approximately 1000 to more than 8800 students, their teachers and schools, took part in the study.

Given the administrational challenges involved, both at international and at national level, and the difficulties of communication across cultures by correspondence the quality of the data collected is extraordinarily good. Most National Centers had little funding for the project and National Research Coordinators in many cases undertook national supervision of the project with minimal resources and with a minimal time allowance.

The wonder is not that a very few of the samples and their consequent data sets were less than flawless but that almost all were of high quality and none was so inadequate that useful information about national mathematics outcomes in relation to those of other countries could be deduced.

Making a judgment about a particular sample requires consideration of the sampling design used, the response rates, achieved sampling fractions, known possible biases, design effects and the level of analysis at which the data is to be used.

Achieving a representative sample is much easier in some systems than in others. In small countries with a relatively uniform school system, such as New Zealand, the task is much easier than in large, highly diverse systems such as the USA or in countries where transport and communications are unreliable. Levels of school and teacher cooperation in studies of this kind also vary between countries. In some countries near perfect samples can be obtained without great difficulty, in others National Centers have to expend huge amounts of time and energy gaining cooperation.

There is no simple answer to the question "Is country X's sample so poor that the data cannot be used?" If there were such an answer it would be "No" for all samples in the study. The more relevant question relates to the various analyses and purposes for which the data is to be used and the extent of the information about the sample, and many other aspects of the study, against which it is to be interpreted.

The study design called for National Research Coordinators to make comprehensive reports to the International Center on the administration of the study in their system.

Part of the NRC report was to be a detailed description of the sampling and data collection phases. In the event many NRCs found themselves unable to complete this task fully. It came at the end of a lengthy and arduous struggle to complete the study so perhaps this is not surprising. Nevertheless, enough information has been gathered from most NRCs to enable considerable confidence to be placed in the quality of the samples. Where there are reservations these have been drawn attention to in the preceding sections.

APPENDIX I

Achieved Sampling Fractions (Student)

Belgium (Flemish)	A	0.035
	B	0.222
Belgium (French)	A	0.031
	B	0.220
British Columbia	A	0.054
	B	0.143
England and Wales	A	0.004
	B	0.029
Finland	A	0.148
	B	0.063
France	A	0.051
Hong Kong	A	0.055
	B	0.181
Hungary	A	0.015
	B	0.056
Israel	A	0.073
	B	0.631
Japan	A	0.005
	B	0.044
Luxembourg	A	0.449
The Netherlands	A	0.025
New Zealand	A	0.086
	B	0.198
Nigeria	A	0.024(est)
Ontario	A	0.038
	B	0.055
Scotland	A	0.015
	B	0.076
Swaziland	A	0.16 (approx)
Sweden	A	0.029
	B	0.211
Thailand	A	0.011
	B	0.036
U.S.A	A	0.002
	B	0.013

APPENDIX II

IEA(MATHS-NZ)/A/149

Revised version of A/122

May 1979

SECOND IEA MATHEMATICS STUDY

SAMPLING MANUAL

Edited by

Malcolm Rosier

on behalf of the

Second IEA Mathematics Study Sampling Committee

160

CONTENTS

TABLES

FIGURES

SECTION A

INTRODUCTION

This Sampling Manual has been prepared by the Sampling Committee of the Second IEA Mathematics Study (SIMS) to help countries intending to partipate in the study to develop a suitable sampling design.

The Sampling Committee has the following members:

Dr Malcolm Rosier, Australian Council for Educational Research (Chairman),

Dr John Keeves, Australian Council for Educational Research,

Mr Ian Livingstone, New Zealand Council for Educational Research, and

Mr Ken Ross, Australian Council for Educational Research.

Correspondence with the SIMS Sampling Committee should be addressed to Dr Rosier at the following address:

Australian Counc·l for Educational Research,
PO Box 210,
Hawthorn,
Victoria 3122,
Australia.

Telephone: (03) 818 1271
Telegraphic address: ROSIER ACERES MELBOURNE AUSTRALIA

Readers seeking further information about sampling, additional to that contained in this Sampling Manual, are referred to four particular texts. The first is a standard reference on sampling by Kish (1965). The next two are statements by Peaker, who was the sampling consultant for the previous IEA studies (Husén, 1967, volume I, chapter 9; Peaker, 1975). The final one is the recent monograph by Ross (1978).

1 Populations for this study

Two populations have been specified by the International Mathematics Committee.

Population A: All students in the grade (year level) where the majority has attained the age of 13.00 to 13.11 years by the middle of the school year.

then the National Center should choose the grade for which the cognitive mathematics tests are most appropriate to the curriculum.

Population B: All students who are in the normally accepted terminal grade of the secondary education system and who are studying mathematics as a substantial part (approximately five hours per week) of their academic program.

2 Aims of the study and sampling designs

The Second IEA Mathematics Study has three major aims:

1 to describe the changes in the mathematics curriculum between 1964 and 1980 and to examine to what extent the achievement of students in 1980 mirrors the changed curriculum,

2 to describe to what extent the students in 1980 achieve the objectives of the 1980 curriculum in mathematics, and

3 to identify the major classroom instruction and curricular concomitants of growth in mathematics achievement over the period of one school year.

The first two aims of the study can be achieved through a cross-sectional sampling design, in which a testing program is administered on one occasion to a sample of students. The results are then generalized to the population from which the sample was drawn to produce 'national estimates' of student mathematics achievement. This requires a probability sample, as discussed later in this Manual. We recognize that the first aim is mainly of interest to the countries that also participated in the first IEA Mathematics Study.

The third aim requires a longitudinal sampling design, in which students are tested on at least two occasions; for example, once near the beginning of the school year and a second time near the end of the school year. This also requires a probability sample if we wish to make any generalizations about the population from which the sample was taken.

At the Population B level, the longitudinal study is a 'national option' since few countries would wish to test near the end of the school year at this population level. As a national option, the country would plan its own study, conduct its own analyses, and prepare its own reports.

As the first step in developing sampling designs, each National Center must choose the population levels at which it wishes to participate. It must then prepare a sampling design or designs to meet the aims which its country wishes to achieve by means of the study. The Sampling Manual describes various sampling designs which differ in terms of the numbers of schools and students, the magnitude of the sampling errors (standard errors of sampling), and the types of analyses that can be carried out. Great care must be taken in selecting sampling designs that minimize the standard errors of sampling while ensuring that the desired analyses can be carried out.

At Population A level, National Centers must choose one of four possible plans for testing:

1 cross-sectional only, using results from one testing program to produce national estimates,

2 longitudinal only, using results from two testing programs (at the beginning and end of the school year) to investigate the effects of classroom and curricular processes on mathematics achievement,

3 cross-sectional and longitudinal together, using results from two testing programs (at the beginning and end of the school year) to produce national estimates and to investigate relationships, and

4 cross-sectional in one year and longitudinal in another year.

At Population B level, National Centers would carry out only a cross-sectional study, unless they undertook a longitudinal study as a national option.

All National Centers are encouraged to carry out both cross-sectional and longitudinal studies at the Population A level, and the cross-sectional study at Population B level.

In most countries, the funds available for the study will be limited. The sampling design has implications for expenditure on:

1 the number of tests and questionnaires to be printed,

2 the amount of secretarial work needed for typing lists of schools and students,

3 the collation and distribution of testing materials,

4 the payment of persons to administer the tests to students, and

5 the sorting, coding, card punching and initial data processing of the completed tests and questionnaires.

In some countries there will be political considerations which influence the type of sampling design; for example, legislation about the collection and archiving of social science data, and possible lack of co-operation from national and/or local educational authorities or teachers associations or school principals.

Each National Center should prepare a sampling design or designs which produces the lowest possible standard errors of sampling, given particular national constraints such as the above. It is important to minimize these standard errors so that sound comparisons can be made across countries at various levels of analysis; for example, between students and between classes.

Later sections of this Sampling Manual describe procedures for preparing a sampling design and drawing a sample. However, before proceeding, some important aspects of the theory of sampling will be discussed.

SECTION B

BASIC SAMPLING THEORY

1 Target and excluded populations

For the IEA educational survey studies, we define a population in which we are interested. From this population we select a sample of persons to be tested. The results from the sample are then generalized to the population.

In most cases the 'elements' of the population are students, and the 'units of analysis' are also students. However, we may also be interested in analyses between classes, or between students within classes, or between schools. The accuracy of the inferences we draw depend on the sampling design. Care must be taken when the units of analysis are not the same as the units of sampling (elements).

For the Second IEA Mathematics Study, the International Mathematics Committee has specified two populations, which we refer to as the 'desired target populations'.

The desired target population for Population A is:

> All students in the grade where the majority has attained the age 13:00 to 13:11 years by the middle of the school year.

Each country must restate this definition in specific terms to meet its own circumstances. This will be the 'defined target population' for that country.

For example, for Australia the defined target population for Population A is:

> All students in normal classes at Year 8 level in all States except the Northern Territory.

It can be seen that we have defined Year 8 as the grade where the majority of students has attained the age 13:00 to 13:11 years by the middle of the school year. This followed an analysis of our national statistics which gives the number of students at each age level on 1 August of each year in each year level (grade) in each State in Australia.

We have also limited the elements in the defined target population by excluding two groups of students:

1 We have excluded students who are not in normal classes, since they are not following the normal mathematics curriculum and would not have been exposed to much of the content of the mathematics achievement tests.

2 We have excluded students in the Northern Territory, since this State has a very high percentage of Aboriginal students undertaking modified curricula which would not cover the content of the mathematics tests.

The difference between the IEA desired target population and the defined target population for a country is the 'excluded population' for that country. The number of students in the excluded population and a description of the character of this excluded population must be clearly specified, and included in the report of design and execution of the sampling for the study.

2 Designed, executed and achieved samples

For the defined target population a sampling design is prepared, which will list the number of schools and students in the 'designed sample'. There will usually be some loss of respondents, so that it is necessary to include in the report a table showing the 'executed sample', which is the number of schools and students who actually participated in the testing program.

Finally, we define the 'achieved sample' as the number of schools and students from whom good data were obtained. This is the same as the executed sample after deletion of the respondents whose data were not suitable for including in the analyses, such as students who left after completing only part of the testing program.

3 Accuracy, bias and precision

There are usually two main objectives involved in the conduct of sample surveys:

a The estimation of certain population values (parameters). In many educational research surveys we are interested in obtaining estimates of the mean level of achievement for the population and various percentile points of the distribution of achievement for the population.

b The testing of a statistical hypothesis about a population. As well as estimates of population parameters we may be interested, for example, in testing the hypothesis that there is no difference between the average achievement of certain subgroups in our sample.

Our capacity to examine sample data with respect to these two objectives depends directly upon our knowledge of the accuracy of sample estimates with respect to population parameters. The accuracy of a sample estimate for a given sample is the difference between the sample estimate and the population parameter. The accuracy is largely determined by two factors: (a) sampling bias, and (b) sampling variability. Bias may result from the use of inappropriate statistical procedures (biased estimators) or from deficiencies in the sampling frame. Sampling variability, described in more detail below, is associated with the statistical relationship between characteristics of a sample and the population from which it has been drawn. The sampling variability, which is usually given by the variance of the sampling distribution of sample means, provides a measure of the precision of any one sample estimate with respect to the corresponding population paramter.

For most well-designed samples in survey research the sampling bias is close to zero. This means that the accuracy of a sample depends largely on the precision as measured by the sampling variability.

In probability sampling each element (person) in the population has a known, non-zero probability of being selected into the sample. The importance of probability sampling for the IEA surveys is that the precision of a sample selected by this method can be calculated from the internal evidence of the sample data; that is, by applying formulae or statistical techniques to the data from one sample we may estimate the sampling variability associated with all possible similar samples. Since we cannot use internal evidence to estimate the accuracy of non-probability samples, such samples are not suitable for dealing with the objectives of estimation and hypothesis testing.

Generally the value of a population parameter is not known, so that the actual accuracy of an individual sample estimate cannot be assessed. Instead, through a knowledge of the behaviour of estimates derived from all possible samples which can be drawn from the population by using the same sample design, we are able to assess the probable accuracy of the obtained sample estimate.

Consider the case of simple random samples of size n drawn from a population of size N. The means of all these samples may be plotted, to give a sampling distribution of sample means. This sampling distribution of sample means has a mean, which is equal to the population mean μ for an unbiased sampling design. The sampling distribution of sample means also has a variance $V(\bar{x})$. The square root of this variance is the standard deviation of the sampling distribution of sample means, and is known as the standard error of the mean $SE(\bar{x})$.

4 Sampling distributions and standard errors

The accuracy of the estimates used in the IEA studies depends principally on precision, which is usually calculated in terms of the standard error of a sample mean. In many pract cal survey research situations the sampling distribution of the sample means is approximately normally distributed. The approximation improves with increasing sample size even though the distribution of elements in the parent population may be far from normal.

From a knowledge of the properties of the normal distribution we can state that, at the 68 per cent confidence level, the range $\bar{x} \pm SE(\bar{x})$ includes the population mean, where $\bar{x}$ is the sample mean obtained from one sample from the population and $SE(\bar{x})$ is the standard error of $\bar{x}$. Similarly we can state that, at the 95 per cent confidence level, the range $\bar{x} \pm 1.96\ SE(\bar{x})$ will include the population mean.

In survey research we are usually dealing with a single sample of data and not with all possible samples from a population, so that we are unable to calculate the value of $V(\bar{x})$ or $SE(\bar{x})$ exactly.

Statisticians have derived some formulae, for certain sample designs, which allow us to make an estimate of $V(\bar{x})$ from the internal evidence of an individual sample of data. For the simple random sample design, each sample element is randomly and independently selected from the population with equal probability of selection. For this design the variance of sampling distribution of sample means may be estimated from a single sample of data by using the formula:

$$\hat{V}(\bar{x}) = \frac{N - n}{N} \cdot \frac{s^2}{n}$$

where $\hat{V}(\bar{x})$ is the estimated variance of the sampling distribution of sample means,

N is the population size,

n is the sample size, and

s^2 is the variance of the sample elements, given by:

$$s^2 = \frac{1}{n - 1} \cdot \Sigma(x_i - \bar{x})^2$$

The value of s^2 is an unbiased estimate of the variance of the element values in the population.

The estimated standard error of the mean $se(\bar{x})$ is given by the square root of the estimated variance:

$$se(\bar{x}) = \sqrt{\frac{N - n}{N}} \cdot \frac{s}{\sqrt{n}}$$

For sufficiently large values of n, we may estimate with 95 per cent confidence that the population mean μ will be in the range $\bar{x} \pm 1.96\ se(\bar{x})$, where $\bar{x}$ is the sample mean of a simple random sample of n elements selected from a population of N elements. The term $(N - n)/N$ is called the finite population correction. For sufficiently large values of N relative to n the finite population correction tends to unity, so that the standard error of the mean may be estimated by:

$$se(\bar{x}) = \frac{s}{\sqrt{n}} \text{ . (for large N)}$$

5 Stratified sampling

One way of increasing the precision of the estimates derived from a simple random sample is to increase the sample size. Another way is to use stratification. Stratification does not imply any departure from probability sampling. It merely requires that, before any selection takes place, the population should be divided into a number of mutually exclusive groups called strata. Following this division, a random sample is selected within each stratum.

Stratification may be used in survey research for reasons other than obtaining gains in precision. Strata may be formed in order to employ different sampling methods within strata, or because the sub-populations defined by the strata are designated as separate domains of study.

Some typical variables used to stratify populations in educational survey research are:

a region (metropolitan/country),

b type of school (government/non-government),

c school size (large/medium/small), or

d sex of school (boys only/girls only/mixed).

Stratification does not necessarily require that the same sampling fraction is used within each stratum. If a uniform sampling fraction is used then the sample design is known as a proportionate stratified sample because the sample size from any stratum is proportional to the population size of the stratum. If the sampling fractions vary between strata then the obtained sample is a disproportionate stratified sample, which is discussed below.

6 Multistage complex sampling designs

A population of elements can usually be described in terms of a hierarchy of sampling units of different sizes and types. For example, a population of school students may be seen as being composed of a number of classes each of which is composed of a number of students. Further, the classes may be grouped into a number of schools.

In the previous discussion we have considered the use of simple random samples in which the students were selected individually from the population. In practice we usually select the individual units of the population as clusters, or in several stages. These modifications in sampling design are often used because they reduce the costs of a research study by minimizing the geographical spread of the sample elements.

Consider the hypothetical population of school students described in Figure B.1. The population consists of eighteen students distributed among six classrooms (with three students per class) and three schools (with two classes per school).

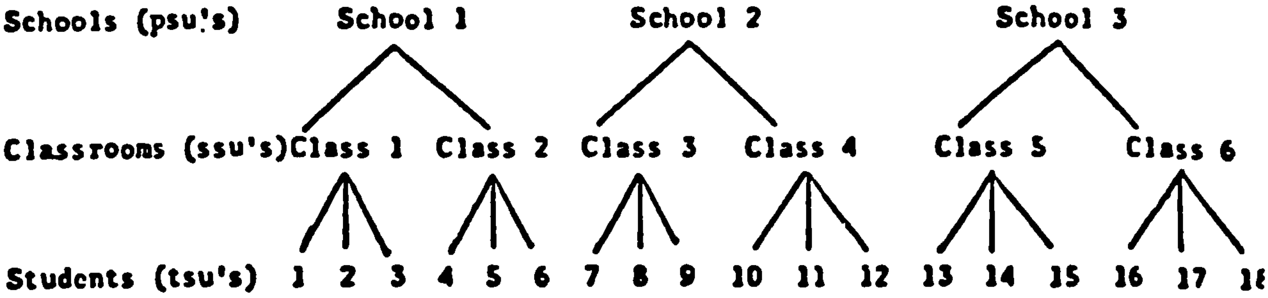

Figure B.1 Hypothetical population of eighteen students grouped into six classrooms and three schools.

From this population we could select a simple random sample of four students or we could employ a multi-stage cluster sample design to select a sample of the same size.

In order to select a multi-stage cluster sample we consider the population to be divided into primary sampling units (schools), secondary sampling units (classrooms) and tertiary sampling units (students). At the first stage of sampling we could randomly select two schools; at the second stage of sampling we could randomly select one classroom from each of the selected schools; and at the third stage of sampling we could randomly select two students from each selected classroom. The procedures required for the selection of sampling units at different stages are discussed later in this Manual.

If we employed either the simple random sample design or the three stage cluster sample design described above to select a sample of four elements, then for both sample designs this would ensure that each population element had an equal chance of appearing in either of the samples. That is, sample estimates of population parameters, such as the population mean, would provide unbiased estimates for both sample designs.

7 Comparison of sampling designs

In the above example we have seen that, for a given sample size, both the simple random sampling design and a three stage cluster sampling design may provide unbiased sample estimates of the population mean. However, the variance of these estimates may vary greatly. In order to compare these two sampling designs we need to examine the stability of the estimates which they provide for samples of the same size.

Kish (1965) suggested the use of the simple random sample design as a baseline for quantifying the efficiency of complex sampling designs, and introduced the term 'deff' (design effect). It may be defined as the ratio of the variance of the sampling distributions of sample means for the complex sampling design to the corresponding variance of a single random sampling design involving samples with the same number of units:

$$\text{deff} = \frac{V(\bar{x}_c)}{V(\bar{x}_{srs})} \text{ (for } n_c = n\text{)}$$

where $V(\bar{x}_c)$ is the variance of the sampling distribution of sample means for complex samples of size n_c, and

$V(\bar{x}_{srs})$ is the variance of the sampling distribution of sample means for sample random samples of size $n = n_c$.

For a simple random sample of elements drawn without replacement . have:

$$V(\bar{x}_{srs}) = \frac{N - n}{N} \cdot \frac{S^2}{n}$$

where N is the population size,

n is the sample size, and

S^2 is the variance of the population elements.

Substituting into the expression which defines deff, we have:

$$\text{deff} = \frac{V(\bar{x}_c)}{\frac{N - n}{N} \cdot \frac{S^2}{n}}$$

$$\text{or } V(\bar{x}_c) = \frac{N - n}{N} \cdot \frac{S^2}{n} \cdot \text{deff} = \frac{N - n_c}{N} \cdot \frac{S^2}{n_c} \cdot \text{deff}$$

Kish (1965: 68, 258) established that s^2 computed from any large probability sample yields a good approximation of S^2. The approximation is quite accurate when deff is near one; in other cases with smaller samples it neglects a term of order $\frac{1}{n}$. By using an estimate of deff, obtained mostly from past experience, and s^2 as an estimate of S^2 the above equation may be used to obtain an estimate of the variance of the sampling distribution of sample means when complex sample designs are used.

In the above section, sampling designs were compared in terms of the variances for samples of equal size. We can also compare sampling designs by equating the variances and examining the relative sample sizes, using the concept of 'effective sample size' (Kish, 1965: 259) or 'simple equivalent sample' (Husén, 1967, Vol.I: 149).

Consider a complex sample of size n_c. The variance of the sampling distribution of sample means for this complex sampling design is $V(\bar{x}_c)$. Consider a simple random sample of size n^* drawn from the same population so that the variance of the sampling distribution for this sampling design $V^*(\bar{x}_{srs})$ is equal to $V(\bar{x}_c)$.

For the simple random sample of n^* elements drawn without replacement:

$$V^*(\bar{x}_{srs}) = \frac{N - n^*}{N} \cdot \frac{S^2}{n^*}$$

But since $V^*(\bar{x}_{srs}) = V(\bar{x}_c)$, we may write:

$$\frac{N - n^*}{N} \cdot \frac{S^2}{n^*} = \frac{N - n_c}{N} \cdot \frac{S^2}{n_c}$$

If N is large compared to n_c or n^*, then the size of the simple equivalent sample (or the effective sample size) is given by $n^* = \frac{n_c}{deff}$.

For many commonly used sample designs and for many commonly used statistics in survey research we find that deff is greater than unity. Consequently, the use of formulae based on the simple random sample model to estimate standard errors may result in gross underestimation of sampling errors.

8 Coefficient of intraclass correlation (rho)

Standard statistical theory has mostly been developed with the assumption that the sample observations are obtained through independent random selection. However, most research in the social sciences has been carried out by using complex sample designs. The main features of complex sample designs are clustering, stratification, unequal probabilities of selection and systematic sampling. Kish (1957) examined the consequences of applying the usual textbook formulae for calculating confidence limits to data obtained by employing complex sample designs. He concluded that:

> In the social sciences the use of srs (simple random sample) formulas on data from complex samples is now the most frequent source of gross mistakes in the construction of confidence statements and tests of hypotheses (Kish, 1957: 156).

The source of this discrepancy in error estimates may be traced to the fact that the researchers find it economical and convenient to use existing clusters as the primary sampling units rather than individual elements. Since individuals within a particular sampling unit tend to resemble each other more than they resemble individuals from other units the basic assumption of independent random selection of observations breaks down and the usual formulae fail to apply.

Kish (1957) points out that this homogeneity of individuals within sampling units may be due to common selective factors, or to joint exposure to the same effects, or to mutual influence (interaction), or to some combination of these. The magnitude of this homogeneity is usually measured by rho, the coefficient of intraclass correlation.

It should be remembered that the value of the coefficient of intraclass correlation has no meaning for the individual except insofar as he is considered to be a member of a group. A high value implies that there is a high degree of homogeneity within the groups of observations.

9 Relationship between rho and simple cluster sampling

When data are gathered in educational survey research with a simple random sample design, the individual selection and measurement of population elements often becomes too expensive. In order to reduce costs by minimizing the geographical spread of the selected sample, survey researchers often employ cluster sampling designs. Cluster sampling involves the division of the population of elements into groups or clusters which serve as the initial units of selection. Sometimes the selection of clusters as the primary units is followed by the selection of a simple random sample of elements within the selected clusters.

When there is more than one stage of selection we refer to the sample design as a multistage sample design. The simplest form of multistage sampling is the simple two-stage cluster sample design. The influence of the selection of elements in clusters on precision may be examined by comparing the simple random sample design with a two stage cluster sample design when the sample size in each design is the same.

Consider a population of N elements divided into equal-sized clusters. Firstly, we can draw a simple random sample of size n from the population. Secondly, we can draw a two-stage sample of the same size from the population by using simple random sampling to select m clusters, and then for each of the selected clusters by using simple random sampling to select $\bar{n}$ elements, so that the total sample size n is given by:

$$n = m.\bar{n}$$

The relationship between the variances of the sampling distributions of sample means for these two sampling designs is given by:

$$V(\bar{x}_c) = V(\bar{x}_{srs}) \; [1 + (\bar{n} - 1).rho]$$

where $V(\bar{x}_c)$ is the variance of the sampling distribution of sample means for the above simple two-stage cluster design

$V(\bar{x}_{srs})$ is the variance of the sampling distribution of sample means for the simple random sample design

$\bar{n}$ is the ultimate cluster size, and

rho is the coefficient of intraclass correlation.

The above expression shows that the sampling accuracy of the simple two-stage cluster sample design depends, for a given ultimate cluster size, on the value of the coefficient of intraclass correlation. When the elementary units within clusters tend to be similar with respect to some characteristic, the intraclass correlation between elementary units within clusters for that characteristic will be high. Conversely, if the elementary units within clusters are relatively heterogeneous with respect to the characteristic, the intraclass correlation will be low positive or, in very unusual situations, even negative (Hansen et al., 1953:260).

In educational survey research rho is generally positive for achievement measures within schools. That is, the homogeneity of students within schools with respect to achievement is greater than if students were assigned to them at random. It is important to remember that the coefficient of intraclass correlation may take different values for different variables, different populations and different clustering units.

Since rho is generally positive for a wide range of characteristics concerning students within schools or students within classrooms, we find that the precision of the simple two-stage cluster sample is less than for a simple random sample of the same size. When contemplating the selection of clusters rather than elements in an educational survey research study, the researcher must balance the losses in precision due to clustering against the advantages of reduced costs arising from the selection and measurement of fewer primary sampling units.

10 Selection of clusters

The selection of classrooms or schools as the primary sampling unit must take account of the fact that these primary sampling units may differ greatly in size. If we choose the primary sampling units with simple random sampling then a self-weighting design would require the use of the same sampling fraction within each selected cluster. By using this procedure the final sample size would depend on which primary sampling units were chosen first.

The following formula indicates a given element's probability of selection for a srs selection of clusters followed by the selection of a fixed proportion of elements per selected cluster.

$$\text{Element probability} = \frac{\left(\begin{array}{c}\text{Number of}\\ \text{clusters}\\ \text{selected}\end{array}\right)}{\left(\begin{array}{c}\text{Number of}\\ \text{clusters in}\\ \text{population}\end{array}\right)} \times \left(\begin{array}{l}\text{Proportion of students}\\ \text{selected from}\\ \text{selected cluster}\end{array}\right)$$

Since all values on the right hand side of the above equation are fixed then the element probability will be constant for all elements. However the final sample size for this method of sample selection will depend both upon the size of the selected clusters and also upon the value of the fixed proportion of students which is to be selected from each selected cluster.

One method of obtaining greater control over the sample size and yet ensuring a self-weighting design is to select the primary sampling units with probability proportional to size (pps), and then select equal sized ultimate clusters from the selected primary sampling units.

The following formula indicates a given element's probability of selection for a pps selection of clusters followed by a srs of a fixed number of elements per selected cluster:

$$\text{Element probability} = \left(\begin{array}{l}\text{Number of}\\\text{clusters}\\\text{selected}\end{array}\right) \times \left(\frac{\text{Cluster size}}{\text{Population size}}\right) \times \left(\frac{\begin{array}{l}\text{Elements selected}\\\text{per selected}\\\text{cluster}\end{array}}{\text{Cluster size}}\right)$$

This formula simplifies to:

$$\text{Element probability} = \left(\begin{array}{l}\text{Number of}\\\text{clusters}\\\text{selected}\end{array}\right) \times \left(\frac{\begin{array}{l}\text{Elements selected}\\\text{per selected}\\\text{cluster}\end{array}}{\text{Population size}}\right)$$

That is, if we have equal sized ultimate clusters then the element probability will be constant for all elements. Further, we have control over our sample size according to the following formula:

$$\text{Sample size} = \left(\begin{array}{l}\text{Number of}\\\text{clusters selected}\end{array}\right) \times \left(\begin{array}{l}\text{Elements selected}\\\text{per selected}\\\text{cluster}\end{array}\right)$$

11 Weighting

The preparation of weighting schemes for participating IEA countries may be undertaken for a variety of reasons:

a A country conducts planned disproportionate sampling within the defined strata of the population. This may occur because separate sample estimates are being prepared for particular strata. For example, a country may require separate estimates of equal sampling accuracy for each of the major administrative regions which taken together make up the country.

b A country suffers loss of data in a particular stratum. This may occur through non-participation of selected sample schools or through loss of data during the transport of questionnaire materials from participating schools to the National Center.

c Students who have been selected into the sample do not attend the testing sessions. This may occur during the cross-sectional or longitudinal phase of the study because a selected student is absent on the day of testing. During the longitudinal phase some students who participated in the pretest may not attend the post-test data gathering stage.

d Some countries may wish to prepare national profiles of teacher characteristics. This will require differential weighting of teachers because we are designing our probability samples around students and not teachers. Certain information will need to be gathered from National Centers in order to calculate appropriate weighting factors for teachers.

e The analysis of data at different levels of aggregation (for example students, classrooms and schools) will require different weighting strategies for each level of analysis.

In order to construct appropriate weighting schemes it will be necessary for each participating country to keep detailed records describing the steps which were taken to select their samples of schools, classrooms and teachers. At a later stage the Sampling Committee will send a questionnaire to all National Centers in order to gather this information.

12 Disproportionate stratified sampling

The simple random sample design is called a self-weighting design because each element has the same probability of selection equal to $\frac{n}{N}$. For this design each element has a weight of $\frac{1}{n}$ in the mean, 1 in the sample total, and $F = \frac{1}{f}$ in the population total, where $f = \frac{n}{N}$ is the uniform sampling rate for all population elements (Kish, 1965:424).

In a disproportionate stratified sample design we employ different sampling fractions in the defined strata of the population. The chance of an element appearing in the sample is specified by the sampling fraction associated with the stratum in which that element is located. The reciprocals of the sampling fractions, which are sometimes called the raising factors, tell us how many elements in the population are represented by an element in the sample. At the data analysis stage we may use either the raising factors, or any set of numbers proportional to them, to assign weights to the elements. The constant of proportionality makes no difference to our estimates. However, in order to avoid confusion for the readers of survey research reports, we usually choose the constant so that the sum of the weights is equal to the sample size.

For example, consider a stratified sample design of n elements which is applied to a population of N elements by selecting a simple random sample of n_h elements from the hth stratum containing N_h elements. In the hth stratum the probability of selecting an element is n_h/N_h, and therefore the raising factor for this stratum is N_h/n_h. That is, each selected element represents N_h/n_h elements in the population.

The sum of the raising factors over all n sample elements is equal to the population size. If we have two strata for our sample design then:

$$\left(\frac{N_1}{n_1} + \frac{N_1}{n_1} + \cdots \text{ for } n_1 \text{ elements}\right) + \left(\frac{N_2}{n_2} + \frac{N_2}{n_2} + \cdots \text{ for } n_2 \text{ elements}\right) = N$$

In order to make the sum of the weights equal the sample size, n, both sides of the above equation will have to be multiplied by a constant factor of n/N. Then we have:

$$\left(\frac{N_1}{n_1}\cdot\frac{n}{N} + \cdots \text{ for } n_1 \text{ elements}\right) + \left(\frac{N_2}{n_2}\cdot\frac{n}{N} + \cdots \text{ for } n_2 \text{ elements}\right) = n$$

Therefore the weight for an element in the hth stratum is $\frac{N_h}{n_h}\cdot\frac{n}{N}$

For the special case of proportionate stratified sampling which was discussed in the previous section we have $\frac{n}{N} = \frac{n_h}{N_h}$ for each stratum. The sample element weight is equal to 1 and we therefore describe this design as a self-weighting design.

13 Other statistics

It should be remembered that, although our discussion has focused on sample means, we could also consider any other population value $\bar{v}$. The confidence limits would take the form $\bar{v} \pm t\sqrt{[V(\bar{v})]}$. The quantity t represents an appropriate constant which usually is obtained from the normal distribution or under certain conditions from the t distribution. For most sample estimates encountered in practical survey research, assumptions of normality lead to errors that are small compared to other sources of inaccuracy.

Although there is general agreement among statistical authors about the formula for estimating the variance of the sampling distribution of sample means for simple random sampling designs, there are minor differences of opinion about the appropriate formulae for calculating the variance of the sampling distributions for more complex statistics. These minor differences generally become insignificant for the typically large population and sample sizes which are associated with survey research.

Table B.1 presents the formulae for calculating the standard error of a statistic from a simple random sample of elements for a range of complex statistics which are commonly employed in educational survey research. For this Manual the formulae were selected from one source (Guildford and Fruchter, 1973).

The formulae in Table B.1 are based on a simple random sample of n elements which are measured on m variables, where variable x has a standard deviation of s. The multiple correlation coefficient $R_{i.jkl}$ refers to the regression equation which uses variable i as the criterion and variables j, k and l as predictors.

The formulae were derived on the assumption that the sample design used to collect the data consisted of a simple random sample of elements. However most social science research, especially survey research, is conducted with data obtained from complex sample designs which employ techniques such as stratification, clustering and varying probabilities of selection. Computational formulae are available for estimating the standard errors of means, aggregate. and differences of means for a wide range of these sample designs (see Kish, 1965). Unfortunately the computational formulae required for estimating the standard error of multivariate statistics such as correlation coefficients, regression coefficients, etc. are not readily available for sample designs which depart from the model of simple random sampling. These formulae either become enormously complicated or, ultimately, they prove resistant to mathematical analysis (Frankel, 1971).

In the past many educational researchers have underestimated the standard errors for multivariate statistics by applying formulae which were appropriate only for data obtained from a simple random sample design although they had used complex sampling designs in their research.

Table B.1 Formulae for Estimating Standard Errors when Data are Gathered with a Simple Random Sampling Procedure

Sample statistic	Estimated se($\hat{v}$)
Mean	$\frac{s}{\sqrt{n}}$ (Guilford and Fruchter, 1973:127)
Correlation coefficient	$\frac{1}{\sqrt{n}}$ (Guilford and Fruchter, 1973:145)
Standardized regression coefficient	$\sqrt{\left[\frac{1 - R^2_{1.234...m}}{(1 - R^2_{2.34...m})(n-m)}\right]}$ (Guilford and Fruchter, 1973:368)
Multiple correlation coefficient	$\frac{1}{\sqrt{(n - m)}}$ (Guilford and Fruchter, 1973: 367)

14 Sampling design tables

Consider the development of student profiles for item difficulty values.

If we select a simple random sample of n_{srs} students from the population in order to estimate the proportion p who have obtained the correct answer to an item, then the standard error of this estimate could be estimated by the following formula (Kish, 1965: 46).

$$se(p) = \sqrt{\frac{p(1-p)}{n_{srs}}}$$

Let us specify that the standard error of p expressed as a percentage should not exceed 2.5 per cent, which gives an estimated population value of p ± 5 per cent for 95 per cent confidence limits if we assume normality. The maximum value of p(1 - p) occurs for p = 0.5. In order to ensure that we could satisfy these error requirements for all items we would require:

$$0.025 \geq \sqrt{\frac{0.25}{n_{srs}}}, \text{ or}$$

$n_{srs} \geq 400$ for a 95 per cent confidence band of ± 5 per cent.

That is, the size of the simple equivalent sample should not be less than 400.

Now consider the estimation of student mean scores on tests and subtests. From previous discussion we have, for the variance of the sample mean:

$$V(\bar{x}_c) = \text{deff} \cdot \frac{s^2}{n_c}$$

$$= \frac{s^2}{\frac{n_c}{\text{deff}}}$$

$$= \frac{s^2}{n^*}$$

Hence: $se(\bar{x}_c) = \frac{s}{\sqrt{n^*}}$,

where s is the value of the standard deviation of student scores on the test.

The calculation of the standard error of the mean for the complex sample can be based on the minimum size of the simple equivalent sample:

$$se(\bar{x}) = \frac{s}{\sqrt{400}} = .05s$$

That is, for $n^* = 400$ the standard error of the sample mean is equal to 5 percent of a student standard deviation. This error limit for sample means is close to the sampling tolerance levels suggested for previous IEA studies.

Now let us consider the size of the two-stage cluster sample which would provide equivalent sampling accuracy to a simple random sample of 400 elements. That is, what numbers of primary sampling units (psu's) and secondary sampling units (ssu's) are required for a two-stage cluster sample which will provide 95 per cent confidence limits for item percentages of ± 5 per cent, and standard errors for test means which are equal to 5 per cent of a student standard deviation score.

The relationship between the size of such a complex sample n_c and the size of a simple equivalent sample n^* may be expressed in the following terms:

$$n^* = \frac{n_c}{deff}$$

$$n_c = n^* . deff = n^* [1 + (\bar{n} - 1).rho]$$

$$= m\bar{n}$$

where rho is the coefficient of intraclass correlation,

m is the number of primary selections, and

$\bar{n}$ is ultimate cluster size.

By using the value of $n^* = 400$, the minimum simple equivalent sample size which will satisfy our error constraints for items, we may rewrite the above formula as:

$$n_c = m\bar{n} = 400 [1 + (\bar{n} - 1).rho]$$

As an example, consider rho = 0.2 and $\bar{n} = 10$. Then:

$$n_c = 400 [1 + (10 - 1) 0.2] = 1120$$

$$m = n_c/\bar{n} = 112$$

In planning a sampling design, the value used for rho should be based on a pilot-testing program or on other prior experience. Table B.2 sets out values for m and n_c for various values of $\bar{n}$ for two particular values of rho, equal to 0.2 and 0.4. Reasons for the selection of these values for rho are discussed below. Each of the sampling designs represented in this table would provide:

a 95 per cent confidence bands of ± 5 per cent for estimated item percentages, and

b a standard error for test means which is equal to 5 per cent of a student standard deviation score.

During previous IEA studies a value of rho = 0.2 was found to be a suitable estimate for two-stage cluster sampling of involving the selection of schools at the first stage followed by the selection of a random cluster of students from these selected schools at the second stage.

There is little hard evidence available to suggest an appropriate value for rho when classrooms are used as the first stage of sampling. The evidence available (Ross, 1978) suggests that students are more alike within classrooms than they are within schools. For this reason

Table B.2 Sampling decision table: 5 per cent tolerance*

	rho = 0.2		rho = 0.4	
$\bar{n}$ Number of students Selected per cluster	m Number of clusters	n_c Complex sample size	m Number of clusters	n_c Complex sample size
2	240	480	280	560
4	160	640	220	880
5	144	720	208	1040
6	134	804	200	1200
8	120	960	190	1520
10	112	1120	184	1840
12	107	1284	180	2160
14	103	1442	178	2492
16	100	1600	175	2800
18	98	1764	174	3132
20	96	1920	172	3440
25	93	2325	170	4250
30	91	2730	168	5040

* Values of n_c and m for a two stage cluster sample design which is required to provide sampling tolerances of ±5% for 95% confidence limits for item percentages, and estimates of means having standard errors equal to 5% of a student standard deviation.

we suggest the use of a value of rho equal to 0.4 for students within classrooms.

Some countries may have suitable data from earlier survey research studies which was gathered by using classrooms as the first stage of sampling. These countries could then calculate their own values for rho and construct their own sampling decision tables. One approach for estimating rho is described in Ross (1978: 178-183).

Consider two countries X and Y which both wish to select a sample of intact classes. In each of these countries there are 24 students in a class at the Population A level. There are four different forms of the test at this level, which are termed the rotated forms. The 'degree of rotation' refers to the number of rotated forms to be completed by each student in the sample. Let us consider that the degree of rotation in Country X is one rotated form per student, and in Country Y it is two rotated forms per student. This means that we will obtain an average of six observations per rotated form from the students in each class in the sample from Country X, and we will obtain 12 observations per rotated form from each class in Country Y.

Let us assume that rho = 0.4 is a fair estimate for the coefficient of intraclass correlation for both countries. Let us now examine the entries in Table B.2 under the heading rho = 0.4. We have $\bar{n}$ = 6 for Country X and $\bar{n}$ = 12 for Country Y. For Country X we would require $\bar{n}$ = 6, m = 200 and n_c = 1200. For Country Y we would require $\bar{n}$ = 12, m = 180, and n_c = 2160.

Note that both of these designs will provide the same error tolerances for both items and rotated form sample means. However, because in Country Y the effective ultimate cluster size is doubled, then we are able to select fewer primary sampling units (180 instead of 200 for Country X).

Also note that the sample means and item percentages derived from core tests for both of these sample designs will be more precise than the planned tolerances because for Country X we will have 200 classrooms with 24 core test responses per class and for Country Y we will have 180 classrooms with 24 core test responses per class.

From Table B.2 a country may choose the sample design which is appropriate for sampling schools as the primary sampling unit (rho = 0.2) or sampling classrooms as the primary sampling unit (rho = 0.4). Consideration must also be given to the 'degree of rotation' which will be used by the National Centers.

The following Table B.3 describes alternative sample designs which will provide 95 per cent confidence limits of p ± 7.5 per cent for item percentages and having sample means with standard errors equal to 7½ per cent of a student standard deviation. This table has been presented because it is recognized that to sample at the recommended precision level may be beyond the administrative and financial resources available for some countries.

Table B.3 Sampling decision table: 7½ per cent tolerance*

$\bar{n}$	rho = 0.2		rho = 0.4	
	m	n_c	m	n_c
Number of students Selected per cluster	Number of clusters	Complex sample size	Number of clusters	Complex sample size
2	107	214	125	250
4	72	288	98	392
5	65	325	93	465
6	60	360	89	534
8	54	432	85	680
10	50	500	82	820
12	48	576	81	972
14	46	644	79	1106
16	45	720	78	1248
18	44	792	78	1404
20	43	860	77	1540
25	42	1050	76	1900
30	41	1230	75	2250

* Values of n_c and m for a two stage cluster sample design which is required to provide sampling tolerances of ±7.5% for 95% confidence limits for item percentages, and estimates of means having standard errors equal to 7.5% of a student standard deviation.

Each of these sample designs will (for the appropriate value of rho) correspond to a simple equivalent sample of 178 elements.

It is important to remember that the use of the designs listed in Table B.3 will diminish the accuracy of sample estimates of item percentages and means. It will also lead to difficulties for the use of between-classrooms causal models because of the need in these types of data analyses for larger numbers of classrooms than are provided in this table. These questions which concern the limitations on the number of sampling units required for multivariate analysis are discussed in the following section.

15 Number of units: multivariate analysis constraints

The longitudinal aspect of the study will be based on classrooms as the unit of analysis and will probably employ regression related techniques to explore the influence of certain independent variables on change in mathematics performance. Sometimes multivariate methods such as regression analysis require large numbers of variables - this may lead to problems of instability if the ratio of the number of cases to the number of variables becomes too small. Although there are no easy solutions to this problem, several authors have provided some rules-of-thumb for the lower bound of the number of cases: Cattell (1952) recommends at least four cases for each variable when using factor analytic methods, Kerlinger and Pedhazur (1973: 46) suggest that between 100 and 200 cases should be required for regression analyses which do not involve large numbers of variables, Tatsuoka (1970: 38) states that the sampling size should preferably be at least three times the number of variables used in discriminant function analyses.

Several regression equations employed in the IEA Six Subject Survey contained more than 25 variables. Considering the advice of the above authors it would seem that if similar numbers of variables are employed in multivariate analyses for this study then at least 100 classrooms will be required to be sampled.

If the analysis procedure employed is path analysis then we may be required to conduct significance tests on the standardized regression coefficients. The standard error of these coefficients will on the average be slightly smaller than the standard error of correlation coefficients (Ross, 1978). Thus a conservative estimate of the standard error of a path coefficient would be $1/(\sqrt{n})$ where n is the sample size. This error estimate is based on the assumption of a simple random sampling of observations. If we use classrooms as the first stage of sampling and employ a stratified systematic selection procedure then we find that this is a safe assumption when applied to between classrooms analyses (Ross, 1978).

For example, from Table B.2 we see that under certain sampling conditions, a sample of 172 classrooms with 20 students per classroom would provide a 95 per cent confidence band of ± 5 per cent for item difficulty values. If we employ a sample of this size and then apply path analysis techniques to the between-classes data then the 95 per cent confidence band for the path coefficients would be $\pm 2/\sqrt{172}$ or $\pm$ 0.2 if we round to one decimal place.

Much published research has usefully employed path coefficients which have magnitudes much less than 0.2. Therefore it would seem that a sample size of 172 classrooms may be too small because it may lead to the deletion of paths which are educationally significant but statistically not significant If we lift the number of classrooms to 200 then, by rounding to one decimal figure, we obtain a 95 per cent confidence band of $\pm$ 0.1. This narrower confidence band would seem to be more in keeping with what experience shows to be the magnitude of a path coefficient which is commonly reported as having educational significance.

16 Some examples in the use of decision tables

Country X wishes to participate in the cross-sectional study at Population B level and also to participate in both the cross-sectional and longitudinal study at the Population A level.

The national data analyses and error constraints for Country X have been stated as:

a Require student profiles on all test items (including core test items and rotated test items) for both populations.

b Require multivariate analyses to be carried out on the data gathered from the Population A level. These analyses are to be carried out at both the between student and between classroom level.

c The error constraints are -

i 95% confidence limits for item difficulties are p $\pm$ 5%

ii 95% confidence limits for means of core and rotated tests are 0.05s (where s is a student standard deviation).

iii Path coefficients greater than 0.1 in causal models employed for the multivariate analyses should be significant at the 95% confidence level.

From the requirements mentioned above Country X would conduct its sampling such that the Population A sample design was a two-stage sample of classrooms followed by students within classrooms (which is approximately equivalent to sampling schools then one class within schools and then sampling students within classrooms).

At the Population B level the sample design would be a two-stage sample of schools, followed by a sampling of students within the selected school (that is, a sampling of students across the school from the appropriate target population level).

Country X would require a sample based on classrooms at the Population A level in order to ensure that between classrooms analyses could be carried out. At the Population B level only a cross-sectional study is required and therefore Country X may employ the more efficient sampling procedure of sampling schools and students within schools. (The procedure is more efficient due to the lower value of rho for students within schools.)

Country X requires student profiles for items in the core test and in the rotated forms to conform to the error bounds stated.

At the Population A level of testing there is 1 core test and 4 rotated forms, at the Population B level of testing there are 7 rotated forms. Let us assume the minimum class size is 24 at the Population A level and the minimum school target population level is 14 at the Population B level.

That is, at any selected school we can expect a minimum of 6 responses per rotated test form for Population A and a minimum of 2 responses per rotated test-form for Population B.

Using the sampling decision table for a simple equivalent sample of size 400 we may select the appropriate sample design for each population.

For Population A (assuming rho = 0.4) the ultimate cluster size (per rotated form) will be 6 and thus we will require the selection of 200 classrooms. Then by taking a total of at least 24 students per class for the testing program we may obtain at least 6 responses to the 4 rotated test forms.

For Population B (assuming rho = 0.4) the ultimate cluster size (per rotated form) will be 2 and thus we will require the use of 240 schools. By taking a total of 14 students per selected school for the testing program we obtain at least 2 responses to the 7 rotated test forms.

The decisions made above are based on the assumption that each student will respond to only one rotated test form.

If it is possible for one student to respond to 2 rotated forms then we may reconsider our sampling plan. For example, when we obtain 2 responses from each student at the Population B level, then our ultimate cluster size per test becomes 4 (since there are at least 14 students per school each of which will respond to two of the possible 7 rotated test forms).

Now, considering the sampling decision table for an ultimate cluster size of 4 we will require 160 schools at the Population B level.

If we could move to a situation at the Population B level in which all 14 students were able to complete all test forms then we would have an ultimate cluster size of 14 which would require only 103 schools (assuming rho = 0.2).

We cannot be so free with our choices for the Population A sample design because of the multivariate constraint in c(iii). From previous discussion we must have around 200 classrooms in order to satisfy the error constraints for the use of path models.

17 Marker variables

In order to check the quality of the sample data obtained in the IEA studies it is useful to compare our samples to some known characteristics of the target populations from which they were selected. Appropriate marker variables may vary from country to country depending on the availability of national statistics describing the population under consideration.

An example of a useful marker variable is sex of student. Table B.4 presents the percentage distribution of male and female students by region in the sample and the target population for a particular study.

Table B.4 Marker Variable: Percentage of Male and Female Students

Region	Population		Sample		
	Males %	Females %	Males %	Females %	Missing %
A	51.6	48.4	53.9	45.4	0.7
B	51.2	48.8	51.6	47.8	0.6
C	52.0	48.0	50.7	49.1	0.2
Country	51.6	48.4	52.1	47.4	0.5

Some other useful marker variables could be the percentages of students in metropolitan and non-metropolitan schools, the percentage of students in different types of school systems and the age distribution of students.

SECTION C

PREPARATION OF SAMPLING DESIGN: CROSS-SECTIONAL STUDY

The preparation of the sampling design and the selection of sample schools and students requires a series of decisions to be made, with action to follow these decisions. The decisions will depend on the circumstances in each country. They depend on the funds available and problems of administration as well as on statistical considerations.

1 Selection of population

National Centers must decide whether to participate in the study at Population A only, Population B only, or at both population levels.

It is then necessary to prepare a statement of the defined target population for each level being tested.

In order to prepare this definition it will be necessary to collect relevant national educational statistics:

a at the Population A level on the distribution of 13-year-old students by age and grade (Year level), and

b at the Population B level on the numbers of mathematics students, proportion of mathematics students in schools of different types, etc.

National Centers should also prepare a statement describing the nature and magnitude of the excluded population.

2 Selection of cross-sectional or longitudinal study

Countries must decide whether they wish to test the students with one or two testing programs.

a One testing program. Countries choosing to undertake the cross-sectional study only would conduct only one testing program, involving the administration to students of one set of instruments (tests and questionnaires) together with associated teacher and school questionnaires. The student instruments would probably be those administered as a post-test in other countries carrying out the longitudinal study as well.

b Two testing programs. Countries undertaking a longitudinal study will require two testing programs, administering the pre-test instruments near the beginning of the school year and the post-test instruments near the end of the school year. For these countries it will also be possible to use the results for cross-sectional purposes if a suitable sampling design is chosen.

If the data collected are to be used only for producing results about relationships between explanatory variables and criteria such as mathematics achievement, it would be possible to use a judgment sample of schools and students instead of a probability sample. If the data collected are to be used at any time for producing national estimates of student, teacher or school characteristics, it is essential that a probability sample be selected. We can only generalize from the sample results to populations if we use probability samples.

Since it is likely that the data from most countries will be used at some stage for producing national estimates, it is recommended that probability samples be selected by all countries. This means that any country which would like to use a judgment sample should discuss this issue with the Sampling Committee.

Table C.1 summarizes a range of common sampling designs, and indicates their suitability for different analysis purposes.

The following list defines the terms used in Table C.1:

pps schools refers to the random selection of schools with a probability proportional to size; that is, a probability proportional to the number of students in the defined target population at that school.

srs schools refers to a simple random sample of schools.

srs fixed cluster of students refers to a group of students of a fixed size (for example, 25) drawn as a simple random sample from all the students in the defined target population in the selected school.

srs variable cluster of students refers to a group of students drawn as a fixed proportion (e.g. one half) from all the students in the defined target population in the selected school; consequently the size of the cluster varies from school to school.

Table C.1 Common Sampling Designs and Suitability for Different Analysis Purposes

	Unit of analysis			
Sampling design	Between students	Between classes	Between students w/i classes	Between classes w/i schools
pps schools				
P1 srs fixed cluster of students	✓	x	x	x
P2 srs variable cluster of students	✓	x	x	x
P3 one class of students	✓	✓	✓	x
P4 more than one class of students	P	✓	P	✓
srs schools				
S1 srs fixed cluster of students	P	x	x	x
S2 srs variable cluster of students	✓	x	x	x
S3 one class of students	P	✓	P	x
S4 more than one class of students	P	✓	P	✓

Key: ✓ This analysis is possible without serious problems.

P Problems are associated with this analysis.

x This analysis cannot be undertaken.

one class of students refers to an intact class of students drawn at random from the selected school.

more than one class of students refers to more than one intact class of students drawn at random from the selected school.

Where the student is regarded as the unit of sampling and analysis, the the designs shown in Table C.1 are known as two-stage sample designs, with schools selected at the first stage (primary sampling units: psu's) and students selected within schools at the second stage (secondary sampling units: ssu's). However, this terminology is often confusing where a sample is designed to enable data to be processed at different levels of analysis, and will not be employed further in this Sampling Manual.

There is no single design which is suitable for providing data at the four indicated levels of analysis. Each country must select the design which is best suited to the analyses in which it is particularly interested.

The following section discusses the eight sampling designs in Table C.1.

Design P1. The simplest design for between students analyses involves a pps selection of schools and a srs fixed cluster of between 20 and 30 students. The resulting sample is self-weighting for all strata which have the same sampling fraction. Where particular strata or super-strata have different sampling fractions, it is relatively easy to construct weighting systems to compensate for these differences. However, this design cannot easily be used for between classes analysis (unless there is an adequate number of students in the cluster who were selected at random from the particular classes identified for the analyses).

This design is suitable for cross-sectional designs at Population A level. It is also suitable at Population B level if a sampling frame (list of schools) can be constructed with good estimates of the number of students in this target population; that is, the number of final-year secondary students undertaking defined mathematics courses.

Design P2. If a variable cluster of students is selected, it is necessary to weight students so that the effective size of each cluster is equal; that is, this design is more complex than P1 without any compensatory advantages. It is also difficult to estimate or control the total sample size.

Design P3. This design selects a single class which can be regarded as an intact cluster of students rather than a randomly selected cluster from within a school. The single class may be selected at random from the set of classes which falls within the target population for that school.

We recommend that a particular class should be selected, as part of the original pps selection of the school. Details of the procedure are set out later in this manual. In this case, the selection of the class may be regarded as equivalent to the simple random selection of a class from the population of classes within the defined target population.

For between students analyses, it is necessary to compensate for the differing number of students in the class by weighting procedures, so that each class has an equal effective size of, say, 20 students. An alternative procedure, which is not recommended, would be to eliminate at random the data for all except 20 students from the class group.

For between student analyses based on intact classes it is necessary to allow for the effects due to clustering by the incorporation of appropriate values for rho (the intraclass correlation). The value of rho will usually be higher for intact classes than for random clusters of students within schools, as we have already noted.

Design P4. For between classes analyses, this sample design is analogous to Design P1 for between students analysis; that is, we have a srs fixed cluster of two classes for each selected school in the stratum (or three classes or four classes, etc).

This design is difficult to execute for between student analyses because of the detailed weighting scheme which would need to be prepared for each school. Further, for many countries, a considerable proportion of target schools may only have one class of students which falls within the defined target population.

Some countries may wish to adopt this design because they intend to examine school effects (between classes within schools). If these countries also wish to undertake between students analyses, these should be based on only one class per school, chosen at random as in Design P3. This would facilitate the preparation of weighting procedures.

In other words, if Design P4 is selected, we recommend that the selection of two or more classes per school be undertaken in two stages:

a Select one class per school as in Design P3. Identify this class carefully for use in the between students analyses.

b Select the additional class or classes per school by an appropriate random selection procedure. The additional class or classes should be used for the between classes analyses but not for the between students analyses.

Design S1. This is an unsuitable design for between students analyses, since national estimates can only be made by means of complex weighting procedures applied to the data from each school.

Design S2. For Design S2 it is necessary to draw a simple random sample of schools from each stratum, and take a fixed proportion of students (constant sampling fraction) from each of the selected schools in the stratum.

Where there is a large range in the size of the target population in each stratum, there will also be a large range in the resulting sample size for each school. In this case it is highly desirable to separate schools into strata prior to the selection process. Each stratum should contain schools of similar size, so that different sampling fractions are applied to each stratum.

This design will probably be the most useful design for Population B, since in most countries it is not possible to obtain estimates of the size of this target population (mathematics students) for each school. Although this design is suitable only for between students analyses, these are likely to be the major analyses at Population B level.

Design S3. This design may be used for explanatory analyses between classes. It is inappropriate for deriving national estimates since this would involve complex weighting procedures as in Design S1.

Design S4. If this design were to be used for deriving national estimates, the weighting procedures are even more complex than for Design S3. In any case, it would be desirable to identify one of the selected classes as the class from which data will be used for national estimates, as in Design P4.

3 Designs omitting initial selection of schools

Some countries may have very detailed national statistics, such that they can draw a one-stage sample; that is, by selecting students or classes directly without first selecting schools.

For example, at Population A level a country may have a central record of all classes at Year 8 (8th grade) level. They could then select classes at random for their sample design.

As a further example, at Population B level a country may have a list of all the students preparing for public examinations at Year 12 (the terminal secondary grade level), together with a list of the courses being taken by each of these students. For this country it would be possible to draw a simple random sample of these students for the Population B sample. Although this would reduce the number of students needed for the sample, it would probably increase the administrative complexity.

4 Designs involving initial selection of regions

Some countries with a large number of administrative regions may wish to limit their sample to a subset of these regions. Where regions or areas are chosen as the first stage in a sampling design, the sampling errors between classes or between students will be large unless an adequate number of regions is selected.

In practice, at least ten regions should be selected at the first stage of such a three-stage sample design.

It is recognized that, for administrative or financial reasons, some countries may select only a small number of regions. It must be carefully noted that the results derived from the samples for these countries should not be generalized to obtain national estimates for the countries.

For a cross-sectional study, regions should be selected at random with a probability proportional to the size of the defined target population in each region. This process corresponds to the selection of schools by pps, which is described in detail below. Countries which do not have suitable education statistics could use the total population of the region as a measure of size.

5 Selection of strata

Before proceeding with the selection of schools it is necessary to specify the strata to be used in the sampling design. These strata should be mutually exclusive, and cover the entire country, or the selected regions within the country; that is, each student in the defined target population in the country, or the selected regions within the country should be in one, but only one, stratum.

As outlined in Section B, strata may be selected where the mean level of mathematics achievement is likely to be significantly different between strata. This may occur if they represent particular types of school or regions.

Where pps sampling is used it is not necessary to develop a stratum for school size. The pps procedure automatically controls for this factor. However, where srs sampling of schools is used, it will generally be necessary to establish a stratum for school size.

It is recommended that the number of strata be kept to a minimum, say six or ten strata. In any case, the maximum number of strata should not exceed 99.

It will be necessary at a later stage to collect information about the size of the defined target population in each of these strata. This information will be used for the development of weighting procedures to compensate for different sampling fractions across strata, and different response rates across strata.

6 Sampling frame

In order to proceed with the selection of schools it is necessary to have a list of schools, which we term the 'sampling frame'. For each school in the sampling frame it is desirable to have basic information for contacting the school; for example, the postal address, the name of the school principal and the telephone number. However, it is strictly necessary to have such contact information only for the schools selected in the sample in order to invite them to participate in the study.

If pps selection of schools is to be used, additional information is needed about each school. This is discussed below.

The sampling frame should take account of the distribution of schools across geographical regions. It is possible to set up separate strata for geographic regions. A more simple solution is to arrange the schools on the sampling frame for each stratum in a systematic way that reflects their geographic distribution. For example, many countries have a numeric area-code (zip-code or post-code) system for their postal system. Schools could be listed on the sampling frame in the order of these numeric codes. Schools with the same area-code could be listed in alphabetic or random order. Selection of schools by the pseudo-random method (random start, constant interval) will result in a geographical distribution of sample schools which matches the overall geographical distribution of schools.

a Sampling frame for pps selection of schools. In order to carry out pps selection of schools it is necessary for the sampling frame to include an estimate for each school of the size of its defined target population.

The accuracy of this estimate will vary from country to country, and will depend on the amount of information available from the authorities who collect educational statistics.

The following list indicates the kinds of information that may be available for the estimates of school size:

i the number of students in the defined target population (say, Year 8) for the current year,

ii the number of students in the defined target population for the previous year,

iii the number of classes of students at the defined target population level for the current year or previous years,

iv the average number of classes of students for schools of this type and size,

v the total enrolment in the school at the secondary school level for the current year or previous years, or

vi a judgment of the size of the school as large, medium or small, in which case the schools are given 'size factors' of 3, 2 or 1 respectively.

The kinds of information have been listed in decreasing order of quality, and the National Center should endeavour to use the best information it can gather. It is not necessary to use the same kind of information for each stratum, although the kind of information should be the same within each stratum.

The schools, with their associated size factors, should be listed by strata. Table C.2 sets out an example of the pps sampling frame for a stratum.

In the following example, the size factor is based on the enrolment of students. These numbers would be lower where based on the number of classes.

The column showing ticket numbers is not strictly necessary. It is included to show how each school is considered to have a set of particular 'tickets' based on its size factor, and derived from the cumulative tally of size factors within a stratum.

Where the number of students in a stratum is large, the stratum may be divided into smaller units to simplify the process of cumulation, and the subsequent selection of schools for the sample.

An alternate example in Table C.3 shows the same schools as in Table C.2 but with the number of classes as the size factor.

b Sampling frame for srs selection of schools. For srs selection of schools, it is necessary only to have a list of schools, but these should be grouped into strata by school size; for example, separate strata for large, medium and small schools.

7 Number of schools and students

The number of schools and students to be included in the selected sampling design for Population A and/or Population B should be calculated by reference to Tables B.2 or B.3. The value of rho to be used in these calculations must be chosen carefully. If typical values for the selected sampling design are not available for the country, it would be highly desirable for the National Center to analyse existing datasets to obtain a range of values of rho to guide their planning.

The same sampling fraction must be applied across all schools within a given stratum. However, it is possible to use a different sampling

Table C.2 Sampling Frame for Stratum 01: Students as Size Factor

School area code	School name	Size factor	Cumulated tally	Ticket numbers
3001	A	50	50	1-50
3002	B	200	250	51-250 *
3002	C	50	300	251-300
3003	D	300	600	301-600 *
3005	E	150	750	601-750 *
3007	F	50	800	751-800
3007	G	250	1050	801-1050*
	etc.	etc.	etc.	etc.
Stratum total	50 (schools)	8700 (students)		

* indicates 'winning' tickets, described later in the manual.

Table C.3 Sampling Frame for Stratum 01: Classes as Size Factor

School area code	School name	Size factor	Cumulated tally	Ticket numbers
etc.	A	2	2	1-2 *
	B	6	8	3-8
	C	2	10	9-10
	D	9	19	11-19*
	E	4	23	20-23*
	F	2	25	24-25
	G	7	32	26-32*
	etc.	etc.	etc.	etc.
Stratum total	50 (schools)	250 (classes)		

* indicates 'winning' tickets, described later in this manual.

fraction for each stratum. In this case, in order to derive the national estimates it will be necessary to apply weighting procedures to the strata to compensate for the different sampling fractions.

8 Selection of schools by pps method

Let us consider the hypothetical Country X from which the data in Tables C.2 and C.3 were obtained. Country X has a defined target population (Population A) of 70,000 students.

Suppose it was decided to draw a two-stage sample involving 224 schools at the first stage and a srs cluster of 25 students from each school at the second stage. If we assume a value of rho = 0.2, then:

$\text{deff} = 1 + (\bar{n} - 1).\text{rho} = 1 + (25 - 1)(0.2) = 5.8$

total sample size $n_c = 224 \times 25 = 5{,}600$

simple equivalent sample $n^* = \frac{n_c}{\text{deff}} = \frac{5{,}600}{5.8} = 966$

standard error $se(\bar{x}) = \frac{s}{\sqrt{n^*}} = 0.03s$

sampling fraction $= \frac{n}{N} = \frac{5{,}600}{70{,}000} = 0.08$

By referring to Table C.2, we see that Stratum 01 has 8,700 students in 50 schools.

If we apply the same sampling fraction of 0.08 to each stratum, we obtain for Stratum 01:

number of students in sample for Stratum 01 $= n_1 = 0.08N_1 = (0.08)(8{,}700) = 696$

Since we take 25 students per school, this leads us to expect to select 696/25 = 27.8 schools from Stratum 01. In practice, this means we will select 27 or 28 schools, and the corresponding number of students in the designed sample will be 675 or 700. We will not know this until we actually select the schools, as described later in the Sampling Manual.

Suppose instead that Country X decided to draw a two-stage sample involving 224 schools at the first stage and a srs cluster of one intact class per school at the second stage.

From Tables C.2 and C.3 we see that the average class size in Stratum 01 is given by:

$$\frac{\text{number of students}}{\text{number of classes}} = \frac{8,700}{250} = 34.8 = 35$$

If we assume a value of rho = 0.4, then:

$$\text{deff} = 1 + (\bar{n} - 1).\text{rho} = 1 + (25 - 1)(0.4) = 14.6$$

$$\text{total sample size } n_c = 224 \times 35 = 7,840$$

$$\text{simple equivalent sample } n^* = \frac{n_c}{\text{deff}} = \frac{7,840}{14.6} = 537$$

$$\text{standard error} = se(\bar{x}) = \frac{s}{\sqrt{n^*}} = 0.04s$$

$$\text{sampling fraction} = \frac{7,840}{70,000} = 0.112$$

If we apply the sampling fraction of 0.112 to Stratum 01, we find from Table C.2 that:

number of students in sample for Stratum 01 $= n_1 = 0.112N_1 = 0.112 \times 8,700 = 974$

Since we assume an average class size of 35 students, this leads us to expect to select 974/35 = 27.8 classes from Stratum 01. This equals 27.8 schools with one class per school. In practice, we will select between 27 and 28 schools (classes) for this stratum.

Alternately, we could apply the sampling fraction of 0.112 for Stratum 01 to the data in Table C.3, where the size factor is the number of classes. We obtain:

number of classes in sample for Stratum 01 $= n_1 = 0.112N_1 = 0.112 \times 250 = 28$

That is, we expect to select 28 classes from Stratum 01, which corresponds to 28 schools with one class selected per school.

9 Selection of schools by srs method

Suppose Country X with 70,000 students in the defined target population decided to draw 100 schools by the srs method, with an average of 35 students per school to give a national sample of 3,500 students.

The sampling fraction for the country overall would be:

$$\frac{n}{N} = \frac{3,500}{70,000} = 0.05$$

For Stratum 01, the expected sample would be:

number of students in sample for Stratum 01 $= n_1 = 0.05 \times 8,700 = 435$

Suppose we chose to select $\frac{1}{5}$ of the schools. Let us refer to Table C.2 (although for srs sampling we would not need to have size factor information in advance).

Suppose our srs selection method chooses School A and School F. We would then select at random $\frac{1}{4}$ of the students in these schools; that is, 12.5 students in each of these schools, rounded to 13 students each.

Alternatively, if we chose School B and School G, we would then select 200/4 = 50 students from School B and 250/4 = 63 students from School G.

Over the whole sample for this stratum, we would hope that the number of students selected for the sample tended to 35, although this number cannot be controlled by this sampling method.

In order to obtain the required sample for Stratum 01 we need to apply the sampling fraction of 0.05 or 1/20. We can do this in various ways.

sampling fraction for Stratum 01 = ($\frac{1}{20}$ of the schools) x (all of the students in each school)

OR = ($\frac{1}{10}$ of the schools) x ($\frac{1}{2}$ of the students in each school)

OR = ($\frac{1}{5}$ of the schools) x ($\frac{1}{4}$ of the students in each school)

In general,

sampling fraction for students = (sampling fraction for schools) x (sampling fraction for students within schools)

Note that this method may be necessary for Population B if we do not have information about the number of defined target population students (terminal year mathematics students) for each school in the sampling frame before we draw the sample.

10 Procedures for selection of schools by pps method

Consider our hypothetical Country X. The calculatic . given above showed that we need 28 schools for Stratum 01. In order to draw these schools at random with a probability proportional to size we allocate a number of 'tickets' to each school. The number of tickets for a school is given by its size factor. In Table C.2, School A has 50 students, and is assigned tickets 1 to 50. School B has 200 students, and is assigned tickets 51 to 250, and so on. In Table C.3, tickets are assigned on the basis of the number of classes. School A has tickets 1 to 2, School B has tickets 3 to 8, and so on.

If we refer to Table C.2 data, the total number of tickets available for Stratum 01 is 8,700. We need to identify the 28 ticket numbers which will select the schools to be included in the sample - the 'winning' tickets.

The winning tickets can be chosen by reference to a table of random numbers, selecting 28 numbers between 1 and 8,700. Alternatively, we can use the pseudo-random method of random start - constant interval. In order to select 28 winning tickets, the constant interval would be given by:

$$\frac{8,700}{28} = 311$$

We then select the random start, which is a number between 1 and 310 chosen from a table of random numbers; for example, let the random start = 93. The winning tickets for Stratum 01 would be:

93, 93 + 311 = 404, 404 + 311 = 715, 1,026, 1,337, etc.

From the sampling frame shown in Table C.2, we see that Schools B, D, E and G had winning tickets, which selected their schools for the sample.

Consider also Table C.3 data, where a different size factor was shown. The total number of tickets for Stratum 01 is 250. The constant interval is given by 250/28 = about 9. Suppose the random start number is 2. The winning tickets are then:

2, 2 + 9 = 11, 11 + 9 = 20, 29, 38, etc.

These winning tickets would select Schools A, D, E, G, etc.

11 Procedures for selection of schools by srs method

From the sampling frame for the stratum, select the required number of schools as given by the sampling fraction for schools.

Suppose this sampling fraction is $\frac{1}{10}$. By the method of random start - constant interval, we selection a random start equal to, say, 3. The schools to be selected are given by:

3, 3 + 10 = 13, 13 + 10 = 23, 33, 43, etc.

That is, we select the 3rd school, the 13th school, etc. from the sampling frame.

12 Invitation to selected schools

Schools selected in the sample must then be invited to participate in the study. Details of this procedure are included in Administrative Manual 1. From each school, information is obtained to enable the National Center to select the classes or students for the sample. These procedures are discussed below.

During the IEA Six Subject Survey, which was limited to cross-sectional data gathering, the sampling losses in the execution of the sampling design were such that ten out of 20 countries had a response rate of less than 80 per cent, and seven of these ten countries had response rates of less than 60 per cent (Peaker, 1975: 36). Since we are attempting a more ambitious data gathering operation, it is very desirable to obtain an excellent response rate. It is difficult to apply powerful analysis to poor data which may have a large and unknown degree of response rate bias.

It is possible that some schools may be selected to participate at both Population A and Population B levels. We suggest that invitations to participate at both levels be sent to these schools. We recognize that such schools may decline at one (or both) levels and will require replacement, as described below. However, this is better than undertaking the replacement at the National Center prior to extending the double invitation to these schools.

13 Replacement of schools

It is likely that some schools selected to participate in the study will decline the invitation to do so. It is necessary to decide on a rule to guide the selection of replacement schools.

Strictly speaking, the use of any replacement schools reduces the quality of the probability sample. If the number of replacements is not large, the effects are not serious in practice. However, if there is a large number of replacements, or if there is a series of replacements for the replacements, the quality of the sample is likely to be reduced. Every effort should be made to encourage a very high response rate from the schools initially selected.

In any case, it is necessary to select a rule for the selection of replacement schools. One system is to draw two independent samples for each stratum, each of which covers the complete sample design. One of these samples is selected at random as the 'main' sample, and the other as the 'replacement' sample. The number of schools in both of these samples will be essentially the same. The rule for replacement would then be:

> If the nth school in the main sample does not agree to participate, it is replaced by the corresponding nth school in the replacement sample.

Another system, which involves less work in the selection of schools, is to return directly to the sampling frame, and apply the following replacement rule:

> If the nth school in the sample does not agree to participate, it is replaced by the next school on the original list of schools (sampling frame) for that stratum.

For schools arranged in the sampling frame according to a systematic geographical distribution, this method ensures that replacement schools are similar to the original schools to the extent that schools in adjacent geographical areas are generally similar.

14 Selection of students: srs cluster

Where a simple random sample of students is to be selected from the school, the school must supply information to enable the National Center to select the students. This applies where a srs cluster of fixed size is to be drawn or where a sampling fraction is to be applied (for example, a half or a quarter of the students).

Tables C.4 and C.5 set out examples of Student Sampling Information Forms for use at Population A and Population B levels respectively.

The structure of Table C.4 assumes that students will be selected on the basis of their birth dates. We suggest the following procedure. Choose into the sample all students born on the 1st day of any of the twelve months covered by the definition of a 13-year-old student. Then choose students born on the 2nd day, 3rd day, etc. until the required number of students is achieved. For the last day needed to complete the sample for each school it will usually be necessary to use random procedures to eliminate the names of some students in order to obtain the required number of students.

When the completed Student Sampling Information Forms are returned to the National Center, they should be checked to eliminate the names of any students with invalid birth dates. When the completed tests and questionnaires are returned to the National Center, the birth date of each sample student should again be checked to ensure that only validly selected students were included in the sample.

The structure of Table C.5 assures that a fixed proportion of students will be selected, as given by the sampling fraction for students within schools.

Suppose the sampling fraction were $\frac{1}{4}$. We suggest that a random start - constant interval method should be used. The constant interval in this case = 4. The random start will be between 1 and 4; say = 2. The selected students will be given by the numbers:

$$2, \; 2 + 4 = 6, \; 6 + 4 = 10, \; 14, \; 18, \; \text{etc.}$$

That is, choose the 2nd student, 6th student, etc. from the list supplied by the school.

In small schools (with fewer than 60 students in the target population, say), the National Center may offer to test all the students taking mathematics at that level, to avoid administrative problems in the schools. This has implications for the number of student test booklets and other instruments to be prepared. In extremely small schools, composite classes may exist. In this case, the principal should be given guidelines to identify the students who belong to the defined target population. If the principal of a small school requests that all the students at the Population A level should be tested, the data for all these students should be returned to the National Center. Only the data from the list of students in the sample should be forwarded to the International Center. If the National Center decides to send feedback information, such as test scores, to the schools it may include the data for all of these students or only for the students in the IEA sample.

If confidentiality of students' names is an important issue, the principal could be requested to keep his own list of classes and students, but assign a three-digit code number to each student. He would then send the list of code numbers to the National Center. The National Center would allocate its own code numbers to the students it selected for the sample.

15 Selection of students: intact class

Some sampling designs will require the selection of one intact class per school. In order to select this class, it is necessary to obtain information about the classes with students in the defined target population in the selected schools. Table C.6 sets out an example of a Class Sampling Information Form which could be used to obtain this information at Population A level.

a srs method. The required class can be selected at random from the list supplied on the Class Sampling Information Form

b Interval method: students as size factor. The particular class selected for the sample can be identified more carefully by the interval method.

Let us suppose School B was selected, and that it had 200 Population B students in 6 intact classes, as shown in Table C.2. The 'tickets' assigned to the school were 51 to 250, and the winning ticket was 93. This winning ticket was the 43rd of the school's 200 tickets (given by 93 - 50 = 43).

Table C.4 Student Sampling Information Form (Population A)

Please enter on this form the name of each student in your school at (grade) level whose date of birth was between (date) and (date).

For each student, please enter the name, number or other identification of the class-group to which each of these students belong, the sex of each of these students, and the date of birth of each of these students.

Name of student	Class name/ number identification	Sex	Date of birth
1			
2			
3			
etc.			
(25 or 30 spaces per page)			

If the space on this form is insufficient, please continue on copies of the form or additional sheets of paper.

Table C.5 Student Sampling Information Form (Population B)

Please enter on this form the name of each student in your school at (grade) level who is studying mathematics in any one of the courses listed in the definition of Population B.

For each student, please enter the name, number or other identification of the class-group to which each of these students belong, and the sex of each of these students.

Name of student	Class name/ number/identification	Sex
1		
2		
3		
etc.		
(25 or 30 spaces per page)		

If the space on this form is insufficient, please continue on copies of the form or additional sheets of paper.

SIMS Sampling Manual, Section C, page 21

Table C.6 Class Sampling Information Form (Population A)

Please enter on this form the name, number or other identification of each class in your school at Year 8 level. For each class, please also enter the name of the teacher with major responsibility for teaching mathematics to this class, and the number of students in the class.

Class name/ number identification	Name of mathematics teacher	Number of students in class
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

We can apply the proportion 43/200 to the number of classes to choose the 'winning' class:

$$\text{selected class ratio} = \frac{93-50}{250-50} \times 6 = \frac{43}{200} \times 6 = 1.29$$

Any ratio between 1.01 and 2.00 would select the 2nd class on the list supplied by the school.

This method of selecting a particular class from a school selected by the pps procedure may be regarded as equivalent to a srs selection of a class from a sampling frame containing all the classes in the defined target population.

c Interval method: classes as size factor. Let us consider the case where the size factor used for assigning tickets to schools was based on the number of classes.

Let us suppose that School D with 9 intact classes was selected, as shown in Table C.3. The tickets assigned to the school were 11 to 19, and the winning ticket was 11. Following the procedure used above:

$$\text{selected class ratio} = \frac{11-11}{19-11} \times 9 = 0$$

Any ratio between 0 and 1.00 would select the 1st class on the list supplied by the school.

d Interval method: poor measures as size factor. Let us consider the case where the size factor used for assigning tickets to schools was based on weak measures of size; for example, large = 3, medium = 2 and small = 1.

For a school with one ticket. Choose one class at random from the list of classes provided by the selected school on the Class Sampling Information Form.

For a school with two tickets. Divide the list of classes into two equal parts (1) and (2). If the winning ticket was the first of the two assigned tickets select a class at random from part (1). If the winning ticket was the second of the two assigned tickets, select a class at random from part (2).

For a school with three tickets. Divide the list of classes into equal parts (1), (2), and (3). If the winning ticket was the first of the three assigned tickets, select a class at random from part (1), and so on.

We recognize that it may be difficult to identify intact classes in schools which use different forms of organization. However, we assume that there will be one teacher with major responsibility for an identifiable group of students within the defined target population who are working together at the time of testing. National Centers in countries where such problems are likely to arise should provide guidance to the schools to assist the identification or formation of 'intact' classes for the purposes of this study.

In some schools the intact classes may contain few students; say, less than 10 students. Such small classes should not be omitted from the sample, but each student may need to complete several of the

Table C.7 Sampling Design Summary

Stratum number	Population		Sample		Sampling fraction
	Schools	Students	Schools	Students	
01	50	8,700	28	700	0.08
02					
etc.					
Total		70,000	224	5,600	0.08

rotated tests in order to provide stable estimates of mean scores on the rotated tests for that class.

16 Selection of students: more than one intact class

Some sampling designs will require the selection of more than one intact class per school. Select one intact class initially by one of the methods suggested in the above section, and identify this class carefully. Then select at random the remaining class or classes required.

17 Sampling design summary

A summary of the sampling design should be set out in the form of a table; for example, as shown in Table C.7.

SECTION D

PREPARATION OF SAMPLING DESIGN: LONGITUDINAL STUDY

The longitudinal study involves the administration of an initial testing program near the beginning of a school year and a final testing program near the end of that year. This means that the selection of schools must be done during the previous year, although the selection of classes may be done very early in school year. A longitudinal sampling design also requires a special effort to ensure that a high proportion of the initial respondents is included in the final testing program.

This section should be read in association with the previous Section C. It will discuss aspects of the preparation of a sampling design for a longitudinal study only to the extent that it differs from a cross-sectional study.

1 Selection of schools and classes

For the longitudinal study, the intact class is the unit of sampling, and also the main unit of analysis.

For most countries this will involve the selection of schools followed by the selection of classes within schools. Some countries may have a complete list (sampling frame) of all the classes in the defined target population, and these classes may be sampled directly. Other countries may wish to sample regions at the first stage, followed by the selection of schools then classes.

Although more care is needed in generalizing results from a judgment sample, the administrative costs involved in using a judgment sample are usually lower. The judgment sample may be selected from schools close to the National Center, which may make it easier for the National Center to encourage teachers to complete their teacher questionnaires.

One approach to the preparation of a judgment sample is to set up a two-dimensional grid. One dimension would list the different types of schools, and the other dimension would list the range of teaching styles used in the country for the teaching of mathematics. It is recognized that some countries may not be able to prepare a classification system for this second dimension, and that the judgment sample will be based on only one dimension.

Countries which are interested only in the relationships between explanatory variables and mathematics achievement may draw a judgment sample of schools and classes. If it is likely that the country will use sample results for the estimation of national population parameters, a probability sample should be used.

For both probability and judgment samples the number of classes should be fairly high to enable multivariate analyses to be undertaken, as discussed in Section B. There should be a minimum of 100 classes; that is, one class from each of 100 schools. Preferably, there should be at least 200 classes, one each from 200 schools.

Schools for the sample will need to be selected during the school year prior to the one in which the testing programs are to be conducted. The agreement of the school principals to participate in the study must be obtained prior to the year of testing. Where necessary replacement schools must also be arranged prior to the year of testing.

Some of the selected schools may be able to complete the Class sampling information form prior to the year of testing so that classes can be selected for the sample prior to the year of testing. For other schools this information may not become available until early in the year of testing. In this case, the National Center should have all their administrative arrangements ready to obtain the information as soon as possible in the year of testing, and to select the classes for the sample.

Where a probability sampling design is being used, the selection of an intact class or classes from the selected schools should follow the procedures given in Section C. For a judgment sample, classes should be selected by judgment, although it is desirable to use classes where the teachers are co-operative about including their classes in the study.

For a two-stage longitudinal study, the selected students fall into one of four categories:

	Pre-test participant	Post-test participant
I	yes	yes
II	yes	no
III	no	yes
IV	no	no

We need to maximize the number of respondents in Category I, since it is only for these students that we can assess growth in mathematics achievement. National Centers should ensure that useful data are obtained from all students in each class for both the pre-test and the post-test. Loss of participants at either stage will reduce the number of Category I respondents.

SECTION E

ACTION SCHEDULE

The preparation of a sample design and the selection of a sample generally takes many months, and is undertaken in parallel with the administrative aspects of the study. It is crucial for each National Center to prepare an action schedule that sets out all the deadlines that must be met for the study.

The following schedule sets out the general range of activities to be undertaken, and the amount of time needed. Each National Center must decide on the deadline dates for each stage or activity. The schedule must also allow time for contact with the Second IEA Mathematics Study Sampling Committee, since at various stages their approval of the sample design is necessary for countries intending to participate in the study.

The following general schedule of activities covers both Population A and B although it will be necessary to prepare separate specific schedules for each population for countries participating at both levels. The schedule assumes that there will be an initial proposed sample design submitted to the SIMS Sampling Committee for its examination. The Sampling Committee may make suggestions for revision of the design so the schedule must allow time for such revision and the submission of the revised design to the Sampling Committee.

As an example, the following schedule shows the deadline dates for a study to be conducted in March 1980. Countries with different testing dates should prepare appropriate schedules.

Table E.1 Action Schedule

Action	Deadline for action
Selection of testing stages and dates for testing (a) one stage (post-test only) (Population A or B) (b) two stage (pre-test and post-test) (Population A only)	April 1979
Definition of target population in specific terms for this country.	April 1979
Preparation of basic national population statistics for this target population (using latest available data).	April 1979
a Number of schools (by administrative strata)	
b Number of students (by administrative strata)	
c Age distributions	
d Grade (Year level) distributions	
(Note: The time needed will depend on the availability of national statistics. Where national statistics are not available, obtain the best possible estimates.)	
Identification of the data which will be available for constructing the sampling frame.	April 1979
Identification of strata available for the sample design.	April 1979
Preparation of proposed sample design.	May 1979
Submission of proposed sample design to SIMS Sampling Committee and return of comments.	June 1979
Preparation of revised sample design.	June 1979
Submission of revised sample design to SIMS Sampling Committee and return of approval.	July 1979
Submission of proposed sample design to national authorities for preliminary approval.	June 1979
Submission of revised sample design to national authorities for approval.	July 1979
Collection of data for the sampling frame.	June 1979
Preparation of the sampling frame.	July 1979

Table E.1 Action Schedule (continued)

Action	Deadline for action
(Note: The preparation of the sampling frame can take a considerable amount of time for typing school names and addresses, and tallying student enrolment data.)	
Selection of schools from the sampling frame.	August 1979
Invitation to selected schools and return of response.	September 1979
Selection of replacement school, invitation to participate, and return of response.	October 1979
Selection of students or classes within schools.	November 1979
Preparation of lists of students within schools.	January 1980
(Note: This may require a considerable amount of time for typing.)	
Despatch of testing materials to schools.	February 1980
Testing date	March 1980

SECTION F

QUESTIONNAIRES

Questionnaire for countries participating at Population A level

1 What are the dates for your testing program(s)?

a one-stage testing (post-test only) date:

b two-stage testing date of pre-test:

date of post-test:

2 Please indicate the types of analyses in which your country is interested.

	cross-sectional (national estimates)	longitudinal (explanatory model)
between students		
between classes		
between students within classes		
between classes within schools		

3 For students in normal schools, what is the number and percentage of students of age 13 in each Year level (grade level)?

Please name the source of this information.

4 What is the official date for the definition of age 13 for the above percentages? That is,

students of age 13 years 0 months to 13 years 11 months inclusive on ____________ (date)?

5 Please express this definition also in terms of actual date of birth. That is,

students born between __________ (date) and __________ (date)

6 What is your proposed defined target population for Population A (the target population)?

7 What students in the IEA general definition of Population A have been excluded from your national definition of the target population for Population A (that is, the excluded population)?

8 What strata do you propose to use for your sampling frame, and hence for your sample?

9 What statistics are available for the construction of the sampling frame; that is, the list of schools together with estimates of the size of the target population in each school?

Please indicate the source of the statistics.

As an example, please send a couple of pages of your proposed sampling frame, including school target population estimates.

10 What marker variables do you plan to use in your country?

Please name the source of the statistics for these marker variables.

11 Please describe your proposed sampling design.

a method for selection of schools,
b method for selection of students (or classes within schools),
c number of schools, and
d number of students or classes.

12 For your proposed sample design, what is your estimated sampling error (for the analyses in which you are interested)? For example:

a between students for the country overall for cognitive total test and sub-test means (national estimates),
b between students for the country overall for individual item percentages, and
c between classes for the country overall for regression coefficients or path coefficients in explanatory analyses.

13 What are the specific deadline dates for your schedule for the sampling design and execution?

Please complete the details in Section F of this Sampling Manual.

14 What is the name of your National Sampling Co-ordinator - the person in your country with whom Dr Rosier will communicate on sampling matters?

Please give name, address, cable/telegraphic address (if applicable) and telephone number (with area/regional codes if applicable).

Questionnaire for countries participating at Population B level

1 What are the dates for your testing program?

2 What is your proposed defined target population for Population B (the target population)?

3 What students in the IEA general definition of Population B have been excluded from your national definition of the target population for Population B?

Note: The following questions may be answered for the country overall, or for separate key strata if there are large differences between these strata.

4 What is the number and percentage of all students at the terminal secondary grade (Year level) at each of the following age levels:

less than age 17, age 17, age 18, age 19, age 20, more than age 20?

Please state the source.

5 What is the official date for the definition of these ages in the national statistics?

6 What is the number of young persons in the total population of the country at the following age levels:

age 16, age 17, age 18, age 19, age 20?

7 What is the percentage of students in the terminal secondary level who are studying mathematics as a substantial part of their academic curriculum (as in the IEA general definition of Population B)?

8 What strata do you propose to use for your sampling frame, and hence for your sample?

9 What statistics are available for the construction of the sampling frame; that is, the list of schools together with estimates of the size of the target population in each school?

Please indicate the source of the statistics.

As an example, please send a couple of pages of your proposed sampling frame, including school target population estimates.

10 What marker variables do you plan to use in your country?

Please name the source of the statistics for these marker variables.

11 Please describe your proposed sample design?

a method for selection of schools;

b method for selection of students (or classes within schools),

c number of schools, and

d number of students or classes.

12 For your proposed sample design, what is your estimated sampling error (for the analyses in which you are interested)? For example:

a between students for the country overall for cognitive total test and sub-test means (national estimates),

b between students for the country overall for individual item percentages, and

c between classes for the country overall for regression coefficients or path coefficients in explanatory analyses.

13 What are the specific deadline dates for your schedule for the sampling design and execution?

Please complete the details in Section 00 of the Sampling Manual.

14 What is the name of your National Sampling Co-ordinator - the person in your country with whom Dr Rosier will communicate on sampling matters?

Please give name, address, cable/telegraphic address (if applicable) and telephone number (with area/regional codes if applicable).

SECTION G

REFERENCES

Cattell, R.B.
1952 Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist. New York: Harper and Row.

Frankel, M. N.
1971 Inference from Survey Samples: An Empirical Investigation. Ann Arbor, Michigan: Institute for Social Research, University of Michigan.

Guilford, J.P. and Fruchter, B.
1973 Fundamental Statistics in Psychology and Education. New York: McGraw-Hill. 5th Edn.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G.
1953 Sample Survey Methods and Theory: Volume I, Methods and Applications. New York: John Wiley and Sons.

Husén T. (ed.)
1967 International Study of Achievement in Mathematics. Stockholm: Almqvist and Wiksell/New York: John Wiley and Sons. 2 vols.

Kerlinger, F.N. and Pedhazur, E.J.
1973 Multiple Regression in Behavioral Research. New York: Holt, Rinehart and Winston.

Kish, L.
1957 'Confidence intervals for clustered samples'. American Sociological Review, 22, 154-165.

Kish, L.
1965 Survey Sampling. New York: John Wiley and Sons.

Peaker, G.F.
1975 An Empirical Study of Education in Twenty-one Countries: A Technical Report. Stockholm: Almqvist and Wiksell/New York: John Wiley and Sons.

Ross, K.N.
1978 'Sample design for educational survey research'. Evaluation in Education: International Progress, 2, 2, 105-195.

Tatsuoka, M.M.
1970 Discriminant Analysis - The Study of Group Differences. Selected Topics in Advanced Statistics, An Elementary Approach, No. 6. Champaign, Illinois: Institute for Personality and Ability Testing.

*U. S. GOVERNMENT PRINTING OFFICE 1987. 175-656/60243