CHAPTER 2

IEA, Populations and Sampling

International Project for the Evaluation of Educational Achievement (IEA)

The data used in this study were collected by the *International Project for the Evaulation of Educational Achievement (IEA)*, and since IEA is the first large-scale international educational research project of its kind, it would seem appropriate to describe briefly its history, structure and mode of operation. A detailed report of the IEA project is given in Husén *et al.* (1967).

In the middle fifties, groups of educators and educational researchers from different countries had met at places like the UNESCO Institute for Education, Hamburg, to examine problems such as those concerned with school structures and organization, selection processes, examinations and failure in school. Two important publications emerging from some of these meetings were edited by Hotyat (1962) and Wall (1962). Throughout these meetings there was a growing awareness of the need to establish evaluation techniques which would be valid cross-nationally. At the same time, more or less independently of each other, several researchers in the United States (Anderson, Bloom and Foshay) began to consider the possibilities of undertaking such research.

In 1958, researchers from several countries came together at a meeting in Eltham, England, chaired by Dr. W. D. Wall of the National Foundation for Educational Research in England and Wales, and also at the UNESCO Institute for Education in Hamburg. At those meetings it was decided to carry out a pilot study to discover if an international research project would be administratively possible and if the results could be expected to be meaningful. Research Centres from Belgium, England, Finland, France, Germany, Israel, Poland, Scotland, Sweden, Switzerland, the United States and Yugoslavia took part. A strategic target population in those countries was the children of age 13:0 to 13:11, since this was the last point where practically all of an age group were still in school in all countries. In most cases, children of schools or areas which were known to be close to the national mean and standard deviation were tested, and thus, there was no strict probability sample. In all, 9,918 students spread over eight languages were administered tests (a total of 120 items) of reading comprehension, mathematics, science, geography, and non-verbal ability. The venture proved to be successful. Foshay *et al.* (1962) have presented some of the results of this study in a monograph.

At a meeting at the Unesco Institute for Education, Hamburg, in June 1960 it was decided to embark on a cross-national study in one subject area, where several populations within secondary education would be sampled using random probability sampling techniques and where specific testing instruments would be specially constructed. This first carefully designed study in one subject area would be known as Phase I and it was hoped subsequently to embark on further phases.

The subject chosen for the first phase of the project was mathematics. The primary reason for this choice was that most countries involved in the project were concerned with improving their scientific and technical education, at the basis of which lies the learning of mathematics. Secondly, many recent national and international surveys (as carried out by the National Science Foundation in the United States and O.E.C.D. in Europe) have re-examined the curricula and the methods of teaching mathematics and various higher branches of mathematics. Thirdly, the so-called "new mathematics" has been introduced to varying degrees in some of the participating countries. Fourthly, since the symbols of arithmetic and mathematics are, with trifling exceptions, international problems of semantics and language would be reduced.

The Research Centres which committed themselves to Phase I at the 1960 meeting were from Belgium, England, Finland, France, Israel, Japan, the Netherlands, Scotland, Sweden and the United States. It was in late 1962 and early 1963 that Research Centres from Australia and Germany entered the project. (The main persons involved from each of the Centres as well as consultants are listed in Table A1 in the Appendix). A research grant from the United States Office of Education was received in the summer of 1962 and this

covered the international costs and the United States national costs only. The representatives of the Research Centres from these twelve countries formed themselves into a Council whose main task was to agree on the overall policies of the research work. On average, they met for a week once a year. They elected a Standing Committee of five of their members and their task was, if necessary, to take major decisions between Council meetings on behalf of the Council. Furthermore a Chairman/Technical Director was elected whose task was to attend to the day to day running of the project. He was assisted by a Project Co-ordinator, who was appointed in 1962 and placed in the UNESCO Institute for Education, Hamburg.

In such a project, the lines of communication were long, and it was very important to set deadlines for various stages of the work and to adhere to them. Several languages were represented, and it was decided that the project should be conducted in English, with occasional French translation. Although there were some misunderstandings, they were fortunately rare. Lessons were learned from experience and improvements in the mode of operation were continually undertaken. A list of "lessons learned" is given in Chapter 2 of Volume I of the international publication.

Consultants were employed in the areas of mathematics education test construction and sampling, and these consultants attended all Council meetings as well as special group-work meetings, which were sometimes held between Council meetings. A great deal of group work was also carried out at Council meetings; thus, for example, further work on mathematics test construction, attitude scale construction, questionnaire construction, formulation of hypotheses and sampling took place in the early meetings. After the full testing, all members helped in writing up the outcomes of the testing of hypotheses.

In its turn, the National Centre, although using most of its own staff on the national work involved in the project, sometimes used sampling consultants. At the content analysis stage at the beginning of the project, the National Centre had to organize national committees of mathematics educators and at the coding and punching stage, they often had to employ extra coders (mostly university students).

The data were put onto magnetic tape at the University of Chicago Computation Center. Needless to say, with approximately fifty million pieces of information, this study could never have been completed without the use of a computer. That the whole project (mathematics phase) was completed within four years, even with the help of a computer, was, in itself, an enormous achievement—the work on content analysis was begun at the beginning of 1962 and the final research reports were completed at the end of 1965; this success was due to the dedication, enthusiasm and ability of all the educational researchers concerned. The data on the master and working tapes at the University of Chicago Computation Center will form a data bank which can be used by qualified research workers. A Data Bank Manual has been prepared by Richard M. Wolf (1967).[1]

The IEA Council has decided to embark on a second major phase where testing in other subject areas will be undertaken, and the frame of reference of the research will be extended in terms of the various psychological, social, cultural and economic forces involved in the process of education.

Populations Tested

One of the most difficult problems in a comparative study of this nature is deciding which populations in the different countries are, in fact, for one's purpose comparable. The pilot project (Foshay *et al.*, 1962) had focussed on the educational attainments of 13-year-olds. This group has the merit of being the highest age level at which, by law, all children are supposed to be attending school in most countries with a tradition of universal education. The 13-year-old group had distinct advantages, therefore, for an assessment of the educational standard reached by an approximation of a total age group of each country and was thus selected. Although this group was chronologically comparable, there were difficulties in that there is a wide variation between countries as to the grades in which 13-year-olds are to be found. In some countries, its members were nearly all in the same grade, while in other countries, because of retardation or acceleration policies, they could be spread over several grades. For example, in England, Scotland and Japan, approximately ninety-nine percent of a year group are to be found within the same grade, whereas in Belgium, for example, twenty-nine percent of 13-year-olds are retarded by one, two or three years. In the latter case, it was thought to be difficult in the testing programme to have all of these children brought from the different classes, and in certain cases, different schools, to the testing session. It was therefore decided to allow Research Centres to award a notional zero score to those children whom they considered to be so retarded as to be unable to attempt any of the questions in the tests. However, in most cases, all students of this age range were, in fact, tested.

A second population, which is the complement of the first population, is that consisting of all students at the *educational* level (grade level) typical of the 13-year-olds in each country. This, then, is an educational level population designed to correspond in general, to the age represented in the first population. The 13-year-old age population was designated Population 1a, and the 13-year-old grade group was designated Population 1b.

The grade group, containing the majority of 13-year-olds will, of course, be different according to the time of year chosen for testing. Take a hypothetical example of two year groups: *a*) 13–14 and *b*) 12–13 at the beginning of the school year. Then, further assume that the school year runs from April to March in the next calendar year. Thus, if testing takes place between April and September, the 13-year-old grade group which will be tested will be group *a*, but after September, will be group *b*. To avoid disparity, it was agreed that the tested group would be the grade where the majority of 13-year-olds were to be found within three months of the end of the current school year. It must be pointed out that in almost no country did Populations 1a and 1b represent students at any terminal point. Therefore, their achievements are *not* to be considered indicative of what has been achieved in a rounded-off course of study. They do, however, provide a more or less hundred percent attendance baseline against which further learning within the system of secondary education can be measured.

Another group of students who seemed of special interest were those who were just completing the pre-college or pre-university level of education. This represents a major transition point in each educational system and also is the termination of formal schooling in each country. It is also a point which can be said to be that where the

[1] A copy of the Data Bank Manual can be obtained upon request to: IEA Coordinator c/o Unesco Institute for Education, 2 Hamburg 13, Feldbrunnenstr. 70, Federal Republic of Germany.

"fruits" of education may be assessed. Obviously, however, there are important differences between countries in the composition of these groups. For example, the average age of completing pre-university education ranges from 17 years 2 months in Australia to 19 years 10 months in the Federal Republic of Germany (cf. Chapter 14 in Vol. I, Husén *et al.* 1967). Again the age at which students begin school varies from country to country, and thus the total length of schooling varies. Secondly, it can be argued that the second and third year sixth-former in an English state school is not the equivalent of an American 12th grader or even of a Swedish *studentexamen* student. Apart from different lengths of schooling, the selection process which has taken place in each of these systems is very different in terms of grade-repeaters and drop-outs, and the number or the percent of a year group in this pre-university year also differs from country to country. Thirdly, the number of subjects studied in the pre-university year ranges from an average of three in England to nine or more in some European countries. Thus, there are differences in the structure of this transition point from one country to another, and this must be borne in mind in the interpretation of the results. However, it was decided that the advantages of working at the pre-university major terminal point appeared to outweigh the disadvantages of lack of comparability, so this population was chosen. It was divided into two sub-populations on the basis of the curriculum being followed. One sub-population consisted of those taking mathematics as a major subject. The second group was made up of those who were not taking mathematics or for whom mathematics was a minor and subsidiary part of their programmes. In most cases the two groups belonged to different sections or tracks of the pre-university school.

Between the 13-year-old level and the pre-university year, there are various major terminal points in the school systems—e.g. end of compulsory school ranging, for example, from 14 years in Germany to 16 years in France, Sweden and the United States, and major examination points such as the G.C.E. "O" level in England. Thus, in some countries these populations represented students terminating their education at the intermediate level, and in other countries they represented a kind of half-way point between the lower and the pre-university populations. It was decided that countries could choose the population(s) they wished to test at these intermediate points. The following are the formulated definitions of the target populations. As indicated above, it was stated that testing should take place within three months of the end of the academic year. The mathematics tests (see Chapter 3) given to the students in each population are given in parentheses:

Population 1a:

All students who are aged between 13:0–13:11 years at the date of testing. This means that all types of schools with students of this age should participate and be represented according to their proportions of students from the population defined.
(These students were to be given Mathematics Tests A, B and C.—See page 42.)

Population 1b:

All students at the grade level where the majority of students of age 13.0–13.11 are found.
(These students were to be given Mathematics Tests A, B and C.)

Intermediate Populations (Optional):

These target populations were defined by the countries testing at these levels. It was desirable, however, that, where possible, these populations should be taken at points which, if terminal, did not lead to universities or similar institutions of higher learning.
(These students were to be given Mathematics Tests 3, 4 and 5.)

Population 3:

All students who are in the grades (forms) of full time study in schools from which the universities of similar institutions of higher learning normally recruit their students. These students, in most countries, were in the grades (forms) from which a qualifying examination for the university of similar institution was taken, e.g. *Abitur, Studentexamen, 2^e partie du baccalauréat, Eindexamen, G.C.E "A" level.*

Qualification—This did not include the small proportion going to universities or similar institutions of higher learning via institutions which came under the heading of "Zweiter Bildungsweg", but the proportion of the population had to be known. Population 3 is divided into two parts:

3a: Those studying mathematics as an integral part of their course for their future training, or as part of their pre-university studies, e.g. mathematicians, physicists, engineers, biologists, etc. or all those being examined at that level. (These students were to be given Mathematics Tests 5, 7, 8 and 9.)

3b (highly desirable, but optional): Those studying mathematics as a part (complementary) of their studies and the remainder. (These students were to be given Mathematics Tests 3, 5 and 6).

Where Centres wished to sub-divide any of the above populations for national purposes, they were, of course, allowed to do so.

For purposes of coding, it was then necessary to create "operational groups". For example, in the following section, it can be seen that Groups 1 and 2 form Population 1a, and Groups 1 and 3 form Population 1b. Populations were thus broken down into operational groups as follows:

Definitions of Groups

Group 1 consists of those students aged between 13.0 and 13.11 on the day of testing in the grade (or year group) which contains the majority of students of this age.

Group 2 consists of those students aged between 13.0 and 13.11 on the day of testing who are in grades (year groups) other than that in which the majority of this age are found.

Group 3 consists of the remainder of students in the grade (year group) from which Group 1 is taken.

Group 4—Level 2(i) as operationally defined by National Centres.

Group 5—Level 2(ii) as operationally defined by National Centres.

Group 6—Level 2(iii) as operationally defined by National Centres.

Group 7—Level 3a as operationally defined by National Centres.

Group 8—Level 3b as operationally defined by National Centres.

Group 9 consists of those students who are tested with Level 3a tests, but who are possibly following a course of mathematics which does not *clearly* place them in Level 3a.

Group 0 consists of those students who are tested with Level 3b tests, but who are possibly following a course of mathematics which does not *clearly* place them in Level 3b.

Since the intermediate populations chosen for testing in the various countries vary so much, it was not thought worthwhile making international comparisons, and therefore these populations were left for national analyses and not included in the international analyses (see e.g. Pidgeon, 1967).

Sampling

Sampling Units and Stratification

The main problem in sampling was to secure a representative sample of the particular target populations in each country. Each national research centre appointed a sampling expert for its country. The IEA, on the other hand, decided that it was necessary to have one person who could devote himself more or less continuously to the task of examining the sampling plans for each target population within each country and who would enter into correspondence with the national sampling expert.

Each target population was divided into a sampled population and an excluded population. It was agreed that where there was a small category of schools that, on the one hand, would be very expensive to sample and, on the other, was so small that the results from it would make little difference to the general picture, it could be reasonably excluded. In all cases, the excluded population was negligible, except in Israel, where students who had recently immigrated from under-privileged areas were excluded.

The procedure used for sampling the "sampled population" was that of stratified random probability sampling. The unique merit of probability sampling is that the standard error of the sample as a whole or of any part of it can be determined from the internal evidence of the sample itself. All of the countries used probability sampling, except for the Federal Republic of Germany (represented by only two of the Länder—Hessen and Schleswig-Holstein) which maintained that if a random process of selection of schools was used, many of them would be unco-operative and that it would be better not to use probability sampling, but to make instead a judgement sample from schools known to be co-operative. This was, of course, for the Germans to decide, but it is clear that the internal evidence, in this case, supplies no guarantee of representativeness.

In the United States, the sampling was in three stages, the first stage being a sampling of communities, the second a sampling of schools within the selected communities, and the third a sampling of students within the selected schools. Elsewhere the sampling was in two stages, with schools as the first and students as the second stage. Multi-stage sampling is needed, because it is impracticable to sample students directly in a single stage. But a multi-stage sample is bound to be larger, in terms of students, than a simple (i.e. a single stage) sample giving standard errors of the same size.

Thus, with two stage sampling, and small sampling fractions, the variance of an estimate is

$$\frac{S}{n}+\frac{P}{nk}$$

where n is the number of schools in the sample, k the average number of students selected within each school, S the variance of school means and P the variance of students within schools. The intra-class correlation—i.e. the measure of the extent to which students in the same school resemble each other more than they resemble students in general—is ϱ where $\varrho = S/S+P$.

Consequently,

$$\frac{S}{n}+\frac{P}{nk}=(\overline{k-1}\,\varrho+1)\frac{P+S}{nk}$$

and $(k-1)\ \varrho+1$ is what Kish (1965) calls the Design Effect (Deff). In other words, it is the ratio of the size of the complex sample, in terms of students, to that of the simple equivalent sample. If the standard errors for the complex sample were calculated by applying simple random sampling (s.r.s.) formulae directly they would be too small. The proper values can be obtained by multiplying the s.r.s. estimates by the square root of Deff.

The Design Effect can be reduced by stratifying the schools, which reduces the intra-class correlation. It could be reduced further by stratifying students within schools. Stratifying schools reduces S, and stratifying students reduces P. In this study schools were stratified but the stratification of students was not attempted. In all countries schools were stratified by sex and type, and in some also by (a) geographical or administrative areas, (b) ethnic and religious groups, and (c) rural-urban locality.

Three principles of random selection of students within the schools were proposed:

1. Working through the registers with a constant sampling interval and a random start.
2. Taking in the students whose surnames begin with certain letters of the alphabet.
3. Taking in the students whose birthdays fall on certain days, spread uniformly around the year.

Research Centres were warned that, when the first principle was used, there is sometimes a strong tendency for schools who draw "unlucky" random numbers to ignore them and to choose, by judgement, a "fairer" sample. Often the headteacher replaces what he considers to be "poor" students by "good" students. This method, in fact, was not used. A warning was also given about the second method—i.e. that there may be an association between the initial letter of surnames and ethnic or other groupings within the society. If this was to produce a bias, it should be avoided. Most Centres used the third principle. This is notionally equivalent to re-defining the population so that it consists only of children with particular birthdays. There is no reason to suppose that the reduced population, defined by birthdays, uniformly spread around the year, differs from the complete population. The size of the samples varied according to the population and the country, but the number of students tested for each population varied from approximately 700 to 6000. All in all, the total number of students tested (including intermediate populations) was about 135,000.

Since the school had been used as the sampling unit, it was decided to deal each population sample into four independent sub-samples. The data were coded in terms of sub-samples and put onto the magnetic tape in this way. The splitting of the population into four independent sub-samples had various advantages. The first was that independent estimates could be obtained from each of the four sub-samples and estimates of error from the comparison of these. The second advantage was an administrative one, namely, that the answer sheets for each sub-sample could be shipped separately to Chicago. Thus, if one were lost, three still remained, whereas if all had been shipped together, all might have been lost.

It turned out that Israel and Australia did not test Population 3b and that France and the Netherlands had to be dropped because of several cases of undersampling of schools. The Federal Republic of Germany and Israel did not test Population 1a.

Weighting[2]

The actual sampling fractions differed somewhat from those suggested in the original sampling design handed in by the national sampling experts. The two main reasons accounting for this disparity were (1) the numbers of schools taken into the sample in each stratum were based on national statistics dating back as far as 1960 or 1961, and in 1964 when the testing took place, there were changes in the figures, and (2) in certain cases it was not possible to test all students drawn within schools which had been sampled. In some cases the school refused to cooperate in the study, and it was too late to take an alternate school in terms of the test programme administration within that country. The differences were not great, however, but it was the actual and not the designed sampling factions which were used to obtain the raising (weighting) factors. The weighting of each stratum sub-sample was carried out in such a way that the weighted number of students in each stratum was in exact proportion to the *total* number of students in each stratum. The estimates of error used in reporting the results in this study are those obtained from the comparison of the estimates of each of the four sub-samples. The formula used for weighting was:

$$_1n_i = {}_2n_i = {}_3n_i = {}_4n_i = \frac{nN_i}{4N} = \frac{n_i}{4}$$

Where N = the number of students in the whole target population
n = the number of students in the whole sample for the target population
N_i = the number of students in the ith stratum of that population
n_i = the weighted number of "students" in the i^{th} stratum of the sample.
$_1n_i$ = the weighted number of "students" in the first subsample.

[2] A full description of the weighting procedures used is given on pages 213 and 214 in Volume I of Husén *et al.*, 1967.

The calculations of means, standard deviations and correlations had to be carried out in terms of weighted N's.[3]

Standard Errors

Peaker in Husén *et al.*, 1967 (Volume I, Chapter 9, p. 154 *et seq.*) has explained in detail the calculations of both the simple random sampling (s.r.s.) standard errors and the complex standard errors (c.s.e.) of sampling.

Suffice it here to give Table 2.1, listing, for Populations 1a, 1b, 3a and 3b, a) factors by which the corresponding s.r.s. estimate should be multiplied to give the complex standard errors and b) complex standard errors for correlations.

The s.r.s formula for the standard error of a correlation coefficient is $(1-r^2)/\sqrt{n}$. The computer obtained the s.r.s. error for each population in each country first by comparing the average correlation coefficients obtained from four replicas (sub-samples) of a 54×54 correlation matrix with the four separate coefficients obtained and then averaging these for each matrix.

The s.r.s. formula for the standard error of a mean is, of course, $\sigma/\sqrt{N}$. To arrive at the c.s.e., the s.r.s. should be multiplied by the factor in the (a) columns in Table 2.1. It will be seen that the

[3] The following formulae for the weighted mean, standard deviation, and correlation were used:

Mean
$$\bar{X} = \frac{\Sigma w_i X_i}{\Sigma w_i}$$

Standard deviation
$$S = \sqrt{\frac{\Sigma((X-\bar{X})^2 w_i)}{\Sigma w_i - 1}}$$

Correlation
$$r_{xy} = \frac{\Sigma((X_i-\bar{X})(Y_i-\bar{Y})w_i)}{\sqrt{\Sigma((X_i-\bar{X})^2 w_i)}\sqrt{\Sigma((Y_i-\bar{Y})^2 w_i)}}$$

where w_i = the weight for the ith student
X_i = the value of the X variable for the ith student
Y_i = the value of the Y variable for the ith student

Table 2.1 *(a). Factors* by which the corresponding s.r.s. estimate should be multiplied to give the complex standard errors and (b) complex standard errors for correlations.*

	Populations							
	1a		1b		3a		3b	
Country	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
Australia	1.7	.03	1.7	.03	**2.0**	.06	—	—
Belgium	1.7	.04	2.0	.04	1.6	.07	1.9	.06
England	1.7	.03	1.7	.03	1.3	.04	1.3	.03
Fed. Rep. of Germany	—	—	**3.3**	.05	1.3	.05	1.0	.04
Finland	1.7	.05	1.8	.05	1.3	.06	1.3	.06
France	2.1	.04	3.1	.05	1.1	.06	—	—
Israel	—	—	1.8	.03	**0.9**	.07	—	—
Japan	**1.4**	.03	**1.4**	.03	1.4	.05	**2.0**	.03
Netherlands	1.7	.08	1.9	.05	1.6	.07	—	—
Scotland	**2.9**	.04	3.1	.04	1.5	.04	1.8	.04
Sweden	2.3	.04	2.5	.04	1.6	.05	**0.9**	.05
U.S.A.	1.7	.02	1.7	.02	1.6	.04	1.8	.04
Mean	1.9	.04	2.1	.04	1.4	.06	1.5	.04

* In each of the factor columns (a) the highest and the lowest factor are in bold type.

average value of the ratios in Table 2.1 is 1.7, and that no ratios are very far from this value. Consequently, the rule of taking two (complex) standard errors as the confidence limits can be replaced by the rule of taking three s.r.s. standard errors.

Summary

This chapter has presented a short account of the history, structure and mode of operation of the first large scale international educational research project—namely that carried out by the IEA in the field of mathematics, from which the data for this study are drawn. It then proceeded to describe and define the target populations chosen for study and the sampling procedures used.

During the fifties there was a growing awareness on the part of some educators, and in particular educational research workers, of the need to establish evaluation techniques which would be valid cross-nationally. Groups of educational research workers from leading research centres in Europe and the United States joined together, and in 1959 undertook a small pilot project to test the feasibility and meaningfulness of carrying out cross-national educational research (see Foshay, 1962). Encouraged by their success, they embarked on a major research in the field of school mathematics education in 1962. They received financial support for their international costs from a grant from the United States Office of Education. National Research Centres were responsible for defraying the national research costs involved in the project. Research Centres from the following countries participated: Australia, Belgium, England, Federal Republic of Germany, Finland, France, Israel, Japan, the Netherlands, Scotland, Sweden and the United States. Each Research Centre had one member on the Council of IEA, whose task it was to agree on the overall policy of the research. Interim decisions were taken by a Standing Committee (elected from the Council), or by the Chairman and Technical Director. Since all persons involved had full-time commitments in their own countries, one full-time co-ordinator was appointed by IEA and placed at the UNESCO Institute for Education in Hamburg. Consultants were also employed and most of the work was undertaken by groups at Council Meetings, but some group work was also undertaken between meetings. Instructions were issued to National Centres in circular letters and special bulletins. There was a continuous two way communication between the research workers in the National Centres and the IEA Secretariat (Chairman, Technical Director and Co-ordinator). The analyses were carried out by computer at the University of Chicago Computation Center.

Four target populations were chosen which had to be sampled and tested by each participating Research Centre. These were

(a) all 13-year-olds (Population 1a)
(b) all students in the grade where most 13-year-olds were to be found within three months of the end of the school year (Population 1b)
(c) pre-university students studying mathematics as a major subject (Population 3a)
d) pre-university students *not* studying mathematics as a major subject. (Population 3b)

It was possible for Research Centres to test major terminal populations at points intermediate to the 13-year-old and preuniversity populations, but this was optional.

Probability sampling was used with the school as the sampling unit. In the United States, three stage sampling was used (community, school and students within schools), and in other countries two stage sampling (school and students within schools). Stratification was employed so as to reduce the intra-class correlation. The factors by which the corresponding simple random sampling (s.r.s.) estimates should be multiplied to give the complex standard errors are given, together with the complex standard errors for correlations.