Postlethwaite, T. Neville

International empirical research in comparative education. An example of the studies of the International Association for the Evaluation of Educational Achievement (IEA)

Tertium comparationis 1 (1995) 1, S. 1-19

Journal für Internationale Bildungsforschung
Vol. 1, No. 1, pp. 1-19, 1995

International Empirical Research in Comparative Education: An Example of the Studies of the International Association for the Evaluation of Educational Achievement (IEA)

T. Neville Postlethwaite

Institute of Comparative Education
University of Hamburg, Germany

Abstract

Two questions are posed: 1) does IEA (International Association for the Evaluation of Educational Achievement) type research fall within the realm of Comparative Education? and, 2) does IEA research produce any generalisations about education? The conceptualisation of IEA studies and the procedures used are outlined and some results of the research are presented. It is left to the reader to decide on whether or not the 'social science empirical approach' can be considered as one of several approaches in comparative education inquiry.

1 Introduction

In 1985, Beauchamp wrote 'my own reading of the substantial body of literature on the nature and methods of Comparative education leads me to the inescapable conclusion that there is no such thing as Comparative Education, that is, Comparative Education as a field of study does not exist'. Writers such as Epstein (1988) and Schriewer (1988) have discursed on the pros and cons of such issues as the positivist approach, the relativist approach, comparison as a universal mental operation, comparison as a social-scientific method and several other approaches. It is basically the epistemological bases of inquiry that they examine. One of the approaches within comparative education is often called the behaviorist or social science approach where the research methods of the social sciences are used. Chambers (1992) has more recently criticised research on teaching in that it has been not scientific but empiricist; second, even as empiricism, such research has had little practical value. Indeed, from many writings, there is the criticism that generalisations produced by the behavioural sciences hold only for a particular period of history and in a particular culture. As time moves on or as one goes from one culture to another the generalisations no longer hold.

The aim of this article is not to enter into a philosophical discussion about the epistemological bases of inquiry but rather to describe, what type of work IEA (The International Association for the Evaluation of Educational Achievement) does as an example of inquiry and to ask the reader to judge the extent to which: 1. what IEA does falls within the realm of Comparative Education; and, 2. IEA research produces generalisations (whether short or long term). The article will briefly describe the history of IEA and its different studies, describe the methodology of IEA analyses, and finally present some results.

2 Brief history of IEA

It was in 1958 that a small group of researchers, most of whom were from the discipline of educational psychology but also including persons from sociology and psychometry, met in Hamburg, Germany to consider the possibility of undertaking a study of measured 'outcomes' in education and their determinants within and between systems of education. In the late 1950s economists, in their studies of education and economic development, tended to use the proportion of an age group graduating from various levels of schooling as an index of the educational productivity of a nation. At the same time, it was recognised that what had been learned by the same proportion of an age group in two different countries might be very different and in this sense the 'productivity' would be different even if the proportion of the age group graduating was the same. Anderson (1961), one of the persons at the Hamburg meeting, also pointed out the need for some objective measure of 'outcomes' using methods of quantitative assessment developed in educational psychology.

The Hamburg group decided to conduct a feasibility 'comparative' study which was carried out in the period 1959-61. The first paragraph in the report (Foshay, Thorndike, Hotyat, Pidgeon & Walker 1967: 7) included the sentence: 'If custom and law define what is educationally allowable within a nation, the educational systems beyond one's national boundaries suggest what is educationally possible'.

The Foshay study showed that such studies were feasible and produced meaningful results. As a result of this study the same group of researchers decided to undertake a 'proper' study using quality probability sampling and rigorous procedures for the development of questionnaires and tests. They began what was to become known as the First International Mathematics Study (FIMS). This was followed by many other studies and in 1995 there were over 50 systems of education that were participating in IEA studies. To be able to receive funding for the international costs of a study, IEA was incorporated under Belgian law in 1966. All national costs were borne by the participating systems of education, typically with funds from their ministries of education.

Four phases of Studies: The IEA studies that ensued can be divided into four phases. These are given below in terms of the title of the studies and the authors of the international reports. As well as the international reports there were many national reports and many articles in journals (Degenhart 1988).

The 1960s	First International Mathematics Study (FIMS) (Husén 1967);
The 1970s	The Six subject survey: Reading (Thorndike 1973), Literature (Purves 1973), First International Science Study (FISS) (Comber & Keeves 1973), English as a Foreign Language (Lewis & Massad 1975), French as a Foreign Language (Carroll 1975), Civic Education (Torney, Oppenheim & Farnen 1976), Technical Report (Peaker 1975), National Case Study Report (Passow, Noah, Eckstein & Mallea 1976), The IEA Six Subject Study (Walker 1976);
The 1980s	Second International Mathematics Study (Travers & Westbury 1989; Garden & Robitaille 1989; Burstein 1993), Second International Science Study (SISS) (Rosier & Keeves 1991, Postlethwaite & Wiley, 1992, Keeves 1992), Classroom Environment Study (Anderson, Ryan & Shapiro 1989), Computers in Education (Pelgrum & Plomp 1991), Written Composition (Gorman, Purves & Degenhart 1988, Purves, Lehmann & Degenhart 1992), Preschool Education (Olmsted & Weikart 1989);
The 1990s	Reading Literacy (Elley 1994), Preprimary Study (Olmstedt & Weikart 1995), Third Mathematics and Science Study (TIMSS) (in progress).

There is also the possibility at the time of writing of a Foreign Language study and a second study of Civic Education being undertaken in the 1990s. It can be seen from the many studies listed above that there was a demand for this kind of work from the part of the ministries of education because without the participation of the many ministries of education there would have been no studies. It is now time to examine what it was that the IEA studies did that was of interest to the various systems.

3 Concepts and procedures

In 1958, the group that met in Hamburg took the position that systems of education represented units that were responsible for the education of young people within their boundaries. They planned education and determined many facets of education such as the age of entry to school, the length of compulsory education, the number of hours of instruction per year that pupils in each grade would receive (and how this would be split into school days and hours per day), the basic curriculum, the way(s) in which teachers would be trained, the form and frequency of inspection of schools, the way in which resources would be allocated to schools and so on. It was expected that these factors would vary between education systems and often within them. There was the hope that it would be possible to identify factors that would vary between systems but be invariant within systems that might account for differences in achievement between systems. At the same time, there would be factors that varied within systems and that it would be possible to identify factors associated with differences in achievement between schools and between pupils within each system of education. There was the hope that if there were factors that were always associated with achievement, that is in every system, then it would be possible to draw a generalisation – at least for the systems of education participating in a particular study. In other words, through comparing relationships in a number of systems it might be possible to come to generalisations. One could think of a table a little like the following:

Patterns of relationships between factors and achievement

	Systems of Education							
Factors	A	B	C	D	E	F	G	H
1.	+	+	+	+	+	+	+	+
2.	0	0	0	0	0	0	0	0
3.	–	–	–	–	–	–	–	–
4.	+	+	+	+	+	0	0	+
5.	+	0	+	0	+	0	+	0
6.	+	0	0	0	0	0	0	0

Assume that regression coefficients have been calculated. The + sign indicates that the coefficient is significant (i.e. more than two standard errors of sampling) and positive, 0 indicates that the coefficient is not significant, and – indicates that it is significant and negative. In the above hypothetical example, it can be seen that Factor 1 is positively significant in each of the systems. In other words, a high value on Factor 1 is associated with a high achievement score and a low value with a low score. In this case, it is possible to state that Factor 1 is positively associated with achievement (a comparison of relationships) for all participating systems in the year X. Factor 2 has no relationship with achievement in any system; this is also of interest because it has probably been included in the study either because previous research indicated that it might be important or because several persons participating in the study believed it to be important. The fact that it turned out not to have a significant relationship with achievement in any country might be grounds for suggesting that it is not important. In education there tends to be much speculation as opposed to evidence and this could be an example of the demythologisation of education. If a study includes 50 national systems of education and the relationship is zero in every country then, assuming that the measures were good measures, there are reasonable grounds for generalisation about education. For Factor 6 there are 5 zero relationships but one positive and significant relationship. This result not only debars generalisation but it also raises the question as to exactly what is happening in Country A. The persons from Country A may have a ready answer but this is rare and they themselves were unaware of the relationship and then conduct more work on the problem in order to discover what is happening.

The first step in any such study is to identify the research questions that are to be answered. In the case of IEA, this depended very much on the interests of the researchers involved in the study as well as the interests of the ministries of education responsible for the education systems in the study. Repeated attempts were made by IEA over the years to produce models of how it was perceived that education systems 'worked' either for learning a particular subject such as mathematics or science, or in general. Super (1967) produced a report of the results of a massive workshop involving many social scientists as well as educators where models were delineated and discussed. The general research questions were specified in more detail as a set of 'specific hypotheses' and each specific hypothesis was operationalised in terms of the measures required to test the hypothesis and the form of analysis that should be undertaken. Examples of specific hypotheses in the first mathematics study were:

- The level of mathematics achievement at age 13 is not related to the age at which compulsory schooling begins.
- The mean level of mathematics achievement in a school will be related to the total enrolment of a school.
- The level of mathematics achievement of those specialising in mathematics in the last grade of secondary school will be higher where the number of subjects studied is smaller.
- In countries retaining larger proportions of an age group in school, higher levels of mathematical achievement will be attained by smaller proportions of those still in school but by a larger proportion of the total age group.
- When the level of instruction is held constant, the total mathematics score will not be related to the number of hours of schooling per week.
- The difference in mathematics achievement between lower and higher occupational status levels will be 1) least when students are in schools which have the greatest variability in occupational status levels, and 2) greatest when students are in schools which are most homogeneous with respect to occupational status levels.

There were 34 such hypotheses. The second step was to select and define the target populations and then produce sampling plans that probability samples with defined errors of sampling (in this case ± 5% for a national percentage or 0.10 of a standard deviation for a national mean) could be drawn. The lower target population drawn was that of 13 years old and the upper was the pre-university year. The lower population was in fact two populations: all 13 year olds and the grade group in which most 13 year olds were to be found. The first population was needed for comparing achievement between systems of education and the second for undertaking multivariate analyses where teachers and school conditions were linked to the pupils tested. One reason for the age sample was that it was of interest to see what different systems of education made of an age group (i.e. so many children born in one year).

The specific hypotheses also indicated the measures needed for the various variables and indicators to be used. Great care was taken to construct the tests; for example, content analyses of syllabi and textbooks were undertaken in each system of education and where there several school types within a system the amount of work was considerable. A matrix was produced for each country with mathematical topics on the vertical axis and cognitive behaviours (Bloom taxonomy) on the horizontal axis, thus having each cell in the matrix representing a learning objective. On the basis of the commonality of what was taught in each country, an international matrix or test blueprint was produced. Items were then written and piloted (about three times as many as would be required in the final test) and a final test or set of tests produced. At the same time background questionnaires for the pupils, teachers, and school principals, as well as some attitude scales for pupils, were developed. All of this developmental work took two years. The aim of it all was to be able to compare like with like in the different school systems and with the same measures. The development of instruments and measures that mean the same thing in the different systems is not an easy matter and many steps over and above producing a national set of instruments is required. However, those interested in the details can read how this was done in the various publications. There was also the problem of translation to ensure that the same difficulty of words is used when translating. Again, much trouble was taken with forward translations and back translations and for test items checks using some statistics were also possible. For the tests in the first mathematics study, multiple choice items were used. The reason for this was that the sample size for any one target population was about 3,500 pupils and there were four target populations in each of twelve countries. There was not money to have items that would have to be scored by people. In recent years systems of education have laid more emphasis on performance-type assessment and IEA has attempted to introduce some of these types of items but it is a costly venture.

In subjects such as French and English as foreign languages the four skills of reading comprehension, listening comprehension, writing, and speaking were assessed. However, for listening tape recorders had to be used in order to have the same stimulus for every pupil and for speaking where the pupils' answers had to be recorded on a tape recorder only a subsample of the total sample could be tested because of cost. In Civic education, the emphasis was much more on pupils' perceptions and attitudes and this entailed much developmental work to ensure that the items used in a scale had the same meaning and psychometric characteristics in each system of education.

Once the measures were finalised, the planned sample had to be drawn and the data collected in every system. The data were then entered into a data base, the data cleaned for inconsistencies and mispunches, the sampling weights calculated and the data analyses undertaken.

The technical work involved in the construction of international instruments and in the planning and drawing of probability samples for each of several target populations in each of the participating education systems was already a major challenge. The analyses of the data constituted a further challenge. It must be remembered that at the time of the first mathematics study and also at the time of the Six subject survey main frame computers had to be used. By the mid 1980s it was possible to use PC networks and this made data processing a somewhat easier task. Statistical thinking and software also made progress and by the 1990s it was possible to undertake many more complex analyses than in the earlier days of IEA.

4 Some examples of Results

The following types of results will be presented. simple comparisons of variables, breakdowns, simple correlations, partial correlations, multivariate analyses and path models.

4.1 Single variables and breakdowns

The first example is taken from the first mathematics study and is for a single variable.

Table 1: Mean enrolments per school per education system (Grade group of 13-year-olds).

System	*Mean*	*Standard deviation*
Australia	542	312
Belgium	412	380
England	499	311
Finland	340	277
France	381	298
Germany*	295	235
Israel	532	254
Japan	985	535
Netherlands	323	224
Scotland	552	414
Sweden	445	300
United States	658	488
All countries	553	435
Range	690	

* In the first mathematics study, Germany was not all states in then West Germany but only the states of Hesse and Schleswig Holstein. In all other countries all regions were included.

The second example is of a breakdown of home language and reading achievement score, for selected countries only, and is taken from the Reading Literacy study.

Table 2: Mean reading achievement score for students speaking a different language at home and for students speaking the school language (Grade group of 14-yr.-olds).

Country	*Non School Language*		*School Language*		*Home Language Achievement Gap*
	% students in sample	*Average score (s.e.)*	*% students in sample*	*Average score (s.e.)*	
New Zealand	5.6	470 (15.9)	94.4	551 (4.1)	81
Germany/W	8.4	455 (10.7)	91.6	530 (3.2)	75
Sweden	5.1	501 (10.8)	94.9	549 (2.3)	48
Switzerland	15.0	497 (6.7)	85.0	544 (2.5)	47
Singapore	74.1	523 (1.2)	25.9	566 (2.3)	43
France	3.9	516 (16.1)	96.1	552 (3.3)	36
Netherlands	9.1	489 (12.6)	90.9	518 (3.7)	29
Greece	2.8	487 (13.4)	97.2	510 (2.3)	23
Spain	11.4	481 (6.8)	88.6	491 (2.4)	10
Germany/E	0.8	521 (32.1)	99.2	527 (2.8)	6

The data for the first mathematics study were collected in 1964 and it is possible that school enrolments have changed since that time. It can be seen that the average enrolments differed widely, the range being 690 students per school. The standard deviations were also different indicating that some systems had less variation in school enrolments than others. However, the data were presented in the first mathematics report simply as a precursor to associating school enrolment with mathematics achievement and other variables. Alone the information is of some interest to see the difference in enrolment but then the reader may well say ‘So what?’. The information only becomes of ‘real’ interest when it is associated with other information. In the second example a relationship is shown. The data for the percentage of pupils not using the same language at home as at school is given together with the mean reading score for those pupils. This is then compared with the percentage of pupils having the same language at home and at school and their mean reading score. The reading score has a mean of 500 and a standard deviation of 100 for all countries. The figure in brackets after a mean score is the standard error of sampling of that score. Standard errors of sampling are important because two standard errors indicate the range within which one can be confident (19 times out of 20) that the mean score lies. In the above table it is also clear that the fewer the pupils in a group the larger the standard error of sampling of the mean score of that group. It is of interest to see the difference between the countries in the proportion of non native speakers and at the same time to see the difference in achievement. In comparative education terms this is the juxtapostioning of values on two variables in a number of countries. Another example of a breakdown is achievement by sex. This is presented for the same countries as in Table 2 above.

Table 3: Reading score means and differences of means for boys and girls in overall scores (Grade group of 14 year-olds).

System	*Average score*				*Difference*	*Stand. score diff.*
	Boys		*Girls*			
	Mean	*ses*	*Mean*	*ses*		
New Zealand	544	(5.9)	549	(5.5)	5	.07
Germany/W	522	(4.4)	526	(4.4)	4	.06
Sweden	540	(3.3)	555	(3.2)	15*	.21
Switzerland	535	(3.5)	538	(3.3)	3	.04
Singapore	534	(1.6)	534	(1.5)	0	00
France	553	(5.0)	549	(4.2)	-4	-.06
Netherlands	511	(4.0)	520	(5.2)	9	.11
Greece	509	(3.3)	510	(3.1)	1	.01
Spain	488	(3.3)	492	(3.1)	4	.06
Germany/E	523	(4.0)	530	(4.0)	7	.10

Again, the standard errors of sampling (ses) are presented in brackets. The difference in scores between boys and girls is significant only in Sweden. Otherwise there is no difference in reading scores between boys and girls. It will be noted that a great deal of effort has gone into measuring the differences in terms of the reading measure as well as into the sampling and the calculation of sampling errors which must be separately calculated. This is a step ahead of many national studies and most other studies examining such differences.

4.2 Simple and partial correlations

When examining relationships it does not take much perspicacity to see that children from 'better-off' homes are often enrolled in 'better-off' schools. 'Better-off' schools will often have more resources in them than 'worse-off' schools and also teachers with more experience or better qualifications. If the researcher is interested in which resources in schools are more conducive to better learning in a particular subject area the question then becomes one of disentangling the relative effects of the home backgrounds of the children from the school resources. Put in another way 'Is it the home background that 'causes' the math achievement or the better resources in the school?'. The pristine correlation can mask other effects. As George Bernard Shaw wrote in the preface to the Doctor's Dilemma:

> „Or, to take another common instance, comparisons which are really comparisons between two social classes with different standards of nutrition and education are palmed off as comparisons between the results of a certain medical treatment and its neglect. Thus it is easy to prove that the wearing of tall hats and the carrying of umbrellas enlarges the chest, prolongs life, and confers comparative immunity from disease; for the statistics show that the classes which use these articles are bigger, healthier, and live longer than the class which never dreams of possessing such things. It does not take much perspicacity to see that what really makes this difference is not the tall hat and the umbrella, but the wealth and nourishment of which they are evidence, and that a gold watch or membership of a club in Pall Mall might be proved in the same way to have the like sovereign virtues. A university degree, a daily bath, the owning of thirty pairs of trousers, a knowledge of Wa g-ner's music, a pew in church, anything, in short, that implies more means and better nu r-ture than the mass of labourers enjoy, can be palmed off as a magic spell conferring all sorts of privileges".

In the first mathematics study one hypothesis dealt with the relationship between mathematics achievement (measured by an international test of mathematics) and educational aspiration (measured by a scale) after holding constant (partialling out) the level of mathematics instruction (what the pupils had actually been taught). This produced figures such as the following:

Table 4: Partial correlation coefficients between mathematics score and educational aspiration – holding level of mathematics instruction constant.

Country	*13 year old grade*	*Preuniversity year specialists*
Australia	.35	.20
Belgium	.34	.22
England	.56	.16
Finland	.32	.15
France	.10	*.05*
etc.		

In the above table only the partial correlation for the pre-university students in France was not statistically significant.

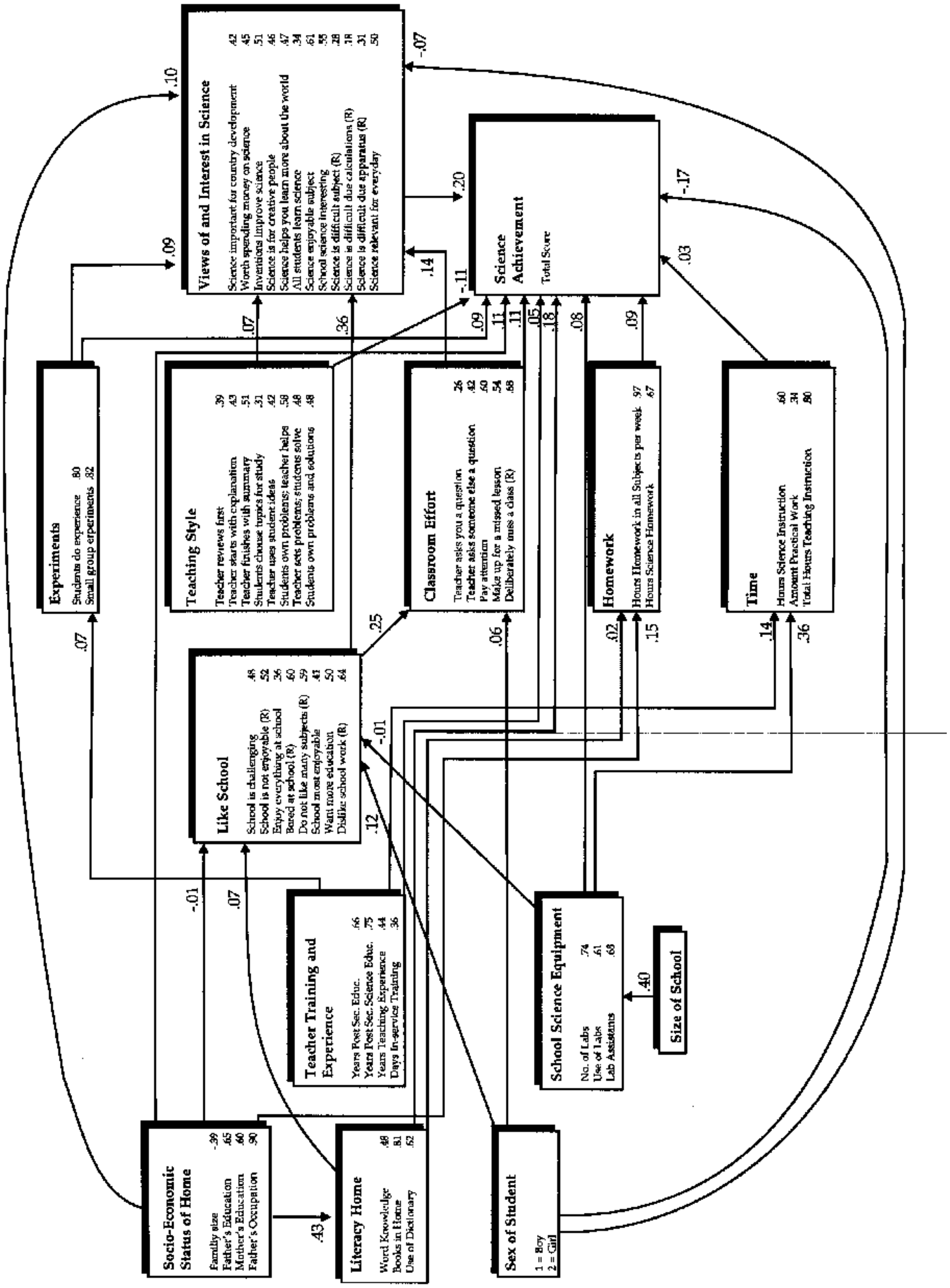
In the Reading Literacy study, the notion of partial correlation was used in a different way. Many systems of education were interested to know if there were any schools, with an intake of pupils from poor home backgrounds, where the average performance of pupils was much higher than expected (meaning higher than their home background would lead one to expect). Using the data from all pupils within a population within a target population in a country the school mean score was adjusted to be one that reflected the mean score the school should have achieved given the home background of the children in it. The expected score was then subtracted from the actual score. If the difference was positive the school was 'doing better than expected'; if negative, the reverse was true. It was then possible to relate many (64) school and teacher variables that had significant partial correlations to the differences between the expected and actual scores and then identify those factors that were associated with the differences. (If a psychometrician is reading this article, he or she should rest assured that the partialling was done at the pupil level and the differences then summed to create the school difference score, thus avoiding the problem of aggregation bias). Postlethwaite & Ross (1992) reported those teacher and school variables where there was more then 0.3 standard score differences between the more effective schools and the less effective schools.

4.3 Multivariate analyses

When testing a hypothesis it is always important to decide which other variables to 'hold constant' or 'partial out' in order to ensure that it is not a variable than the one of particular interest that is 'causing' achievement. This is known as the problem of '*ceteris paribus*' (other things being equal). The question always is how much of *ceteris* there should be and researchers can debate this point for a particular hypothesis for a long time. A basic common approach to this problem has already been hinted at. There are many variables that have been included in a study that will not be significantly related to the criterion (usually some sort of achievement, but it could be a perceptual measure or an attitudinal measure), significance being that the relationship should exceed two standard errors of sampling. The non-significant variables are then dropped from further analyses but it is of interest to note the variables that are dropped as part of the demythologisation of education issue referred to earlier. For correlations of school and teacher variables with pupil achievement, it may also be desirable to partial out the effect of home background in order to see if it is worth looking further at such variables. The remaining variables are then examined in terms of their regression coefficients in a multiple regression analysis which indicates the relative 'effect' of each of the variables or cluster of variables on the criterion. This was the approach taken in the final analysis of the first mathematics study and the six-subject study. An alternative approach was that of postulating path models and testing them. This was done in the six-subject area and later studies. In some cases. an Ordinary Least Squares approach was used and in other cases either Partial Least Squares (PLS) or LISREL. In the Third Mathematics and Science Study, Hierarchical Linear Modelling (HLM) will be used. This allows the effect of a variable to be assessed at two levels (for example, the pupil and school levels) at the same time. This approach assumes that there is reasonable agreement on the model to be tested. The testing of the model is an iterative process with paths being dropped that do not work until a final model is arrived at. There can be disagreement about models and this is part of the research process.

An example of a path model (Postlethwaite & Wiley 1992: 126) from the Second International Science Study is presented in Figure 1 on the next page. Each box represents a construct that is made up of one or more variables. The coefficient next to each variable is the relative weight of that variable in the construct. The arrowed lines denote the postulated path and the value on the line is the path coefficient showing the strength of the effect. The model may look complicated but then the process of education is complicated and many would argue that it is even more complicated than shown in the model below. The construct 'Socio-economic Status of the Home' is made up of four variables. The highest weight is for Father's Occupation (.90); parental education is represented by mother's and father's education having weights of .60 and .65 respectively, and family size has a weight of -.39 (in this case meaning that smaller family size 'goes with' higher occupation and education). Five direct paths were postulated from the 'Socio-economic status of the home'. The first is to 'Views of and Interest in Science', the second to 'Science Achievement', the third to 'Like School', the fourth to 'Homework', and the fifth to 'Literacy of the Home'. It can be seen that all had path coefficients of .10 (normally taken as significant in this type of analyses with this type of sample) or more, except for 'Like school' where it was about zero. It can be seen that the constructs most 'influencing' 'Science Achievement' were 'Views of and Interest in Science' (.20), 'Teacher training and experience' (.18), 'Teaching style' (.11), 'Ses of the home' (.11), and 'Classroom effort' (.11). If 'Views of and Interest in Science' has the most effect on the criterion, then it becomes of interest to see what is affecting 'Views of and Interest in Science'. It will be seen that the constructs having the largest coefficients were 'Like school', 'Clasroom effort' and the 'Ses of the home'. It can therfoe be inferred from this figure that the 'Ses of the home' has both direct and indirect effects on 'Science Achievement'. It is then possible to look up in tables in the research report the strength of the direct effect and the total effect, the difference between the two effects being the indirect effect. In other words, it is possible to see not only the direct effect of a construct but how it can work through other constructs.

Figure 1: Path model of the Philippines (14 year-olds grade) (from Postlethwaite & Wiley 1992).



This is an example for one country. It is also possible to present the results of such models run in various countries. Table 5 below (Keeves & Dryden 1992: 198) presents the results of a PLS analysis. The direct effects are the regression coefficients for the postulated paths. The total effects are the direct plus indirect effects. Space limitation does not allow a full explanation of the table and readers are referred to the original book.

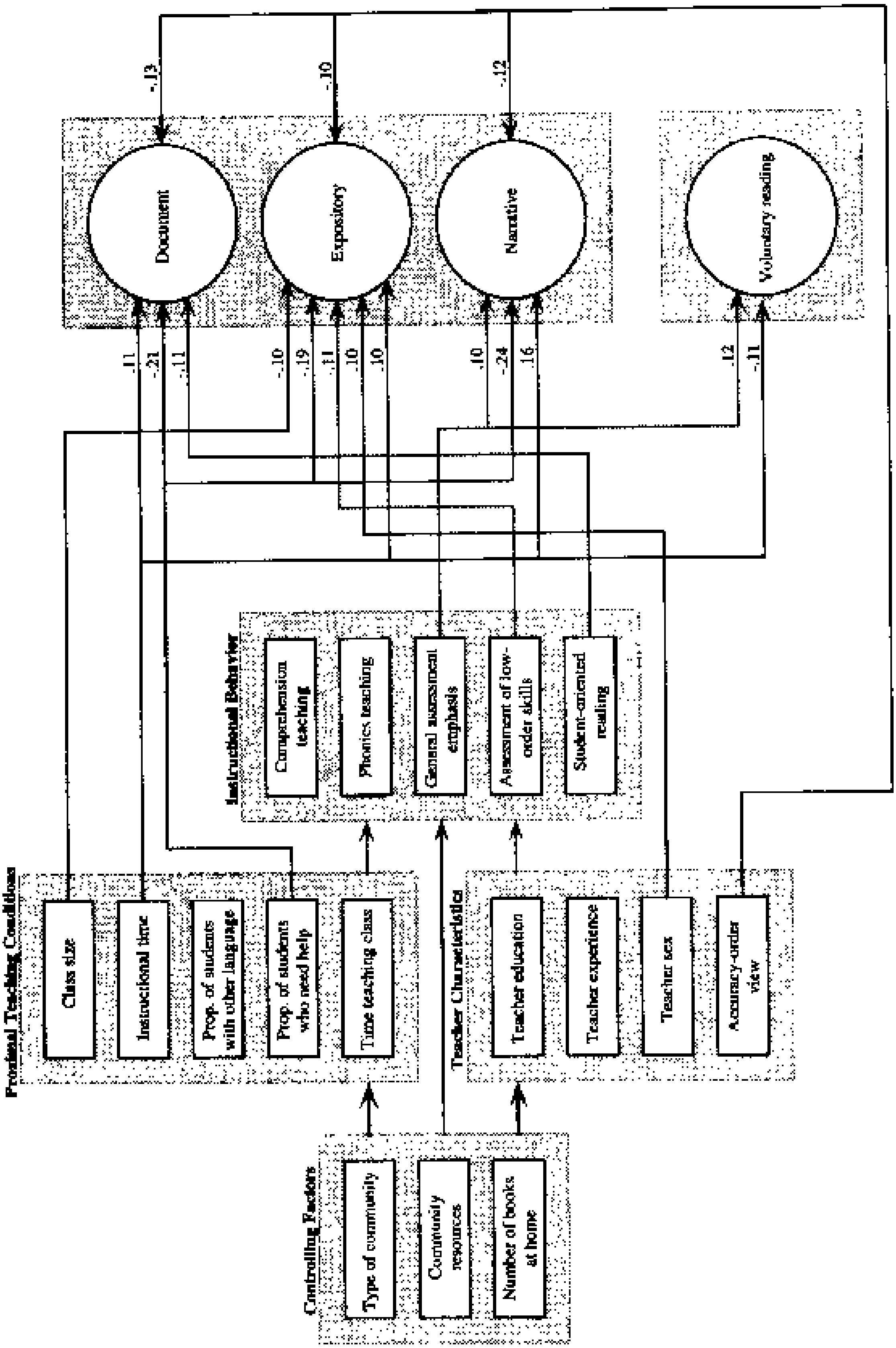
Table 5: Reported effects from PLS analyses at the between school level for 14-year-old grade Pop. 2.

Direct Effects	*Aust*	*Eng*	*Finl*	*Hun*	*Italy*	*Jap*	*NL*	*Swe*	*Thai*	*US*	*Sign* +	*Sign* −	*Mean*
Social Context	17	19	10	0	0	0	26	(8)	19	0	5	0	10
Home background	65	31	40	54	33	32	15	42	0	44	9	0	36
Science values	0	18	20	9	0	0	0	20	0	10	5	0	8
Sch Sci support	-22	11	0	0	15	14	0	26	36	13	5	1	9
Science teachers	18	0	0	13	0	0	25	0	0	0	3	0	6
Sci tchg emphasis	0	-12	0	0	21	9	0	0	14	0	3	1	3
Time and OTL	18	0	0	0	11	11	0	22	9	0	5	0	6
Views of Sci Tchg	0	25	26	0	15	0	9	0	11	19	6	0	11
Motivation	12	14	0	0	21	25	16	0	35	18	7	0	14
Sci attitudes	0	0	0	20	11	0	17	0	14	9	4	0	6
Total Effects													
Social Context	34	23	34	40	45	20	57	11	25	36	10	0	33
Home background	68	71	60	55	65	47	29	57	35	70	10	0	56
Science values	0	22	20	19	0	16	28	16	0	22	7	0	14
Sch Sci support	-11	17	0	11	19	22	0	26	63	11	7	1	16
Science teachers	15	0	0	14	0	0	23	0	28	12	5	0	9
Sci tchg emphasis	0	0	0	0	24	12	0	0	19	0	3	0	7
Time and OTL	25	11	0	0	21	25	0	22	18	0	6	0	12
Views of Sci Tchg	0	25	27	10	15	0	11	0	22	23	7	0	13
Motivation	12	14	0	0	22	25	13	0	41	18	7	0	15
Sci attitudes	0	0	0	20	11	0	17	0	14	0	4	0	6
Var expl Achievem.	64	68	51	42	59	39	79	46	73	62			58
Av % var. expl.	33	34	26	32	33	25	51	19	25	37			32
N. of schools	233	146	90	99	224	199	224	137	96	88			

It is the total effects that are important in Table 5 and it can be seen that although the social context and home background have large effects in each and every country, the other variables also have considerable effects but not in every country or system of education.

Both Figure 1 and Table 5 represent examples of analyses in single countries. Figure 1 is for one country and Table 5 presents the analyses for each of ten countries. It is also possible to undertake what are known as pooled analyses. This is where all, say, schools from all countries in a study are put together for the purpose of analysis. Figure 2 is an example of such an analysis (Lundberg & Linnakylä 1992: 84).

Figure 2: A path model of the relationship between taeching factors and reading achievement.



This kind of analysis produces information on those variables that have an effect in all countries after certain factors (first column) have been controlled for. This is important information because it is based on all schools (in this case 7,800) from all countries in the study.

The procedures and types of analyses used by IEA have been presented. At the same time some results have been given.

5 Major Results

Keeves (1995) has summarised the major results of IEA studies. He divided them into results that are generalisable across subject areas and those that are generalisable within a subject area. Below, only the generalisable results across subject areas are reproduced:

- There are marked differences in average levels of achievement between the students in school in the more developed countries (MDCs) and those in the less developed countries LDCs). This occurs in spite of the fact that in the LDCs less then 100 percent of the relevant age group is enrolled at school.
- The average level of achievement within a country at the terminal secondary school stage is inversely related to the proportion of the age group enrolled at school or participating in the study of the subject under survey.
- At the terminal secondary stage, when equal proportions of the age groups are compared, there are no differences in levels of achievement, irrespective of the proportions of the age groups retained at school. The best students do not suffer with increased retention rates.
- Student achievement in Mathematics, French as a Foreign Language and Science is positively related to the time given to the study of the subject at school, both in comparisons across countries and between students within countries.
- The achievement of students is related to the time spent on homework after other factors influencing achievement have been taken into account.
- The average level of student achievement across countries is positively related to the opportunity that students had to learn the content of the items tested.
- In developing countries (LDCs) the use of a textbook has an effect on student learning. However, the same effects have not been reported from the studies in more developed countries (MDCs).
- The level of reading resources of the home is positively related to student achievement, as are other indicators of language usage in the home, such as use of a dictionary in the home and whether or not the language of the home is the language of instruction.
- Measures of the socioeconomic status of the home are positively related to student achievement in all countries, at all age levels and for all subject areas.
- Although the effects of home background variables are similar across subject areas, the effects of the learning conditions in the schools differ between subject areas, and in some subject areas are equivalent or greater in the size of their influence than the effects of the home.

- Differences are found between the sexes in achievement which vary in size and direction across countries, school subjects and over time. They are argued to be socially based.

Again, space limitations prevent the presentation of many results but these are intended to give a flavour of the kind of results that have emerged from some of the studies.

6 Problems

No studies of the magnitude of the IEA studies are without problems. They take a long time. As has been seen, the development of the hypotheses and the instruments take a long time – at least two to three years. Attempts at short cuts would be disastrous. The *exact* comparability of the target populations is somewhat of a problem. If an age group is the population then in some systems of education these are spread across various school grades whereas in others they are in one or two grades. Thus the coverage of the target population will differ from about 85 to 99 percent. If the target group is a grade group then the coverage is complete except for pupils in special education where this can vary from one to nine percent (Postlethwaite 1994, Schleicher 1994). The planning and design of the probability sampling can take some time. Probability sampling has however been one of IEA's strong points. The data collection can be spread over nearly one year because of the different times of beginning and ending a school year in different systems. The data entry (and cleaning and weighting) can also take a good six months. The analysis, especially the more complicated analyses, can take from one to two years.

There has been a tendency for some systems to be much more interested in the simple comparison of test scores only to see if their system scores better or worse than other systems of education. Whereas this can be of some interest, the ensuing question is 'So what?'. Why is it that one system is 'better' than another and this returns the researcher to the identification of variables that do account for such differences. It is also true that the media highlight the system score differences. One gains the impression that the media is particularly interested when the United States score poorly compared with other countries. Thus, the low scores US score in Mathematics and Science were highlighted but when the US scored relatively well in Reading Literacy, very little mention of this occurred. OECD (1992, 1994) has also published the distribution of single variables (including test scores) in tables without attempting to account for such differences. Much of this has resulted in 'push' to have IEA act as a monitoring rather than research organisation. Funding is surely more readily forthcoming for monitoring studies than for research studies. But, at the time of writing the semantics of monitoring and research are still being debated and it remains to be seen what transpires.

7 Conclusions

It is a fact that the knowledge and skills required for this type of research require a great deal of investment in terms of time and energy. The skills are conceptual and technical. Many working in the area of education in general and comparative education in particular never do learn them. As can be seen from what has been presented in this brief article the conceptualisation and procedures are quite complicated. Education is complicated. It is not purely didactics and it is important that many specialists are involved in such studies including economists, political scientists, historians, sociologists, and comparative educators.

To return to the two aims of this article it is this author's contention that this kind of study constitutes *one* of many approaches to the study of educational systems in a comparative way and hence IEA does fall within the realm of comparative education. There are those who discuss about IEA being within or not within the realm of comparative education. In this sense comparative education is somewhat reminiscent of sex: „Those who have sex do it and don't talk about it". Secondly, it is possible to attempt to make generalisations concerning education although some may be time bound, some subject matter bound, and some national system bound. Finally, it should be added that many educational systems use the results of IEA studies: some examine the learning objectives that are poorly achieved and revise the curriculum and textbooks; others look at resource allocation where particular resources are shown to be related to achievement; yet others revamp parts of their teacher training programs where particular teaching methods or strategies are shown to affect achievement.

There are many challenges for IEA. Can it reduce the time scale of its work without reducing the quality of its studies? Can it extend the coverage of the target populations that are age groups and still have test coverage that is good for all children? Should it enter into adaptive testing whereby systems of education can have different but overlapping tests but in such a way that all children in the world can come onto one continuous scale? Given that ministries of education are rather like hydra-headed monsters where different divisions within a ministry rarely, if ever, talk to each other, is it possible for IEA to discuss the implications of the results of its studies with ministry personnel and then produce publications that indicate what short-term, medium-term, and long-term actions that ministries might undertake at low, medium and high cost.

Theory in education has relied a great deal on speculation in the past. As theory is tested, so it should improve. This requires a lot of effort by all. in the various areas of specialisation in education and in the social sciences. The 'social science empirical approach'is useful. Although 'armchair reflection' can be useful in the process of postulating hypotheses or when having to interpret results, it is essential to collect data because without data collection and the analyses of the data there would be no check whatsoever on the reflection and speculation, however realistic or unrealistic it might be.

Bibliography

Anderson, C.A. (1961). Methodology of Comparative Education. *International Review of Education* 7(1), 1-23.

Anderson, L. W., Ryan, D.W. & Shapiro, R.J. (Eds.). (1989). *The IEA Classroom Environment Study. International Studies in Educational Achievement*, (Vol.2). Oxford: Pergamon.

Beauchamp, E.R. (1985). Some Reflections on Comparative Education. *East West Education* 6 (2), 12. [quoted from Epstein, E.H. (1988). The Problematic Meaning of 'Comparison' in Comparative Education. In J. Schriewer & B. Holmes (Eds.), *Theories and Methods in Comparative Education* (pp. 3-23)]. Frankfurt: Lang].

Burstein, L. (Ed.). (1993). *The IEA Study of Mathematics*, Vol. 3. Oxford: Pergamon.

Carroll, J.B. (1975). *The Teaching of French as a Foreign Language in Eight Countries. International Studies in Evaluation,* Vol. 5. New York: Wiley.

Chambers, J.H. (1992). *An Essay-Review of Empiricist Research and Practical Critique if its Scientific Pretensions.* Boston: Kluwer.

Comber, L.C. & Keeves, J.P. (1973). *Science Education in Nineteen Countries. International Studies in Evaluation*, Vol. 1. New York: Wiley.

Degenhart, R.E. (1990). *Thirty Years of International Research: An Annotated Bibliography of IEA Publications* (1960-1990). The Hague: IEA.

Elley, W.B. (1993). *Reading Literacy in Thirty Countries*. Oxford: Pergamon.

Epstein, E.H. (1988). The Problematic Meaning of 'Comparison' in Comparative Education. In J. Schriewer & B. Holmes (Eds.), *Theories and Methods in Comparative Education* (pp. 3-23). Frankfurt: Lang.

Foshay, A.W., Thorndike, R.L., Hotyat, F., Pidgeon, D.A. & Walker, D.A. (1962). *Educational Achievement of Thirteen-year-Olds in Twelve Countries*. Hamburg: UNESCO Institute for Education.

Garden, R.A. & Robitaille, D.F. (1989). *The IEA Study of Mathematics 11: Contexts and Outcomes of School Mathematics*. Oxford: Pergamon.

Gorman, T.P., Purves, A.C. & Degenhart, R.E. (Eds.). (1988). *The IEA Study of Written Composition 1: The International Writing Tasks and Scoring Scales*. Oxford: Pergamon.

Husén, T. (Ed.). (1967). *International Study of Achievement in Mathematics: A Comparison of Twelve Countries*, Vols. 1-2. Stockholm: Almqvist & Wiksell.

Keeves, J.P. (1992). *The IEA Study of Science: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon.

Keeves, J.P. & Dryden, M. (1992). The Teaching and Learning of Science at Population 2: 1983-84. In J.P. Keeves (Ed.), *The IEA Study of Science: Changes in Science Education and Achievement: 1970 to 1984*. Oxford: Pergamon.

Keeves, J.P. (1995). *The World of School Learning: Selected Key Findings from 35 Years of IEA Research.* The Hague: IEA.

Lewis, E.G. & Massad, C.E. (1975). *The Teaching of English as a Foreign Language in Ten Countries. International Studies in Evaluation*, Vol. 4. Stockholm: Almqvist & Wiksell.

Lundberg, I. & Linnakylä, P. (1992). *Teaching Reading Around the World*. The Hague: IEA.

OECD (Ed.). (1992). *Education at a Glance: OECD Indicators*. Paris: OECD.

OECD (Ed.). (1994). *Education at a Glance 2: OECD Indicators*. Paris: OECD.

Olmsted, P.P. & Weikart, D.E. (1989). *How Nations Serve Young Children: Profiles of Child Care and Education in 14 Countries.* Ypsilanti: High/Scope Educational Research Foundation.

Olmsted, P.P. & Weikart, D.E. (Eds.) (1995). *The IEA Preprimary Study. Early Childhood Care and Education in 11 Countries.* Oxford: Pergamon.

Passow, A.H., Noah, H,J., Eckstein, M.A. & Mallea, J.R. (1976) *The National Case Study: An Empirical Comparative Study of Twenty-One Educational Systems. International Studies in Evaluation*, Vol. 7. Stockholm: Almqvist & Wiksell.

Peaker, G.F. (1975). *An Empirical Study of Education in Twenty-One Countries: A Technical Report. International Studies in Evaluation*, Vol. 8. Stockholm: Almqvist & Wiksell.

Pelgrum, W.J. & Plomp, T. (1991). *The Use of Computers in Education Worldwide: Results from the IEA 'Computers in Education' Survey in Nineteen Systems of Education*. Oxford: Pergamon.

Postlethwaite, T.N. (1994). Monitoring and Evaluation in Different Education Systems. In A.C. Tuijnman & T.N. Postlethwaite (Eds.), *Monitoring the Standards of Education (pp. 23-46).* Oxford: Pergamon.

Postlethwaite, T.N. & Ross, K.N. (1992). *Effective Schools In Reading: Implications for Educational Planners*. The Hague: IEA.

Postlethwaite, T.N. & Wiley, D.E. (1992). *The IEA Study of Science 11: Science Achievement in Twenty-Three Countries*. Oxford: Pergamon.

Purves, A.C. (1973). *Literature Education in Ten Countries: An Empirical Study. International Studies in Evaluation*, Vol. 2. Stockholm: Almqvist & Wiksell.

Purves, A.C., Lehmann. R. & Degenhart, R.E. (1992). *The IEA Study of Written Composition 11: Education and Performance in Fourteen Countries*. Oxford: Pergamon.

Rosier, M.J. & Keeves, J.P. (1991). *The IEA Study of Science 1: Science Education and Curricula in Twenty-Three Countries*. Oxford: Pergamon.

Schleicher, A. (1994). International Standards for Educational Comparisons. In A.C. Tuijnman & T.N. Postlethwaite (Eds.), *Monitoring the Standards of Education (pp. 229-247).* Oxford: Pergamon.

Schriewer, J. (1988). The Method of Comparison and the need for Externalisation: Methodological Criteria and Sociological Concepts. In J. Schriewer & B. Holmes (Eds.), *Theories and Methods in Comparative Education* (pp. 25-88). Frankfurt: Lang.

Super, D.E. (Ed.). (1967). *Toward a Crossnational Model of Educational Achievement in a National Economy.* New York: Teachers College Press.

Thorndike, R.L. (1973). *Reading Comprehension in Fifteen Countries: An Empirical Study. International Studies in Evaluation*, Vol. 3. Stockholm: Almqvist & Wiksell.

Torney, J.V., Oppenheim, A.N. & Farnen, R.F. (1976). *Civic Education in Ten Countries: An Empirical Study. International Studies in Evaluation*. Vol. 6. Stockholm: Almqvist & Wiksell.

Travers, K.J. & Westbury, I. (1989). *The IEA Study of Mathematics 1: International Analysis of Mathematics Curricula*. Oxford: Pergamon.

Walker, D.A. (1976). *The IEA Six Subject Survey: An Empirical Study of Education in Twenty-One Countries. International Studies in Evaluation*, Vol. 9. Stockholm: Almqvist & Wiksell.