What can Planners Expect from International Quantitative Studies?

Rosine Lambin

Institute of Comparative Education
University of Hamburg

Abstract

International quantitative research in comparative education should be an instrument employed to improve systems of education around the world and therewith the system of education in individual countries as well. This article will therefore try to answer two questions: 1. Can international quantitative research in education directly help policy makers and educational planners around the world improve their system of education, and, if so, 2. how is this likely to occur? The answers will be based on the main findings of the *International Association for the Evaluation of Educational Achievement* (IEA) and, to a lesser degree, on other sources.

1 Introduction

The most interesting part of research on educational achievement in an international perspective is the question of the practical use ministries of education, policy makers and educational planners can make of the findings provided by the data analysis of such studies. This is so important because these agents are increasingly being asked to base their administrative decisions on information and findings provided by research in education. The main question that this article addresses is therefore: Do the international studies which test student achievement around the world and search for its common causes really help policy makers and educational planners make concrete choices concerning their system of education? In other words, do these studies answer the following questions: Is the information produced by such studies sufficiently generalizable to base upon it good policy making with far-reaching consequences? and: What can one country do concretely to improve its system of education by looking at the findings of international quantitative studies and by discovering how other countries are successful or not?

The debate about the relevance of such studies has been going on for years and many questions are repeatedly posed (Wolf 1993). Many important questions and critiques about the problems inherent to these studies have emerged and are worthy of discussion. Among others, one could propose: are not the costs of such studies too high for the results they provide? are these studies not taking too much time (sometimes ten years) to be completed? are the tests relevant in all cultures and can the instruments be valid for all countries: is it possible, for example, to really compare Iran with the United States or Kuwait with Ukraine like the last IEA[1] study does? As a matter of fact, the Third International Mathematics and Sciences Study (TIMSS) started in 1993, compares 50 countries and states[2]: this is the highest number of countries to ever participate in an IEA Study. Given the enormous amount of data provided by all these different countries, would not data analysis become an impossible task? How can one make all target populations comparable (Schleicher 1994)? Another question which has been raised is how to fairly compare achievement between the sexes: Do the tests discriminate against one of the sexes and, if so, are the findings derived from them really reliable?

Provided that these problems can be solved, there is still the question of ranking countries. If, for example, students in the participating countries have been subject to very different curricula and learnt at very different levels the materials covered by the tests, how can one fairly rank countries from the best to the worst achievement level? As a matter of fact, a ranking would only be meaningful if all countries included had given to their pupils the same - or about the same - opportunity to learn what was tested (OTL), if they were equally selective, if they employed the same criteria of selection, and if they had the same percentage of enrollment. When one country has an enrollment rate of 80 percent - likely a developing country - and another an enrollment rate of 100 percent, how can the comparison be really meaningful? Does the coincidental availability of funds and resources in a given country constitute a reason for it to be included in the study by the survey institution? Is it useful, both for the country itself and for the international consortium implementing the study? Should there not be a limit to generalizations made about systems of education and should not countries be compared only to a narrowly defined group to which they also belong? All this has to be taken into account, and only if positive answers can be justified, the tests really can rank countries efficiently.

Some other, rather abstract and, as it seems, unproductive questions about quantitative international studies will be left out of the present discussion. The very new debate about whether international studies belong to the field of comparative education or if comparative education is a well-defined field at all

is hardly relevant when considering the work to be done to improve systems of education and to achieve education for all in the world. It seems advisable, therefore, to remain as true as possible to the practical theme described above: Do schools and systems of education benefit from quantitative international studies? How researchers can further improve education around the world must be the main focal point.

As a first step, the goals of IEA studies as the leading example for international survey research are briefly delineated. This has to be done before one can decide if these goals have been achieved or not. Secondly, the main findings from IEA studies are summarized in an attempt to sort out which ones are important for countries if they want to implement them to improve their systems. Thirdly, it will be argued that international quantitative studies in education are, indeed, very important for national systems of education if countries are ready to learn from each other and evolve toward an international system of communication in education.

2 What Are the Goals of International Studies of Educational Achievement?

Theoretically, the goal of this type of studies is to measure, with the help of more and more elaborated techniques and instruments, achievement - known as 'outcomes' - and to ascertain the causes of these different outcomes between and within the participating countries' systems of education. The basic method is to use the differences between these systems - in curriculum, for example - as a carefully controlled variation in "one big educational laboratory" (Husén 1973) in order to determine what works and what does not: "We conceived of the world as one big educational laboratory where a great variety of practices in terms of school structure and curriculum were tried out. We simply wanted to take advantage of the international variability" (Husén 1973, p. 32).

Data sets which are structurally identical across countries - first and foremost cognitive scales considered by the IEA researchers as being objective[3] measures of students' achievement - are expected to provide new results which could not be found in limited national studies. Internationally firmly established findings should then help policy makers and educational planners - as well as teachers - to see what works and what does not work in their own setting. With the help of these findings, they should be able to revise their system with greater effectiveness and efficiency. Thus it is the explicit goal of IEA type studies to directly advise policy makers and educational planners (Husén 1973). This

should be kept in mind each time new data are analysed. Three major categories of findings are expected to emerge from such IEA international studies:

1. A description of the differences among countries.
2. A discovery of the factors causing these differences, and
3. an understanding on how these factors work to cause these differences (Husén 1973).

It goes without saying that this is not a light task to achieve. But, is it an easy task for planners to use the results of such research in decision making? By looking at what works and what does not in other countries, by comparing what can be compared, a country can create a review mechanism which may help its own system to progress steadily as judged by internationally monitored standards. This is probably the major advantage of quantitative international research in education, which will be discussed below in more detail. In the absence of such standards it is difficult, if not impossible, to make a judgment of one's own system of education, and, it may require some humbleness to be willing to learn from others. It is also difficult to see the world as a communication network where national systems of education depend on each other in order to progress. To facilitate some understanding of the impact such studies can have on various country systems, some major findings provided by IEA data in the last 40 years are reviewed below.

3 Classifying and Evaluating Generalisable Findings from the IEA International Surveys

One should first keep in mind what a finding is not: it is not the Truth with a capital "T". It might change over time and it might even be wrong (Wolf 1993). It is not fixed forever and not unchangeable. Major generalizable findings from the studies IEA has undertaken can be found in the respective series of international reports (see References below). Most of those discussed here are also listed in Keeves' summaries (Keeves 1992, 1995). These findings can be classified into three categories:

1. Findings which confirm what has been known by educators for years through mere observation and/or intuition.
2. New findings which can give very little help to policy makers and educational planners.
3. New findings which give concrete answers to policy makers and educational planners.

3.1 Findings which Confirm what has been Known by Educators for Years through Mere Observation and/or Intuition.

The majority of these findings consists in confirmations of what has been known already for years by most specialists in education in most countries, but had not been established scientifically. One may have different opinions on whether the price for mere confirmations of practical knowledge was justified, and the same applies to results which do not appear to have immediate applications. It is argued here, however, that findings from the third group are neither small in number, nor do they lack substantial and quite often far-reaching implications. Some of these can, indeed, help policy makers and educational planners considerably in making better choices. They can also remind countries of old needs in education which are still waiting to be implemented.

Time spent on task is positively related to achievement.

The time spent at school on the study of Mathematics, Science, and French as a foreign language - and by extention to other foreign languages -, has a positive influence on learning and achievement (Carroll 1975; Keeves 1992). In other words, the more time a student spends learning a subject at school, the more she or he learns in the field. This finding, which seems plausible enough, unfortunately does not help policy makers and educational planners with their decisions on the number of hours which should be allocated to each school subject. Since this finding concerns all subjects, they all have to compete for more time. The question is then: which subjects are the most important to teach? And this question cannot be answered on the basis of achievement data alone. If, for example, a policy maker wants students to learn more math, she or he should then increase the number of hours spent on this particular subject, obviously at the expense of other subjects if teaching and learning time are constrained. That leads to the question of the limits of increasing the overall time spent at school. How much time should a student spend at school? But when one takes into account the IEA findings about the marginal influence schools have on achievement (see below), the answer to this question appears difficult to elaborate. Moreover, the SISS data, for instance, do not even answer the question of when science should be introduced in the curriculum (Keeves 1992): thus, the time of introducing a subject into curriculum is also a problem to be discussed.

Opportunity to learn (OTL) makes a difference.

The greater the opportunity students have to learn the content of the items tested, the better they do on the tests (Husén 1967; Comber & Keeves 1973; Postlethwaite & Wiley 1991; Keeves 1992). Again, the finding appears to be rather obvious, but it points to a major problem. It actually shows that the tests cannot cover everything learned by students at school in all countries and therefore, strictly speaking, the tests cannot simply be taken as generalizable measures for student achievement around the world. If OTL is appropriately controlled, the respective analyses show that rankings are relative (Travers & Westbury 1989) and that it would be naive to consider quantitative international comparisons as "horse races" between educational systems. However, these analyses can help national researchers and policy makers to balance the results of the tests in their country by taking into account what is actually taught in their schools.

Selectivity results in better (mean) results on achievement tests.

The greater the dropout rate is and/or the smaller the proportion of an age group participating in the study, the better the average performance of those who are taking the test (Husén 1967; Comber & Keeves 1973; Postlethwaite & Wiley 1991; Keeves 1992). In other words: Eliminate the weak students and you will rank higher. This is very fundamental choice for policy makers and educational planners to make (particularly at the secondary level). A country must choose either to include as many pupils as it can in its system of education and to rank low, or to opt for a more selective system and to rank high. Education for all is likely to make the average level of achievement of a country lower than it could have attained by having only good students in schools, but is the latter solution the best? The fact that the very best students do equally well under both conditions (see below), suggests that it does not really matter if a country ranks low on the test because a change toward less selectivity does not necessarily affect the highest achieving students.

Curricula reflect what policy makers want.

The SISS data show that the curriculum is uniform if determined centrally and varies if determined by teachers (Postlethwaite & Wiley 1991). How a country can produce the best possible curriculum can only be determined by conducting more specialized studies about the subject. Whether or not centralized

responsibility for curriculum development is better than a decentralized one cannot be derived from the IEA data. Efficiency may be an important issue here, however. Lewy (1991), for example, has shown that decisions about the curriculum which attained the most efficiency were made through a mixed approach using both central authority and decentralized democracy.

Developing countries have problems.

Developed countries enrolling virtually all of the relevant age cohort rank higher on tests than developing countries, despite the fact that enrollment in developing countries is often considerably less than 100 percent. As was stated above, selectivity is usually associated with higher levels of achievement within a single country or between comparable countries. This is to be expected because selection keeps the higher performing students and leaves problematic students behind. A developed country which enrolls 100 percent of an age group, may display effects of internal selectivity, but its level of achievement, as ranked in the IEA Studies, is the mean of all test results, high and low. Careful consideration is, indeed, given to age and/or grade level coverage. The selection process in developing countries does not follow the same pattern: it is usually an economic selection. Children who work cannot afford to go to school. Therefore, the achievement level of these countries is often the reflection of a significant loss of intellectual potential - the loss attributable to children who could go to school and reach high levels of achievement, but are precluded from doing so. This means: something has to be done in developing countries. Unfortunately, the IEA data do not provide many clues as to what exactly has to be done in developing countries in order to offset economic selectivity and, at the same time, reach the level of achievement found in developed countries. But the IEA data do provide findings which can and ought to function as cries of alarm.

Textbooks help.

In developing countries, the availability of textbooks has a positive influence on learning. This finding cannot be confirmed in more developed countries with certainty because there is usually little variation within these countries, and there may also be a threshold effect beyond which additional textbook contribute little to achievement. But even if this only underlines what everybody knows: developed countries should help developing countries to get school supplies and textbooks. So, this result should also be taken as a sign of alarm.

3.2 New Findings which can Give Little Help to Policy Makers

The list of findings in this category is rather long. Results about home resources, socio-cultural background, or language spoken at home can hardly be turned into practical action by policy makers except, perhaps, indirectly by way of encouraging parent-school dialogue. However, these variables have to be taken into account in the analysis of test results in order to determine the real influence of schools and to see how schools could improve (Postlethwaite & Ross 1992). Beyond this function as a means of statistical control, they give few indications on how to deal with the shortcomings so identified.

Schools have little influence on achievement.

The First and Second Science Studies have shown that the socio-cultural background of the home has a far greater influence than schools on achievement among students in all countries, at all age levels and for all subject areas (Comber & Keeves 1973; Postlethwaite & Wiley 1991). Similarly, the study on Reading Literacy has demonstrated that the level of reading resources of the home is positively related to student achievement (Postlethwaite & Ross 1992). Clearly, this finding is linked to the socio-cultural background, although its specificity is quite significant. Yet, some skepticism seems to be in order. First of all, these findings, in and of themselves, are unlikely to teach something concrete to policy makers and educational planners. Obviously, the institutions responsible for education in a country cannot controll everything. They can at best encourage teachers and school principals to take these factors into account accordingly. Teachers and school principals can then try to advise students and their families and do their best. Secondly, these findings show the schools' limited influence. In its apparent generality, this is, in many ways, a very depressing conclusion despite some reservations expressed by the IEA reports. In fact, these findings lead directly to questions about the usefulness of international studies about schools and about the wisdom of putting so much effort into educational reforms. If home and socio-cultural background variables have such a strong influence on student achievement, to the extent that schools cannot effectively compensate inequalities associated with them, where is the solution? The data as such do not provide solutions to these enormous problems. Yet, they do raise the awareness of social conditions not being what they should. A similar challenge arises with the next finding.

Females are still behind in science.

Gender differences in science achievement increase from the 10-year-old to the 14-year-old and on to the terminal secondary school level and increase across fields of science (biology, chemistry and physics) (Keeves 1992). Boys perform better in science than girls, however with some variation between countries and over time (Comber & Keeves 1973; Postlethwaite & Wiley 1991; Keeves 1992). Boys are also more interested in science than girls. This is believed to be a primarily social phenomenon (Harding 1992). So, if indeed the roots are again beyond the range of (substantial) school influence, this finding is another one to classify in the category "signs of alarm". If girls are really inferior to boys in science achievement, policy makers and educational planners should act. But how? Once more, the data do not provide an answer to this question (Keeves 1992) and cannot tell policy makers and educational planners concretely how they should deal with the problem in the field. The same remark applies to the cluster "interest", influenced by background variables and school variables: how could girls be motivated more effectively in the field of science? How can countries, especially the ones presenting a large gap between boys and girls, reduce the gap?

Some grades have a greater influence than others.

In science (SISS), grades 11 and 12 add to student learning but grade 13 apparently does not (Postlethwaite & Wiley 1991). This finding is very odd, firstly because it is based on the fact that one country which had only 10 grades achieved poorly compared to the others and, secondly, because countries which had 13 grades did not appear to achieve significantly above the level of grade 12. If the report suggests that countries which have only 10 grades should add two more grades to their years of schooling in order to raise achievement, it implies that countries which have 13 grades have, in fact, invested time and ressources in a way unlikely to raise their performance levels (on the items tested). What should policy makers do with this finding? Should they conclude that 12 grades must always be the norm? And what if grade 13 added something substantial which happened to be outside the content covered by the tests?

Some countries have faster students.

In science (SISS), students learn faster in some countries than they appear to do in others (Postlethwaite & Wiley 1991). But why? And what can countries with slower students do to help them learn faster?

Positive attitudes toward learning helps.

Attitudes and values favourable to school learning in general and the subject under investigation in particular are positively related to achievement outcomes. The effect is unique in the sense that it stands out even after other, possibly confounding, factors have been taken into account. It could be demonstrated for different aspects of math and science (Husén 1967; Comber & Keeves 1973; Postlethwaite & Wiley 1991; Keeves 1992), and in French as a foreign language, that a positive attitude toward learning has also been shown to have a positive effect on achievement (Carroll 1975). How to motivate children has always been a question for teachers. There is work to be done. Anderson (1991) suggests a revision of pre-service-training for teachers, more efficient and systematic in-service-training programmes and more communication between teachers and planners.

Class size makes no difference.

There is no linear relationship between class size and achievement consistent across countries at any level of schooling (Husén 1967; Comber & Keeves 1973; Postlethwaite & Wiley 1991; Keeves 1992). This finding, which is clearly opposed to the common wisdom of teachers, is the source of considerable irritation. If the relationship is curvilinear, the optimal class size (associated with maximum achievement) is still to be found … if there is any. But if there is none, can policy makers enlarge class sizes with no limits?

3.3 New Findings which Give Concrete Answers to Policy Makers and Educational Planners.

The teaching of science should include between 3 and 10 subjects.

Separate subjects in science appear to produce better results than integrated science (SISS: Postlethwaite & Wiley 1991), but it is not quite clear just how

many subjects should be taught in science. The degree of specialisation in grade 12 ranges between countries from 3 to 9 or 10 subjects, and, in this range, it did not seem to have any influence on the subjects tested. This leads the IEA report to advise that grade 12 students should cover 5 or 6 subjects which is somewhere in the middle between 3 and 10 subjects (Postlethwaite & Wiley 1991). But, knowing that the number of subjects between 3 and 10 does not make much difference, one could also argue that policy makers should choose what suits them best, as long as they avoid integrated sciences. This is a finding which can help policy makers.

Homework should be given.

Across countries, more homework results in more learning and better achievement (Husén 1967; Comber & Keeves 1973; Postlethwaite & Wiley 1991; Keeves 1992) and more reading at home has a positive influence on reading achievement (Postlethwaite & Ross 1992). The IEA data do not say how much homework should be given per subject to reach a good balance and what must be regarded as the upper limit for the aggregated assignments. But the general implication for policy makers is ambiguous: homework should be retained as a regular component of school-related learning because it enhances achievement. The Reading Literacy survey suggests, however, that not all types of homework are equally effective. Anderson (1991) argues that teachers who structure well and supervise homework are likely to influence positively the achievement of their students. In reading, teachers and parents should try to encourage students to read frequently and intensively in their leisure time, to use libraries and to borrow books. Policy makers and educational planners should look for ways to institutionalize support and to engage teachers and parents in an effort to guide and counsel children and adolescents along this path (Postlethwaite & Ross 1992).

Streaming or tracking students according to academic ability is not always necessary.

The highest-achieving students tend to perform well under (nearly) all conditions: they do as well or nearly as well in less selective systems as they do in more selective ones (Comber & Keeves 1973). In science, they do well regardless of their immediate peers' performances (Postlethwaite & Wiley 1991), and, at the end of the secondary stage, they do not suffer from increased retention rates. A policy implication might be to focus attention on how to

increase the academic level of the problematic students: they are the ones who might need special programmes or more attention in regular school settings.

4 Conclusion

As one can well see, quantitative international comparisons in education describe and explain quite adequately many aspects of education around the world but they give few indications as to how to deal with the probelmatic aspects identified. One cannot expect these studies to produce fail-safe recommendations for national/regional policy making and for educational planners in each country tested as a result of the way they are currently designed and implemented. If a country wants to participate in an IEA Study - or a similar international exercise - with such unrealistic goals in mind, it would be better advised to withdraw from the study. On the grounds of such arguments, the question has been raised whether these international studies were not too expensive, considering the benefits they have provided for policy makers and educational planners. If one shares the assumption embedded in this question, one will have to agree. But, despite of the limitations of quantitative international studies (note the reservations advanced in many of the international reports), their main advantage is that they provide clear evidence in problems which would be very diffcult or even impossible to analyze in strictly national studies. Moreover, the comparisons force the participating countries to consider their own policies on the basis of information on a whole range of options whose effects can be investigated only internationally. The dynamic of this line of research lies in the fact that surveys are self-correcting over time and can be challenged or verified constantly by other research. To some extent this is already an essential part of internationally comparative studies, as they are explicitly designed as "multiple replications", with the same design implemented in each participating country. Another aspect in the repetition of IEA studies over time should be mentioned: "time series" are by no means restricted to descriptive data and can be extended to the even more interesting field of relational findings. So, the major results of IEA (or similar) cross-national studies seems to lie in their contributions to an expanded awareness of the available options and to the kinds of effects associated with them. And, as was argued here, this has direct effects on policy making.

Notes:

1. "*International Association for the Evaluation of Educational Achievement*". The IEA Studies are: First International Mathematics Study (FIMS); Six Subject Study (6SUB): FISS, Reading, Literature, French, English, Civics; Second International Mathematics Study (SIMS); Second International Science Study (SISS); Classroom Environment Study (CES); Computers in Education Study (COMPED); Written Composition Study (WC); Preschool Education Study (PS); Reading Literacy Study (RL); Third Mathematics and Science Study (TIMSS).

2. Belgium, Bulgaria, Canada (Alska) (BC) (NB) (ON) (Ontario), China, Chinese Taipei, Columbia, Cyprus, Czech Republik, Denmark, England-Wales, France, Germany, Greece, Hong-Kong, Hungary, Iceland, Indonesia, Iran, Ireland, Israel, Italy, Japan, Korea, Kuwait, Latvia, Lithuania, Mexico, Netherlands, New-Zealand, Norway, Philippines, Portugal, Romania, Russia, Scotland, Singapore, Slovakia, Slovenia, South-Africa, Spain, Sweden, Switzerland, Thailand, Tunesia, Ukraine, United States.

3. The question of what is objective or not is still a question.

Bibliography

Anderson, L.W. (1991). *Increasing Teacher Effectiveness*. Paris: UNESCO.

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

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

Harding, J. (1992). *Breaking the Barrier: Girls in Science Education*. Paris: UNESCO.

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

Husén, T. (1973). Foreword. In L.C. Comber & J.P. Keeves, *Science Achievement in Nineteen Countries* (pp. 13-24). New York: Wiley.

Keeves, J.P. (1992). *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.

Lewy, A. (1991). Decisions of Educational System Level. In A. Lewy (Ed.), *The International Encyclopedia of Curriculum* (pp. 349-354). 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. (1991). *The IEA Study of Science II: Science Achievement in Twenty-Three Countries*. Oxford: Pergamon.

Schleicher, A. (1994). International Standards for Educational Comparisons. In A.C. Tuijnman, & Postlethwaite, T.N. (Eds.), *Monitoring the Standards of Education* (pp. 29-247). Oxford: Pergamon.
Travers, K.J. & Westbury, I. (1989). *The IEA Study of Mathematics I: 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.
Wolf, R.M. (1993). *Judging Educational Research Based on Experiments and Surveys* Paris: UNESCO.