MATHEMATICS IN MIDDLE SCHOOLS IN WESTERN EUROPEAN COUNTRIES

W. J. PELGRUM

University of Twente, Department of Education, The Netherlands

Abstract

The International Association for the Evaluation of Educational Achievement has conducted a number of cross-national studies in which Western European countries participated. Results from the Second International Mathematics Study regarding the content and outcomes of this study in some Western European countries, Japan and the U.S.A. are presented and compared. The results show that comprehensive systems do not differ systematically from non-comprehensive systems with respect to achievement measures of students. However, measures of the implemented curriculum show differences which are consistent with system characteristics. Consequences for the comparative methodology are discussed.

Introduction

Education is in most societies a matter for serious discussions; its deficiencies are considered and improvements suggested. Many people and agencies are engaged in this process, which according to Coleman (1984) can be characterized by a pluralistic model of societal decision-making. If the decision-making is to be effective, it is important to have information about the functioning of education. Output indicators are particularly important as a basis for (re)directing measures. Banathy (1973) categorically states that systems cannot survive if these measures are not available. In middle school discussions the need for information about possible effects is often expressed. Whatever the claims of proponents and opponents may be, the ultimate "proof" of anticipated effects should be empirically determined [possibly Hofstee's (1985) betting model] in terms of the output the system is expected to realize. Opponents of an alternative system do not usually accept a long period of disturbance in the system, when the only goal is to check whether it was possible to demonstrate the expected effects. We must therefore look for other ways of collecting information about the effects of certain systems. Effect studies in the strict sense of cause and effect are not possible when issues are concerned which deal with the organization of the total system. It is, however, possible to try to exploit natural variation in order to study the potential effect of certain organizations. One possible way is the

conducting of international research in order to compare the outcomes of education in educational systems which are differently organized. The studies of the International Association for the Evaluation of Educational Achievement (IEA) are designed in such a way that they can be used for this goal. In this article I present results from the Second International Mathematics Study (SIMS), which are relevant for the middle school discussion. First there is a short description of the design of the study and overall statistics for all countries as a background for discussion of the comparative methodology and a more detailed presentation of results from selected Western European countries. In the final discussion special attention will be paid to the question of how a more integrated approach to national and international monitoring could help increase our insight into the functioning of educational systems.

The IEA Mathematics Study: Background Information

The Second International Mathematics Study (SIMS) was conducted in the following Western European countries: Belgium (Flemish and French), England, Finland, France, Luxembourg, The Netherlands, Scotland and Sweden. Outside Western Europe the following countries participated: Canada (Ontario and British Columbia), Hong Kong, Hungary, Israel, Japan, New Zealand, Nigeria, Swaziland, Thailand and the U.S.A.

In this article I discuss only the results from the Western European countries, Japan and the U.S.A. in order to have some contrastive reference. Belgium (French) and Scotland are also not discussed because for the purpose of this article no complete data were available for these systems. In SIMS two study components were distinguished: a cross-sectional study with data collection at one moment in time, and a longitudinal component in which data were collected twice, at the beginning and end of the school year. An important aspect of the design of the SIMS is that the mathematics curriculum was conceived from out of three different perspectives. A distinction was made between the *intended, implemented* and *attained curriculum*. The intended curriculum was defined as the curriculum as is prescribed by educational authorities in terms of formal documents or textbooks. The implemented curriculum is the curriculum as it is put into practice by teachers. The attained curriculum consists of outcomes (cognitive and affective) at student level. Not only may countries differ with respect to one or more of these curriculum definitions but there may also be variations within a single country.

The attained curriculum was measured with a multiple choice test of respectively 176 and 180 items for the cross-sectional and longitudinal study, distributed over subtests called Arithmetic, Algebra, Geometry, Statistics and Measurement. The overlap between cross-sectional and longitudinal study components consists of 157 items. The implemented curriculum was measured with a so-called Opportunity to Learn (OTL) questionnaire and a teacher questionnaire. The teacher questionnaire is not relevant for this article and will not be further described. The Opportunity to Learn questionnaire involved for each cognitive test item, a judgement of the teacher with respect to the estimated difficulty of the item and a rating indicating whether the corresponding subject matter had been taught to the students participating in the study. The intended curriculum has been registered by letting experts from each country judge the importance of cells in a grid consisting of content and behavioural categories. Two populations of students were distinguished:

students in their final examination year in secondary education (population B) and 13 year old students (population A). The actual population definitions were more exact. In practice participating countries did not choose exactly comparable populations. In Table 6.1 there is a list of the actual year levels in which the population A students in the selected countries were located and the mean age of these students.

Table 6.1
International Year Level of Population A Students in SIMS and Mean Age

Countries	Year	Age
Belgium-Flemish	8	14.2
England	9	14.1
France	9	14.1
Luxembourg	8	14.5
Netherlands	8	14.4
Sweden	7	13.9
Japan	7	13.5
U.S.A.	8	14.1

Source: data archives SIMS.

Table 6.1 shows that there are differences between countries in the choice of the year for population A. It is important to take this into consideration in interpreting the results and differences between countries. The achievement test results from the selected countries are presented in Table 6.2. Although the table gives a first impression of differences between countries it is quite difficult to interpret without additional information, as I show in the next section.

Table 6.2
Percentages Correct on Total Test and Subtest for Each Selected System

Countries	Total	Arit	Algb	Geom	Stat	Meas
Belgium-Flemish	52	57	51	42	58	58
England	47	48	39	44	60	48
France	53	58	55	38	57	60
Luxembourg	39	48	34	26	39	52
Netherlands	58	60	52	53	67	63
Sweden	44	43	34	40	60	52
Japan	62	60	60	57	71	69
U.S.A.	45	51	43	38	58	41

Arit: arithmetic; Algb: algebra; Geom: geometry; Stat: statistics; Meas: measurement.
Source: Pelgrum, Eggen & Plomp (1986).

Systems Output Evaluation and Curricular Variation

Most authors would agree that the ultimate outcome of an educational system is seen in terms of students' attainments, but it is much less clear how these outcomes should be measured and valued. Although it is common practice to measure student outcomes by means of standardized achievement tests, there is growing doubt about the usefulness of this practice. The reason for this doubt is that standardized achievement tests are based on

the assumption that tests can be developed which are equally fair for situations which in many respects may differ. There is a growing body of literature showing that educational opportunities are not equally distributed or implemented across different local situations (Madaus, 1983; Leinhardt & Seewald, 1981). McLean (1985) states that there is curriculum diversity everywhere, even in countries where it was thought that there was a strong and effective central control. Therefore in order to understand the outcomes at student level it is important to take into account the local implementation of the curriculum as measured by tests. This is in fact what is done with the Opportunity to Learn measures. However, the OTL measures are not only interesting as a way of controlling for the match between test and curriculum (Anderson, 1985), but are in themselves valuable. Analogous to the way an achievement test is conceived as a representative sample of items from an infinite domain, the OTL measure may also be conceived as a representative estimate of the curriculum as implemented by teachers. This conceptualization of OTL gives interesting possibilities for empirical study of the way that the curriculum is organized in a country. Table 6.3 illustrates the use of OTL in describing the global curricular differences between school types in the Netherlands with respect to mathematics and science. The science data in Table 6.3 were taken from the Dutch participation in the Second International Science Study of IEA.

Table 6.3
Opportunity to Learn with Respect to Mathematics and Science in Dutch School Types (Average Percentage Items Taught)

Subject	S1	S2	S3	S4
Mathematics	72	70	59	52
Biology	74	67	56	69
Chemistry	79	71	25	28
Physics	68	64	69	44

S1: senior secondary education; S2: junior secondary education; S3: technical vocational education; S4: domestic science education (+ for Science: administrative education).

Table 6.3 clearly shows the differences between school types which are consistent with the curricular guidelines that are provided for these school types. The fact that Opportunity to Learn differs so much within a system affects the interpretation of test scores. Opportunity to Learn may be quite unevenly distributed across different ability levels of students or (sub)groups of students or may have differential effects depending on ability level. Consequently, the interpretation of test scores in terms of system outcomes or effects may interact with ability levels. This may be illustrated with Figure 6.1 in which for different intervals of total score on the mathematics test the p-value of instructed and non-instructed groups (OTL and NOT OTL) of students has been plotted. The item in question deals with the Pythagorean theme. From Figure 6.1, one may conclude that the effect of education interacts with the general mathematical ability of students. This illustrates that OTL contains information which is relevant for interpreting test scores. Therefore I use OTL in this article as an important variable to describe differences between countries with respect to mathematics education.

```
1.00
0.9                              I ——————→ OTL
0.8                         I
0.7                      I
0.6               I
P 0.5
0.4         I  I           O
0.3     I  I  O  O  O  O      O ————→ Not OTL
0.2  O  O  O
0.1
0.0
     0  I  2  3  4  5  6  7  8  9
            Total score
```
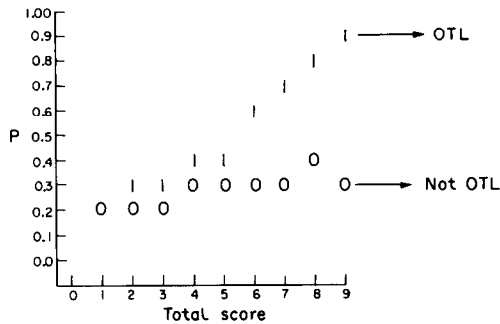
Figure 6.1 P-values of an item dealing with the Pythagorean theme in instructed (OTL) and non-instructed (NOT OTL) groups of students in different ability groups (intervals of the total percentage correct: 0 = 0–9%; 1 = 10–19%, etc.).

Mathematics Education in Western European Countries: Some Indicators

As a background for the presentation of results from the mathematics study we first will describe some of the characteristics of the selected systems with respect to the organization of the system and mathematics education. Table 6.4 contains a brief statement for each system characteristics of the school system at the level of population A and how many mathematics courses exist at that level.

Table 6.4
Description of System Characteristics at Population A Level and the Number of Mathematics Courses

System	Description
Belgium-Flemish	General secondary, technical secondary, vocational secondary, traditional and non-traditional school types.
England	Mostly comprehensive. Three different mathematics courses.
France	Mostly comprehensive. One course.
Luxembourg	Secondary general and secondary technical with and without streaming. Three different mathematics courses.
Netherlands	General secondary and vocational secondary. Four different courses.
Sweden	Comprehensive. Two different courses.
Japan	Comprehensive. One course.
U.S.A.	Junior high, middle school and elementary school. Three different courses.

The overview in Table 6.2 shows that France and Japan can be called "true comprehensive" systems for mathematics. Other "comprehensive" systems like England and Sweden still differentiate between students by offering different courses.

In Table 6.5 the distribution of total test score and OTL are given for the selected Western European countries which participated in SIMS and for which achievement and OTL data were available, as well as for Japan and the U.S.A., while Figures 6.2 and 6.3 show a plot of the attained and implemented maths curriculum for some of these countries.

From Table 6.5 and Figures 6.2 and 6.3, a number of observations can be made. First it is clear that there are large differences in the central tendencies of the distributions between countries, not only in the level of attainment of students, but also in the amount of implementation of mathematics content at population A level. The fact that across

Table 6.5
Relative Percentages of Students Per Interval of the Total Percentage Correct (TEST) and of Teachers for
Intervals of the Total Percentage OTL Weighted According to Stratum Weights

System		0	1	2	3	Intervals* 4	5	6	7	8	9
Belgium-Flemish	TEST	0	2	7	9	14	20	18	17	10	1
	OTL	0	1	4	3	9	18	36	25	4	0
England	TEST	0	8	18	19	16	13	10	9	5	1
	OTL	0	3	4	7	10	16	21	20	16	3
France	TEST	0	1	7	17	24	24	27	8	3	0
	OTL	0	0	0	1	0	6	34	46	12	1
Luxembourg	TEST	0	4	15	26	27	18	7	3	0	0
	OTL	3	0	2	6	20	28	35	7	0	0
Netherlands	TEST	0	1	6	10	13	17	18	21	12	2
	OTL	0	1	2	2	6	14	26	29	19	1
Sweden	TEST	0	6	22	24	23	15	8	2	0	0
	OTL	0	3	6	15	30	34	11	1	0	0
Japan	TEST	0	2	4	7	10	16	22	24	14	1
	OTL	0	0	1	0	1	3	13	60	22	1
U.S.A.	TEST	0	8	19	17	16	15	12	9	4	0
	OTL	0	1	1	3	6	16	27	24	17	6

* 0 = 0–9%; 1 = 10–19%; etc.

countries the population A students are located in different year levels, may be an important explanatory variable for some of the inter-country differences. Besides the differences in central tendencies, it appears that especially the implemented curriculum in some countries is more homogeneous than in other countries. The systems with a comprehensive system offering one mathematics course to population A students (Japan and France) are the most homogeneous in this respect. With the exception of Sweden and Luxembourg the range of the total test score distribution is the same for all systems, although the relative size of high and low ability groups differs considerably between systems. This also means that, at least from a cross-national perspective, the comprehensiveness of a school system is not necessarily associated with major differences at the extremes of the distribution.
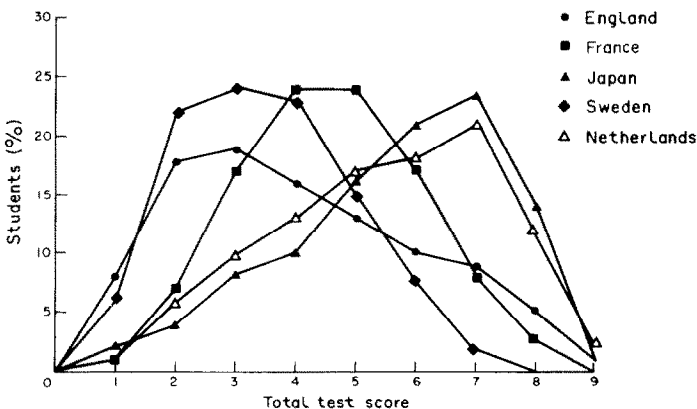

Figure 6.2 Percentages of students (weighted) in intervals of the total percentage correct on the mathematics test
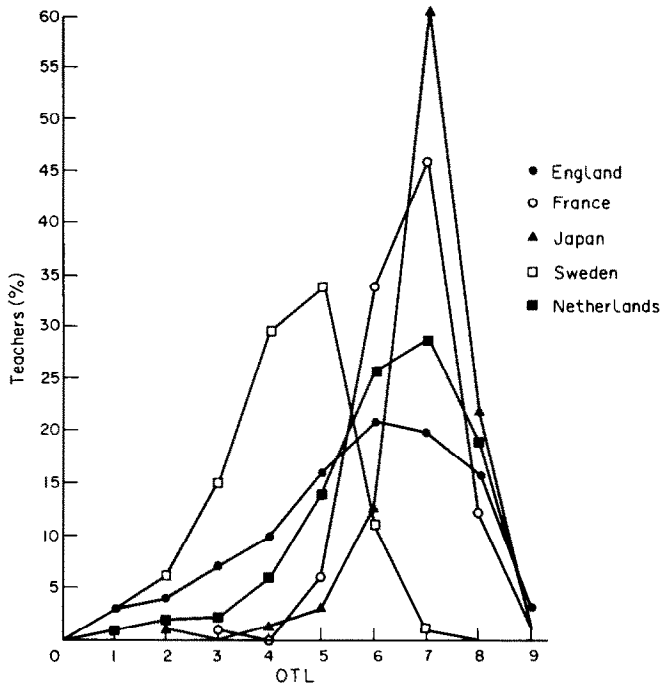(0 = 0–4%; 1 = 5–14%; etc.).

Figure 6.3 Percentage of teachers (weighted), in intervals of the total percentage items taught (0 = 0–4%; 1 = 5–14%; etc.).

In the Netherlands the curricular arrangements in terms of OTL differ considerably between school types. The lower ability students are generally confronted with less subject matter than the higher ability students. An interesting question is how different systems deal with the implementation of the mathematics curriculum in different ability groups. We calculated the percentage of OTL in different ability groups of students; in Figure 6.4 the results are presented. Figure 6.4 shows that there are some systems in which Opportunity to Learn is fairly stable across ability levels, while in other systems the mathematics curriculum for the high ability students is much richer than for low ability students. Especially in Japan and France, the amount of OTL is equally distributed across ability levels. This finding is consistent with the fact that both systems are comprehensive and that there is only one mathematics course at this population level. In the Netherlands the Opportunity to Learn of high and low ability students differs considerably. In Sweden, with its comprehensive system, there are also quite large differences, probably due to the two courses. It should be noted that the same amount of OTL not necessarily reflects comparability of the implemented curriculum. More detailed analyses would be needed to determine which differences exist in relative emphasis on certain (sub)domains.

Figure 6.4 is also interesting in that it shows that more OTL is not always associated with higher achievement, even though from the curves of some countries (e.g. the Netherlands and Sweden) one might gain this impression. The results of the analyses above show that total test score comparisons between and within countries may be unfair, because of the differences of implementation of the mathematics curriculum as measured by the tests. It
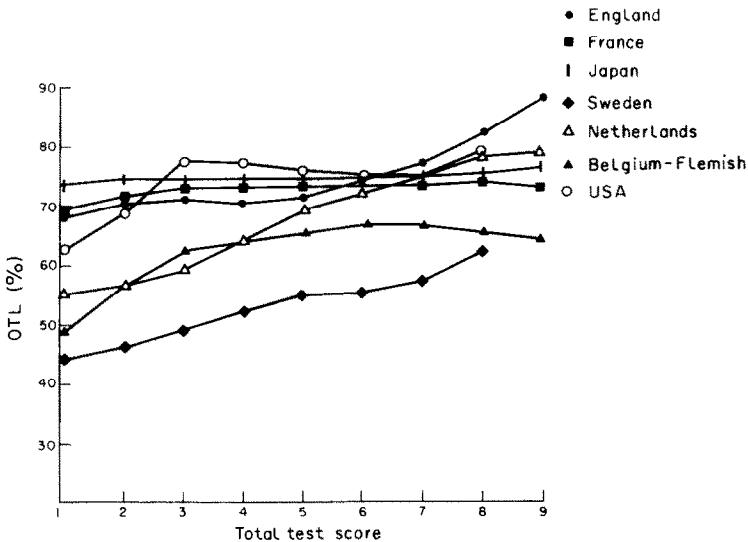
Figure 6.4 Percentage of OTL per ability group according to the total test score (0 = 0–4%; 1 = 5–14%; etc.).

is therefore necessary to take OTL into account when comparing test results across and within countries. Suitable ways of doing this are in a very early stage of development (Muthen, 1987). Item difficulties and student abilities need to be estimated independent of OTL and problems of equivalence of total test and subtests between and within countries need to be resolved.

We followed a simple item-based procedure for making fair comparisons of achievement results between countries. This procedure consists of estimating the student's mathematical ability by first calculating total scores of students (in terms of a percentage correct score) based upon the items which were taught to each student. Next, each student was allocated to a low, medium and high ability group (corresponding with respectively the intervals 0–40%, 41–60% and 61–100%). For each country and per item the proportion correct (given that the item was taught) in each of these groups was calculated. These proportions constitute the basis for comparisons between countries and to find the set of items for each country in which under- or over-achievement occurs. As an example of the outcomes of this procedure we have compared Japan (which on total test and subtests scores very high) with other countries and identified in Table 6.6 those items on which the different ability groups of Japanese students score considerably lower (i.e. differences greater than 0.20) than comparable groups in other countries. The results show that, even though in the overall outcomes Japan ranks very high, there are still areas in which considerable under-achievement occurs.

Discussion

Although in the previous sections only some results from the IEA Second International Mathematics Study could be presented, the tables and figures give a reasonable overview

Table 6.6
Comparison of Japan with Other Countries

	0	10	20	30	40	50	60	70	80	90	100
POSITION TEST SCORES											
Other countries				1	1	11	11	11			
Japan							j				
POSITION CURRICULUM											
Other countries					11	21		11	1		
Japan									j		

Comparison by item of Japan with other countries. Criterion: 0.20

Group	Valid items	Item number	Content	N country
LOWAB	134			
		14	Which of the following sequences of numbers is i	3
		30	In which diagram below is the second figure the	6
		32	Here is a table that shows the number of trees p	2
		42	Which of the following is equal to a quarter of	3
		46	20% of 125 is equal to	4
		93	Which of these is a correct statement for this t	2
		99	In the graph, rainfall in centimetres is plotted	3
		110	In a school election with three candidates, Joe	2
		140	7 3/20 is equal to	2
		164	According to the scale shown, the length of side	6
		166	What is the surface area of this solid rectangul	2
MEDAB	139			
		6	Alexandra walked from Riverview to Bridgeport, w	3
		8	In a school of 800 pupils, 200 are boys. The rat	2
		9	30 is 75% of what number?	7
		13	The air temperature at the foot of a mountain is	2
		14	Which of the following sequences of numbers is i	3
		29	One of the following points can be joined to the	5
		30	In which diagram below is the second figure the	4
		47	If the ratio of 2 to 5 equals the ratio of N to	3
		87	The Davis family took a car trip from Anabru thr	2
		88	A/15–B/5 is equal to	2
		93	Which of these is a correct statement for this t	4
		99	In the graph, rainfall in centimeters is plotted	2
		122	(Triangle with 2 angles given) X is equal to	2
		126	The straight line joining the points (2,3) and (	4
		133	How many pieces of pipe each 20 metres long woul	2
		140	7 3/20 is equal to	3
		166	What is the surface area of this solid rectangul	2
HIGAB	133			
		6	Alexandra walked from Riverview to Bridgeport, w	3
		8	In a school of 800 pupils, 300 are boys. The rat	4
		29	One of the following points can be joined to the	3
		126	The straight line joining the points (2,3) and (	2

of the system and curricular characteristics for one school subject. The results show that system comparisons are not only revealing with respect to the attained curriculum, but especially that the implemented curriculum is an important aspect of cross-national system comparisons. What the data regarding OTL showed, is that between but also within systems, there are great differences in the degree to which mathematics is implemented. It is interesting that a global implementation indicator in terms of the total amount of OTL

reflects more clearly certain system characteristics (e.g. comprehensiveness) than the achievement test results. For those who expect that comprehensiveness leads to more equal opportunities between students, the implementation measures could constitute interesting evidence. However, it should also be pointed out that the achievement results seem to provide contrary evidence.

The implementation measures in terms of OTL are not only interesting for this descriptive purpose, but may also lead us to rethink the use and interpretation of test scores in (international as well as national) comparative research. The fact that OTL differs so much between and within countries means that, although the achievement tests may be a good indicator for the abilities of students in the total mathematics domain, unless there is careful control the tests are less suited as an indicator of the effectiveness of the mathematics curriculum. In theory it is possible that low OTL is associated with higher effectiveness than high OTL.

In experimental designs the effects of certain curricula may be measured by comparisons of achievement before and after instruction. For survey assessments at one point in time OTL may be an interesting variable to help identify the baseline test scores for uninstructed groups (as was done in Figure 6.1). Effect measures could then be derived from the differences between instructed and non-instructed groups, while comparisons are then useful to compare (inter)national (sub)populations with respect to these effects. Besides that, comparisons between countries could also be based on the achievement results of the instructed groups of students only, as was illustrated in Table 6.6. Procedures still need to be developed and psychometric models would need to be modified in order to follow such an approach. But ultimately it would produce more valid baseline information concerning student achievement and the curriculum implementation of teachers.

References

Anderson, L. W. (1985). Opportunity to learn. In T. Husen & T. N. Postlethwaite (Eds.), *The International Encyclopoedia of Education*. Oxford: Pergamon Press.

Banathy, B. B. (1973). *Developing a systems view of education: The systems-model approach*. Belmont, California: Lear Siegeler Inc./Fearon Publishers.

Hofstee, W. K. B. (1985). A betting model of evaluation research. In B. P. M. Creemers (Ed.), *Evaluation research in education*. The Hague: SVO.

Leinhardt, G., & Seewald, A. M. (1981). Overlap: what's tested, what's taught? *Journal of Educational Measurement*, **18**(2), 85–96.

Madaus, G. F. (Ed.) (1983). *The courts, validity and minimum competency testing*. The Hague: Kluwer-Nijhoff Publishing.

McLean, L. (1985). *Drawing implications for instruction from item, topic and classroom-level scores in large-scale science assessment*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Muthen, B. O. (1987). *Using item specific instructional information in achievement modeling*. Paper presented at the 1987 AERA meeting in Washington D.C.

Pelgrum, W. J., Eggen, T. J. H. M., & Plomp, Tj. (1986). *The implemented and attained mathematics curriculum: a comparison of eighteen countries*. Enschede: University of Twente, Department of Education.