Die Stiftung für liberale Politik

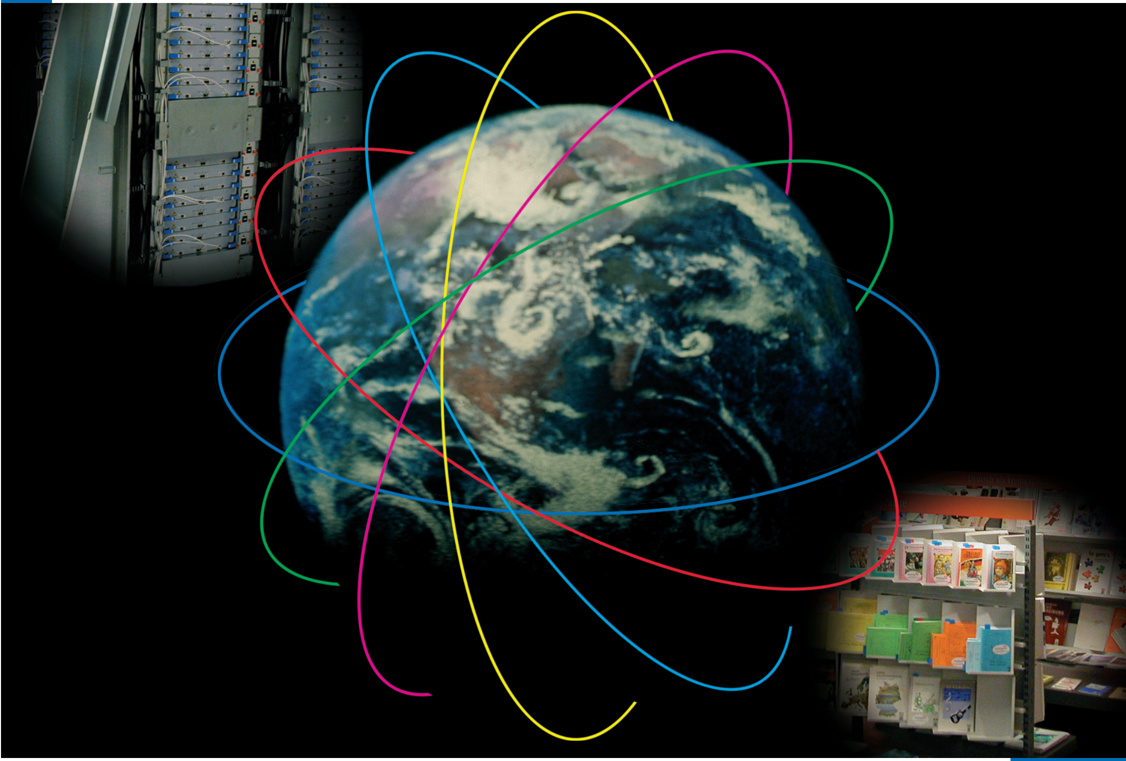
JEDER MENSCH BRAUCHT
FREIHEIT, UM SEINE
ANLAGEN UND FÄHIGKEITEN
ENTFALTEN UND
VERWIRKLICHEN ZU KÖNNEN.
OHNE FREIHEIT ERMÜDET
DER MENSCHLICHE GEIST,
VERFALLEN KULTUR UND
WISSENSCHAFTEN, STAGNIERT
DIE WIRTSCHAFT.
GEISTIGES LEBEN BRAUCHT
FREIHEIT GENAUSO, WIE DER
KÖRPER DIE LUFT ZUM ATMEN.

Liberales Institut

Friedrich-Naumann-Stiftung

Jaap Scheerens

The Use of International Comparative Assessment Studies



OccasionalPaper 12

Imprint:

Published by
The Liberal Institute of the
Friedrich Naumann Foundation
Truman House
Karl-Marx-Straße 2
D–14482 Potsdam

Phone +49 (3 31) 70 19 - 2 10
Fax +49 (3 31) 70 19 - 2 16
libinst@fnst.org
www.libinst.de

Production
COMDOK GmbH
Office Berlin
Reinhardtstraße 16
D–10117 Berlin

Printed by
ESM Satz und Grafik GmbH
Wilhelminenhofstraße 83-85
D–12459 Berlin

First Edition 2006

The Use of International Comparative Assessment Studies to Answer Questions About Educational Productivity and Effectiveness

Jaap Scheerens

Paper presented at the conference
"Liberal Education – International Perspectives" organised
by the Liberal Institute of the Friedrich-Naumann-Stiftung,
Potsdam 2-4. September 2005.

Introduction

International comparative assessment studies can be used to rank countries on the basis of their mean performance on a particular test. This practice has often been ironically compared to a „horse race" between countries or to the Olympic Games. Yet, it is the most prevalent way in which the data from international studies are used and published. It is also to be seen as a perfectly sensible approach in a situation where countries are increasingly concerned about performance „standards" in education. While the setting of norms in order to operationalise standards is likely to be a politically (and to a certain extent also technically) complex matter (Scheerens, 2004), the comparative performance of other countries on the same criterion provides a reasonable solution. The more so, if the performance levels on the test, have clear interpretations in terms of proficiency levels (PISA) or content covered (TIMSS). This approach to standard setting is generally indicated as international benchmarking.

In this paper four different types of uses of international comparative assessment studies will be briefly described and illustrated:

- comparison of country mean scores on a particular achievement test;
- analyzing between schools, between classes and between student variation;
- separating the effects of „given" conditions and malleable factors;
- answering questions about the effectiveness of specific school, context and classroom characteristics.

These applications can be seen as specific interpretations of educational quality, namely in terms of productivity, effectiveness and equity. These terms will be clarified first:

According to the productivity interpretation of educational quality outcomes, either in the sense of achievement on performance tests, educational attainment or societal impact of schooling, matter most. In the case of effectiveness the question about the instrumental association, between context, input and process indicators on the one hand, and outcomes on the other is the central issue. Effectiveness goes beyond establishing productivity by addressing the why behind performance differences. It is therefore a scientifically more ambitious quality perspective as well as a potentially more policy relevant one, because it would offer handles to improve education. Equity focuses at the variability of performance between different units (e.g. students, home background, geographical regions).

In the final section of the paper conclusions will be drawn about the possibilities and limitations of international studies to answer questions about educational productivity, school effectiveness and equity.

The comparison of country average scores as a measure of productivity

The most frequent use that is made of the results of internationally comparative assessment studies, as those carried out by the IEA and the OECD, is to compare the country mean scores on a particular achievement test. When standard errors are presented with these averages, differences between countries that are over 2 times the standard error indicate statistical significance.

The table below (Table 1), representing the results of the OECD PISA study in reading literacy, carried out in 2000, is presented as an illustration. When countries differ about 8 points from one another the differences are statistically significant.

The PISA results illustrate the substantial difference between the higher and lower scoring countries.

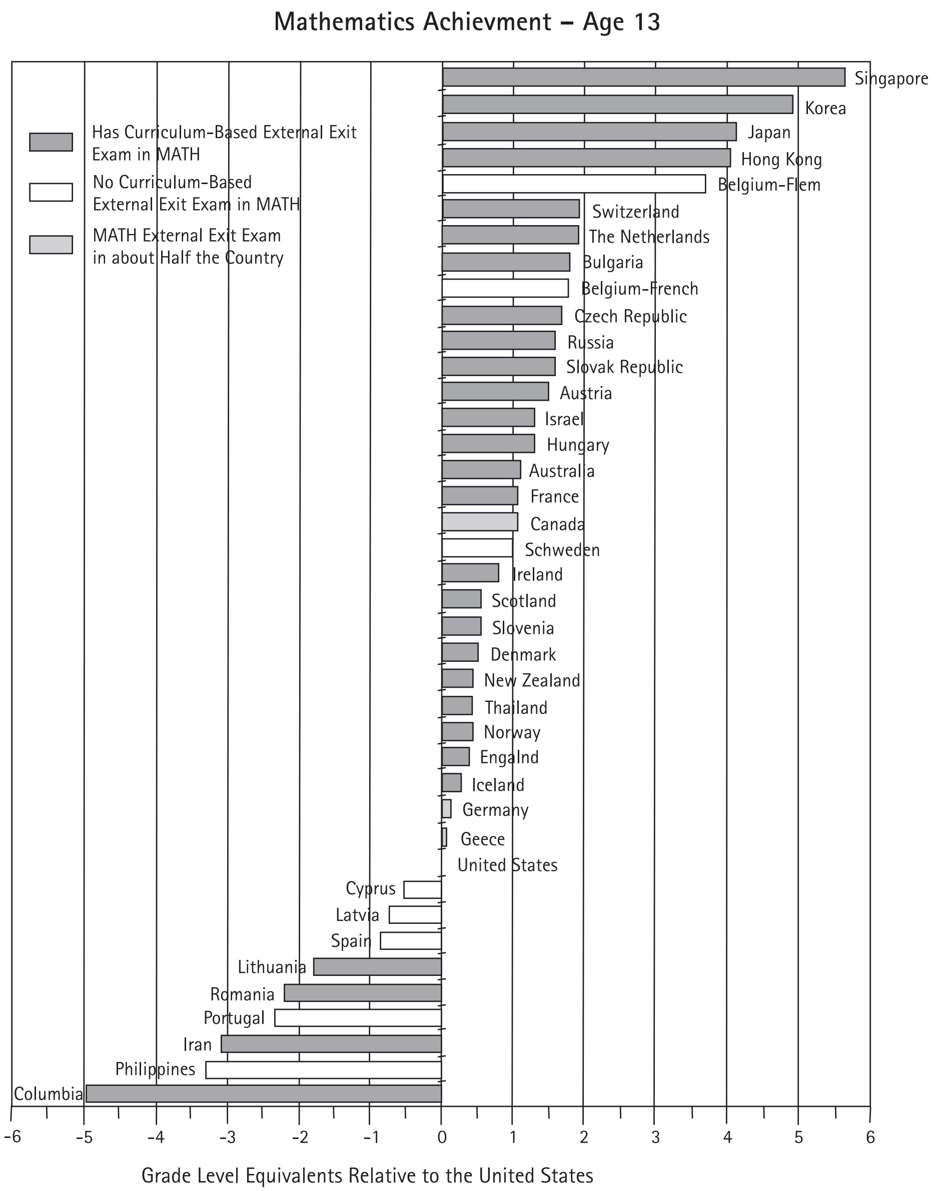
Data as presented in Table 1 could be used as targets or benchmarks. Countries, for example might take the international average as a target for a future comparison.

Table 1: Average achievement in reading literacy (Source: PISA 2000 and PISA plus).

Countries are ranked according to their average reading literacy score.

	Reading literacy score**
Finland	546
Canada	534
Netherlands	532
New Zealand	529
Australia	528
Ireland	527
Korea	525
Hong Kong	525
United Kingdom	523
Sweden	516
Austria	507
Belgium	507
Iceland	507
France	505
Norway	505
United States	504
Denmark	497
Switzerland	494
Spain	493
Czech Republic	492
Italy	487
Germany	484
Hungary	480
Poland	479
Greece	474
Portugal	470
Russian Federation	462
Latvia	458
Israel	452
Luxembourg	441
Thailand	431
Bulgaria	430
Mexico	422
Argentina	418
Chile	409
Brazil	396
Macedonia	372
Indonesia	371
Albania	349
Peru	322
Average across countries	473

Table 2: Grade level equivalents relative to the United States (source: John H. Bishop, 1997)



A perhaps more insightful way to present comparisons in achievement is to express differences in mean scores in terms of grade equivalents. A grade equivalent difference is defined as the difference in mean score between students at the beginning and the end of a particular grade level, in a particular country. In the table above, cited from Bishop, grade equivalents were defined on the basis of data from the USA. His table, by the way, contains also information on countries having a standard based examination system or not. The data are from TIMSS, 1995.

As the number of international comparative assessments has risen, particularly now that both OECD (PISA) and IEA (TIMSS) regularly assess mathematics and science performance, it becomes possible to examine the consistency of countries being ranked high on various subjects and on various occasions. In Table 3, an overview is given of countries that were ranked among the „top ten", in at least 1 of 8 studies. The studies are: PISA 2003 mathematics, PISA 2003 Reading Literacy, PISA 2003 science, PISA 2000 Reading Literacy, PISA 2000 mathematics, PISA 2000 science, TIMSS 2003 mathematics and TIMSS 2003 science.

Table 3: Ranking of the „top ten" countries in 8 international assessments; the last column shows the proportion of being in the top ten out of all studies in which the country participated (Sources: OECD, 2001 and 2004, IEA, 2004, a and b).

	PISA 2003 Math.	PISA 2003 Read.	PISA 2003 Science	PISA 2000 Read.	PISA 2000 Math.	PISA 2000 Science	TIMSS 2003 Math.	TIMSS 2003 Science	
Hong Kong	1	10	3	8	2	3	3	4	8/8
Finland	2	1	1	1	6	4			6/6
Korea	3	2	4	7	4	1	2	3	8/8
Netherlands	4	9	9	3	1	7	7	9	8/8
Lichtenstein	5	5	5						3/3
Japan	6		2	10	3	2	5	6	6/7
Canada	7	3		2	8	6			5/6
Belgium	8						6		2/8
Macao China	9								1/3
Switzerland	10				9				2/6
Australia		4	6	5	7	9		10	6/8
New Zealand		6	10	4	5	8			5/8
Ireland		7		6		10			3/6
Sweden		8							1/8
Czech Rep.			9						1/6
Macao China			7						1/3

PISA	PISA 2003 Math.	PISA 2003 Read.	PISA 2003 Science	PISA 2000 Read.	PISA 2000 Math.	TIMSS 2000 Science	TIMSS 2003 Math.	2003 Science	
UK				9	10	5		7	4/8
Singapore							1	1	2/2
Chinese Taip							4	2	2/2
Estonia							8	5	2/2
Hungary							9	8	2/2
Malaysia							10	9	2/2
USA								10	1/8

Consistency of country averages across studies, as well as in relative position with respect to other countries can give rise to interesting interpretations. Arguments that are likely to be addressed in this are: the agreement of the test with the national curriculum, possible cultural biases, selectivity aspect of the sampling procedure, the strengths and weaknesses of age-based or grade-based samples and comparability aspects of national school systems (Prais, 2004; Adams, 2004).

Analyzing between school, between classes, and total between student variation

International comparative studies allow for comparisons of the patterns of variation in student achievement scores. The between school, within country variation, is the percentage of total between student variation that is explained by the factor school, i.e. it expresses the difference it would make for the average student to be enrolled in one school as compared to the next. The between classes, within school variation expresses the difference it would make for an average student in that school to be in one parallel classroom (at the same grade level) as compared to the next. A large total between student variation in a country shows that there is much heterogeneity in student performance, which is likely to be interpreted as low equity. A large between school variation expresses „implicit“ or „explicit“ segregation. The term implicit segregation could be used when, in a formal structural sense all schools are equal, as in the case of a comprehensive secondary education system. When there is still a large between school variation in a comprehensive system this could be caused by large school autonomy, or by selection policies of the schools or the parents that choose a school for their children. Explicit segregation appears when countries have a categorical school system, most common at secondary level. In categorical school systems students go to schools that cater to different ability levels.

Large between classes variation within schools, is indicative of within school tracking or streaming.

The table below, Table 4 shows between school and between classes variation patterns in countries that took part in the Second Mathematics Study of the IEA.

Table 4: Estimates of the Variance Explained by Schools and Classes (Source: Scheerens, Vermeulen & Pelgrum, 1989).

Country		Classroom variance component	School variance component
15	Belgium (Flemish)		.50
16	Belgium (French)		.64
22	Canada (British Columbia)		.27
25	Canada (Ontario)	.18	09
39	Finland	.45	.002
40	France	.17	.06
43	Hong Kong		.51
44	Hungary		.30
50	Israel	.22	.10
54	Japan		.08
59	Luxembourg	.29	.15
62	Netherlands		.67
63	New Zealand	.45	.01
72	Scotland	.34	.12
76	Sweden	.45	.00
79	Thailand		.39
81	USA	.46	.10

Note: Estimated of the variances expressed in terms of the intra-class correlation coefficient, for all countries, assuming schools are sampled at random within countries and classrooms are sampled at random within schools.

Since in 8 countries only one class per school was selected, classroom variance could not be separated from school variance in these cases. When looking at the results in Table 4 four groups of countries can be distinguished. First of all there are countries (Belgium Flemish, Belgium French, and The Netherlands) where there are vast differences in the mean achievement of students across schools: in this situation we have to do with vertically organized, strongly differentiated school systems. Secondly there is a group of countries with relatively small differences between schools but with large differences between classes within schools: the USA, Sweden, New Zealand, and Finland: this pattern indicates homogeneous grouping of pupils within a horizontally organized, integrated system of secondary schools. Next, there is a group of countries (Canada, France, Israel) where differences both between schools as well as between classes within schools are relatively small, probably because of (partially) mixed ability grouping within an integrated schooling system. Then of course there are countries that do not have a tracked, vertically organised system, but where de facto there are large quality differences between schools (most notably in Hong Kong and Thailand).

In a re-analyses of the PISA-2000 data set it appeared that different patterns can be discerned in countries having high versus low between school and total between student variation (Scheerens & Visscher, 2004). The following patterns appeared to occur:

High between school and high between students variation, e.g. **Germany**	Low between school and high between students variation, e.g. **New Zealand**
High between school and low between students variation, e.g. Korea, the **Netherlands**	Low between school and low between students variation, e.g. **Sweden**

The precise information is presented in Table 5 below.

Table 5: The total variance and the proportion of variance at the school level in reading literacy scores of the students based on an empty model. The 95% confidence interval (CI) of the proportion of variance at school level* is also presented. (source Scheerens & Visscher, 2004)

	Total variance	Proportion of variance at school level	Lower limit 95% CI	Upper limit 95% CI
OECD countries				
Australia	11407.42	0.21	0.16	0.27
Austria	9798.47	0.54	0.46	0.62
Belgium - French	12662.11	0.57	0.46	0.68
Belgium - Flemish	8601.80	0.52	0.43	0.61
Canada	9617.84	0.20	0.18	0.22
Czech Republic	8918.52	0.55	0.48	0.62
Denmark	9297.15	0.16	0.11	0.20
Finland	7640.28	0.07	0.04	0.10
France	8232.26	0.47	0.38	0.55
Germany	11761.27	0.60	0.53	0.66
Greece	9905.62	0.51	0.43	0.59
Hungary	8478.65	0.60	0.53	0.68
Iceland	8642.28	0.10	0.04	0.16
Ireland	8545.64	0.16	0.10	0.21
Italy	8373.22	0.53	0.45	0.61
Korea	5144.63	0.40	0.32	0.47
Luxembourg	9510.25	0.27	0.09	0.45
Mexico	7090.29	0.53	0.45	0.61
New Zealand	12057.62	0.17	0.12	0.23
Norway	10200.14	0.07	0.04	0.11
Poland	8975.21	0.59	0.50	0.68
Portugal	9068.37	0.36	0.28	0.44
Spain	7213.15	0.22	0.16	0.28
Sweden	8122.29	0.07	0.04	0.11
Switzerland	10423.20	0.42	0.36	0.49
United Kingdom	10017.19	0.31	0.26	0.36
United States	10826.89	0.27	0.20	0.34

Total	Proportion of variance	variance at school level	Lower limit 95 % CI	Upper limit 95 % CI
Non-OECD countries				
Brazil	7586.31	0.46	0.39	0.52
Latvia	10264.05	0.29	0.21	0.38
Russian Federation	8170.58	0.33	0.26	0.39
PISA-plus countries				
Albania	10286.00	0.41	0.33	0.49
Argentina	10507.14	0.44	0.36	0.53
Bulgaria	11394.92	0.57	0.49	0.65
Chile	8366.54	0.51	0.42	0.59
Hong Kong	7349.67	0.50	0.42	0.59
Indonesia	5043.70	0.46	0.40	0.52
Israel	12094.30	0.47	0.39	0.56
Peru	11706.09	0.63	0.55	0.70
Thailand	6667.43	0.35	0.28	0.41
Macedonia	9044.18	0.46	0.34	0.58
Netherlands**	6973.92	0.47	0.37	0.57

** response rate is too low to ensure comparability

As stated above, at the beginning of this section high between school variance indicates the degree of selectivity or segregation in a school system. The total between student variation on an achievement test in a particular country can be read as a measure of inequality of education. A large total between student variation indicates that an education system produces a lot of dispersion in actual learning outcomes; usually implying that a large proportion of students achieves at the low end of the score distribution. Further analyses of this distribution, for example by indicating which part of the student population is in the lowest percentile or quartile of the distribution can clarify this issue of inequality further. Most segregated and „unequal" are systems, which combine high total between student variation and high between school variation. Systems, such as Korea and the Netherlands, characterized by relatively low total between student variation and high between school variation, are selective in grouping students in schools, but manage to keep the total variation in achievement between limits. From an equity perspective school systems with high between school variation are still undesirable, since the selectivity is likely to be based on the socio-economic status of the students. School systems that have low between school variation and high total variation are probably systems with high degrees of internal tracking or streaming, leading to high between classes variation. Most „equal" are school systems that combine low total between student variation and low between school variation. Sweden is a case in point.

These examples show that studying patterns of variance based on international comparative assessments provide additional information to comparing average achievement levels. More particularly these patterns provide information on the internal segregation and differentiation of school systems and corresponding implications for the distribution of learning outcomes.

Separating the effects of „given" background conditions and malleable school variables

In school effectiveness research it is standard practice to adjust student achievement scores for background conditions, preferably prior educational achievement in the same subject matter area or scholastic aptitude and, as a second choice, by adjusting on the basis of socio-economic status or minority background. Only after these adjustments have been made are the impact of malleable school variables tested. It appears, however, that even after these adjustments have been made, the „aggregates" or „composites" of the student background characteristics, like the school's average socio economic status, still explain a sizeable part of the between school variation.

The „net" effects[1] of schooling can therefore be attributed to two categories of variables: student composition and malleable school variables like leadership styles, school climate and instructional strategies. Compositional effects are likely to be thought of as „given" factors, while the factors that are malleable in the sense that they are seen as „handles" to improve the primary processes of schooling, teaching and learning. On further reflection, however, it is clear that composition is also malleable, namely on the basis of overt admission or selection policies of the school. Or, influenced by selection processes by parents and students in the case of free school choice.

[1] „Net" effect in the sense of student achievement adjusted for student background characteristics at the individual level.

Databases of international comparative assessment studies provide an interesting source for assessing the relative impact of malleable school variables and school composition.

In Table 6, cited from Scheerens and Visscher, 2004, the relative impact of these two categories of factors is shown for the three subject matter areas that were covered in the PISA 2000 study, reading literacy, mathematics and science.

Table 6: Partitioning of the between-school variance in reading, mathematical and scientific literacy by student background characteristics, school context variables, all school variables, and each of three groups of malleable school characteristics (Source: Scheerens & Visscher, 2004).

			Malleable school characteristics			
	Student back-ground (%)	School context Variables (%)	All school variables (%)	School resources variables (%)	School climate variables (%)	School process variables (%)
Reading						
Overall	10.7	47.8	7.8	1.5	6.1	1.2
OECD	12.7	48.1	8.1	1.5	6.1	1.7
Mathematics						
Overall	23.2	30.9	7.8	1.5	6.0	1.3
OECD	26.3	29.6	8.9	1.4	6.7	1.9
Science						
Overall	25.9	29.8	7.4	0.9	6.1	1.2
OECD	28.6	29.5	8.2	1.1	6.4	1.7

In the figure, also cited from Scheerens and Visscher, 2004, these patterns are visualized for all the participating countries.

It is interesting to note that the between school variation is explained for a very large part by the student background variables and the school context variables (mainly the average socio-economic status of a school) and that only a relatively small part is explained by malleable school variables.

When we compare these results with similar analyses carried out in school effectiveness research studies, the balance between the impact of background conditions and composition on the one hand and malleable school conditions on the other is less extreme as in the case of the PISA data. In school effectiveness studies one is likely to find about 10 % of the *total between student variation* explained by

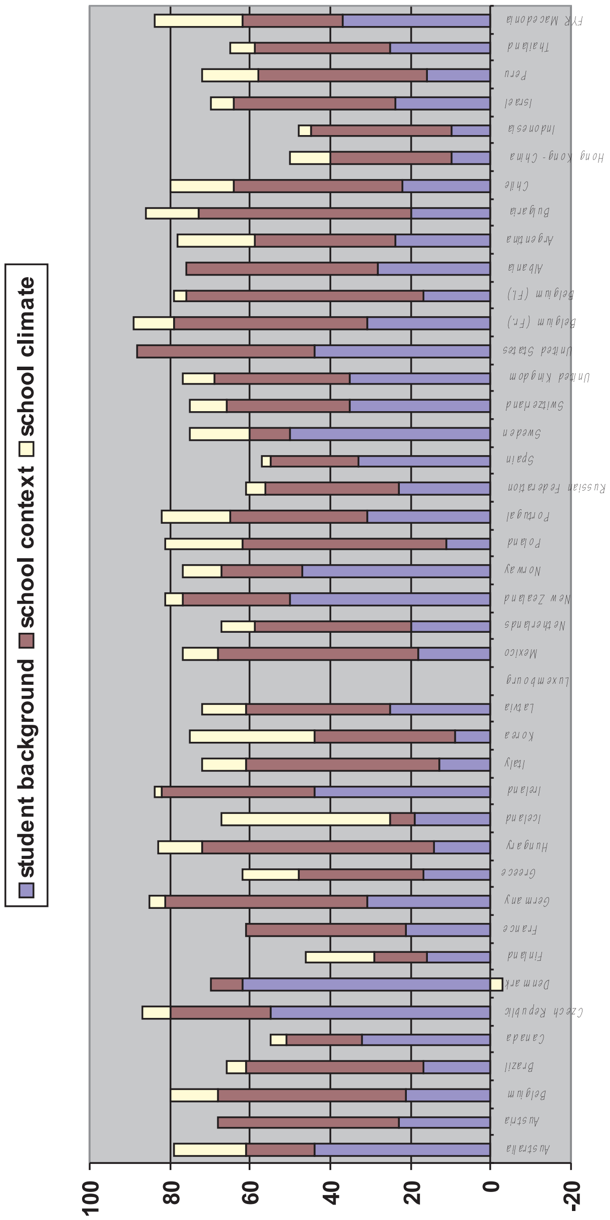


Figure 1: Percentages of between school variance in reading literacy explained by student background variables, school context variables and malleable school variables

measured school variables, which, given a total between school variance component of 30, would be equal to explaining 30 % of the between school variance.

The low estimates of the effects of malleable school variables that one usually finds in international comparative assessment studies, in comparison to school effectiveness research studies can be attributed to two causes:

- the usually rather superficial way of measuring school factors, namely by just using one or two questionnaire items to cover complex concepts;
- the lack of pre-test data in practically all international comparative assessment studies.

The impact of individual student level background conditions and their averages (compositional effects) on performance can be interpreted as indicators of (in)equity. Conditions of schooling are seen as less equitable to the extent that performance depends on the socio-economic background or the ethnicity of the students. The PISA data show that countries with a categorical system of secondary schools show larger impact of ses-related variables than do countries with a comprehensive system (Scheerens & Visscher, 2004).

A difficulty in making unequivocal comparisons between compositional effects on the one hand, and malleable school variables on the other is that these two categories of variables „overlap" in their impact on performance. Scheerens and Visscher (2004) estimated this „overlap" as a joint effect of both categories of variables. For many countries the joint effect is larger than the two unique effects. Interestingly enough countries differ considerably in the magnitude of this joint effect. Since the joint effect comes down to favorable conditions of schooling „going together" with a student population of the school that has favorable background characteristics the joint effect can be interpreted as another indicator of (in) equity. The cross-sectional nature of the data in a study like PISA 2000, precludes a sharper identification of the kind of selection processes that give rise to the joint effect of malleable an school composition variables. It might be the case that schools with better teaching conditions attract „better" students or that favorable characteristics of the students attract better teaching conditions.

The relative importance of the „joint effect" of background and malleable school variables - school climate in this case-, is illustrated in Figure 2 below (Source: Luyten, Scheerens, Visscher and others, 2005).

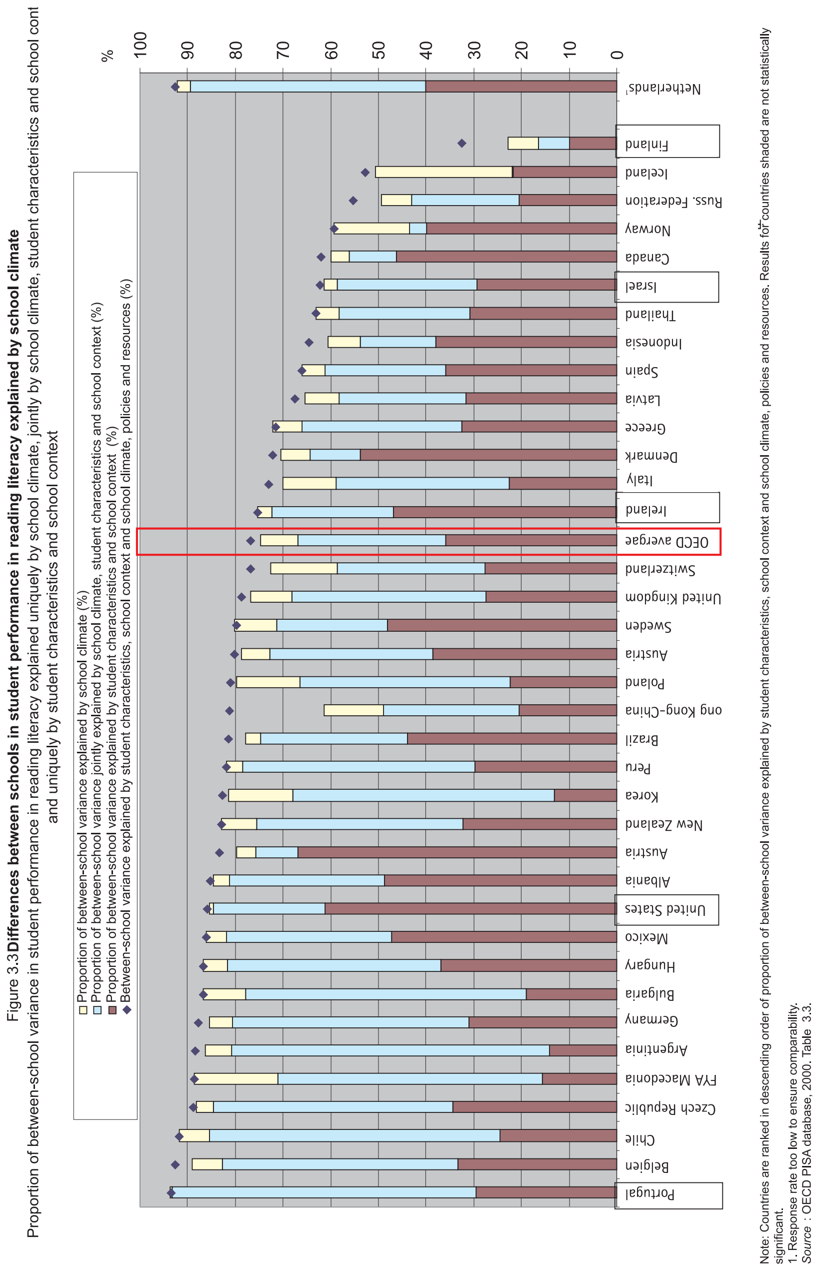


Figure 2: Differences between schools in student performance in reading literacy explained by school climate.

Assessing the effectiveness of specific school, context and classroom characteristics

The effectiveness question considers the impact of specific school, school context and classroom characteristics on performance. International comparative assessment studies can address the effectiveness question to the extent that school and classroom variables are actually measured, usually on the basis of questionnaires administered to school directors, teachers and/or students. The fact that effects of these variables are being assessed in a multitude of countries provides the interesting possibility to establish whether „what works" in one country also works in the next. Stated in less popular terms this question refers to the generalizability of effectiveness enhancing conditions across countries.

A few illustrations will be provided, based on SIMS (the Second International Mathematics Study of the IEA, the IEA Reading Literacy Study, and PISA 2000).

In Table 7 significant associations of a set of school variables and mathematics achievement from SIMS are shown for 17countries (Source: Scheerens, Vermeulen & Pelgrum, 1989).

Table 7: Predictor variables with significant positive (+) or negative (-) associations (5 % level) with mathematics achievement, when the variance component model is analyzed by means of the VARCL-Programme

Country	Belgium (Fl.)	Belgium (Fr.)	Canada (B.C.)	Canada (Ont.)	Finland	France	Hong Kong	Hungary	Israel	Japan	Luxembourg	Netherlands	New Zealand	Scotland	Sweden	Thailand	U.S.A.		Total
	15	16	22	25	39	40	43	44	50	54	59	62	63	72	76	79	81		
Predictor variable																			
Fathers'occupation *(yfocci)*	m	m		+	+	+		+	m				+	+	+	+	+		9
Fathers' education *(yfeduc)*									+										1
Level of expected further education *(umoreed)*	+	+	+	+	+	+	+	m	+	+	+	+	+	+	+	+	+		16
Homework *(yahwkt)*			-				+	+	+	+		-	+		+	+			9
Teacher experience *(texpmth)*						+	+				+	+	+						5
Time spent keeping order *(tordert)*		-			-	-							-					-	5
Time spent on teaching *(tlistt)*	+						+					+			+				4
Teacher expectations *(ttop)*			+	+	+	+	+	+	+		+	+	+		+		+		13
Use of published tests *(tpubtst)*			+		+			+											3
Use of own tests *(towntst)*							+												2
Opportunity to learn *(totl)*	+	m		+		+	m		+	+		+	+	m	+		+		9
Class size *(klgrt)*	-	+			+	+						+	+	+	+				8
Urbanization *(saera)*		+					m	+				-							3
Number of woman teachers *(ssommf)*	+										-				-				3
Number of male teachers *(sallmm)*	+							+	-										3
Number of meetings *(smeet)*																			0

Note: School and classroom predictor variables are corrected for father's occupation or father's education – when a predictor variable was not measured in a country this is indicated by the letter m.

The authors conclude that only a small number of school/classroom characteristics show a consistently positive association with mathematics achievement. These factors are: positive expectations of pupils‘ achievement (the variables *umoreed* with an average association of .19 with achievement and *ttop*, average association of .22), and opportunity to learn (average association of .15). The authors (ibid. p. 797) go on to critically analyze these few positive associations. „The educational significance of the positive results might be challenged on conceptual and statistical ground. One could argue that associations of variables such as „positive expectations“ and „opportunity to learn“ with achievement, are something like a tautology. In the worst case, opportunity to learn could reflect the purposeful training of test items. „High expectations“ might just as well be seen as the *effects* of high achievement rather than one of its *causes*.“ They also conclude that variables that have received empirical support in the international research literature on school and instructional effectiveness, like frequent evaluation of students‘ progress, teachers‘ experience and „time on task“ were found to have weak and/or inconsistent effects across countries.

Postlethwaite and Ross (1992) followed a different approach in their analysis of the data from the IEA Reading Literacy Study. In each country they identified variables that significantly discriminated between the 20 % highest and the 20 % lowest scoring schools in the country. In this way they could produce a list of those variables that discriminated high and low effective schools in a certain number of countries. The relevance of the variables could thus be judged in terms of the number of countries in which a particular variable discriminated. The results are summarized in table 8.

Table 8: Teacher and school indicators discriminating effective and ineffective schools (top 15) (Source: Postlethwaite & Ross, 1992)

Rank	Indicator	No. of countries
1	degree of parental cooperation	16
2	reading in class	17
3	no serious problems	18
4	urban-rural	14
5	school size	12
6	community resources	14
7	reading materials in schools	13
8	comprehension instruction	11
9	percent female teachers	14
10	classroom library	10
11	total teaching experience	11
12	school resources	13
13	student-teacher ratio	12
14	sponsor reading initiatives	13
15	literature emphasis	9

Scheerens and Bosker (1997) re-analyzed this data set using multi-level analyses. Their results with respect to overall effects across countries, using the total data set, were summarized as follows.

„Both context indicators *public/private* and *rural/urban* show a positive association with adjusted school effects in reading, showing advantages for private and urban schools. From the input indicators class size has a small, and meaningless, positive effect, and parental involvement has a clear positive effect (.08).

From the school process variables two achievement press variables (*focus on higher order problem solving* skills and *focus on reading*) have significant but small (.02) positive effects. The *consensus & cooperation* indicator has a significant but small (-.02) negative effect. The climate indicator shows a somewhat higher association (.04).

The other school process variables have estimated effects that are, statistically speaking, not discernable from zero. Of all teacher/classroom process variables only one has (an unexpected) negative effect: -.02 namely is the effect of time for reading“

And they conclude:
„All in all the model for the international data does poorly, with only 9 percent of unique variation between schools accounted for by the educational effectiveness variables“ (Ibid p. 260).

A final illustration is based on PISA, 2000, source Scheerens and Visscher, 2004. After student achievement in reading literacy had been adjusted for student background conditions the following school variables appeared to have a significant association with performance when the whole data-set was used:

Figure 3: School variables significantly related to reading literacy performance, after adjustment for student background characteristics (Source: Scheerens & Visscher, 2004).

SCHOOL RESOURCES VARIABLES	
*	school size
*	index of the quality of schools' educational resources
*	proportion of teachers with a third level qualification
SCHOOL CLIMATE VARIABLES	
*	index of disciplinary climate
*	index of teacher support (-)
*	index of teacher-student relations
*	index of students' sense of belonging tot the school
*	index op principals' perception of teacher-related factors affecting school climate (-)
*	index of principals' perception of student-related factors affecting school climate
SCHOOL PROCESS VARIABLES	
*	students' performance is considered for school admission*
*	transfer of low achievers to another school*

*) significant for OECD-countries only

When associations with unadjusted performance scores are considered (see the initial OECD report on PISA) considerably more school variables, such as school autonomy appear to be significantly associated with performance; effects that disappear when the proper adjustments are being made. Wößmann, (2000), incidentally reports a significant effect of school autonomy on the basis of an analysis of the TIMSS data set, using a country level model.

Willms and Somers, (2001) report findings that are more in line with the knowledge base on school effectiveness. Their analyses are based on UNESCO's *Primer Estudio Internacional Comparaivo* (PEIC) on 13 Latin American countries.

These authors conclude that the most effective schools are those with:

„1) high levels of school resources, including a low pupil-teacher ratio, more instructional materials, a large library, and well-trained teachers;

2) classrooms which are not multigrade, and where students are not grouped by ability;

3) classrooms where teachers are tested frequently"

4) classrooms and schools with a high level of parental involvement; and

5) classrooms that have a positive classroom climate, especially with respect to classroom discipline" (ibid. p. 439)

In conclusion it can be said that generally the results of associating school and classroom variables in international comparative assessment studies have been somewhat disappointing as far as the „global" studies or IEA and the OECD are concerned. Consistently smaller associations are found as in the case of national empirical school effectiveness studies. Moreover, consistency in certain variables being associated with performance across countries is also relatively disappointing.

The same methodological explanations could be given as the ones that were presented in the previous section: lack of longitudinally measured performance and relatively weak operationalizations of the process variables. At the same time part of the results might also be due to genuine differences between countries, or cultures. The PISA-re-analysis appears to point out that the school effectiveness variables that are known from the literature „work best" in traditionally English speaking countries. In these countries most of the empirical school effectiveness studies have been carried out as well. Nordic countries generally do very well in these international assessments but probably due to a somewhat different set of conditions, like the esteem for the teaching profession and the value education has in the society. Climate variables, also part of the school effectiveness heritage, work well in the Nordic countries as in countries with an Anglo-Saxon tradition.

Conclusion: making up the balance on the usefulness of international comparative assessment studies for answering questions about educational productivity and effectiveness

International comparative assessment studies are particularly useful for assessing the productivity of education systems, in terms of average achievement in a specific subject matter area or literacy domain. Countries can pick and choose the benchmarks they would like to use, to compare themselves: the international average, the score of a neighbor country or the average of the highest scoring country. As the illustrative data from PISA and TIMSS have shown large differences exist between the highest and the lowest scoring countries. For resources poor countries this might be problematic, because students might feel discouraged in not being able to do a substantial part of the items. At the same time it could be seen as important that such international comparisons can be made. A possible solution might be to expand the difficulty range in the sense of including sets of easier items for countries that are expected to score relatively low. If tests confirm to the assumptions of item response models, these easier item sets could then be vertically equated to the general international tests. International assessments for specific regions, like the PASEQ and SACMEC studies in Africa, and the *Primer Estudio Internacional Comparativo* in thirteen Latin American countries, have the advantage of being able to choose a more adapted difficulty level of the achievement tests, and include perhaps more ecologically valid items on the context of schooling in resources-poor countries.

Not only achievement *levels* such as the country averages are useful but also the patterns of *variability* that the score distributions of international comparative assessments show. As has been illustrated interesting conclusions can be drawn on the basis of the total between school variance, the proportion of the variance that is between schools, (usually indicated as the between school variance), and sometimes also the variance between parallel classes in one school. If the data can be broken down according to regions within countries, such analyses of the patterns of variability gain in relevance. Variability measures provide indications about the inequality among students in their achievement results, about the degree of segregation of the system of schools, and into practices like ability grouping and streaming within schools. In theory it would also be feasible to set benchmarks for keeping the different types of variability of and within schools systems within limits. Such benchmarks would speak to the equity interpretation of educational quality.

Comparing the impact of malleable school variables on the one hand and student background conditions and composition factors on the other, indicate the margins of control and change in education. On further reflection these different effects can be related to two different strategies to influence outcomes: productivity improvement on the one hand, and selection and admission policies on the other. The size of the composite effects, as was illustrated on the basis of the PISA data set, may give rise to pay more attention to „selection management" and establishing fixed „quota" of students with specific background characteristics. It cannot be excluded that the impact of student background and compositional factors is overrated in international comparative assessment studies, because of weaknesses in the operationalization of the school variables. Besides, as was illustrated as well, the two types of factors overlap in their impact on achievement, which further complicates interpretation. In any case do international assessments provide the occasion to globally examine the margins of „malleability" in schooling, as well as the degree of dependency of results on student background characteristics and their aggregates. The latter providing an additional interpretation relevant for the equity perspective, implying that systems in which achievement results depend to a larger degree on „given" student background conditions like their socio-economic status, are considered to be less equitable than systems for which this association is lower.

The global international assessment studies from IEA and OECD have yielded relatively disappointing results with respect to confirming the effectiveness enhancing factors that are part of the school effectiveness knowledge base. This applies both to the size of the association of these variables with performance, after controlling for student background conditions, as to the weak consistency of the significance of the effects of these variables across countries. Regional studies, like the Latin American PEIC, however, do show results that are more in line with results of school effectiveness research studies. One way of improving the relevance of international comparative assessment studies to answer questions about educational effectiveness would be to invest more in measuring school factors and processes, using more extensive scales and perhaps also direct classroom observations. Another alternative would be to consider stand-alone school surveys and classroom observation studies to yield information on effectiveness enhancing process indicators. An example is the school and teacher survey in the countries united in the World Education Indicator Project of UNESCO, OECD and the World Bank.

References

Adams, R. J., (2003) Rejoinder to „Cautions on OECD's Recent Educational Survey (PISA)" *Oxford Review of Education*, Vol 29, No. 3 2003

Bishop, J. (1997) *The effect of National Standards and Curriculum-Based Exams on Achievement*. Cornell University. Center for Advanced Human Resource Studies.

Luyten, H., Scheerens, J., Visscher, A.J., Maslowksi, R., Witziers, B., and Steen, R. (2005) *School factors Related to Quality and Equity. Results from PISA 2000*.

Prais, S.J. (2003) Cations on OECD's Recent Educational Survey (PISA). *Oxford Review of Education, Vol. 29, No. 2*

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

Scheerens, J., & Bosker, R.J. (1997). *The Foundations of Educational Effectiveness.* Oxford: Elsevier Science Ltd.

Scheerens, J., Vermeulen, C.J.A.J., & Pelgrum, W.J. (1989). Generalizability of school and instructional effectiveness indicators across nations. In: B.P.M. Creemers &

J. Scheerens (eds.), *Development in school effectiveness research. Special issue of the International Journal of Educational Research*, 13(7), 789-800. Oxford: Pergamon Press.

Scheerens, J., & Visscher, A.J. (2004). *School factors related to quality and equity.* Draft version of PISA thematic report; available at the University of Twente.

Willms, J.D., & Somers, M.-A. (2000). *Schooling outcomes in Latin America*. Report prepared for UNESCO-OREALC and the Laboratorio Latinoamericano de la Calidad de la Educación [The Latin American Laboratory for the Quality of Education].

Wößmann, L. (2000). *Schooling resources, educational institutions, and student performance: The international evidence*. (Kiel Working paper No. 983). Kiel, Germany: Kiel Institute of World Economics.

Jaap Scheerens is a Professor of Educational Organization and Management at the University of Twente in the Netherlands. His main publications are in the domain of educational effectiveness, international comparative education indicators and school self-evaluation. His latest publication is „Educational Evaluation, Assessment, and Monitoring - a systemic approach" together with C. Glas and S.M. Thomas, Swets & Zeitlinger Publishers, Lisse, The Netherlands, 2003.