CHAPTER 8

GENERALIZIBILITY OF INSTRUCTIONAL AND SCHOOL EFFECTIVENESS INDICATORS ACROSS NATIONS

JAAP SCHEERENS, C. J. A. J. VERMEULEN and W. J. PELGRUM

Department of Education, University of Twente, Enschede, The Netherlands

Abstract

In a secondary analysis of the data base of the IEA* second mathematics study, the association of school and instructional characteristics with mathematics achievement within and across 17 countries at the school and student level was investigated. The results indicate consistent positive associations of the factors 'high expectations of pupils' progress' and 'opportunity to learn' with achievement. Other factors, known from the literature on school and instructional effectiveness, showed more moderate and less consistent effects. The results of this study further indicate that, in each country, considerably more variance in pupils' mathematics achievement lies between classes than between schools when school and class are treated as independent variance components. It was established that specific variables defined at the class level explain more variance than specific variables defined at the school level.

Introduction

Despite research in various disciplines (psychology, sociology, economics) there is as yet no established causal model of school effectiveness (see e.g., Hanushek, 1986). Although there is a growing consensus regarding a set of potential predictors of school outcomes, critics still have every reason to question the consistency, stability and applicability of this set of indicators (cf. Ralph & Fennessey, 1983; Kyle, 1985).

The best-known formulation of effectiveness predictors is the so-called 5-factor model of school effectiveness, first formulated by Edmonds (1979):
— strong educational leadership;
— emphasis on basic skills achievement;
— safe and orderly climate;
— high expectations of pupils' achievement;
— frequent evaluation of pupils' progress.

Other major school effectiveness studies (Rutter *et al.*, 1979; Brookover *et al.*, 1979; Madaus, Kellaghan, Rakow, & King, 1979; Mortimore, Sammons, Stoll, Lewis, & Ecob,

---

*IEA = International Association for the Evaluation of Educational Achievement.

1988) have yielded more elaborate sets of predictors, though generally in line with the '5-factor model'. Scheerens and Stoel (1988) propose a multilevel conceptual model of school effectiveness, in which school level variables are seen as supportive conditions for instructional measures at the class level that have been found to be associated with high achievement. An important, recently started, line of research on school effectiveness, looks at the generalizibility of school effectiveness predictors across contexts. 'Context' can be defined either as type of school (Firestone & Herriott, 1982; Stoel & Scheerens, 1988) or as student-body composition (Wimpelberg, Teddlie, & Springfield, 1987). These studies show there is reason to believe that the set of significant effectiveness predictors will vary across contexts. However, there is also evidence that certain class and school factors are valuable effectiveness predictors in several contexts, for example structured teaching and time on task (cf. Kyle, 1985).

This chapter discusses the results of a study aimed at exploring the generalizibility vs. context-specifity of school and instructional effectiveness predictors, across nations. In a secondary analysis, data from the IEA second mathematics study* were used to compare degrees to which schools 'made a difference' in different countries, and the consistency in school level and class level variables that could explain these differences between schools. The fact that a secondary analysis was carried out implies that certain limitations of the available data with respect of the research aim had to be taken for granted. These were: predictor variables that from the literature appear important could only be partially covered, and the fact that in about half of the countries only one class per school was investigated, so that school effects could not be distinguished from class effects.

General Description of the Original Study and its Resulting Data Base

The Second International Mathematics Study (hereafter referred to as SIMS) was conducted in the following countries: Belgium (Flemish and French), Canada (Ontario and British Columbia), England, Finland, France, Hong Kong, Hungary, Israel, Japan, Luxembourg, The Netherlands, New Zealand, Nigeria, Scotland, Swaziland, Thailand, Sweden and the U.S.A. In this study a multilevel approach was used to investigate students' achievement in mathematics, and its determinants, at the level of the second grade of secondary education.

The study contained a cross sectional and longitudinal part. The cross sectional part involved the collection of data at one point in time (near the end of the second year), whereas the longitudinal part consisted of collecting pre and post test data at the beginning and end of the school year. Data were collected about student achievement in mathematics, with a multiple choice test of 154 items, common to the tests used in the cross sectional and longitudinal part of the study, distributed across subtests called Arithmetic, Algebra, Geometry, Statistics and Measurement. A multiple matrix sampling design was used to collect the test data of students. This consisted in giving each student a core booklet of 40 items and one out of four rotated forms (containing 34 or 35 items for the cross-sectional and longitudinal part of the study respectively). Furthermore, additional data were collected at the level of students, teachers and schools. All data were collected by means of questionnaires and/or paper and pencil tests. The documentation accompanying

---

*The data were made available on tape by the International Coordinating Centre for that study in Wellington, New Zealand.

the tapes contains a description of the variables for which data were collected at each level.

Table 8.1 contains an overview of numbers of the various units (students, teachers, schools and classes per school) for those countries included in the analyses presented in this chapter, as well as some descriptive statistics on the outcome variable.

We shall now describe which variables were selected from the data bases for investigating the research problem.

Variables

The dependent variables of the IEA second mathematics study were multiple choice tests in arithmetic, algebra, geometry, statistics and measurement. In all secondary analyses, the total score across these sub-tests was used. Total scores for each student were determined by calculating the percentage of items administered to a particular student, that was correct. Although this might introduce some error variance due to the difference in difficulty between rotated forms, for the purpose of our analyses this effect is assumed not to be important. First of all a sub-set of available predictor variables was selected, categorized as school variables, class variables, teacher variables and pupil variables, that had some correspondence to relevant school effectiveness predictors, as appears in the literature. Since this sub-set was still as many as 73 variables, a further reduction procedure was used. Working from the correlation matrix of all independent variables and the dependent variable, where student and class scores were aggregated up to the school level, the following selection criteria were used:

— the independent variable had to correlate .10 or more with the dependent variable in at least 5 countries;

Table 8.1
Number of Students, Teachers, Schools and Classes Per School for Selected Countries, Means* and Standard Deviations on Mathematics Achievement

Country	Students	Teachers	Schools	Classes per school	Mean	Standard deviation
15 Belgium (Fl.)	3282	150	150	1	53.36	26.50
16 Belgium (Fr.)	1558	83	82	1	53.23	19.88
22 Canada (Br. Col.)	2228	87	87	1	52.07	20.66
25 Canada (Ontario)	4597	170	106	>1	47.26	20.62
39 Finland	4484	206	98	>1	43.60	19.52
40 France	8230	338	174	>1	48.83	18.98
43 Hong Kong	5548	125	125	1	48.43	19.69
44 Hungary	1754	70	70	1	58.65	20.67
50 Israel	2540	93	68	>1	44.10	22.56
54 Japan	8091	212	212	1	61.58	18.56
59 Luxembourg	2106	107	42	>1	36.63	15.61
62 Netherlands	5500	236	236	1	54.96	21.66
63 New Zealand	5252	193	100	>1	43.77	21.16
72 Scotland	853	274	58	>1	50.90	22.67
76 Sweden	3571	186	97	>1	36.53	16.48
79 Thailand	3806	98	98	1	42.56	17.77
81 U.S.A.	6792	277	157	>1	44.71	21.63
Total	70192	2910	1960			

*Means were computed without applying sampling weights.

— any particular independent variable should not correlate over .20 with any other independent variable;

— the total number of missing cases should not be higher than 30 per cent in more than three countries;

Where predictor variables failed on any one of these criteria, but were seen, considering the literature, as particularly relevant, these variables were nevertheless included in the analysis. These variables were: number of meetings of mathematics teachers(*smeet*), estimate by teachers of number of pupils belonging to the top in mathematics achievement (*ttop*), degree of urbanization of the school area (*sarea*) and the use of teacher-made tests (*townst*).

Using these selections criteria the following list of 15 predictor variables was arrived at:
— Pupil characteristics
    father's occupation (*yffocci*),
    father's education (*yfeduc*).
— Teacher characteristics
    experience as a mathematics teacher (in years — *texpmth*),
    time spent on keeping order (in min per week — *tordert*),
    time spent on teaching (in min per week — *tlisst*).
— Opportunity to learn
    items to test covered in tuition (*totl*).
— Expectations
    pupil's expectation of the number of years he/she will follow formal education (*ymoreed*),
    estimate by teacher of the number of pupils who belong to the top in mathematics (*ttop*).
— Instructional characteristics
    total time (hours) spent on homework (*yahwkt*),
    the use of published tests (*tpubst*),
    the use of teacher made tests (*towntst*).
— School characteristics
    the number of women teachers in mathematics (*ssommf*),
    the number of men teachers that teach only mathematics (*sallmm*),
    the number of meetings of mathematics teachers (*smeet*).
— Contextual characteristics
    degree of urbanization of the school area (*saera*),
    class size (*klgrt*).

Analyses

Apart from providing some basic descriptive statistics on the data set, our analyses aimed to answer two questions:
    (1) do the variances between schools differ between countries?
    (2) which school and instructional characteristics are stable predictors of achievement across countries?

Variance component analysis, using the VARCL-computing programme (cf. Aitkin & Longford, 1986) — was carried out to estimate variances between school per country and

to disentangle school and classroom effects, for those countries where data on more than one class per school were available. This type of analysis was also used to obtain estimates of the effect of specific school and class variables in terms of proportions of explained variance (in answer to question (2) stated above).

Variance component analysis was chosen because it belongs to a class of techniques (along with Raudenbush & Bryk's HLM procedure, Raudenbush & Bryk, 1986) that is most efficient for analyzing multilevel data (cf. Aitkin & Longford, 1986). By allowing for the coefficients for slopes and intercepts to be considered as random effects, this technique also offers possibilities for testing interesting substantive hypotheses about the equality of education in a particular school (cf. Brandsma & Knuver, 1989) — when slopes are treated as outcomes.

The formulation of the model with random intercepts and random slopes is as follows:

$$y_{ij} + (\alpha + \alpha_j) + (\beta + \beta_j)x_{ij} + \delta z_j + e_{ij}$$

where:

y_{ij} = dependent variable (score of pupil $_i$ in school $_j$),
α = intercept,
α_j = random part of intercept,
β, δ = regression coefficients,
β_j = random part of the regression coefficient,
x_{ij} = pupil score on co-variate,
z_j = score on school or classroom level variable,
e_{ij} = individual error term.

Analysis of more or less restricted versions of this model makes it possible to partition the total variance into school (and/or classroom) and pupil components; to examine the 'fit' of the model when specific sets of school (or class) level variables are added; to investigate (under the assumption of random effects) the significance of variances in slopes and intercepts among units, and, finally, to determine the effect-size of sets of predictor variables defined at the school and/or classroom level and individual predictor variables in terms of percentage of variance explained. Further details on the use of this particular technique for multilevel analysis are provided in the references stated above.

Results

School and Class Effects

One problem in using the SIMS data set for secondary analyses with respect to school effectiveness is the confounding of school and class effects for those countries where only one class per school was sampled. This is the case for 8 of the 17 countries (see Table 8.2). By investigating the so-called 'null-model' (excluding all predictor variables), variance component analysis by means of the VARCL programme yields intra-class correlations as

a measure of the proportion of total variance explained by schools and classrooms respectively.

The results summarized in Table 8.2 indicate: (a) the degree to which classrooms differ within schools; (b) the degree to which schools still differ in average achievement when the between-classroom effect has been accounted for — at least for those countries where classroom and school effects could be separated; and (c) differences between countries in the pattern of between-class and between-school variances.

It is evident from Table 8.2 that important differences exist between countries in the degree to which school and class effects account for total variance. The general trend is that greater variance lies between classes than between schools. In some countries (notably Sweden, Finland and New Zealand) the simultaneous estimation of school and class effects shows that school effects are virtually non-existent when the classroom effect has been accounted for.

Another pattern, found in the data on Ontario and France is that both school and classroom effects are relatively small, indicating an educational system that shows considerable uniformity between and within schools. A further combination of class and school effects found in countries such as the U.S.A., Luxembourg and Scotland is a relatively large class effect together with a moderate school effect. In these countries there is both heterogeneity at the classroom and the school level. The very large estimates of the school variance components for Belgium (both Flemish and French) and The Netherlands are indicative of a strongly differentiated (as opposed to an integrated) organization of the national school system. This can be demonstrated by calculating the rho's within school types. For the four school types in The Netherlands, senior secondary, junior secondary, vocational technical and vocational domestic science the rho's are .27, .26, .42 and .30

Table 8.2
Estimates of the Variance Explained by Schools and Classes

Country	Classroom variance component	School variance component
15 Belgium (Flemish)		.50 (.48)
16 Belgium (French)		.64 (.62)
22 Canada (British Columbia)		.27 (.27)
25 Canada (Ontario)	.18 (.17)	.09 (.09)
39 Finland	.45 (.41)	.002
40 France	.17 (.16)	.06 (.05)
43 Hong Kong		.51 (.50)
44 Hungary		.30 (.27)
50 Israel	.22 (.21)	.10 (.08)
54 Japan		.08 (.07)
59 Luxembourg	.29 (.29)	.15 (.15)
62 Netherlands		.67 (.66)
63 New Zealand	.45 (.42)	.01 (.004)
72 Scotland	.34 (.31)	.12 (.05)
76 Sweden	.45 (.45)	.00
79 Thailand		.39 (.38)
81 U.S.A.	.46 (.45)	.10 (.09)

Note: Estimates 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; the coefficients shown between brackets are the intra-class correlation coefficients are controlling for fathers' occupation (*yfocci*).

respectively. The estimates of classroom and school effects were only slightly smaller when father's occupation, as a student background variable, was controlled for (figures shown between brackets in Table 8.2). The overall conclusion from the results presented in Table 8.2, with respect to the phenomenon of school effectiveness, is that the expression 'schools can make a difference' (Brookover *et al.*, 1979) means something quite different in different countries.

Effects of Specific School and Class Variables

To examine the effects of specific school and classroom variables we used the VARCL programme to estimate the association (in terms of a standardized regression coefficient) of each predictor variable with mathematics achievement, after controlling for father's occupation (when this variable was not measured in a particular country we used father's education as the concomitant variable). The results are summarized in Table 8.3, where positive associations, significant at the 5 per cent are indicated by a plus, and negative significant associations by a minus sign.

Predictor variables that show the strongest and most consistent (i.e., consistent across countries) positive association with mathematics achievement are educational expectations as expressed by pupils (*ymoreed*) and by teachers (*ttop*).

Other predictor variables that show consistent positive associations with achievement are total time spent on homework (*yahwkt*) and opportunity to learn (*totl*). As was expected, the variable time spent on keeping order (*tordert*), has a consistent negative association with achievement, though the coefficient reaches significance in only 5 countries. Class size (*klgrt*) is mostly positively associated with achievement, indicating that larger classroom generally do better; the coefficient is significant in eight cases (7 positive and 1 negative). Other variables that are well known from the literature on school effectiveness, such as teacher experience (*texpmth*) and use of tests (*tpubst* and *townst*) generally show small and/or inconsistent associations with mathematics achievement. Classroom variables were generally more strongly associated with mathematics achievement than were school variables.

Finally, it was examined how well selected sets of predictor variables explained between-classroom and between-school variances respectively, after controlling for father's occupation. For each country those predictor variables were selected that were significantly associated with achievement (see the results in Table 8.3). The results are summarized in Table 8.4.

The results in Table 8.4 indicate that in most cases (7 countries) the selected sets of predictor variables operated at the classroom level to a larger degree than at the school level, although in two countries (Israel and Luxembourg) the selected sets of variables did better in explaining between-school variance than in explaining between-class variance. Again the overall impression is that there are large differences between countries in the way correlates of school effectiveness work. It should be noted, however, that the figures shown in Table 8.4 are the result of two kinds of sources of differences between countries: (a) the amount of variance that lies between classrooms vs. schools (Table 8.2); and (b) the differences among the sets of predictors that are significantly associated with mathematics achievement (Table 8.3).

Table 8.3

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

Predictor variable	Belgium (Fl.) 15	Belgium (Fr.) 16	Canada (B.C.) 22	Canada (Ont.) 25	Finland 39	France 40	Hong Kong 43	Hungary 44	Israel 50	Japan 54	Luxembourg 59	Netherlands 62	New Zealand 63	Scotland 72	Sweden 76	Thailand 79	U.S.A. 81	Total
Fathers' occupation (yfocci)	m	m		+	+	+		+	m				+	+	+	+	+	9
Fathers' education (yfeduc)									+									1
Level of expected further education (ymoreed)	+	+	−	+	+	+	+	m	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)	−	+				+	m					+	+	+	+			8
Urbanization (saera)	+	+						+				−				+		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.

Table 8.4
Effects of Selected Sub-Sets of Predictor Variables in Terms of Percentage of Explained Between-Classroom and
Explained Between-School Variance

Country	% of between classroom variance explained by selected variables	% of between school variance explained by selected variables
15 Belgium (Flemish)		8
16 Belgium (French)		11
22 Canada (British Columbia)		9
25 Canada (Ontario)	3	1
39 Finland	41	4
40 France	8	3
43 Hong Kong		13
44 Hungary		8
50 Israel	1	4
54 Japan		2
59 Luxembourg	3	7
62 Netherlands		22
63 New Zealand	25	0
72 Scotland	17	6
76 Sweden	20	0
79 Thailand		3
81 U.S.A.	20	2

Note: For those countries where data on only 1 classroom per school were collected, there was no way to separate class components and school variance components.

Discussion

The results of this secondary analysis suggest that a small number of school/class characteristics show a consistently positive association with mathematics achievement. These factors are: positive expectations of pupils' achievement (the variables *ymoreed* with an average association of .19 with achievement and *ttop*, average association of .22), and opportunity to learn (average association with achievement of .15).

Other variables that, in the literature on school and instructional effectiveness, have repeatedly shown positive associations with achievement such as frequent evaluation (here represented in the variables *tpubst* and *towntst*), teachers' experience (*texpmth*), and indicators of 'time on task' (*texplm*) were found to have weak and/or inconsistent effects.

The educational significance of the positive results might be challenged on conceptual and statistical grounds. One could argue that associations of variables such as 'positive expectations' and 'opportunity to learn' with achievement, are something of a tautology. In the worst case, opportunity to learn could reflect the purposeful training of test items. 'High expectations' — it has been said before (Ralph & Fennessey, 1983) — might just as well be seen as the *effects* of high achievement rather than one of its *causes*. Statistical objections to the effects that were found in this study are that the effects are small and that background factors of pupils such as social economical status and intelligence were not sufficiently accounted for. As to the first type of criticism, it should be pointed out that correlations of about .10, however low in an absolute sense, may still represent educationally meaningful effects (cf. Bosker & Scheerens, 1988). The objection of insufficient control of student background variables cannot be countered: this is one of the limits imposed by the fact that an existing data base was used. The one student background

variable that was used, 'father's occupation', is a proxy-variable for social economic status; as appeared from the analysis, it did not explain much variance in achievement.

The results further show many differences between countries as far as average achievement and between-school variance is concerned (Tables 8.1 and 8.2). Differences in the between-school variances can be interpreted as consequences of more integrated and centralized, as compared with more differentiated and decentralized, national educational systems. In those cases where the available data allowed us to disentangle school and class variance, it was found that generally more variance is explained by classrooms than by schools. Moreover, specific patterns of school and class variances were discerned, where some countries had both a sizable between school variance and a large between-classroom variance (U.S.A., Luxembourg and Scotland), other countries showed almost no between-school variance but large between-class variances (Sweden, Finland, New Zealand) and finally some countries (France, Canada Ontario) had both a relatively low between-class variance and a relatively low between-school variance (see Table 8.2).

A final conclusion concerns the relative importance of specific predictor variables, defined either at the school or classroom level. The question whether indicators (i.e., predictors) of school effectiveness are genuine school characteristics, or aggregated class characteristics, has repeatedly been posed in the literature on school and class effectiveness. Our data support the position that the case for 'effective classrooms' is somewhat strong than that for 'effective schools'. First, because it was found that in most countries where this could be analyzed, the between class variance was larger than the between school variance and secondly, because the predictor variables defined at the school level generally explained less variance than the predictor variables defined at the class level. Yet one cannot be definite about this last conclusion since important school level variables such as 'educational leadership' and 'achievement oriented school policy' were not included in the data set and classroom effects may well be inflated due to streaming.

References

Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *The Journal of the Royal Statistical Society*, Series A (General), **149** (1), 1–43.

Bosker, R. J., & Scheerens, J. (1989). Criterion definition, effect size and stability, three fundamental questions in school effectiveness research. In B. P. M. Creemers & B. Reynolds (Eds.), *School effectiveness and school improvement*. Lisse: Swets & Zeitlinger.

Brandsma, H. P., & Knuver, A. W. M. (1989). De invloed van school- en klaskenmerken op rekenprestaties in het onderwijs. In J. Scheerens en J. Verhoeven, (Eds.), *Schoolorganisatie, beleid en onderwijskwaliteit*. Lisse: Swets & Zeitlinger.

Brookover et al. (1979). *School social systems and student achievement: schools can make a difference*. New York: Praeger.

Edmonds, R. (1979). Some schools work and more and more can. *Social Policy*, **9**, 28–32.

Firestone, W. A., & Herriott, R. E. (1982). Prescriptions for effective elementary schools don't fit secondary schools. *Educational Administration Quarterly*, **18** (2), 39–59.

Hanushek, E. A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, **24**, 1141–1177.

Kyle, R. M. J. (1985). *Reaching for excellence. An effective schools sourcebook*. Washington: US Government Printing Office.

Madaus, G. F., Kellaghan, T., Rakow, E. A., & King, D. (1979). The sensitivity of measures of school effectiveness. *Harvard Educational Review*, **49**, 207–230.

Mortimore, P., Sammons, P., Stoll, L., Lewis, D., & Ecob, R. (1988). *The junior school project; technical appendices*. London: ILEA, Research and Statistics Branch.

Ralph, J. H., & Fennessey, J. (1983). Science or reform: some questions about the effective schools models. *Phi Delta Kappan,* **64** (10), 689–695.

Raudenbush, S. W., & Bryk, A. S. (1986). A hierarchical model for studying school effects. *Sociology of Education,* **59**, 1–17.

Rutter, M. *et al.* (1979). *Fifteen thousand hours: secondary schools and their effects on children.* Cambridge MA: Harvard University Press.

Scheerens, J., & Stoel, W. G. R. (1988). Development of theories on school-effectiveness. Paper presented at the AERA convention. New Orleans.

Stoel, W. G. R., & Scheerens, J. (1988). The stability of school effects across contexts and nations. Paper presented at the AERA convention. New Orleans.

Wimpelberg, R. V., Teddlie, C., & Springfield, S. (1987). Sensitivity to context: the past and future of effective schools research. Washington: AERA paper.