DOCUMENT RESUME

ED 376 559 EA 026 247

AUTHOR Luyten, Hans

TITLE School Size Effects on Achievement in Secondary Education: Evidence from the Netherlands, Sweden and the USA.

PUB DATE Apr 94

NOTE 35p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 4-8, 1994).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Academic Achievement; Effective Schools Research; Foreign Countries; Institutional Characteristics; *Mathematics Achievement; School Demography; *School Effectiveness; *School Size; Science Tests; Secondary Education

IDENTIFIERS *Netherlands; Science Achievement; *Sweden; *United States

ABSTRACT

This paper reports the results of an investigation into the relationship between school size and achievement. The study examined the impact of school size on mathematics achievement in Dutch, Swedish, and American secondary education and on science achievement in the Netherlands. The following research questions were explored: (1) Is school size related to achievement independently of student background characteristics, such as sex, achievement motivation, socioeconomic status, and cognitive aptitude? (2) Is the effect of school size related to any of the aforementioned background characteristics? (3) Does the effect of school size on achievement differ among the educational systems of the Netherlands, Sweden, and the United States? and (4) Is the effect of school size the same for different measures of student achievement (mathematics versus science)? Datasets from two international studies sponsored by the International Association for the Evaluation of Educational Achievement were analyzed--the Second International Mathematics Study (SIMS) and the Second International Science Study (SISS). The findings found little empirical evidence for the existence of school-size effects on achievement in any of the three countries, possibly because school size and curriculum comprehensiveness are not strongly related in these countries. Some useful additional information regarding the robustness of the detected relationships between the five covariates and student achievement is presented. Five tables are included. Contains 39 references. (LMI)

SCHOOL SIZE EFFECTS ON ACHIEVEMENT IN SECONDARY EDUCATION

Evidence from the Netherlands, Sweden and the USA

Hans Luyten
University of Twente, Department of Education

A paper presented to the American Educational Research Association
Annual Meeting in New Orleans, Louisiana, April 4-8, 1994

SCHOOL SIZE EFFECTS ON ACHIEVEMENT IN SECONDARY EDUCATION
Evidence from the Netherlands, Sweden and the USA

Hans Luyten, *University of Twente, Department of Education*

In this paper the results of an investigation into the relationship between school size and achievement are reported. The findings relate to mathematics achievement in Dutch, Swedish and American secondary education and to science achievement in the Netherlands. The analyses sought to provide an answer to the following questions: (1) Is school size related to achievement independently of student background characteristics such as sex, achievement motivation, socio-economic status and cognitive aptitude? (2) Is the effect of school size related to any of the aforementioned background characteristics? (3) Does the effect of school size on achievement differ between the educational systems of the Netherlands, Sweden and the USA? (4) Is the effect of school size the same for different measures of student achievement (mathematics versus science)? It was hypothesized that school size would be most strongly related to achievement in the USA. The analyses, however, revealed little empirical evidence for the existence of school size effects on achievement in any of the three countries, possibly because school size and curriculum comprehensiveness are not strongly related in these countries.
Because the investigations involved the analysis of five separate datasets, the research outcomes revealed some useful additional information with respect to the robustness of the detected relationships between the five covariates and student achievement.

1. INTRODUCTION

The notion of economies of scale indicates, a priori, that large schools are preferable to smaller ones in at least two respects. First of all, one would expect the per student expenditures of larger schools to be lower than those of smaller schools. Secondly, school size would be expected to reveal a positive relation with achievement, because larger schools can offer their students broader curricula and better support to their teachers (Conant, 1959; 1967). They can also invest more easily in expensive facilities, such as libraries, computers and science equipment. The relations between school size and curriculum comprehensiveness, between school size and expenditures and between school size and student achievement as established in empirical research, however, are certainly not as straightforward as might be inferred from economic theory at first sight. Available evidence suggests that small schools are still able to offer a solid basic curriculum and that only a restricted number of students profits from the more extensive course offerings in the larger American High Schools (Monk, 1987; Haller et al., 1990; Fowler, 1992). Several restrictions to the relation between school size and efficiency have emerged from educational research as well (Guthrie, 1979; Fox, 1981; Bell & Sigsworth, 1987; Bray, 1988). Since the costs of schooling are largely consumed by teachers' salaries, factors like the student-teacher ratio and the height of the salaries determine the per student expenditures to a considerable extent. In most educational systems the number of teachers and the number of students at a school are not linearly related. A school is usually entitled to employ an extra teacher if its number of students exceeds a certain cut-off point. It may be possible, e.g., that a school with 50 pupils is entitled to employ only two teachers, whereas a school with 51 pupils is allowed to engage an extra teacher. Moreover, the costs per student in a school are also determined by certain teacher characteristics, as their salaries tend to vary considerably according to their age, experience and qualifications. The nature of the school buildings will also have an impact on the costs. Another important factor with respect to economies of scale is presented by the costs of transportation. Large schools in sparsely populated areas may give rise to high transportation costs. It might even become necessary to supply boarding facilities, which would have not only financial but also serious social implications.

Large school size is believed to produce a number of undesirable side-effects that are difficult to express in monetary terms. Schools are generally considered to be important centres for social development, especially in rural regions where alternative centres are largely absent. This notion has had a significant influence on educational policy in such countries as Australia, Finland, Norway and the United Kingdom (Husen & Postlethwaite, 1990, p. 542). Large school size might entail some undesirable consequences in more densely populated areas as well, as it might impede competition between schools. Small school size should be expected to provide a better opportunity for competition among schools, because there will be more alternatives to choose if there are many (small) schools. Increasing competition between schools, however, might also involve certain negative effects, e.g. opportunistic behaviour on the part of schools, such as rejecting low ability students or lowering examination standards (Brown, 1992; Ball, 1993).

The internal environment of large schools is often thought to be rather impersonal and relatively frequently suffering from discipline problems, whereas small schools are believed to offer a more cooperative climate stimulating both teacher commitment and student achievement. This view has been corroborated in a number of American studies reporting beneficial effects of small school size on student participation, satisfaction and dropout rates (Barker & Gump, 1964; Lindsay, 1982; 1984; Pittman & Haughwout, 1987; Schoggen & Schoggen, 1988). In this way small schools might very well be able to

compensate for any disadvantages of scale. Dutch research, on the other hand, has failed to reveal a clear (linear) association between secondary school size and student satisfaction (Stoel, 1982). The research literature does provide some empirical support for the hypothesis that the climate in small schools compensates for scale disadvantages, as several researchers claim to have found a negative relation between school size and achievement. If this relation is indeed negative, policy makers will be forced to weigh the (potential) financial advantages of increasing school size against the disadvantages of lower achievement. If the relationship is not negative, policy makers would have one dilemma less to solve.
In this paper only the relationship between size and achievement in secondary education will be dealt with. The research to be reported was not supposed to establish the relationships between size and efficiency, size and school climate, size and curriculum comprehensiveness or the social implications of school size. The analyses aimed to answer the following questions:

1/ Is school size related to achievement independently of student background characteristics such as sex, achievement motivation, socio-economic status and cognitive aptitude?
2/ Is the effect of school size related to any of the aforementioned background characteristics? In other words: does the effect of school size interact with any of these background characteristics?
3/ Does the effect of school size on achievement differ between the educational systems of the Netherlands, Sweden and the USA?
4/ Does the effect of school size vary to any extent between two different achievement measures (mathematics and science)?

In section 1.1 the results of previous research with respect to school size and achievement will be discussed. Next the results of an original study regarding the effects of school size on achievement in the Netherlands, Sweden and the USA will be reported.

Presently school size is a topic of administrative concern in the Netherlands, especially with respect to secondary education. In 1992 a bill was passed which forces secondary schools with less than 240 students to close down. One of the motives underlying this policy is the cost reduction that is believed to result from increasing the size of schools. Another important reason is the government's intention to create schools that are able to offer a comprehensive curriculum to the students. The Dutch system of secondary education is divided into several curriculum tracks, between which there is little mobility and for which students are selected at the age of twelve on the basis of their (presumed) scholastic aptitude. Most secondary schools in the Netherlands cover only a limited number of these tracks, usually no more than one. Schools covering the whole range of curriculum tracks are still very rare. The present government policy, however, is to stimulate the creation of broad multi-track schools, so that children who differ with respect to their cognitive aptitude will still attend the same school, which is already the case in the more integrated educational systems of Sweden and the USA. The research outcomes reported in section 3 allow for a comparison of the effects of school size on achievement in the less integrated educational system of the Netherlands to the effects in the more integrated systems of Sweden and America. The effects of school size on achievement were expected to be stronger in the USA than in the Netherlands or Sweden. American High Schools have to deal with students who differ considerably in cognitive aptitude. In many schools, however, students of similar ability are grouped together into homogeneous classes. Teachers seem to prefer this practice because they find homogeneous classes easier to teach (Kulik & Kulik, 1982, p. 416). Homogeneous grouping can be more easily applied in larger schools, because in those schools there will be more classes per grade. In small schools the classes will be more heterogeneous, which presents the teachers with a more difficult task. This may result in slightly lower achievements of the students (Kulik & Kulik, 1982). The small schools in the Netherlands, on the other hand, are hardly ever faced with such problems as they generally cover only one or two curriculum tracks. In the Netherlands at least four different curriculum tracks can be distinguished, into which students are grouped on the basis of their cognitive ability. As a result Dutch teachers generally work with quite homogeneous classes. Classroom heterogeneity and school size can therefore be assumed to be unrelated in Dutch secondary education. The same is true in the case of Sweden. The Swedish system presents two curriculum tracks, but students are not selected into separate schools, which often occurs in the Netherlands. Secondary schools in Sweden are nearly always large enough to allow for the grouping of students into homogeneous classrooms. Very small schools are extremely rare in Sweden (see section 3.2).

In the USA the size of secondary schools has increased steadily for the past few decades, but the controversy with respect to the relationship between school size and student achievement seems to be limited to academic circles in America. When a community tries to prevent the closing down of a small school, the importance a school presents for a community is usually emphasized. Both policy makers and the general public seem to treat economies of scale as established facts (Haller et al., 1990, p. 110), whereas the evidence that has resulted from empirical research is far from conclusive in this respect.

The costs of education in Sweden have faced this country's government with a major financial problem. The Swedish expenditures per student in primary and secondary education are among the highest in the world, which is at least partly due to the low student-teacher ratio (Husen & Postlethwaite, 1988, pp. 4958-4966; OECD/OCDE, 1992, pp. 56-59). The Swedish government, however, would prefer to keep the student-teacher ratio at a low level.

1.1. Research literature on the relation between school size and achievement

Although in the past few decades (from the late fifties until now) numerous studies on the relationship between school size and student achievement have been conducted, much uncertainty about the effects of school size in secondary education still remains. The research on school size effects has predominantly dealt with elementary education. Moreover, the vast majority of this research relates to the educational system in the USA. In recent reviews dealing with the effects of school size on achievement little support can be found for the hypothesis that large schools exert a positive influence on student achievement, but the opposite view that school effectiveness is enhanced by small school size doesn't receive unqualified support either.

The conclusion with which Fowler and Walberg (1991) summarise their review sounds quite firm:

> "A number of studies conducted during the past 20 years, particularly at the elementary-school level, have found small school size to have an independent, positive effect upon student achievement." (p. 191).

However, this statement is mainly based on research findings pertaining to elementary education. Of the ten studies reviewed only two relate to the relationship between school size and achievement in secondary education. The confounding influence of socio-economic background is reported to be taken into account in only one of both studies. Fowler and Walberg contend that their own research findings corroborate the assertion that small school size affects student achievement in a positive way (Fowler & Walberg, 1991), but it should be noted that statistically significant effects of school size were only found for six out of fifteen achievement measures. These effects are not very strong either. The practical significance of school size effects on achievement seems rather limited[1].

Fowler (1992) offers a review of empirical studies with respect to school size effects on students' attitudes, achievements and voluntary participation which is explicitly focused on American High schools. Four studies dealing with the relationship between size and achievement are discussed, including the one by Fowler and Walberg. In each study the influence of students' background characteristics was taken into account. In three studies a positive effect of small school size was reported on achievement, but in one study higher achievement was found in the larger schools. Fowler's conclusion with respect to the relationship between school size and achievement is therefore more cautiously formulated than the one reached by Fowler and Walberg:

> "The finding that student achievement is enhanced by small high school size was supported by the fewest studies, and so must be considered less robust than the findings for student attitudes, attendance, participation and satisfaction. In addition, it was the one area where contradictory findings occurred." (p. 16).

It is both surprising and disappointing to find that Fowler's systematic search, which covered a twenty-one year period (1971-1992), yielded no more than four studies that can be considered to present some valid evidence with respect to the effects of school size on achievement.

Haller, Monk and Tien (1992) present a brief discussion of the research literature on school size and achievement in both elementary and secondary education which they summarise as follows:

> "Overall, it seems safe to conclude that small school size does not lead to noticeable decrements in student achievement." (p. 6).

This conclusion is based on seven studies, but not much information about these studies is provided. Haller et al. mention that in five studies a negative effect or no effect at all was detected of school size on achievement. One study revealed a slightly positive effect. In

[1]It can be inferred from the figures provided by Fowler and Walberg that the percentage of students passing the "High School Proficiency Mathematics Test" drops 2.4 % when a school's total enrollment increases with 500 students, which is the strongest school size effect they report. The (statistically significant) effects on other measures are considerably weaker. E.g., the percentage of students passing the "Minimum Basic Skills Reading Test" drops only 0.3 % with an increase of 500 students. The average school size in their sample is 1070 students, while the standard deviation equals 519.

the other one the size-achievement relationship was reported to be dependent on the socio-economic background of the school population. It remains unclear whether such background characteristics were taken into account in the other studies.

According to Stoel (1992) no general conclusion can be drawn from the various research reports that deal with the relationship between school size and student achievement in secondary education. Sometimes a positive correlation is found, sometimes a negative correlation and sometimes no correlation at all. This conclusion is based on a considerable number of studies, but these are not described in any detail. It is only mentioned whether the studies revealed a positive, negative or zero correlation. No information about the use of control variables is provided. The studies reviewed by Stoel cannot be expected to present a picture that is really up-to-date, because twelve out of the nineteen studies mentioned were published before 1972 and only three were published after 1982. Stoel's review, however, is the only one that is not entirely based on research dealing with the educational system in America. Two studies relating to Dutch secondary education are mentioned. In both cases no relation between size and achievement was detected. This finding has been corroborated by a recent study with respect to the effects of school size on achievement in Dutch secondary education (Kleintjes & Kremers, 1992). These outcomes are in line with the idea that school size effects are not very strong in the Dutch system of secondary education.

It is clear that the conclusions in the four reviews diverge to a considerable extent, although in none of them it is concluded that large school size exerts a positive effect on student achievement. Fowler's review seems the most reliable even though he only discusses four research reports. The conclusion reached by Fowler and Walberg is mainly based on findings with respect to elementary education, while the reviews by Haller et al. and Stoel offer no more than a very condensed description of previous research, which does not enable the reader to assess the validity of the reported outcomes to any extent, especially because no information about the use of control variables is provided.

The fact that until recently no techniques of analysis were at hand that could take into account the hierarchical structure typical of most educational datasets, renders the available research outcomes even more unreliable. Statisticians have rather heavily criticized the research on school effects for inadequate statistical modelling, as they have pointed out that analyzing data which are in some way hierarchically structured by means of a single-level technique (such as multiple regression or analysis of covariance) can result in potentially serious misinterpretations (e.g. Aitkin & Longford, 1986; Bosker, 1990, pp. 37-47; Paterson & Goldstein, 1991). What kind of misleading results might be obtained when hierarchically structured data are analyzed by means of a single-level technique will be outlined in the next section. In section 3 the results of an original investigation into the relationship between school size and student achievement will be reported. The data were analyzed using suitable multilevel software (VARCL; Longford, 1986), so that the inherent hierarchical structure of the data could be taken into account.

1.2. Shortcomings of aggregation and disaggregation

In order to be able to investigate relationships between student level and class-room or school level variables researchers usually either aggregated student level characteristics to a higher level or disaggregated higher level data to the student level. In both cases the researcher runs a serious risk of obtaining misleading results.

If one aggregates data from an individual level to some higher level, the meaning of the data is altered. E.g. if the individual sympathies of voters for a racist political party are aggregated to the level of voting districts, the meaning shifts from individual political sympathies to a measure of the political climate in voting districts. The same goes for the geographical origin (e.g. native or foreign) of the voters. When these characteristics are aggregated, one obtains a measure for the cultural climate in the voting districts. In this example not only the meaning of the variables changes, but also their relationship. At the individual level there will be virtually no sympathy among voters of foreign origin for a racist party (at least not in the Netherlands), but at the level of voting districts a rather strong positive correlation will be found between the percentage of voters of foreign origin and the percentage of votes for racist political parties.

This is of course an extreme example and no researcher would conclude from the correlation at the aggregated level that voters of foreign origin vote for racist parties. However, this example demonstrates that a correlation between two aggregated variables can differ drastically from the correlation between the original, non-aggregated variables. So, if a researcher wants to control for certain student level background variables (e.g. initial achievement or socio-economic status) when investigating the relationship between school size and achievement, the use of aggregated data is bound to produce results that are only valid at the aggregated level. In such an analysis one only controls for initial achievement or socio-economic status at the aggregated level, but the relationship at the individual level may be very different.

Another shortcoming of aggregated data is that any detection of cross-level interactions will be impossible. So, for instance, the relationship between sex and achievement might differ from school to school. If the analysis is confined to aggregated data, such a phenomenon can never be detected.

If the analysis is conducted at the student level using disaggregated class-room or school level data, one is faced with the problem that in those cases standard tests for statistical significance are usually not applicable (Cheung et al., 1990, p. 221). This is due to the fact that the data in educational research are nearly always collected by means of a cluster sample, while the statistical software packages routinely used in educational data-analysis (e.g. SPSS, SAS) produce standard errors that are only valid if the data originate from a single random sample. In educational research, however, usually a sample of schools is taken. Sometimes all the students in the selected schools are included in the sample, sometimes a sample of classes and/or students within the schools is taken. Only in exceptional cases, when the differences between classes and schools are very small compared with the differences between students, can such samples be considered equivalent to single random samples and are standard tests for statistical significance appropriate. In general, however, will the statistical reliability of cluster sample data be considerably lower than the reliability of data originating from a single random sample of the same size (Moser & Kalton, 1971, pp. 201-209).

Multilevel analysis, nevertheless, enables us to produce correct estimates of standard errors of school and class-room effects on individual achievement. Group characteristics can thus be easily incorporated into models of individual behaviour. Multilevel analysis can be considered as a generalization of ordinary multiple regression. The effects of the explanatory variables are expressed as regression coefficients that should be interpreted in the same way as the familiar regression coefficients, the important difference being that in multilevel analysis the coefficients refer to specific levels in the hierarchical structure of the data. E.g. individual cognitive aptitude might explain differences in achievement within classes and between classes. In multilevel analysis different regression coefficients are estimated for each level.

If one wants to check whether a more elaborated model fits the data significantly better than a more parsimonious one, the difference between the goodness-of-fit measures of both models (usually called deviance) should be computed. The distribution of this statistic approaches a χ^2-distribution, so that it can easily be checked whether the difference is statistically significant.

2. DATA AND STRATEGY OF ANALYSIS

The analyzed datasets were derived from two international studies sponsored by the IEA (International Association for the Evaluation of Educational Achievement): the Second International Mathematics Study (SIMS; Travers & Westbury, 1989; Robitaille & Garden, 1989) and the Second International Science Study (SISS; Postlethwaite & Wiley, 1992). The analyses were conducted on Dutch, Swedish and American SIMS-data and on Dutch SISS-data. The data originating from SIMS were collected between May 1980 and June 1982. The SISS-data were collected in May and June 1984. The criterion variable in the analyses of the SIMS-data is formed by the score on a 75 multiple choice item mathematics test. In the SISS-file the criterion variable is formed by a test score which relates to 61 items in the domain of physics, chemistry, biology and earth science. School size was treated as a categorical variable. Thus possible non-linear relations between school size and achievement could easily be detected. The following five school size categories were used in the analyses:

- schools with less than 240 students enroled
- school with at least 240 students but less than 360
- school with at least 360 students but less than 500
- school with at least 500 students but less than 1000
- schools with 1000 students or more

This categorization has also been applied in the research report of the Dutch Social and Cultural Planning Agency on the effects to be expected as a result from school size increases (Blank et al., 1990) and in the research by Kleintjes & Kremers (1992) into the relationship between school size and achievement. It is in line with the prevalent regulations in the Netherlands. The minimum enrolment allowed for single-track schools is 240; for multi-track schools the minimum is 360. The school size categorization applied in the present study thus originates from Dutch research. As a result the categorizations may be somewhat less appropriate for the Swedish and American systems[2]. However, the alternative, employing different categorizations for each country, would entail other

[2]In the analysis of the second American sample (see section 3.1) the second and the third category were combined into one category, because otherwise the number of schools in either category would become too small. In the Swedish sample no schools were found that belonged to the first (< 240 students) or the fifth category (≥ 1000 students).

undesirable consequences. In that case school size would become a rather equivocal concept. A "big" school in Sweden might then be exactly as big as a "medium-sized" school in the Netherlands, which would render the outcomes of the analyses quite confusing. For the sake of clarity it was decided to employ the same categorization across all three countries.

In the analyses it was investigated to what extent student achievement is related to school size if one controls for the following covariates:

a/ Sex
b/ Achievement Motivation
c/ Socio-Economic Status of the Family (SES)
d/ Cognitive Aptitude
e/ Curriculum track

ad b: This variable was measured by means of an index composed of nine items. The achievement motivation measures used in the analyses of the SIMS-data were exactly identical. In the Dutch SISS-study another set of items was used to operationalize this variable. Cronbach's α exceeded .70 for all four scales.

ad c: This variable was measured by four items relating to the profession and the education of the student's parents. In the analysis of the Dutch SISS-data only the parents' education could be taken into account.

ad d: In the American SIMS-study the students had to complete two mathematics tests. The first one at the beginning of the school year and the other at the end. In the analysis of the American SIMS-data the pretest score served as a covariate.

In the Dutch SISS-study the students were supposed to complete either a mathematics test or a word knowledge test apart from the 61 item science test. The scores on these tests were used as covariates. Because a substantial amount of the students didn't complete the mathematics nor the word knowledge test the question whether either one of these two tests had been completed was taken into account as a covariate as well.

In the Dutch and Swedish SIMS-files no direct indicators for cognitive aptitude can be found. In the analyses of the data from these files the curriculum track served as a covariate. Students are selected into these tracks on the basis of their (presumed) scholastic aptitude.

ad e: In the Dutch SIMS-file four curriculum tracks were distinguished: "HAVO/VWO", "MAVO", "LTO" and "LHNO". Both "HAVO/VWO" and "MAVO" offer a general secondary education, "HAVO/VWO" being the more advanced. "LTO" and "LHNO" both offer a lower vocational training. "LTO" stands for lower technical education and "LHNO" for lower domestic education. "LTO"-classes contain mainly male students and "LHNO"-classes mainly female students. In the SISS-file the grouping is somewhat different. Five types are distinguished: "HAVO/VWO", "MAVO", "LTO", "LEAO/LHNO" and "LAO". "LEAO" stands for lower economic and administrative education, "LAO" for lower general education. Most "LEAO/LHNO"-students are girls.
In the Swedish SIMS-file two curriculum tracks are distinguished. The Swedish students in the grades 7,8 and 9 can choose among two mathematics curricula: an advanced and a less advanced curriculum. Usually classes are composed in such a way that all students in a class are in the same track.

Cognitive aptitude was thus measured in various ways. With respect to the American SIMS-data a measure was used that was partly identical to the achievement measure. The analysis of the American data in fact revealed the effects of school size on achievement *gain* with respect to mathematics within one school year. The use of curriculum track as an indicator for cognitive achievement has different implications, however. On the one hand it is a somewhat crude measure as it distinguishes only a few categories, on the other hand it expresses a student's general cognitive aptitude rather than a particular type of academic achievement. It should also be taken into account at what age students are selected into the curriculum tracks and at what age their achievement was measured. In both the Netherlands and Sweden students are selected into the curriculum tracks after six years of elementary education. Since the Swedish data were collected at the end of the first year in secondary education, the variable "curriculum track" can be considered to reflect a student's general aptitude of one year ago in the case of Sweden. The Dutch SIMS-data pertain to students in their second year and the SISS-data to students in their third year of secondary education. So in these cases the curriculum track expresses a student's general aptitude of two and three years ago.
In the analysis it was also checked whether interaction effects on achievement could be discerned between school size on the one hand and sex, achievement motivation, SES or cognitive aptitude on the other. An interaction effect would imply that the effect of school size on achievement is related to any of the aforementioned covariates, e.g., that school size does affect the achievement of students with a low socio-economic background more strongly than those from a higher socio-economic background.

The data were collected by means of a multi-stage sample. The primary sampling units were the schools. Within schools classes were sampled. In both Dutch studies only one class per school was sampled, which means that in the Dutch datasets the school level coincides with the class-room level. The analyses were performed using suitable multilevel software, so that the inherent hierarchical structure of the data could be taken into account (compare Bosker & Snijders, 1990). The independent variables were centered around their group means[3]. Thus the effects at the individual, class-room and school level could be distinguished from each other as clearly as possible. Student characteristics (e.g. the individual pretest scores) were expressed as deviations from the class mean, class-room characteristics (e.g. the class mean pretest scores) as deviations from the school mean and school characteristics as deviations from the grand mean. The only exception being the variables "Curriculum track" (in the Swedish and Dutch datasets) and school size because centring these higher level categorical variables would not reveal any useful information. Sex, however, was treated as a numerical variable. Classes with high scores on this variable are classes with relatively many male students. An individual score higher than zero means that the student is male in a class which doesn't exclusively contain boys. A zero score at the individual level is only possible in classes that are either exclusively male or exclusively female. A score on this variable which is extremely high implies that the student is male and that the vast majority of his class-mates is female. Using this approach it was possible to distinguish several types of sex differences with respect to mathematics and science achievement. The statistical significance of within class differences, between class and between school differences could thus be established.
Several models were examined. Each analysis started with a so-called "zero model". These models establish what percentage of the total variance in the individual achievement can be attributed to differences between classes and/or schools and what percentage can be attributed to individual differences. In the next step a model was examined in which the five covariates served as the explanatory variables. Finally it was examined if the model could be improved by taking school size into account. It was also investigated whether statistically significant interaction effects could be discerned. As a rule only regression coefficients significant for $\alpha < .01$ were allowed in the models. Deviations from this rule are always explicitly reported. The American dataset was randomly split up into two subsamples, so that two separate analyses could be conducted. Subsequently the results of both analyses were compared. This approach was chosen because the character of the

[3]In multilevel models with heterogeneous slopes the interpretation of the intercept variance is facilitated by centering the predictor variables; e.g. around their group mean (Bryk & Raudenbusch, 1992; pp. 25-29)

research was somewhat explorative. The analysis of the first subsample should be considered as a first exploration of the American data. The second analysis served as a test of the validity of the results from the first analysis. In this way the chance that the results are biased because of random fluctuations was further reduced.

As a result five separate datasets were analyzed: two American files, two Dutch and one Swedish. Thus it was possible to compare the results of five independent investigations into the effects of several variables on student achievement. The percentages in achievement variance attributable to schools, classes and individual students could also be compared across the five datasets. The fact that the data originate from three different countries, allowed for an evaluation of cross-national differences. The two Dutch datasets, which relate to two different types of student achievement (mathematics versus science) provided the opportunity for a cross-subject comparison and the splitting up of the American dataset into two subsamples produced useful information about the impact of random fluctuations on the outcomes.

3. RESULTS

3.1. United States

The analyses of the American data relate to students in grade 8 in mainstream public and private schools. For the majority of the students this was their second year of secondary schooling. The average age of the students was 14 years and one month at the time they completed the posttest. The data originate from the Second International Mathematics Study (SIMS).

A 50% sample was taken from the complete dataset, the schools being the sampling units. This part of the file was used to establish in a first analysis which variables affect the achievement level of the students and especially to what extent school size is of any importance in this respect. Next the other half of the dataset was analyzed in an identical fashion. Only the students who at least partially completed both the pretest and the posttest were included in the analyses.

This approach was chosen, because in the analyses the statistical significance of a large number of regression coefficients was examined. The estimation of the school size effect yielded four regression coefficients, because this variable had been operationalized as a five category variable. Regarding the covariates, separate coefficients had to be computed for each level of analysis. The detection of interaction effects involved the computation of more than a dozen of coefficients. As a result the risk of chance capitalization had to be considered. Still this cross-validation approach was not applied in the analyses of the other datasets, because in the American case the results of the two separate analyses were quite similar, at least with respect to the school size effect and the interaction effects. Both analyses revealed no statistically significant effect at all.

The results that were found in both datasets are shown in table 3.1.1. Although both samples produced roughly the same outcomes two remarkable differences need to be mentioned. The first one relates to the percentages of variance in achievement that can be attributed to differences between classes and to differences between schools. In the first sample the zero-model revealed no significant differences in achievement between the schools, while in the second sample 26% of the variance could be attributed to differences between schools. The other remarkable difference refers to the effect of achievement motivation. In the first sample a significant effect was found only at the student level, but the second sample revealed significant effects at the school and class-room level as well.

Table 3.1.1 shows an increase (or negative reduction) of variance at the school level for the first sample when the covariates were included in the analysis. The explanation for this rather counter-intuitive result is that, although no statistically significant differences were found with regard to achievement between schools in the first sample, a non-zero variance did emerge when a number of relevant background characteristics were taken into consideration (especially the pretest score). In other words: schools did not differ significantly with respect to the scores on the mathematics test, but they did differ with respect to the progress of their students. Across all three levels, however, the covariates accounted for a substantial reduction in variance, both in the first and the second sample. The absence of school differences in the first sample as established by the zero-model is in a certain sense misleading because differences do emerge as soon as attention is paid to the differences in the students' background.

Four covariates were included in the analyses of both samples (pretest score, SES, achievement motivation and sex). Only the pretest score revealed significant ($\alpha < .01$) regression coefficients at all three levels in both samples. This implies that:

- the students that got higher scores on the pretest than their classmates also scored higher on the posttest;
- the classes with pretest averages higher than the school mean also got higher posttest averages;
- and finally that the schools with a higher pretest average also got higher posttest averages.

TABLE 3.1: *School size effects on mathematics achievement in the USA.*

	first sample	second sample
Number of students:	2212	2295
Number of classes:	104	107
Number of schools:	58	58

	Model 0		Model 1		Model 2	
	first sample	second sample	first sample	second sample	first sample	second sample
FIXED PART: regression coefficients						
1/ student level						
pretest score	..	..	.82	.88	.82	.88
SES	..	..	.48	.49	.48	.49
achievement motivation	.	..	.31	.21	.31	.21
2/ class-room level						
pretest score	..	..	1.15	.94	1.15	.94
achievement motivation	.	..	not significant	1.14	not significant	1.13
3/ school level						
pretest score	.	..	1.14	1.08	1.14	1.08
achievement motivation	..	.	not significant	1.27	not significant	1.27
school size	..	..	..	..	not significant	not significant
Grand Mean	48.19	48.55	47.96	48.48	47.38	48.80
RANDOM PART: variances of regression coefficients						
1/ class-room level						
pretest score	..	..	.02	.02	.02	.02
	VARIANCE		**VARIANCE EXPLAINED compared with Model 0**			
student level	39.3 %	43.5 %	57.3 %	57.4 %	57.3 %	57.4 %
class-room level	60.7 %	30.2 %	94.6 %	92.6 %	94.6 %	91.4 %
school level	0.0 %	26.2 %	negative	90.0 %	negative	90.0 %
total	100.0 %	100.0 %	78.0 %	76.6 %	78.1 %	76.6 %
Deviance	17894.73	18783.82	15901.64	16719.05	15899.76	16718.73
Difference in degrees of freedom	..	..	7	9	4	3
Model improvement (p)	..	..	< .001	< .001	> .750	> .950

That the student-level pretest coefficient appeared to vary significantly between classes indicates that the effect of the pretest score on the posttest score differed from class to class. SES revealed a statistically significant effect only at the individual level, while the significance of the achievement motivation was not identical in both samples. Sex did not show any significant relation with the achievement of the students.

The most important finding is that the inclusion of school size in the models did not amount to any significant model improvement at all. The same is true for the interaction terms of school size with the pretest score, school size with SES and school size with sex.

3.2. Sweden

The analysis dealt with in this section refers to Swedish students in grade 7 of the 9-year compulsory school. The mean age of these students was 13 years and 9 months. The data originate from the Second International Mathematics Study. The results are summarised in table 3.2.

The analysis of the Swedish data revealed a phenomenon similar to the one discovered in the analysis of the first American subsample. Initially no variance between schools with respect to the (unadjusted) achievements of their students could be detected, but after controlling for a number of relevant covariates school differences did emerge. The most striking differences between the Swedish and the American results were presented by the effects of achievement motivation. Not only did the achievement motivation effect differ significantly between both classes and schools, but also was the sign of the regression coefficient negative at the school level. This is an unexpected result indicating that in schools with a high average achievement motivation the students performed relatively low on the mathematics test. Whether this observed negative correlation reflects a causal relationship is dubious, however. In this particular case achievement motivation might just as well reflect a reaction to achievement in stead of explaining it. Schools that in the past got poor results could be trying to improve their students' achievement by creating a more achievement oriented school climate, which might result in a relatively high achievement motivation for such schools. But, if the achievement levels do not improve, a negative correlation between motivation and achievement will be observed[4]. It should also be borne in mind that the achievement motivation effect displayed a considerable amount of variance at the school level. In approximately a quarter of the schools the relation between the mean school motivation score and achievement was positive.

The Swedish and American data revealed similar findings with respect to sex (no significant effect) and SES (only significant at the student level). The Swedish dataset did not contain any information about initial achievement, but a strong effect of the curriculum track variable on mathematics achievement was found. The students in the classes where the advanced course was offered got higher scores.

The results with respect to school size were identical to those found in both the American samples. No significant effect of school size could be detected. Interaction terms of school size with SES, achievement motivation en sex revealed no statistically significant effects on achievement either.

3.3. The Netherlands

In this section the research outcomes derived from the Dutch SIMS and SISS-data will be dealt with. The SIMS-data relate to students who were in their second year of secondary education. Their average age was 14 years and 4 months. The SISS-data relate to students who were in their third year of secondary education. The average age of these students was 15 years and 6 months. The criterion variable in the SIMS-file is the score on a 75 item mathematics test. In the SISS-file the criterion variable is a 61 item test score referring to physics, chemistry, biology and earth science.

3.3.1. Mathematics (SIMS)

The outcomes of the analysis of the Dutch SIMS-data are shown in table 3.3.1. Curriculum track and achievement motivation revealed significant effects on mathematics achievement.

---

[4]A similar argument can, of course, be made about positive correlations between motivation and achievement. The possibility that motivation is the effect rather than the cause of achievement can not be ruled out.

TABLE 3.2: *School size effects on mathematics achievement in Sweden*

Number of students:	3500
Number of classes:	182
Number of schools:	95

	Model 0	Model 1	Model 2
FIXED PART regression coefficients			
1/ student level			
SES	--	.90	.90
achievement motivation	--	3.76*	3.75*
2/ class-room level			
achievement motivation	--	8.57	8.59
curriculum track	--	16.01	15.93
3/ school level			
achievement motivation	--	-10.30	-11.01
school size	--	--	not significant
Grand Mean	35.01	26.99	27.94
RANDOM PART: variances of regression coefficients			
1/ class-room level			
achievement motivation	--	12.16	12.17
2/ school level			
achievement motivation	--	318.29	329.31
VARIANCE		**VARIANCE EXPLAINED compared with model 0**	
student level	55.6 %	6.7 %	6.8 %
class-room level	44.4 %	93.6 %	93.8 %
school level	0.0 %	negative	negative
total	100.0 %	39.9 %	39.9 %
Deviance	27959.79	27543.53	27542.60
Difference in degrees of freedom	--	9	2
Model improvement (p)	--	< .001	> .500

*This regression coefficient was not significant for $\alpha<.01$ in the models 1 and 2. It has been maintained in these models because the analysis revealed a significant variance of this coefficient between classes. The coefficient was significant for $\alpha<.05$ in both models.

With respect to sex and SES the results diverged from those found in America and Sweden. No significant effect of SES on achievement could be detected, while a significant effect of sex was detected at the student level. Within classes the Dutch boys got higher scores on the mathematics test than their female class-mates.
The analysis of the Dutch data again yielded no significant effects of school size on achievement ($\alpha >.10$). One of the interaction terms of school size with sex, however, did reveal a statistically significant effect ($t = 3.96$; $\alpha < .0001$). The negative regression coefficient of this interaction term should be interpreted as follows: In the schools with at least 360 but less than 500 students the girls got higher scores on the mathematics test than their male classmates. The interaction terms of school size with SES and achievement motivation revealed no significant effects ($\alpha > .05$).
The interaction effect of school size with sex is almost completely due to the fact that the female "MAVO"-students in schools with 360 up to 500 students outperformed their male classmates. The effect was not confirmed in the analysis of the science achievement data (see section 3.3.2), but in another study into the relationship between school size and student achievement in Dutch secondary education a similar though not identical phenomenon was detected. It appeared that students in schools with 360 up to 500 students got better scores on a mathematics test, but not on tests for biology, Dutch and English language (Kleintjes & Kremers, 1992). A convincing explanation for these overachievements of (female) students in Dutch secondary schools of medium size with respect to mathematics has not yet been offered.

3.3.2. Science (SISS)

The results for science achievement are summarised in table 3.3.2. Curriculum track revealed a strong effect on science achievement. The same goes for the variables referring to the mathematics and the word knowledge test. Significant coefficients were found both at the student level and the class/school level. The regression coefficient of the variable "one of both tests completed" differed significantly between classes. Achievement motivation and sex also showed significant effects on both levels. The regression coefficient of sex differed significantly between classes.

TABLE 3.3.1: *School size effects on mathematics achievement in the Netherlands.*

Number of students: 5313
Number of classes/schools: 228

Class and school levels coincide, because only one class per school was sampled.

	Model 0	Model 1	Model 2
FIXED PART: regression coefficients			
1/ student level			
achievement motivation	..	4.21	4.20
sex (male high score)	..	2.19	2.64
interaction terms of sex with school size - < 240 students*sex - 240-359 students*sex - 360-499 students*sex - 500-999 students*sex			 not significant not significant -5.35 not significant
2/ class/school level			
achievement motivation	..	12.79	12.53
curriculum track - VWO/HAVO - MAVO - LTO - LHNO		 .00 -17.49 -31.39 -41.95	 .00 -15.48 -29.62 -39.51
school size	..	..	not significant
Grand Mean	54.08	73.06	72.86
VARIANCE		**VARIANCE EXPLAINED compared with model 0**	
student level	33.1 %	4.6 %	5.1 %
class and school level	66.9 %	82.4 %	83.1 %
total	100.0 %	56.7 %	57.3 %
Deviance	42770.91	42154.70	42121.68
Difference in degrees of freedom	..	6	8
Model improvement (p)	..	< .001	< .001

BEST COPY AVAILABLE

23

TABLE 3.3.2: *School size effects on science achievement in the Netherlands*

Number of students: 4286
Number of classes/schools: 194

Class and school levels coincide, because only one class per school was sampled.

	Model 0	Model 1	Model 2
FIXED PART: regression coefficients			
1/ student level			
mathematics test score	..	.24	.24
word knowledge test score	..	.47	.47
one of both tests completed? ("yes" high score)	..	1.85*	1.84*
parents' education	..	-.42	-.42
achievement motivation	..	.36	.36
sex (male high score)	..	5.76	5.76
2/ class/school level			
mathematics test score	..	.29	.29
word knowledge test score	..	.46	.44
one of both tests completed? ("yes" high score)	..	12.49	12.22
achievement motivation	..	1.94	1.75
sex (male high score)	..	7.24	7.88
curriculum track - VWO/HAVO - MAVO - LTO - LEAO/LHNO - LAO		 .00 -5.23 -12.09 -11.70 -7.28	 .00 -5.72 -12.73 -11.75 -7.98
school size	..	..	not significant
Grand Mean	57.79	61.48	61.53
RANDOM PART: variances of regression coefficients			
1/ class/school level			
one of both tests completed? ("yes" high score)	..	22.74	22.81
sex	..	14.22	14.37

24

	Model 0	Model 1	Model 2
VARIANCE		VARIANCE EXPLAINED compared with Model 0	
student level	45.0 %	15.9 %	15.9 %
class and school level	55.0 %	84.5 %	84.8 %
total	100.0 %	53.6 %	53.8 %
Deviance	32740.01	31795.52	31791.54
Difference in degrees of freedom	..	19	4
Model improvement (p)	..	< .001	> .250

*This regression coefficient was not significant for $\alpha<.01$ in the models 1 and 2; It has been maintained in these models, because the analysis revealed a significant variance of this coefficient at the class/school level. The coefficient was significant for $\alpha<.05$ in both models.

A remarkable outcome was represented by the negative effect at the student level of the variable "parents' education" on science achievement. This variable referred to the amount of secondary schooling received by the parents and to the amount of education in addition to their secondary schooling. The negative sign of the regression coefficient implies that within classes the students whose parents received little schooling got better scores on the science test. An explanation for this phenomenon might be that in the Netherlands the parents who received a lot of a schooling themselves send their children more often to schools which offer the more advanced curriculum tracks even when their children are not so bright.

No significant improvement of the model was realized when school size was included. The same is true for the interaction terms of school size with the first six explanatory variables in the model (mathematics and word knowledge test score, one of both tests completed, parents' education, achievement motivation and sex). The interaction effect reported in the previous section was not confirmed in the analysis of the science achievement data.

3.4. Size of the effects

It can not be concluded on the basis of the research outcomes reported thus far that student achievement is independent of school size. The analyses started from the assumption that there is no relation between school size and achievement. It has only been shown that the data did not reveal any results that allow for a rejection of this null hypothesis. The null hypothesis would only have been rejected if an effect of school size had been found that could not be attributed to mere coincidence.

This is a very common approach in social scientific research: one hypothesizes that there is no relation and this null hypothesis will only be rejected if the empirical analysis reveals a result that would be quite unlikely if the actual relation is non-existent. Researchers in the social sciences tend to neglect the possibility that the null hypothesis might *not* be rejected, when in reality a relationship *does* exist (Cohen, 1988). In the present case it would have been a serious omission not to pay any attention to the chance of wrongly concluding that school size has no (negative) effect on achievement. The samples that were investigated in the present research all contained a large number of students, but at the school level the sample sizes were much more modest. While in large samples even small effects can be statistically significant, the opposite applies to samples which contain only a limited number of units. This is why the magnitude of the school size effects needs to mentioned in addition to their statistical significance, before valid answers can be given to the questions formulated in section 1.

The school size regression coefficients that were found, however, all revealed very modest effects of school size on achievement. The largest negative effect was found in the Dutch mathematics sample. In schools with at least 240 but less than 360 students the test scores appeared to be 1.87 points lower than in the smallest schools (less than 240 students). The largest positive effect appeared in the same sample. In schools with at least 1000 students the test scores were 3.34 points higher than in the schools from the smallest category. This effect is really quite modest considering that the maximum score students could achieve was 100 points and the minimum score 0 points. The standard deviation of the test score frequency distribution equalled 21.6. The existence of more than very modest school size effects in any of the three countries included in the research seems thus not very likely. This conclusion is corroborated by the fact that the observed non-significant effects[5] did not reveal any clear pattern whatsoever. In the first American sample, e.g., high achievement scores were found in the schools with 500 up to 1000 students, while in the second sample these schools revealed low test scores. The two Dutch datasets revealed similar contradictions. Nor could any cross-national pattern of school size effects be detected. Furthermore the percentages of additional variance explained by the models containing school size as an explanatory variable were very low in every case. In the Dutch mathematics sample this percentage equalled 0.6%, but this is mainly due to the

[5]These non-significant effects were not presented in the tables.

interaction effect of school size with sex. The percentages found in the other samples ranged from 0.0% to 0.2%.

With respect to the interactions between school size and the student background characteristics one statistically significant effect could be detected. Girls got relatively high scores on the mathematics test in the Dutch schools of medium size, but it should be mentioned that even this effect was still quite modest in terms of explained variance (0.6%). Moreover this interaction term was the only one to show a statistically significant correlation with achievement, while the effect of several dozens had been examined.

The four research questions (see section 1) can therefore be answered as follows:

(1) The systems of secondary education in the Netherlands, Sweden and the USA did not reveal any statistically significant or practically meaningful relationship between school size and achievement that was independent of student background characteristics;

(2) The effects of school size on achievement did, in general, not interact with the student background characteristics that were taken into account, the only exception being the interaction effect of school size with sex on mathematics achievement in the Netherlands;

(3) As the effect of school size on achievement appeared to be absent in all three countries, no differences of school size effects between countries were detected; the supposition that school size would have a stronger effect in America than in the two European countries was not corroborated;

(4) The main effect of school size on mathematics and science achievement turned out to be identical, namely zero. However, with respect to mathematics achievement an interaction effect of school size with sex was found.

3.5. Robustness of the research outcomes

Although the absence of school size effects was a consistent finding in each of the five samples, the analyses did reveal a number of contradictory results as well. Table 3.5 presents an overview of the effects of school size and the five covariates on achievement. The "zero model" percentages of variance in achievement attributable to schools, classes and individual students are listed in this table as well. The fact that the research produced information about three different countries, allowed for an evaluation of cross-national differences. The two Dutch samples, which relate to two different types of student achievement (mathematics versus science) provided an opportunity for a cross-subject comparison, although we should bear in mind that apart from the differences in subjects these datasets also refer to different kinds of students. In the case of mathematics achievement the investigations related to students in their second year of secondary education, whereas the analyses with respect to science achievement pertained to students in their third year. The two American samples yielded useful information about the possible impact of random fluctuations on the research outcomes.

If we consider the cross-national differences that have emerged from the analyses, it can be concluded from table 3.5 that the variables cognitive aptitude and curriculum track revealed similar effects on achievement across all three countries. Achievement motivation appeared to be positively related to achievement in most instances, although a contradictory outcome was found in Sweden at the school level. In the Netherlands some diverging outcomes with respect to sex and SES were found. A significant effect of sex on achievement could only be detected in the Dutch educational system, while the positive effect of SES found in Sweden and America was not confirmed by the results in the Netherlands.

A comparison of the outcomes found in the two Dutch samples shows that the kind of achievement to be explained can lead to different results as well. The effects of sex and SES were not exactly identical for mathematics and science achievement. It should be borne in mind, however, that SES was not measured in exactly the same way in both instances. In the Dutch SISS-sample the SES-measure relates only to the parents' education. In the other samples it relates to both their education and their profession. The percentages of student level and class/school level variance did also differ to a certain extent.

The two American datasets yielded virtually identical outcomes with respect to the student level. At the class-room and school levels some diverging results were found. The percentages of class-room and school level variance differed in both samples as well as the effect of achievement motivation. The impact of random fluctuations appeared to be much stronger a the higher levels than at the student level. This is not very surprising, because the research outcomes at these higher levels are based on much smaller numbers of units. Both samples contain over 2200 students each, but they only comprise a little more than 100 classes and no more than 58 schools each. Such sample sizes inevitably render the results of the analyses statistically less reliable.

Thus, the analyses dit reveal some divergence with respect to the effects of the covariates on achievement across the datasets. This lack of consistency can be attributed to at least three factors:

TABLE 3.5: *Comparison of variances and effects across the five datasets*

	Variance	Sex	Achievement motivation	SES	Cognitive aptitude	Curriculum track	School size
USA: first sample							
student level	39.3 %	0	+	+	+'	not applicable	not applicable
class-room level	60.7 %	0	0	0	+	not applicable	not applicable
school level	0.0 %	0	0	0	+	not applicable	0
USA: second sample							
student level	43.5 %	0	+	+	+'	not applicable	not applicable
class-room level	30.2 %	0	+	0	+	not applicable	not applicable
school level	26.2 %	0	+	0	+	not applicable	0
Sweden							
student level	55.6 %	0	+'	+	not applicable	not applicable	not applicable
class-room level	44.4 %	0	+'	0	not applicable	+	not applicable
school level	0.0 %	0	-	0	not applicable	not applicable	0
Netherlands: mathematics							
student level	33.1 %	+	+	0	not applicable	not applicable	not applicable
class/school level	66.9 %	0	+	0	not applicable	+	0
Netherlands: science							
student level	45.0 %	+'	+	-	+'	not applicable	not applicable
class/school level	55.0 %	+	+	0	+	+	0

The signs that are marked with an ' refer to effects that differed from class to class or from school to school.
A positive effect of sex on achievement means that boys got higher test scores. A positive effect of type of education means that the more advanced types of education produced higher test scores.

- National differences: in many instances the size and direction of a relationship between achievement and an independent variable is probably influenced by context characteristics which vary between countries. E.g., the absence of a positive effect of SES on achievement in the Netherlands might be due to the existence of a considerable number of curriculum tracks. Parents with a high SES might be more eager than other parents to get their children into one of the more advanced tracks, even if the children are not very talented.
- Different kinds of achievement: which variables are related to student achievement will also depend to some extent on the kind of achievement that is to be explained; the partly contradictory results produced by the analyses of the two Dutch files support this idea.
- Random fluctuations: the fact that the two American subsamples yielded some conflicting outcomes at the school and class-room level proves that even data from samples containing an impressive number of students can present a biased picture because of random fluctuations.

Apart from the lack of consistency across datasets, some divergence within datasets was detected as well. Two types can be distinguished:

- Differences of effects between class-rooms and schools: the effect of several covariates on achievement varied significantly across class-rooms or schools in a number of instances.
- Differences of effects between levels: covariates with a positive effect at the student level often revealed a zero effect or even a negative effect at the class-room or school level[6].

Clearly, the interpretation of research outcomes in the field of education requires a great deal of caution. In many cases the impact of at least some of the five mentioned sources of inconsistency will be obscured. E.g., it is usually not possible to compare the effects of an explanatory variable on achievement across countries; often the research outcomes are based exclusively on data at the school level. As long as researchers are aware of the limited validity of the outcomes, it will be feasible to obtain useful information on the basis of sound educational research. However, we should be careful not to draw any

---

[6]This illustrates in fact that the meaning of variables, and consequently their relationship with other variables, might alter when they are aggregated to some higher level. This possibility was already discussed in section 1.2.

unwarranted conclusions. The research reported in this paper has demonstrated, e.g., that it would be foolish to assume that a variable that appears to have an effect on student achievement in one country, will reveal an identical effect in other countries. Conclusions about relationships between variables at the student level, based on research which only involved the analysis of aggregated data would be unjustified for similar reasons. In both cases the research outcomes do not allow such general conclusions.

4. DISCUSSION

The discussion of the four reviews in section 1.1 showed that contemporary and reliable research with respect to the effects of school size on achievement in secondary education is (surprisingly) scarce. Fowler's systematic search for American reports published after 1970 yielded no more than four studies in which the effects of school size were controlled for socio-economic background. These studies did not produce a consistent picture, as both positive and negative relationships between school size and achievement were found. Research into the effects of school size in Dutch secondary education has not revealed any effect on achievement at all. The fact that until recently educational researchers did not have the techniques of analysis at their disposal that can take into account the hierarchical structure typical of many educational data, renders most of the findings that have resulted from previous research somewhat questionable.

It was pointed out that based on the notion of economies of scale a positive relation between school size and student achievement can be expected. It was hypothesized that such a positive effect would be stronger in the United States than in Sweden or the Netherlands. The possibilities to group students into homogeneous classes are limited for the small American High Schools, while classroom heterogeneity has been found to have a (moderately) negative effect on student achievement (Kulik & Kulik, 1982). Small secondary schools in the Netherlands do not have to deal with a heterogeneous student population, while the Swedish schools are usually large enough to group students into homogeneous classes.

In section 3 the results of an investigation into the relationship between school size and student achievement in three different countries have been reported. The analyses revealed little empirical evidence for the existence of any school size effects on achievement. It must be emphasized, though, that the investigations in the cases of the USA and Sweden only pertain to mathematics achievement and in the Dutch case to mathematics and science achievement. It should also be taken into account that the data on which the research outcomes are based were collected some time ago, in the early eighties and only refer to the relationship between school size and achievement in Sweden, the Netherlands and the USA. Nevertheless, the outcomes demonstrated that there is no apparent reason for policy makers to fear for detrimental effects of school size increases on student achievement, nor to hope for beneficial ones.

Economic theory did not seem very fruitful in the present study. The absence of a positive school size effect on achievement may be explained (partly) by the hypothesis that small schools exhibit a more favourable climate which compensates for any disadvantages of scale (e.g. Barker & Gump, 1964; Lindsay, 1982; 1984). This argument becomes even more plausible, when we acknowledge that the available research indicates that strong school size effects on achievement are not very likely anyway. The effects of school size on curriculum comprehensiveness appear quite modest and probably affect only a small number of students (Monk, 1987; Haller et al., 1990). The effects of homogeneous grouping cannot be expected to be very strong either (Kulik & Kulik, 1982) and the effects of material facilities, such as libraries, computers and science equipment are very questionable (Scheerens, 1992; pp. 36-37). However, the fact that the hypotheses derived from economic theory were not confirmed in the present study does not seriously undermine its validity. The absence of a clear association between school size and achievement, independent of student background characteristics is most probably due to the weak relation between school size and curriculum comprehensiveness. For each of the three countries it can be explained why school size and curriculum comprehensiveness are not strongly related. In Sweden schools are usually large enough to offer a comprehensive curriculum and to group students into homogeneous classes. In the Netherlands small schools are not supposed to deal with a heterogeneous student population (with respect to their cognitive aptitudes), nor to offer a comprehensive curriculum. In America most small schools are still large enough to offer a basic curriculum. The size of secondary schools has increased steadily in the past few decades, so that very small schools which are unable to offer a basically comprehensive curriculum are largely extinct in the United States.

It should be borne in mind that in the present study schools of different sizes were compared. The investigations were not focused on schools that had actually undergone an organizational change resulting in a larger school size. Although no differences in achievement were found between students from schools of different size, this does not guarantee that school size increases will not affect the achievement level of students. An operation aimed at increasing the size of schools is bound to entail certain transition effects, such as the need to invest in new school buildings, or a rise in unemployment among teachers. It is not known if transition effects with respect to student achievement are to be expected. To be able to reach a more clear-cut conclusion, schools should be studied that have actually experienced a considerable change in size. The Dutch system of secondary education provides an interesting opportunity to study such processes, because many schools have actually undergone such changes quite recently and many others will experience them in the near future. The changes that are taking place will probably result in a system of larger schools offering a broader range of curriculum tracks than is presently the case. This implies that the secondary schools will have to deal with students of more diverging backgrounds. How the schools will cope with such changing circumstances and how this will affect student achievement would be an interesting topic for future research.

The analyses did also reveal some useful information about the robustness of the statistical relationships that were detected in the analyses. The following sources of inconsistency could be identified on the basis of the research outcomes:

- Differences of effects between countries;
- Differences of effects due to the kind of achievement to be explained;
- Differences of effects between class-rooms and schools;
- Differences of effects between levels;
- Random fluctuations.

When it comes down to drawing conclusions from research the restrictions imposed on the generalizability of the outcomes by the possible impact of these factors (and many others) should be taken into account.

REFERENCES

Aitkin, M. & Longford, N. (1986). Statistical Modelling Issues in School Effectiveness Research. *Journal of the Royal Statistical Society, Series A (General) 149*, 1-42.

Ball, S.J. (1993). Education Markets, Choice and Social Class: the market as a class strategy in the UK and the USA. *British Journal of Sociology of Education, 14(1)*, 3-19.

Barker, R.G. & Gump, P.V. (1964). *Big School, Small School. High School Size and Student Behavior.* Stanford, Stanford University Press.

Blank, J.L.T., Boef-van der Meulen, S., Bronneman-Helmers, H.M., Herweijer, L.J., Kuhry, B. & Schreurs, R.A.H. (1990). *School en schaal.* Rijswijk/Den Haag: Sociaal en Cultureel Planbureau/VUGA

Bell, A. & Sigsworth, A. (1987) *The Small Rural Primary School. A Matter of Quality.* London: Falmer Press.

Bosker, R.J. (1990). *Extra kansen dankzij de school?* Nijmegen: ITS/OoMO.

Bosker, R.J. & Snijders, T.A.B. (1990). Statistische aspecten van multiniveau onderzoek. *Tijdschrift voor Onderwijsresearch, 15(5),* 317-329.

Bray, M. (1988). Small size and unit costs. International evidence and its usefulness. *Research on Rural Education, 5(1),* 7-11.

Brown, B.W. (1992). Why Governments Run Schools. *Economics of Education Review, 11(4),* 287-300.

Bryk, A.S. & Raudenbusch, S.W. *Hierarchical Linear Models: applications and data analysis methods.* Newbury Park [etc.]: Sage Publications.

Cheung, K.C., Keeves, J.P., Sellin, N. & Tsoi, S.C. (1990). The Analysis of Multi-level Data in Educational Research: Study of Problems and Their Solutions. *International Journal of Educational Research, 14(3),* 217-319.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences, second edition,* Hillsdale, New Jersey: Lawrence Erlbaum Associates, Publishers.

Conant, J.B. (1959) *The American High School Today.* New York: McGraw-Hill Book Company.

Conant, J.B. (1967). *The Comprehensive School.* New York: McGraw-Hill Book Company.

Fowler, W.J. (1992). What do we know about school size? - What should we know? *A paper presented to the American Educational Research Association Annual Meeting in San Francisco, California, April 22, 1992.*

Fowler, W.J. & Walberg, H.J. (1991). School Size, Characteristics, and Outcomes. *Educational Evaluation and Policy Analysis, 13(2),* 189-202.

Fox, W.F. (1981). Reviewing economies of size in education. *Journal of Education Finance, 6(4),* 273-296.

Guthrie, J.W. (1979). Organizational scale and school success. *Educational Evaluation and Policy Analysis, 1(1),* 17-27.

Haller, E.J., Monk, D.H., Spotted Bear, A., Griffith, J., & Moss, P. (1990). School size and Programme Comprehensiveness: Evidence From "High School and Beyond". *Educational Evaluation and Policy Analysis, 12(2),* 109-120.

Haller, E.J., Monk, D.H. & Tien, L.T. (1992). Small schools and higher order thinking skills. *A paper prepared for presentation at the annual meeting of the American Educational Research Association, San Francisco, April 1992.*

Husen, T. & Postlethwaite, T.N. (eds.) (1988). *The International Encyclopedia of Education, Research and Studies, Volume 8,* Oxford [etc.]: Pergamon Press.

Husen, T. & Postlethwaite, T.N. (eds.) (1990). *The International Encyclopedia of Education, Research and Studies, Supplementary Volume Two,* Oxford [etc.]: Pergamon Press

Kleintjes, F.G.M. & Kremers, E.J.J. (1992). School size, characteristics and outcomes in Dutch secondary education. *Paper presented to the European Conference on Educational Research, Enschede, the Netherlands, June 22-25, 1992.*

Kulik, C.C. & Kulik, J.A (1982). Effects of ability grouping on secondary school students: a meta-analysis of evaluation findings. *American Educational Research Journal, 19(3)*, 415-428.

Lindsay, P. (1982). The effect of high school size on student participation, satisfaction and attendance. *Educational Evaluation and Policy Analysis, 4*, 57-65.

Lindsay, P. (1984). High school size, participation in activities, and young adult social participation: Some enduring effects of schooling. *Educational Evaluation and Policy Analysis , 6(1)*, 73-83

Longford, N.T. (1986). *Variance component analysis: manual.* University of Lancaster.

Monk, D.H. (1987). Secondary school size and curriculum comprehensiveness. *Economics of Education Review, 6(2)*, 137-150.

Moser, Sir C. & Kalton, G. (1971). *Survey methods in social investigation.* London: Heinemann Educational Books.

OECD/OCDE (1992). *Education at a Glance: OECD Indicators/ Regards sur l'éducation: Les indicateurs de l'OCDE.* Paris: OECD/OCDE.

Paterson, L. & Goldstein, H. (1991). New Statistical Methods for Analysing Social Structures: an introduction to multilevel models. *British Educational Research Journal, 17(4)*, 387-393.

Pittman, R.B. & Haughwout, P. (1987). Influence of high school size on dropout rate. *Educational Evaluation and Policy Analysis, 9(4)*, 337-343.

Postlethwaite, T.N. & Wiley, D.E. (1992). *Science achievement in twenty-three countries.* Oxford [etc.]: Pergamon Press.

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

Scheerens, J. (1992). *Effective Schooling.* London: Cassel.

Schoggen, P. & Schoggen, M. (1988). Students voluntary participation and high school size. *Journal of Educational Research, 81(5)*, 288-293.

Stoel, W.G.R. (1982). De grootte van scholen voor voortgezet onderwijs en het welbevinden van de leerlingen. *Pedagogische Studiën, 59*, 493-506.

Stoel, W.G.R. (1992). "Schoolgrootte, kosten en kwaliteit; een literatuuronderzoek". In: R.J. Bosker (Ed.):*Schoolgrootte, effectiviteit en de basisvorming (SVO-1032).* Enschede: University of Twente, OCTO.

Travers, K.J. & Westbury, I. (1989). *The IEA Study of Mathematics I: Analysis of Mathematics Curricula.* Oxford [etc.]: Pergamon Press.

BEST COPY AVAILABLE