DOCUMENT RESUME

ED 340 721 TM 017 740

AUTHOR	Kreft, Ita G. G.
TITLE	Using Multilevel Analysis To Assess School Effectiveness: A Study of Dutch Secondary Schools.
INSTITUTION	Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	CSE-TR-303
PUB DATE	Feb 90
CONTRACT	G0086-003
NOTE	34p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
AVAILABLE FROM	CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, CA 90024-1521.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Academic Achievement; Admission Criteria; *Educational Environment; Educational History; Educational Policy; Effective Schools Research; Foreign Countries; Holistic Approach; Mathematical Models; *Outcomes of Education; Private Schools; Public Schools; *School Effectiveness; School Role; Secondary Education; *Secondary School Students
IDENTIFIERS	Homogeneity of Variance; *Multilevel Analysis; *Netherlands; United Kingdom; United States

ABSTRACT

Separate but related developments in the new generation of school effectiveness research are discussed. The first part of the report presents a short historical overview that traces the change from a largely individualistic research approach to one that is more holistic. Whether schools make a difference is explored, and, if so, what makes them effective. Two competing theories are described that offer different explanations for student effects: the student body hypothesis and school climate hypothesis. These two theories are also the basis for a discussion about the merits of private versus public education in the United States. The second part of the paper reviews the accompanying search for an appropriate data analysis model. The third part describes a new way of analyzing hierarchically nested data by using a large data set (5,310 students in 70 secondary schools in Amsterdam) that represents the different forms of secondary education existing in the Netherlands (comprehensive, administrative, college and university preparatory, gymnasium/lyceum, and miscellaneous). The effects of selection policies in the Dutch educational system are clearly indicated through a statistical tool, homogeneity analysis, that is especially suited for analysis of multilevel data. Results are compared with those from studies of school systems in the United Kingdom and the United States. The selection policies of schools have consequences for student achievement. Eight tables, four figures, and a 44-item list of references are included. An appendix describes the Dutch school system. (SLD)

CRESST

Center for Research on Evaluation, Standards, and Student Testing

USING MULTILEVEL ANALYSIS TO ASSESS SCHOOL EFFECTIVENESS: A STUDY OF DUTCH SECONDARY SCHOOLS

CSE Technical Report 303

Ita G.G. Kreft

UCLA Center for the Study of Evaluation

USING MULTILEVEL ANALYSIS TO ASSESS SCHOOL EFFECTIVENESS: A STUDY OF DUTCH SECONDARY SCHOOLS

CSE Technical Report 303

Ita G.G. Kreft

UCLA Center for the Study of Evaluation

February, 1990

The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred

This paper was written as part of a research project sponsored by the UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST). It was prepared for presentation at the Annual Meeting of the American Educational Research Association, San Francisco, March, 1989.

Please address inquiries to: CSE Dissemination Office, UCLA Graduate School of Education, 405 Hilgard Avenue, Los Angeles, California, 90024-1521

Abstract

This paper is divided into three parts to show separate but related developments in the new generation of school effectiveness research. The first part presents a short historical overview which traces the change from a largely individualistic research approach to one that is more holistic. We explore a question that has had prominence in the last decades of educational research: Do schools make a difference? and if they do, what makes them effective? We examine two competing theories that offer different explanations for school effects. These two theories are also the basis of a discussion about the merits of private versus public education in the United States. The second part of this paper reviews the accompanying search for an appropriate analysis model. The last part describes a new way of analyzing hierarchically nested data by using a large data set that represents students from the different forms of secondary education that exist today in the Netherlands. The effects of selection policies in the Dutch educational system are clearly indicated through use of a statistical tool that is especially suited for the analysis of multilevel data. We compare these results with those found in studies that examine school systems in the United Kingdom (UK) and the United States of America (USA) and generalize our findings. We conclude that the selection policies of schools indeed have consequences for student achievement.

Introduction to the Problem

What makes schools effective? If some schools, or types of schools, have a better effect on student achievement, to what can we attribute their success?

Although the research problem that these questions present is basically a theoretical one, it has methodological aspects as well. Until recently, no analysis method was available that could incorporate the variables that are integral to finding reliable answers to these important questions. The development of the new, so-called "second generation" of school effectiveness research has provided the theory as well as the methodology to deal with these variables, some of which are measured at the student level, some of which are measured at the school level. The resulting models for multilevel data are sensitive to differences that exist among students and types of schools, whether these differences occur as a result of natural individuality or of social or political constructions. We developed such a model to study the effects of school selection policies on student achievement.

To find support for the theory that school selection policies do influence student outcome, we compared research from different fields and different countries. We examined research that explores the effects of "tracking" and looked at related research in the Netherlands and the UK that explores the effects of different school types on secondary education. We also looked at researchers' recent interest in larger school systems and their resulting studies into the effects of school organization (for instance, the comparison between public and private schools in the United States). We then designed our model and applied it in a multilevel analysis of a data set collected in the Netherlands.

Before presenting our method of analysis and the results of our study, we will look at the evolution of school effectiveness research since the mid 1960s. A reading of the literature clearly shows that political climate and the perception of society have had an influence on research in education. In our conclusion we will emphasize this link between policy and research.

Developments in School Effectiveness Research after the First Coleman Report

Educational research into the effect of school characteristics as an influence beyond individual student attributes was interrupted after the devastating conclusions of the first Coleman report (Coleman et al., 1966), which stated that schools do not make a difference in student achievement. Following the publication of that report, educational research was concerned mainly with identifying the student characteristics that influence student achievement, since school characteristics reportedly did not. Finding the important relationships between individual achievement and individual background variables became the main objective, and as a consequence, analysis focused on the student level. The empirical model was used primarily to examine the mobility of individuals in society. In later models the school did play a role, but only as it included information concerning "significant others" such as teachers and peers (see Hauser, Sewell, & Alwin, 1976; Hauser, Tsai & Sewell, 1983).

The models used in the analyses cited above were based on the work of Blau and Duncan (1967), and were variations of what is known widely as the "Wisconsin Model." Researchers who pursued this status attainment approach based their theory on this one particular empirical model, which was described in a path analysis framework. The model could not accommodate school characteristics since they could not be disaggregated conveniently to the individual level. This was not a problem, however, since it had already been "proven" that schools did not matter. The model made the socio-economic milieu of the students the emphasis of the study of school success, an emphasis prompted by the Coleman report (1966), which reported that socio-economic factors explained more variance in the achievement of students than did the combination of the amount of money allocated to schools, the quality of the teachers, and the appearance of the school building.

School effectiveness research in the Netherlands was also dominated by the Wisconsin Model approach (see Dronkers & De Jong, 1979; Dronkers, 1983). Jencks' book *Inequality* (1972) also had a clear influence on Dutch research during this period. Jencks concluded on the basis of his individual-level analysis model that amount of schooling was not the important factor in determining later income and success in society: Individual characteristics were more important determinants. Two decades of research in the Netherlands indicated that the socio-economic status of the parents was an important factor, and school factors also seemed to play a role. However, since the school could not be described in either the theoretical or the analytic model as an organization with its own characteristics, there was no way to determine how schools reproduced this inequality observed in society. As a result, not much progress was made in establishing if or how school environment influences student achievement. The urge to investigate the influence of school environment remained low for another, more philosophical reason: Researchers in the Netherlands considered single schools as part of a large selection system in society as a whole. All schools were considered to function in more or less the same way in a selective society. But these ideas changed rapidly.

The Introduction of a New Generation of School Effectiveness Research

In the period that followed, the period of second generation school effectiveness research, an association with a particular analysis model was made again. The theory of what determines school success changed, and emphasis was placed on the school as an organization that can influence student learning. Consequently, methods of data analysis had to change as well.

The development of this new generation of research was strengthened by developments in the analysis of hierarchically nested data. Such data are very common in the field of education, where students are nested in classrooms, classrooms in schools, and schools in school systems, regions, counties, and countries. The transition to this new type of research was motivated to a large extent by recent research in the fields of economics and sociology that considered the merits of the public versus the private sector in the USA; this policy-oriented research has renewed interest in the influence of schools and their organization on learning and academic achievement. The research has renewed to some extent the notion that homogeneity of school populations may have a differential effect on student achievement. In an excellent and complete overview of the developments in Dutch educational attainment research, Dronkers noted, "The study of these school effects is one of the most promising topics of the current Dutch educational attainment research" (1989, p. 13). As Dronkers said, "The introduction of multilevel analysis models [allows] the precise estimation of these school effects together with the individual effect" (1989, p. 13).

In the American literature the emphasis shifted from instructional hardware (see Coleman, et al., 1966) to the organization of instructional activities within educational institutions. An example is research into the effects of curriculum enrollment and its consequences for student achievement and self-respect (see Alexander & McDill, 1976; Alexander, Cook, & McDill, 1978; Anderson, 1982; Oakes, 1985; Lee & Bryk, 1988). Although the results of this empirical research were not conclusive, several suggestions were made that are worth investigation. Some results indicated that school bureaucracy had a minimal effect on students' aspiration and that the type of school had little influence on student achievement. Rather, the curriculum or type of program in which the student was enrolled had a high influence on aspirations and achievement. Other research reported that there was no relationship at all between homogeneous grouping and achievement (see Camoran & Mare, 1989, for an overview).

Research into the effects of selection, as well as research into the private and public sectors, by Coleman, Hoffer and Kilgore (1982), opened up renewed interest in the investigation of school effects. It is remarkable that the same researcher who figured prominently in the demise of the first generation of school effectiveness research (see Coleman et al., 1966) became an important stimulant for the second generation of school effectiveness research with a similar large national study (Coleman, Hoffer, & Kilgore, 1982). The 1982 study investigated the differences in achievement between students in public and in private schools. The discussion and reanalysis of the same data that followed this study (e.g., the discussion in the issues of *Sociology of Education*, 1982, 1983, and 1985; and the secondary analysis of Raudenbush & Bryk, 1986; and Lee & Bryk, 1988) set off the new type of research.

What these studies left largely unattended, however, were the possible determinants of the observed differences in school climate between, for instance, public and private schools. In the remainder of this paper, we address this problem in a series of four steps: First, two theories are considered, with competing explanations of what causes effective school climates. By comparing results of research into selective school systems in the UK and research into the sector differences (public versus private schools) and the effects of tracking in the USA, a hypothesis is developed. Second, the recent developments in the analysis of hierarchically nested data then are described in relation to the usefulness for school effectiveness research. Third, the analysis of Dutch school data is reported; the analysis supports the theory that selection mechanisms do contribute to student achievement. Last, a generalization is made by comparing the outcome of the Dutch analysis with results from studies that examine school systems in the USA and the UK.

What Makes Schools Effective?

Within-school stratification such as that produced by tracking can be compared with the between-schools stratification that exists between private and public schools in the USA. In the Netherlands and the UK this between-school variation is demonstrated in a secondary school system that uses achievement test scores as a basis for placing students into different types of schools. Although higher achievement has been connected to private schools in the USA, it still has not been determined whether the results are due to a better school climate that is based on consensus and shared morality within the school (as was suggested by Salganik & Karweit, 1982; Coleman et al., 1982; and Bryk, 1988) or to the selection process itself, as research on tracking suggests (Camoran & Mare, 1989). If the first theory is true, then a better organizational design could help public schools create an equally stimulating school climate by incorporating more order and discipline into the schools and encouraging more consensus between teachers and parents and among teachers themselves. However, if the latter theory is true and the selection mechanism itself creates a better climate by selecting better motivated and smarter students, stricter rules and more discipline in the school cannot be expected to change the school climate. Some selection processes can stimulate learning, but others can hinder it, as research on tracking shows.

Figures 1 and 2 illustrate the two competing theories that focus on the differences among schools in the USA. In these figures the direction of causality is vertical, from top to bottom.

Figure 1
Student Body Hypothesis

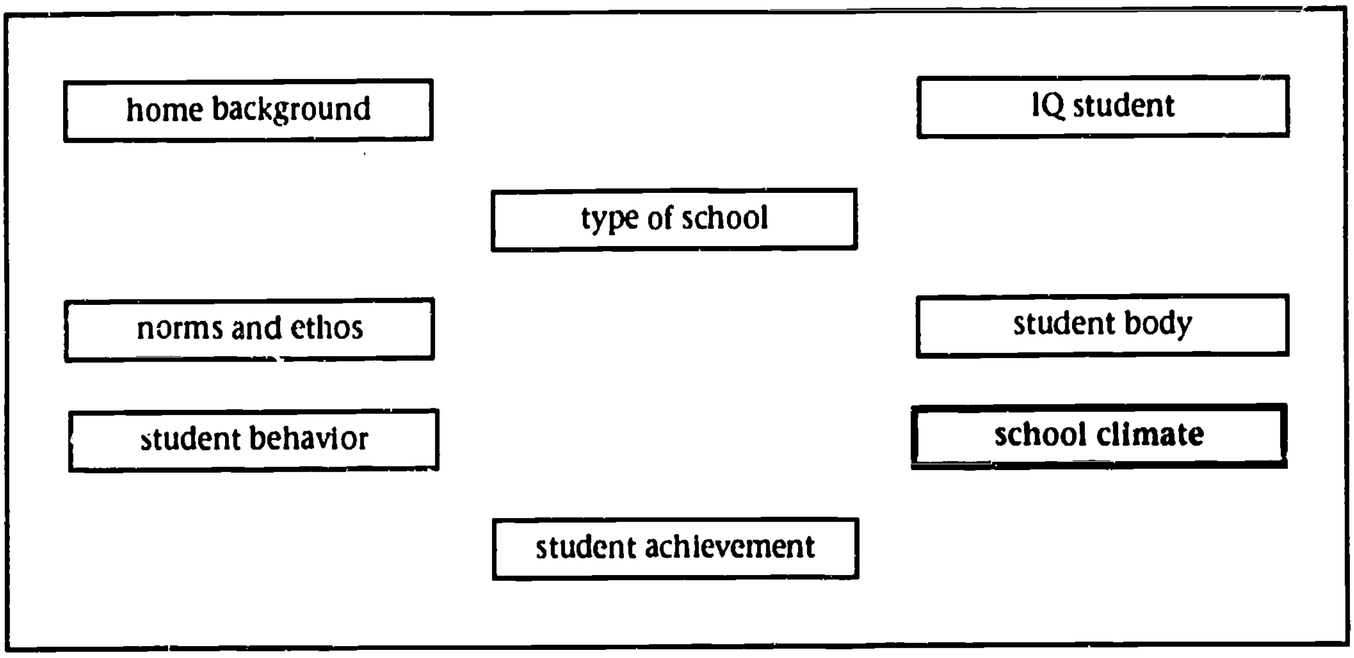
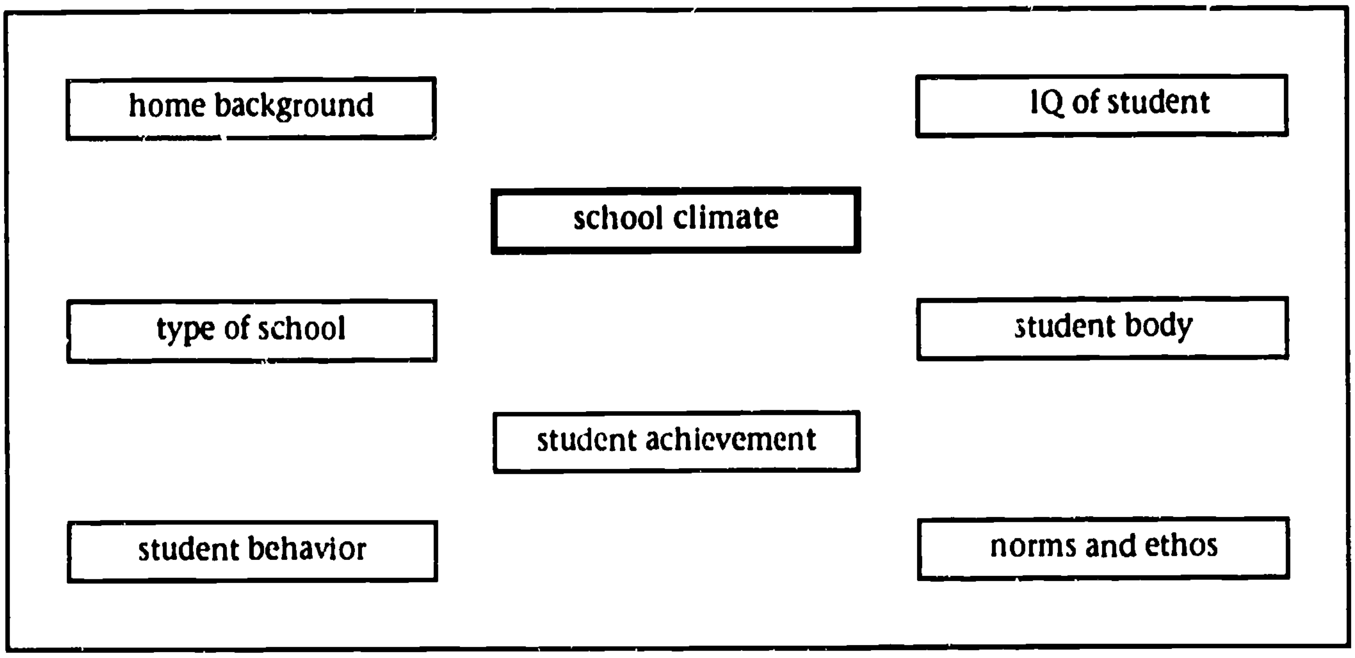


Figure 2
School Climate Hypothesis



Direction of causality is from top to bottom in both figures

Figure 1 is the student body hypothesis. This theory purports that with different student populations, the type of school will determine the climate, and hence will have an effect on norms and ethos and on individual student behavior and achievement. This theory assumes that the tendency of private schools to select students who are more motivated and have higher socioeconomic status (SES) in itself generates the better results shown by students in those schools. Since private schools are relatively expensive in the USA, it is quite likely that the parents who send their children to private schools are greatly concerned about the quality of their children's education; this extra consideration may influence the achievement of their children. As a result of selectivity, the student populations in private schools will be different from those in public schools.

Figure 2 shows a different pattern. The school climate hypothesis suggests a direct effect of home background and school climate on student behavior and achievement. This theory assumes that school climate is fostered by the morals and attitudes of parents and teachers in private schools, whom often are considered to be more outspoken and in agreement (particularly in schools with religious affiliation) than parents and teachers in public schools.

While in theory the distinction between these two models is clear, they cannot be separated in a data analysis without further information. Determining the causal ordering of the variables, in the way the two models do, is a theoretical, not a methodological, problem.

Research in the USA and the Netherlands shows that student achievement and the socio-economic status of the parents are closely related in selection processes. In the USA, Alexander et al. (1978) found evidence that curriculum placement is, for a large part, subjective, since only 40% of the variance of this placement is explained by ability, former achievement, and aspiration of the students. A disproportionate number of students with higher status are channeled into college preparatory curricula, a situation quite similar to that in the Netherlands. Placement into the several forms of secondary education in the Netherlands is based on achievement test scores and also on the recommendation of the elementary school principal (via the advice given by teachers to parents to help them choose suitable types of secondary education for their children). Students whose parents have a higher socio-economic status are more often advised to attend a secondary school with higher status, irrespective of their achievement scores. This advice is important because approximately 87% of the parents follow it (Herpen & Smulders, 1980) and because the admitting schools value it. The overall conclusion is that selection in both countries takes place on the basis of the achievement of the student and also on the basis of the socio-economic status of the parents.

A similar relation shows up when research in the UK is explored. Steedman (1983) reported that 57% of the students in the highly selective and high-achieving grammar schools are from parents with non-manual occupations; 27% of these students attend comprehensive and secondary modern schools. Gray, McPherson, and Raffe's (1983) research in Scotland reports an interaction effect of sectors and student ability. They conclude that selective schools increase the distance in achievement between working-class and middle-class students.

In the USA, in a different type of study, Lee and Bryk (1988) showed that achievement in (selective) Catholic private schools is indeed higher—72% of these students end up in academic tracts, compared to 38% of students in public schools. The background characteristics of the student populations also differ: the mean family income is higher in Catholic schools; fewer Blacks and Hispanics (13% in Catholic schools, compared to 23% in public schools) are enrolled; and SES (measured by parents' education) is also higher (l.c. p.81).

The Netherlands and the UK have comparable school systems. In both countries selectivity is used actively to restrict the opportunities of students who are in the lower parts of the achievement continuum: These students are kept out of the grammar schools in the UK and out of the Gymnasium/Lyceum in the Netherlands (these most selective forms of Dutch secondary education serve only 7.6% of the total student population). The way selectivity is activated in the USA is different, but the effects may be similar. Here SES and income are the prime selection factors. The role of the school in the selection process appears to be relatively passive in the USA when compared to the same phenomenon in the Netherlands and the UK. However, Catholic schools serve only a small part of the total U.S. student population in secondary education (7.7%, as reported in Catterall, 1983). Since socio-economic status and achievement are interacting factors, we expect that the passive selectivity in the USA has a comparable effect as the active selectivity in the Netherlands and the UK.

By examining the selectivity of the Dutch secondary education system, we are able to investigate the effects of selection on the student population in the Netherlands. In fact, the Dutch system allows us to investigate both theories described above (see Figures 1 & 2). Research into the differences between religious and non-religious schools, comparable with Coleman et al.'s (1982) investigation, may test the effects of differences in morals and school climate, as influenced by religious or secular attitudes. Dutch results of this type of investigation were reported by Van Laarhoven, Bakker, Dronkers, and Schijf (1987). The results of their study showed that the distinction in achievement between schools with different philosophies is not so clear cut in the Netherlands as it is in the USA. An explanation may be that religious and private schools are free of charge in the Netherlands, which makes them less selective in the direction of income compared to the situation in the USA. For the same reason, the concept of "private" versus "public" means something else in the Netherlands since all schools are publicly funded. This explains why 62% of Dutch secondary students are in (publicly funded) private and religious schools (Van Laarhoven et al., 1987), compared to 7.7% of high school students in the USA (Catterall, 1983).

Multilevel Analysis as a Useful Tool in School Effectiveness Research

The second generation of school effectiveness research also looks for new ways to analyze data. The widely used multiple regression and individual level path models are being replaced by new ways of analyzing the hierarchically nested data that researchers deal with in this type of study. Coleman et al. (1982) introduced the idea of multilevel analysis by using separate multiple regressions for different grades and sectors and finding a way to compare these different analyses. An elegant solution was provided by Raudenbush and Bryk (1986) when they analyzed the same data set. This multilevel approach has several advantages. It offers a means of investigating Cronbach and Webb's (1975) claim that aptitude can be measured by treatment interactions. The multilevel approach also can deal with the question: Do some schools have good results for special groups of students and not for others? Some of the results of the comparisons between public and private schools seemed to point in this direction: Lower SES and black students profited more than others from the stimulating climate in the private school (Lee & Bryk, 1989), which is a clear interaction effect.

When researchers in education investigate how bureaucracy or school organization affects student learning, they use data that are, in general, hierarchically nested, as is the case of Coleman et al.'s data. The most commonly used data are from students nested in schools or classes; schools also may be nested in sectors (public versus private schools) or in states. Most data in educational research have these multilevel characteristics.

In the first generation of school effectiveness research, methodologists had to overlook the problems inherent in the analysis of multilevel hierarchical data when using traditional statistical analysis methods such as multiple regression or path analysis. These single-level models are problematic when inferences to other levels are made. Discrepancies occur when the data are either disaggregated (for example, school characteristics are disaggregated to student level) or aggregated (student characteristics are aggregated to school level). The first Coleman report (1966) used regression analysis with school characteristics disaggregated to the individual level. Problems with this type of analysis occur when researchers make inferences—either from student level to school level or the other way around—that are based solely on student-level analysis; this problem was illustrated by Robinson as early as 1950. In the High School and Beyond study (Coleman et al., 1982), the final conclusions were based on an analysis at the sector level (public versus private schools). Even if, as this study demonstrates, Catholic schools in general show better results than public schools, it still may be misleading to advise parents to choose a private school over a public school because this cross-level inference from sector to school level may be not valid.

The disappointing findings of the first generation of school effectiveness studies (e.g., Coleman et al., 1966) may be partly a result of the way multilevel data were analyzed with traditional single-level models. Multiple regression and path analysis were and still are the most widely used methods. Using analysis of (co)variance instead of single-level methods has disadvantages too; for instance, it cannot deal with the interaction between individual- and group-level data. Raudenbush and Bryk (1986) succinctly summarized the problems with traditional models when they concluded that school research data were analyzed in ways that concealed more than they revealed. The obvious mismatch between the way data are collected and organized and the analysis techniques used was recognized by many researchers. The mismatch worried them enough to spur a search for better tools and new ways of analyzing multilevel data (Goldstein, 1984; Longford, 1987; Bryk, Raudenbush, Seltzer, & Congdon, 1988). The analysis methods that have been developed for multilevel data promise that the results generated by the second generation of school effectiveness research will be meaningful.

The Development of Methods for Multilevel Analysis

Although appropriate methods for analyzing hierarchically structured data in education have been available for some time, the absence of computationally efficient algorithms has prevented an application of these methods. The development of an estimation procedure for hierarchically nested data that would be based on the early work of Lindley and Smith (1972) was problematic because of the mathematical complexity of their proposed Baysean estimation procedure. Dempster, Laird, and Rubin (1977) composed a numerical approach to maximum likelihood estimation of covariance components using the EM (expectation maximization) algorithm. This method produces maximum likelihood estimates for variance components with known large sample properties.

Use of the new hierarchical linear (random coefficient) model rather than the traditional linear model has several advantages. The best way to describe this new technique is to say that it performs multiple regressions within each context (school or class) separately. Different models then are fitted for different contexts in the first step and compared over contexts in the second step. For example, in Coleman et al.'s study (1982) several multiple regressions were performed for grades and sectors. The means of one sector were fitted into the equation for the other sector in an attempt to find significant differences between sectors. The intercepts and/or regression coefficients also were compared between sectors, to examine differences in discipline, homework assignments, and student behavior. In a sense this also was aggregation, but it was model-based aggregation.

The newly developed multilevel data analysis models are conceptually based on the same principles, but instead of estimating different fixed effect models for sectors or schools, a single estimation procedure is employed that can deal with variables from both levels simultaneously. These procedures compare the aggregated coefficients while taking the standard errors of the coefficients into account. Another improvement in this analysis method is the weighting procedure for unequal numbers of students in the classes contained in the sample (or the schools in the sample). The presence of unequal numbers makes debatable the assumption of equal error variances for all groups. The fact that students within classes (or schools) are not independently sampled, and as a result have correlated error terms, is another problem that is solved in the new models. The error term in regression models is assumed to consist of measurement error and of unmeasured variables that influence the dependent variable in a random way. For students within the same group we may assume that the unmeasured variables (school characteristics, for instance) are not random but structural. This follows from the consideration that "natural" groups are close to each other in space and time, and as a result, students in one school are more alike than are students in diverse schools.

In addition to their link with multiple regression, random coefficient multilevel models are also related to the analysis of variance. In AN(C)OVA, group means (or corrected means, as in analysis of covariance) are compared. The result of such an analysis is simple: There is or is not a significant effect of school (second level) or sector (third level). With this kind of analysis it is possible to test whether schools (or sectors) differ significantly from each other, but it is not possible to determine what causes them to be different. This analysis model has additional problems. In ANCOVA models it is assumed that the estimated slope for the within-group effect is equal over all groups; thus, the slope is fixed or constant. As a result, in cases where slope heterogeneity exists, the estimated regression coefficients are imprecise and inefficient. Tate and Wongbundit (1983) were among the first to recognize this. They argued that for multilevel analysis in educational research, random coefficient models with random slope and intercept are more appropriate than random coefficient models with only random intercepts. The assumption of random effects allows the researcher to make inferences to groups (schools or classes) not included in the sample.

Random Coefficient Models for the Analysis of Hierarchically Nested Data

In a multilevel analysis all levels are recognized and analyzed in relation to one another. We will demonstrate that it is possible to analyze students within schools and schools within sectors together without losing the distinction between the levels. Inferences then can be made to school and student levels. We need models that can separate the processes within different groups and at the same time compare the results over groups.

For example, if we have data with two levels such as the Dutch data set discussed below, the first level could be the student level. Independent variables would include gender and achievement test scores; the dependent variable would be school career. The second level could be the school level; school type would be the second level variable. Our first analysis might be an analysis using a random slope for gender, for test scores, and for the interaction of gender with test. For this example, the equation is of the following form (we use bold face notation for random variables):

$$Y_{ij} = \mathbf{a}_j + \mathbf{b}_{j1}X_{i1} + \mathbf{b}_{j2}X_{i2} + \mathbf{b}_{j3}X_{i3} + \mathbf{e}_{ij}, \tag{1}$$

where: the index i is used for individuals; the index j is used for schools; X_{i1} is test score; X_{i2} is gender; X_{i3} is the interaction between gender and test. The random variables e_{ij} are the individual error terms; a_j is the random intercept; and b_{j1} to b_{j3} are random slopes. We consider all these coefficients not only as functions of the individuals, but also as functions of the different types of schools in the following way:

$$a_j = a_t + g_j \tag{2}$$

and

$$b_{jr} = b_{tr} + h_{jr} \tag{3}$$

Here, index t is used for the type of school (the Dutch data set includes five types). Index r is used for variables; thus r is between one and three. Equation 3 says, for instance, that the random slope b_{j1} for the achievement test scores is composed of a fixed type specific slope b_{t1} and a disturbance h_{j1}. Thus g_j is the school level-disturbance of the random intercept, and h_{jr} is the disturbance of the random slope for variable r on school j. If we substitute (2) and (3) in (1), we obtain the mixed linear model:

$$\begin{aligned} Y_{ij} &= (a_t + g_j) + (b_{t1} + h_{j1})X_{i1} + (b_{t2} + h_{j2})X_{i2} + (b_{t3} + h_{j3})X_{i3} + e_{ij} = \\ &= (a_t + b_{t1}X_{i1} + b_{t2}X_{i2} + b_{t3}X_{i3}) + (g_j + h_{j1}X_{i1} + h_{j2}X_{i2} + h_{j3}X_{i3} + e_{ij}). \end{aligned} \tag{4}$$

Equation 4 shows how the fixed and random parts of the model are combined. (For more details, see Aitkin & Longford, 1986; De Leeuw & Kreft, 1986; Raudenbush & Bryk, 1986). The equation shows that the model allows us to test the effects of a second level (e.g., school type) on relations between within-unit variables (e.g., gender and/or test scores). Two separate effects can be estimated: an overall effect (an effect on the intercept of the within-unit model), and an interaction effect (an effect on the slopes of the within-unit model). The model also provides a test to determine whether the variance of the parameters at the second level differs significantly. A deviance score allows the researcher to ascertain if a simpler model within a full model is appropriate.

A Multilevel Analysis with Dutch Data

The Dutch organization of secondary schooling resembles the English school system. Vigorous arguments over the merits of selective versus non-selective schooling started shortly after the Second World War in both countries, and their current systems are the result of laws created in the sixties to change secondary education from a very restricted and selective system to one that is less selective: The Mammoth Law (*Mammoet wet*) has been in effect in the Netherlands since 1968; Circular 10/65 was adopted in England in 1965. In both countries the aim was to provide equal educational opportunities for all social classes of students. We see this real life experimentation as an opportunity to test the theory of the effects of selectivity.

A Description of the Data

In general terms our study of the Dutch secondary school system examines whether there is a relationship between the school careers of students and the structure of the secondary school system as it was organized after implementation of the Mammoth Law. School organization is defined by considering school selection (i.e., the way schools recruit students, in either a selective or nonselective way). Selective recruitment takes place in the first year of secondary schooling and is based on a student's ability as determined by scores on achievement tests. Non-selective recruitment means that the school admits all students to the first year of secondary education (usually around the age of twelve) independent of their achievement scores. We want to know if the nonselective system fosters greater equality of opportunity than does the selective system. Since all schools in the Netherlands are free of charge (including Catholic, Protestant, Montessori, and other kinds of private schools), no distinction between private schools and public schools exists in that respect. Even the poorest parents can send their children to a private school. Since the Mammoth Law came into effect, the non-selective system, or comprehensive system (*scholengemeenschap*), exists in conjunction with the old selective system.

Our sample contains 5,310 students in 70 secondary schools in Amsterdam; these schools are divided into five types, based on selectivity (see Appendix for a description of school types). The comprehensive (*scholengemeenschap*) schools are the only non-selective schools in Amsterdam. These schools are open to all students and offer either an administrative (MAVO), a college prep (HAVO), or a university prep (VWO) diploma at the end of four, five, and six years, respectively. These schools are non-selective in the way they recruit their students, which does not mean that no selection within the system takes place. The number (m) of comprehensive schools is 6; they have a total enrollment (n) of 1,060 students, or 19.5% of Amsterdam's population of secondary students. The other four types of schools are selective schools. These schools are from the original system; they existed before the Mammoth Law came into effect. They are: (a) administrative schools (MAVO), m=42 and n=2038, or 38% of the population; (b) college and university prep schools (VWO/HAVO), m=13 and n=1440, or 27% of the student population; (c) Gymnasium/Lyceum m=5 and n=397, or 7%; and (d) a miscellaneous group of schools whose curricula range from vocational programs to a five-year course of study (MAVO/LBO and MAVO/HAVO), m=4 and n=375, or 7%.

The longitudinal data were collected over consecutive years from 1975 to 1981. A comparison between the student population in Amsterdam and a representative sample of the student population of the whole country in 1975 shows that there was not much difference in the percentage of students within each type of school (see Table 1).

Table 1
Percentage of Students in the Five Types of Dutch Secondary Schools

	Amsterdam	The Netherlands		
	1975	1975	1983	1986
Comprehensive	19.5%	20%	24%	28%
Administrative (MAVO)	36.5%	30%	30%	27%
College and university prep (HAVO/VWO)	29.0%	30%	30%	30%
Gymnasium/Lyceum	7.6%	3%	3%	3%
Miscellaneous	7.4%	4%	4%	8%

Table 1 shows how students were distributed into the different types of schools. The first two columns show percentages for the population under study and the percentages for students in the Netherlands in the same year. The next two columns show the change that took place in the Netherlands in the years 1983 and 1985. (Because some schools are not considered, the percentages do not always add up to 100%.) When we compare the two columns for 1975, it is clear that the percentages of students in the five school types in Amsterdam were almost equal to the percentages that represented the total Dutch student population during that year. When we compare between years, we see that the greatest changes occurred in the percentage of administrative students, which diminished, and the percentage of students that attended comprehensive schools, which increased.

The Dependent Variable: School Career Over Six Successive Years

Since parents in the Netherlands do not have to send their children to a designated neighborhood school, they are free to choose between the selective and non-selective systems. The type of secondary school that students enter within each system depends on the choice of the parents, if the school is non-selective, or on the results of the achievement test that is administered at the end of elementary school and the recommendation of the principal of the elementary school, if the school is selective. In Amsterdam, the consideration of the distance to either type of school is rarely a factor. Therefore, the composition of the student body is more a function of the interaction of school and student background characteristics and less of accidental circumstances such as neighborhood and/or tuition costs.

The quality of schooling differs according to the curriculum that is offered. In each type of secondary school, all students at the same level follow the same curriculum. Differences among curricula can be found in pacing, in intensity, and in the extent to which course content is covered. The more a curriculum demands in terms of quality and quantity, the larger the total number of years of schooling is, varying from three to four years (in administrative schools) to six years (in pre-university schools, Lyceum, and Gymnasium). A student's educational status can be clearly identified by an indication of his school level and system. MAVO3, for example, indicates that the student has attained the third level of the MAVO system. Schools of all types require their graduates to take a centrally organized exam that is administered nation-wide at the end of the course of study. Students who fail this exit exam must return to school for an entire year before they may retake it.

The selective schools offer a single curriculum, either administrative (MAVO), college prep (HAVO) or university prep (VWO or Gymnasium/Lyceum). The non-selective comprehensive schools are more or less modeled after the American high school, except that no vocational training is offered.[2] Comprehensive schools offer more than one curriculum within a school. For example, a student enrolled in a comprehensive school can choose to study the college prep (HAVO) curriculum or the university prep (VWO) curriculum. The main difference between selective and non-selective schools is that, in the latter, selection takes place in a later stage. In addition, switching from one curriculum to another, whether more or less difficult, is easier within non-selective schools. A student who wants to change curricular focus within the selective system must go to a different school in a different location. In general, students are most likely to move to a less demanding curriculum before the exit examination. Moving to a more demanding curriculum takes place primarily

[2] Vocational training is offered in a totally different system (LBO, see Appendix). Changing from schools in this system to any other is hardly possible, unless the student starts in the other system from the beginning. In this study vocational training is not considered.

after a student completes the exit exam for one type of school. Students who change to a higher level curriculum must complete one more year of education than do students who start at this level from the beginning of their career in secondary education.

Since the path taken through education is so clearly marked by level and system in the Dutch secondary schools, we can use the data for individual students measured over six successive school years as the indicator variable "school career" for the differences among schools and the differences between the selective and non-selective systems. These measurements of school career will indicate the kind of curriculum each student was studying for each of the six years. To give some idea of what this means, we report the school career for the entire cohort.

In the first year (1975) of the study, information was available for 7,444 students. After the fourth year (1979), when most of the students were 16 years of age and schooling was no longer obligatory, 2,358 of these students (32%) left secondary education, some to start working, some to go to post-secondary training. After the six year (1981), 77% of the students were no longer in secondary education. Some students entered post-secondary training (colleges or universities), others started working. Table 2 summarizes the school careers of students in each type of curriculum as they neared the end of the last required year of coursework.

Table 2
Status of Students' School Careers
During the Final Year of Each Curriculum Type

	Students in Administrative Schools—Year 4	Students in College Prep Schools—Year 5	Students in University Prep Schools—Year 6
Enter exam class without delay	38.0%	13.8%	33.0%
Lag behind	33.0%	56.0%	32.7%
Drop out	18.5%	16.6%	18.8%
Move to lower curriculum	9.5%	11.5%	14.0%
Move to higher curriculum	1.0%	2.0%	—

Administrative schools (MAVO) require four years of education before the exit exam. In Table 2 we see that after the four years only 38% were enrolled in the exit exam class. Of the rest, 18.5% dropped out and 9.5% went to vocational training (LBO). Only 1% continued school at a higher level of secondary education (HAVO, VWO, or Gymnasium/Lyceum). College-prep schools (HAVO) had a lower success rate. Only 13.8% were in the exam class after completing the mandatory five years of schooling. Most students (56%) lagged behind. Of these, 16.6% dropped out, 11.5% moved down to administrative schools (MAVO), and 2.0% moved up to university prep schools (VWO or Gymnasium/Lyceum). The course of study in university prep schools spans six years. The dropout rate after six years was not much different from the other types of schools: 18.8%. We see that 32.7%

lagged behind and 13.5% moved down to the college prep schools (HAVO); only 0.5% moved to the administrative schools (MAVO). Thirty percent of the students in the sample who started in administrative schools (MAVO) went on to a college or university prep program after obtaining an administrative diploma, and 18% of the students who started in college-prep schools (HAVO) went on to university prep schools (VWO or Gymnasium/Lyceum) after finishing their initial education.

The Use of Homogeneity Analysis to Construct the Dependent Variable

Since the school career of an individual student may differ insofar as the type of school the student attends (or type of curriculum within the school) and the number of years the student needs to finish a certain type of program, many combinations are possible. The path taken through the different school types in combination with the time needed provided us with over 800 possible variations.

We must construct a single dependent variable from all this information in order to compare the outcomes among the five types of schools and between the selective and non-selective schools. This variable must contain all information about success or failure in the secondary schooling of the student. Degree of failure is defined in different ways: dropping out, ending at a level lower than that started, or lagging behind. Success is defined as earning a diploma without delay or ending in a higher level than originally started. Different diplomas have different meanings, because the highest diploma—for university prep schools (VWO, Gymnasium/Lyceum)—gives entrance to all colleges and universities, and the lowest diploma—for administrative schools (MAVO)—leaves the student with less choice for post-secondary education.

The constructed variable "school career" has a high score for a university prep diploma and a low score for a administrative diploma. In general, the highest possible score represents receiving the highest type of diploma in the fewest possible years. The lowest possible score represents receiving the lowest type of diploma in the maximum number of years or leaving school before earning a diploma. In terms of comparison, this means that the schools with the highest scores for the dependent variable are those that are successful in keeping their students in school (lower drop-out rates) and at the same time are successful in letting their students pass the highest exit examination in the shortest possible time.

To analyze our data we used the homogeneity program HOMALS (distributed by SPSS, 1989) described in Gifi (1989). Homogeneity analysis, or multiple correspondence analysis, is quite useful for quantifying school careers. The main advantage of this program is that only a relatively small number of prior decisions have to be taken, although it remains true, of course, that the results are dependent on the coding of the states that is chosen. But, given the coding, the quantification follows automatically. Another advantage is that the average scores of persons in state k at time t can be used as the state k quantification at time t. This means that we can replace the original data matrix, which has a row for each of the n students and a column for each of the m time points, with a quantified data matrix in which each label is replaced by its quantification at that particular time point. The scores for the individual careers are row averages of that matrix, and we obtain perfect homogeneity if the elements within each row are the same (i.e., if the career score of each individual is constant over time). The results of homogeneity analysis are based almost completely on the structure of the student's career, and actually on general tendencies in the structure of all careers in the cohort. The quality of the career is not taken into account: The program uses the labels of the categories, so no prior ordering of these categories is necessary. The ordering, and thus the quality, is dictated by the structure in the data. The eigenvalue of the first dimension is 5.64, which means that the first factor explains 95% of the variance.

The correlation between variables measured in years 1 through 6 ranges from .90 to .98. The highest correlations are between succeeding years, and the lowest are between variables several years apart (for instance, the correlation between the first and the sixth year is, after transformation, .92). For more details on the construction of the variable "school career" we refer to Kreft (1987).

The Explanatory Analysis Variables

To analyze the data we use a random coefficient model. We start with a multilevel analysis in which the first-level individual variables are: (a) standardized achievement test scores for language and arithmetic (CITO); (b) gender; and (c) the interaction between gender and CITO test scores. CITO is a standardized test on achievement in reading, writing, and arithmetic (see Appendix for more information). It is the primary predictor, and criterion as well, for the selection of a secondary school. The second-level variables are the means for the CITO test scores for each of the five types of schools.

In Table 3 we see the differences among the school types when means and measures of spread are compared. The mean and the variance of school career is given with the mean and standard deviation of CITO test scores for each of the five different school types. The table also shows the number of schools and the number of students in each type.

Table 3
Differences in CITO Scores and School Careers for
Students in the Five Types of Dutch Secondary Schools

	CITO Mean	Career Mean	CITO SD	Career SD	School n	Student n
Comprehensive	.04	-.09	.70	1.13	6	1060
Administrative (MAVO)	-.60	-1.02	.70	.22	42	2038
College and university prep (HAVO/VWO)	.65	.77	.32	.97	13	1440
Gymnasium/Lyceum	.81	1.45	.32	.42	5	397
Miscellaneous	-.28	-.43	.70	.71	4	375
Total	-.006	.12	.88	1.30	70	5310

The mean CITO test score is shown in the first column of Table 3; the standard deviation of the same test is shown in the third column. Adding and subtracting two times the standard deviation from the mean gives the interval in which 95% of the population of each type of school lies (assuming normal populations). The results of this computation show that comparable students with the same test scores can be found in all types of schools (although in unequal numbers). In the comprehensive schools 95% of the population has CITO z-scores between -1.68 and 1.72. Compared with the administrative schools, in which most students score between -2.28 and 1.08, it is clear that the comprehensive schools serve a higher range of the ability continuum. The college prep schools serve a higher but also smaller range of students than administrative and college prep schools (95% of these scores fall between -.49 and 1.79). The pre-university schools serve the students with the highest range of test scores; the majority of the students have scores between -.33 and 1.95. Table 3 also shows that schools in the non-selective system, the comprehensive schools, serve the most representative student population by having a mean z-(test)score of around zero.

A Multilevel Analysis

The analysis starts with a multiple regression in each of the seventy schools, with gender and CITO scores, the independent variables, and the status of students' school career as the dependent variable. The contextual variables are the type of school and the mean CITO score for each school. After completing the individual-level analysis within the 70 different schools, the contextual effect is tested on both the intercept (overall effect) and on the slopes (e.g., the interaction effect type by CITO scores). Use of the intercepts from the within schools analysis as the dependent variable in the second-step contextual analysis is the same as investigating whether the overall effect among the types of schools or among individual schools is significantly different. Use of the slopes from the within-schools analysis as dependent variables is the same as investigating, for instance, whether an interaction exists between the type of school (or type of system) and the variable that is related to that slope, which is in our case the CITO test score. (Compare equations 1 through 4 presented earlier in this paper). Using the slopes for CITO scores as the dependent variable in the second step is the same as testing for interaction effects between type of school and type of student. Through this multilevel analysis we can determine whether schools are more effective for students who have higher scores on the CITO than for those who have lower scores.

Results

The first variance component analysis is the most extensive one. By using all the available variables at both levels a model is fitted with random slopes for gender, CITO score, and the interaction of gender with CITO. The two second-level variables, the mean CITO score for each school and school type, are tested for a total effect on intercepts and for an interaction effect on the slopes of gender and CITO. The mean CITO score is included here to determine if the effect of selectivity can be explained by school type only, or if the school mean of the achievement score also has an effect. The results of this first analysis with both second-level variables (see Table 4) show that the effect of the variable "type of school" is large, but that the mean CITO score has no significant effect as a second-level variable. The coefficient is .0123, with a standard error of .0418 (see Model 3 in Table 4).

Table 4
Random Coefficient Model:
Fixed Effects with Standard Errors,
Using Slopes for CITO

Variables	Model 1 Fixed Slopes Without CITO Mean		Model 3 Fixed Slopes With CITO Mean	
	Estimates	SD	Estimates	SD
Grand Mean Score	-0.1355		-0.1400	
Gender	0.0963	0.0177	0.0975	0.0176
CITO Mean Score			0.0123	0.0418
The Interaction between CITO and Type of School (Fixed)				
CITO x Type 1 (Comprehensive)	0.7281	0.0240	0.7257	0.0371
CITO x Type 2 (Administrative)	0.0808	0.0180	0.0811	0.0190
CITO x Type 3 (College and university prep)	0.7635	0.0315	0.7680	0.0609
CITO x Type 4 (Gymnasium/Lyceum)	0.2975	0.0572	0.2988	0.0591
CITO x Type 5 (Miscellaneous)	0.4066	0.0427	0.0411	0.0538
Effect on the Random Intercept				
Type 1: (Comprehensive)	0.0000	0.0000	0.0000	0.0000
Type 2: (Administrative)	-0.9853	0.0792	-0.9826	0.0855
Type 3: (College and university prep)	0.2714	0.0899	0.2599	0.0975
Type 4: (Gymnasium/Lyceum)	1.1825	0.1186	1.1830	0.1268
Type 5: (Miscellaneous)	-0.3598	0.1210	-0.3483	0.1300

The lack of an effect of the school CITO mean, as a contextual effect next to school type, also can be shown as the difference in fit of the model with and without school mean. This is shown in Table 5. The difference between Model 1 (without school mean) and Model 3 (with school mean) has a chi-square of 0.13 with one degree of freedom. From this we conclude that including the CITO mean does not provide a significantly better fit of the model.

Table 5
Test of Models 1, 2, and 3 for Goodness of Fit

Model	Context	Intercept	Slope	Number of Parameters	Deviance
Model 1	Type	Random	Fixed	13	10354.21
Model 2	Type	Random	Random	23	10325.95
Model 3	Type + School CITO Mean	Random	Fixed	14	10354.08

Chi-square difference for Models 1 & 2: 28.26 with 10 dfr

Chi-square difference for Models 1 & 3: 0.13 with 1 dfr

An analysis with random slopes for gender and CITO shows that, next to test scores, gender has a significant effect on school success: Girls with the same test scores as boys have more successful school careers. This effect is the same for all school types, since the gender has no significant random effect. The interaction between test scores and type is highly significant at the fixed level and barely significant at the random level (i.e., the b_{t1} for CITO score are significantly different from each other and from zero). The difference in fit between Models 1 and 2 in Table 5 above shows that the model with random slopes is not a large improvement over the simpler model with only a random intercept.

As the table indicates, a deviance of 28.26 with ten degrees of freedom shows a (barely) significant difference between models. With 5,310 pupils and 70 schools, this difference is so small that we can assume that the fit of the model does not suffer much if we use the simpler model. As the outcome of the first analysis suggests, a model can be fitted with only the intercept random and with the slope of CITO fixed but variable. The model with fixed but variable slopes for CITO and an interaction effect of type of school provides a good model. Since the interaction effect is significant, it results in different slopes for each type of school for CITO test score on career, as is shown in Table 6 and Figure 3.

In sum, we found that a model with a random intercept, a fixed but variable slope for test scores, and a fixed constant slope for gender is an acceptable model. In formula notation, the equations 5a and 5b Model 1 (the fixed slope model) and Model 2 (the random slope model), respectively:

$$Y_{ij} = a_t + b_{t1}X_{i1} + b_2X_{i2} + \mathbf{g_j} + \mathbf{e_{ij}}. \tag{5a}$$

and

$$Y_{ij} = a_t + b_{t1}X_{i1} + b_2X_{i2} + \mathbf{g_j} + \mathbf{h_j}X_{i1} + \mathbf{e_{ij}}. \tag{5b}$$

When we look at the two models, we see that equation (5b) with a random slope for CITO scores is more complicated, especially in the random part (shown in bold face) of the formula.

The results of the simpler model, Model 1—which has only test scores as individual indicators, a random intercept, and fixed, but variable slopes as functions of individuals and schools—are given in Table 6 and Figure 3.

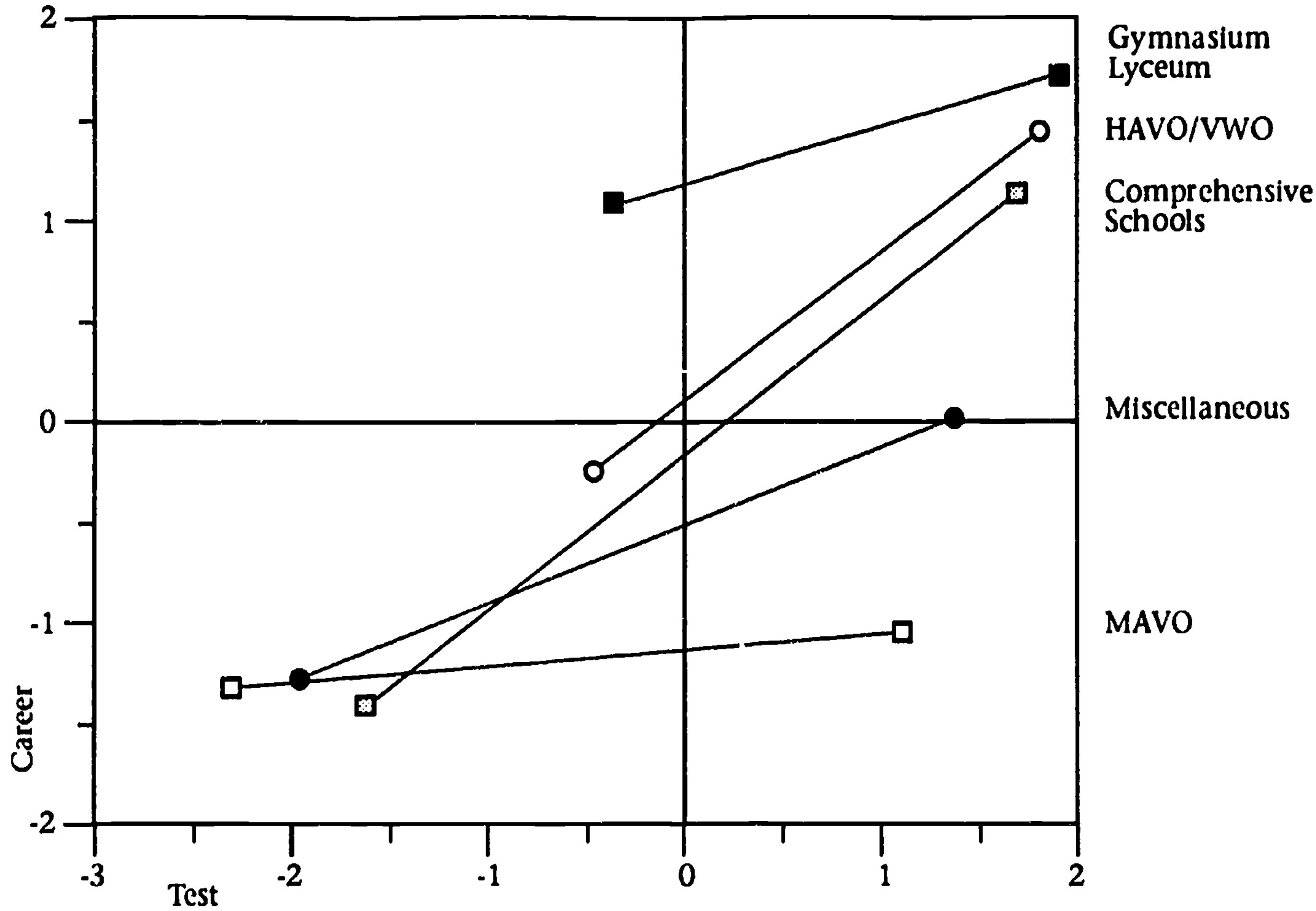
Table 6
Regression Lines for Girls:
Findings for the Five Types of Dutch Secondary Schools

Comprehensive	Y = -0.13 + .73X
Administrative (MAVO)	Y = -1.11 + .08X
College and university prep (HAVO/VWO)	Y = +0.14 + .76X
Gymnasium/Lyceum	Y = +1.15 + .30X
Miscellaneous	Y = -0.48 + .40X

Y = school career X = CITO test score

Table 6 and Figure 3 are composed by adding the effects of the two levels (see the third column of Table 4) and ignoring the coefficient for gender by giving only the slopes for girls. This way we find five different regression lines, one for each type of school. Looking at Figure 3 we see that the largest distance between regression lines is found between administrative schools (MAVO) and Gymnasium/Lyceum; this difference is seen mainly in the intercepts. It is not surprising that both types of schools have a very similar effect of achievement scores on career, since both are quite selective (even though they are selective at opposite ends of the scale) and have, as a result, more homogeneous populations. The flatter slopes for both types of schools indicate that CITO scores have less predictive power for school career here than in the other types. As soon as the student has entered MAVO or Gymnasium/Lyceum, prior achievement does not have much effect on later educational career. This may be an effect of the lack of options that these schools offer, which serves to lock a student into a certain educational career. One explanation for a flatter slope is to assume that these schools are more egalitarian; another is that students seem to be "locked up" in the selective schools. In the lowest level of secondary education, administrative schools,

Figure 3
Regression Lines for Girls:
Findings for the Five Types of Dutch Secondary Schools



this is a disadvantage, since the figures clearly show that students in other types of schools who have the same test scores or the same ability are higher in terms of career by the end of secondary school. Being locked into a curriculum at the highest type of school, the Gymnasium/Lyceum, is an advantage, since students do better here than in any other type of school.

On the other hand, the regression lines for the less selective college prep schools (HAVO) and the non-selective comprehensive schools show a different pattern. The two regression lines of these two types of schools behave in a similar way, by being steeper than the others, which means that achievement scores have more predictive power. This is not surprising, since comprehensive schools, and to a lesser extent college prep schools, offer a choice between more or less demanding tracks. What is surprising is that more students seem to choose higher level careers, and this seems to stimulate higher achievement. This effect shows up in our next analysis as well, where we compare the non-selective comprehensive schools with all the other types. Both analyses show that primary school achievement test scores are better predictors of secondary school success in non-selective (comprehensive) and less selective (college prep) schools than in the purely selective schools (administrative and Gymnasium-Lyceum).

Interpretation of the Results

Given the outcome of this research it seems reasonable to conclude that parents in Amsterdam should send their children to a selective type of school that is the most highly valued: Gymnasium/Lyceum. Since this type serves only 7% of the population and is highly selective, this choice is not always possible. If this means that parents have to make a second best choice, such as a choice between administrative (MAVO) and the non-selective comprehensive schools then the comprehensive schools seems to be the best choice. Another possibility is the choice is between comprehensive schools and college prep schools (HAVO), with the latter being the best choice (compare the lines in Figure 2).

These results are also in accord with findings in the UK, which has a comparable secondary school system. British grammar schools, which are comparable to the Dutch Gymnasium/Lyceum, select students with high ability scores; British secondary modern schools, which are comparable to Dutch administrative schools, receive the ones who fail to be admitted to grammar schools. As a result, both types of schools have homogeneous populations. Comprehensive schools admit all students without a selection process. Steedman (1980, 1983) reported that English grammar schools were better for the more able students and that comprehensive schools were better for the less able students when compared to secondary modern schools. Steedman also reported that, in a comparison with the non-selective comprehensive schools, the highly selective grammar schools had, on average, higher attainment; secondary modern schools showed lower attainment, even after adjusting for the differences in intake.

When we compare the effects of Gymnasium/Lyceum to those of Catholic schools in the USA (we can do this since both serve a small and select portion of the total student population), we see similar results. The slope of SES on achievement is flatter in Catholic schools than in public schools, leading some researchers to conclude that the Catholic school is more egalitarian. But these positive effects may be caused by the same processes that play a role in Dutch Gymnasium/Lyceum and British grammar schools. Like Gymnasium/Lyceum and grammar schools, the private or Catholic schools in the USA serve only a few low-scoring students, but these students seem to be better off than are their counterparts in public schools. In the USA the admission of students to private schools is determined by social class and race (e.g., Raudenbush & Bryk, 1986; Bryk & Lee,1988); in the Netherlands and in the UK it is determined by achievement scores. It remains to be seen if the higher achievement evidenced in private schools in the USA will continue if these schools become funded by the state and, as a result, become less selective, a situation that undoubtedly would alter current enrollments, which usually favor white students and students from higher SES groups. It is also doubtful that the Dutch Gymnasium/Lyceum and the English grammar schools would produce the same high results if they admitted all students, irrespective of their achievement scores.

A Second Multilevel Analysis

The English results also show that grammar schools and secondary modern schools combined (both are selective types of schools, but are on opposing ends of the continuum) perform better than do the comprehensive schools only in relation to truancy, behavior in school, and parental satisfaction--not in achievement. In our next analysis we replicate this study by grouping all Dutch school types that are more or less selective (all schools other than comprehensive schools) and comparing them with the non-selective comprehensive schools.

To analyze the differences between the non-selective comprehensive schools and the more or less selective schools, we analyzed the data again with the same models, but using a different contextual variable with only two categories. We know from our previous analyses (see Figure 3 and Tables 4 and 6) that the four selective types differ greatly. It is not surprising that the fit of a model in which all four selective types are put together in one category is much worse than the fit of a model which treats each category separately. This is the effect of the (untenable) assumption that all four selective schools behave in the same direction. In Table 7 the difference in goodness of fit between the two models with a random intercept but a fixed CITO slope and the same model with a random CITO slope is given for the analysis that now has system as the contextual variable.

Table 7
Test of Models 4 and 5 for Goodness of Fit:
Findings for Selective and Non-Selective Schools

Model	Context	Intercept	Slope	Number of Parameters	Deviance
Model 4	System	Random	Fixed	7	10897.80
Model 5	System	Random	Random	9	10609.76

Chi-square difference for Models 4 & 5: 288.04

From this table it is clear that fixed slopes for regression of test scores on school career is not a reasonable assumption. The best fitting model for selective versus non-selective schools is a model with random intercepts and a random slope. This is Model 5 in Table 7. The difference in deviance between the fixed and random slope models is clearly significant with a difference of 288.04 and two degrees of freedom. In Table 8 and Figure 4 we show the solutions for the non-selective versus the selective systems, constructed in the same way as before.

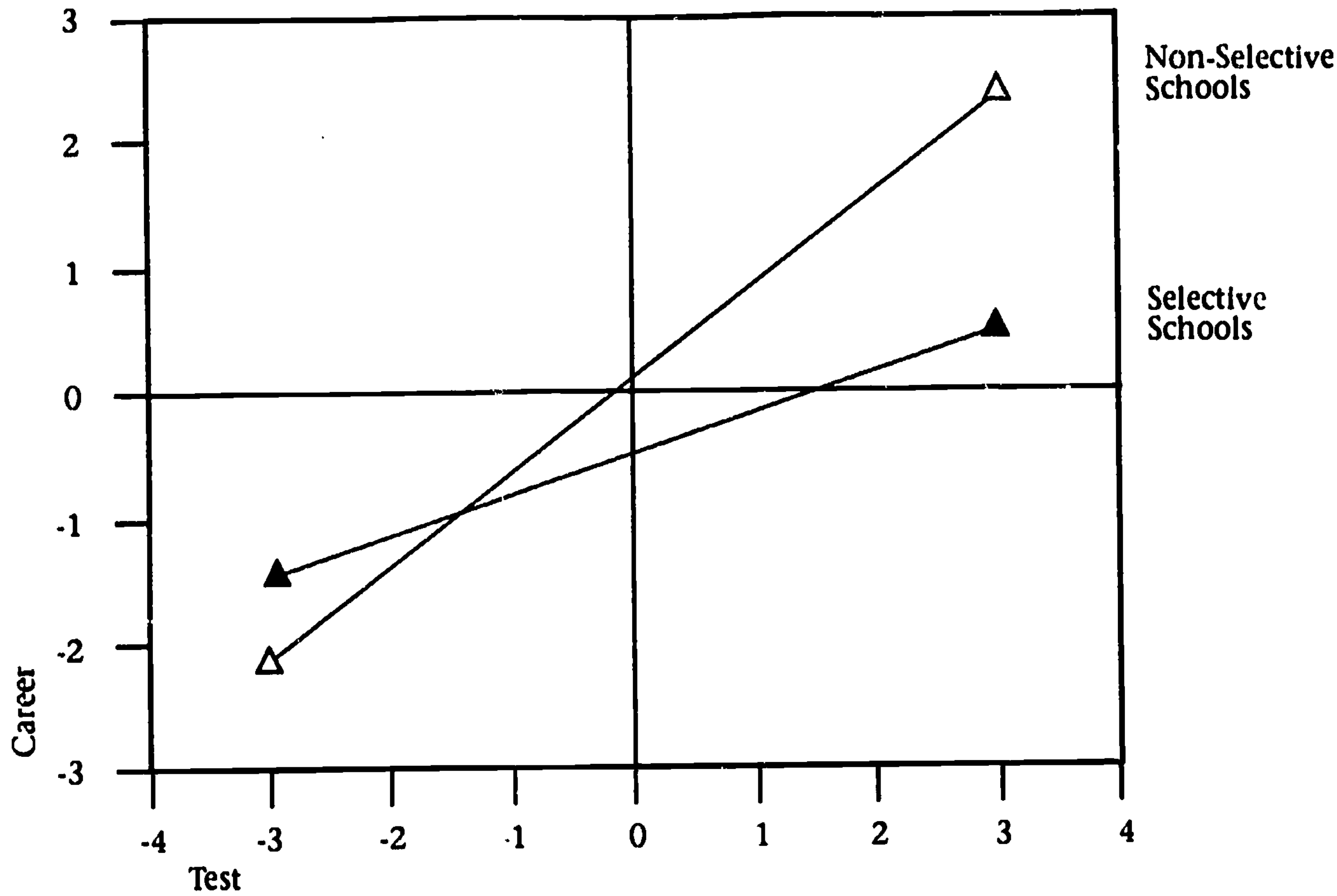
Interpretation of Results

It is clear from Table 8 and Figure 4 that test scores are less important for future success in the selective system than they are in the non-selective system. Thus, the comprehensive system appears to be meritocratic, at least more so than is the selective system. Within systems the effect of prior achievement is strongest in comprehensive schools. Here again we see that the schools that serve a more homogeneous population (the selective schools) have a flatter slope. If we follow the same reasoning of Lee and Bryk (1988), we conclude that the better defined a school population is and the fewer the number of choices students have within the system, the flatter the slope of prior achievement and school career will be. We know that Catholic schools do offer a smaller variety of courses and have more homogeneous demands for all students (Lee & Bryk 1988).

Table 8
Regression Lines for Girls:
Findings for Selective and Non-Selective Schools

Comprehensive schools	Y = +.13 + .74X
Selective schools	Y = -.44 + .28X

Figure 4
Regression Lines for Girls:
Findings for Selective and Non-Selective Schools



From the analysis of the two systems we learn that the selective schools together produce poorer overall results than the non-selective comprehensive schools. The non-selective system has a slightly lower intercept but a steeper slope. These results are in agreement with those found for the English schools (Steedman, 1980, 1983). The steeper slope for the non-selective comprehensive schools indicates that prior achievement better predicts later school career for this type of school than it does for the more selective types. Once students are selected for certain selective types of secondary education, prior achievement no longer has as important a role in determining their progress through school.

Summary and Conclusion

Our examination of research studies that focused on the effects of selection policies within and among schools uncovered not only a need for a better theoretical model but also a need for a suitable analysis model. School systems are devised for teaching individuals; thus, to determine if school organization has an effect on the success of individual students and the schools in which they are taught, we must have analysis models that combine student-level variables with school and sector variables. The interaction among these levels is reflected nicely in analysis models that can accommodate hierarchically nested data.

The analysis of the Dutch data set showed that the reorganization mandated by the Mammoth Law changed students' opportunities in the city of Amsterdam, indicating that the organizational structure of education can effect a student's school career. The comparison of British and American school systems with the Dutch system found similarities that support this finding. The results from American schools showed that certain students will be better off in a private school (specifically, a Catholic school) than in a public school. The Dutch results showed that a student is better off in a non-selective school than in the lowest selective type of school; however, these results also indicated that a student is better off in the highest type of selective school—a school that makes great demands of its students—than in a non-selective school. Together, these results led us to the conclusion reached by Lee and Bryk (1988): A student is always better off in a school that demands high levels of effort and achievement.

Our analysis still leaves crucial questions unanswered. Would selective schools in the Netherlands and the UK produce the same good results if they did not select the most able students, but admitted students representing a wider range from the continuum of achievement? Does the makeup of the student population in selective schools cause higher achievement? Are the high demands made by selective schools the cause or the effects of better results?

Before we can make an educational system more productive, we need to know what causes the academic success we found in selective schools. Several explanations are suggested in the literature. School climate, influence of the peer group, better curricula, and better teachers all may have an influence, as might the fact that the most successful types of schools attract the best students. Several studies have suggested that the morale of teachers, students, and parents is the primary cause of higher academic achievement; this suggestion is elaborated in Salganik and Karweit (1982), Coleman et al. (1982), and Bryk (1988). The arguments of these authors are philosophical—they are not supported by quantitative data analyses that can distinguish between causes and effects. Moreover, questions such as those above cannot be answered with non-experimental data: Researchers who work with observational data can make only educated guesses. We still cannot predict with what will happen if parents of school children in the USA are provided with vouchers that allow them to choose between a public or private school.

The success of a school may be most closely related to the makeup of its student population, which in certain types of schools may be determined by factors such as test scores or the cost of tuition. This situation may heighten possible peer group effects, such as those suggested in the first Coleman study in 1966. School climate may be the result of a certain student population, or it may be the result of an interaction between student performance and the expectations of teachers and parents. The combination of factors that lead to high achievement must still be determined: As yet, we cannot choose between the two theories illustrated in Figures 1 and 2. Although certain students seem to be better off in an educational system that makes high demands, researchers are faced with the task of understanding what "high demands" are and how they effect individual students. Achievement cannot be increased by simply making greater demands of students, as Bryk (1988) illustrated so clearly when he cited a student's letter about suicide.

Our evaluation of the Dutch school system does present one clear finding: The effects of education are influenced by the way school administrations implement the criteria for selection and the allocation of resources to individual students by way of their placement in a particular curriculum. Further, although institutional and organizational changes may not be direct causes of higher achievement, laws such as the Mammoth Law do change opportunities to learn because they alter social and educational structures. Our study indicates that legislative changes create an educational structure that facilitates access to educational opportunities and more effective learning environments by giving parents and students more choices and by delaying the effects of practices such as tracking. Individuals who doubt that legislation has the power to change educational opportunity should consider the results of this study.

References

Aitkin, M.A., & Longford, N.T. (1986). Statistical modeling issues in school effectiveness studies. *Journal of the Royal Statistical Society, 149,* 1-43.

Alexander, K.L., Cook, M., & McDill, E.L. (1978). Curriculum tracking and educational stratification: Some further evidence. *American Sociological Review, 43,* 47-66.

Alexander, K.L., & McDill, E.L. (1976). Selection and allocation within schools: Some causes and consequences of curriculum placement. *American Sociological Review, 41,* 963-980.

Anderson, C.S. (1982). The search for school climate: A review of the research. *Review of Educational Research, 52,* 368-420.

Blau, P.M., & Duncan, O.D. (1967). *The American occupational structure.* New York: Wiley.

Bryk, A.S. (1988). Musing on the moral life of schools. *American Journal of Education, 96,* 256-290.

Bryk, A.S. (1981). Disciplined inquiry or policy argument? *Harvard Educational Review, 51,* 479-509.

Bryk, A.S., Raudenbush, S.W., Seltzer, M., & Congdon, R.T. (1988). *An introduction to HLM: Computer program and users' guide.* Chicago: University of Chicago.

Camoran, A., & Mare, R.D. (1989). Secondary school tracking and educational inequality: Compensation, reinforcement, or neutrality? *American Journal of Sociology, 94,* 1146-83.

Catterall, J.S. (1982). Tuition tax credits: Issues of equity. In T. James & M. Levin (Eds.), *Public dollars for private schools.* Philadelphia: Temple University Press.

Clifford, P., & Heath, A. (1984). Selection does make a difference. *Oxford Review of Education, 10,* 85-99.

Coleman, J.S., Campbell, E.Q., Hobson, C.F., McPartland, J., Mood, A.M., Weifeld, F.D., & York, R.L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Education and Welfare.

Coleman, J.S., Hoffer, T., & Kilgore, S.B. (1982a). *High school achievement: Public, Catholic, and private schools compared.* New York: Basic Books, Inc.

Coleman, J.S., Hoffer, T., & Kilgore, S.B. (1982b). Cognitive outcomes in public and private schools. *Sociology of Education, 55,* 65-76.

Cronbach L.J., & Webb, N. (1975). Between class and within class effects in a reported aptitude x treatment interaction: A reanalysis of a study by G.L. Anderson. *Journal of Educational Psychology, 67,* 717-724.

De Leeuw, J., & Kreft, I.G.G. (1986). Random coefficient models for multilevel analysis. *Journal of Educational Statistics, 11,* 57-86.

Dempster, A.P., Laird, N.M., & Rubin, D.B. (1977). Maximum likelihood for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B, 39*, 1-38.

Dronkers, J. (1989). The yield of Dutch educational attainment research: A retrospective view of the last decennium. In B.F.M. Bakker, J. Dronkers, & G.W. Meynen (Eds.), *Educational opportunities in the welfare state: Longitudinal studies in educational and occupational attainment in the Netherlands*. Nijmegen: ITS.

Dronkers, J. (1983). Have inequalities in educational opportunity in the Netherlands changed? A review of empirical evidence. *The Netherlands Journal of Sociology, 19*, 133-150.

Dronkers, J., & De Jong, U. (1979). Jencks and Fägerlind in a Dutch Way: A report on research on the relationship between social background, intelligence, education, occupation, and income in the Netherlands. *Social Science Information, 18*, 761-782.

Gifi, A. (1989). *Nonlinear multivariate analysis*. New York: Wiley.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrik, 73*, 43-56.

Goldstein, H. (1984). The methodology of school comparisons. *Oxford Review of Education, 10*, 69-75.

Gray, J., McPherson, A.F., & Raffe, D. (1983). *Reconstructions of secondary education. Theory, myth, and practice since the war*. London: Routledge and Kegan Paul.

Hauser, R.M., Sewell, W.H., & Alwin, D.F. (1976). High school effects on achievement. In W.H. Sewell, R.M. Hauser, & D.L. Featherman (Eds.), *Schooling and achievement in American society*. New York: Academic Press.

Hauser, R.M., Tsai, S.L., & Sewell,W.H. (1983). A model of stratification with response error in social and psychological variables. *Sociology of Education, 56*, 20-46.

Jencks, C. (1972). *Inequality. A reassessment of the effect of family and schooling in America*. New York: Basic Books, Inc.

Kreft, I.G.G. (1987). *Models and methods for the measurement of school effect* (Doctoral dissertation). Amsterdam: University of Amsterdam.

Lee, V.E., & Bryk, A.S. (1988). Curriculum tracking as mediating the social distribution of high school achievement. *Sociology of Education, 61*, 78-95.

Lindley, D.V., & Smith, A.F.M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B, 34*,1-41.

Longford, N. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika, 74*, 817-827.

Oakes, J. (1985). *How schools structure inequality.* New Haven, CT: Yale University Press.

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

Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review, 15,* 351-357.

Salganik, L.H., & Karweit, N. (1982). Voluntarism and governance in education. *Sociology of Education, 55,* 152-161.

Sociology of Education (1985). Special issue, *58,* 74-13.

Sociology of Education (1983). Special issue, *56,* 170-234.

Sociology of Education (1982). Special issue, *55,* 65-182.

SPSS. (1989). *SPSS-Categories.* Chicago: Author.

Steedman, J. (1983). *Examination results in selective and nonselective schools.* London: The National Children's Bureau.

Steedman, J. (1980). *Progress in secondary schools.* London: The National Children's Bureau.

Tate, R.L., & Wongbundit, Y. (1983). Random versus nonrandom coefficient models for multilevel analysis. *Journal of Educational Statistics, 8,* 103-120.

Van Herpen, L.W., & Smulders, R.H.M. (1980). Sociale beroepsgroep en schoolkeuze [Social status and school choice]. *CBS Select, 1,* 117-133.

Van Laarhoven, P., Bakker, B., Dronkers, J., & Schijf, H. (1987). Achievement in public and private secondary education in the Netherlands. *Zeitschrift fuer Internationale Erziehung und Sozialwissenschaftliche Forschung, 4,* 335-356.

Appendix

Types of Schools in the Dutch School System

CITO: National Institute of Educational Measurement. This institute provides nation-wide reading and arithmetic tests for students who are finishing elementary school. The test results together with the recommendation of teachers are used to distribute students over the different types of schools. Selective schools such as HAVO/VWO and Gymnasium/Lyceum do not permit the enrollment of students with low CITO scores.

Comprehensive schools: See Scholengemeenschap.

Gymnasium/Lyceum: A six year curriculum. The Lyceum is basically the same as the VWO, except that is has a higher social status as a result of being a highly selective school that enrolls the higher SES students. The distinction between the Lyceum and Gymnasium is that the latter offers a mandatory six-year curriculum in Greek and/or Latin, while the first does not.

HAVO: Pre-college education. It is comparable to an American senior high school, but it gives entrance only to colleges, not to the university. A five year curriculum. The HAVO curriculum can be studied within the non-selective comprehensive school; it is also found in combination with a MAVO (placing it in the category of miscellaneous schools) or VWO curricula.

LBO: Vocational training. A four-year curriculum, educates for specific vocations such as construction worker, house painter, mechanic, etc.

MAVO: Administrative training, comparable to an American junior high school. A three- to four-year curriculum.

Scholengemeenschap: Comprehensive schools; the only non-selective type of school. Within the first two years of the curriculum, students choose a curricular emphasis that leads to either an administrative degree (MAVO) after four years, a college prep degree (HAVO) after five years, or a university prep degree (VWO) after six years.

VWO: Pre-university education. A six-year curriculum. VWO is always part of a larger type of school, either a college prep school (HAVO—the combination is HAVO/VWO) or is part of the non-selective system (comprehensive schools). VWO has the same status as Gymnasium/Lyceum and gives entrance to any type of college or university.