



Antero Malin

School Differences and Inequities in Educational Outcomes

PISA 2000 Results of Reading Literacy in Finland



INSTITUTE FOR EDUCATIONAL RESEARCH
UNIVERSITY OF JYVÄSKYLÄ

Institute for Educational Research
Research Reports 15

School Differences and Inequities in Educational Outcomes

PISA 2000 Results of Reading Literacy in Finland

Antero Malin

Esitetään Jyväskylän yliopiston kasvatustieteiden tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston vanhassa juhlasalissa (S212)
syyskuun 23. päivänä 2005 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education of the University of Jyväskylä,
in Auditorium S212, on September 23, 2005 at 12 o'clock noon.



INSTITUTE FOR EDUCATIONAL RESEARCH
UNIVERSITY OF JYVÄSKYLÄ

THIS PUBLICATION
CAN BE OBTAINED FROM:
Institute for Educational Research
Customer services
University of Jyväskylä
P.O. Box 35
FIN-40014 Jyväskylä, Finland
Phone + 358 14 260 3220
Fax + 358 14 260 3241
E-mail: ktl-asiakaspalvelu@ktl.jyu.fi
<http://www.ktl-julkaisukauppa.fi/>

© Antero Malin and the Institute for Educational Research

Cover & Photo: Vesa Huhtala
Graphic design: Martti Minkkinen
Layout: Kaija Mannström

ISSN 1455-447X
ISBN 951-39-2247-2 (nid.)
ISBN 978-951-39-3041-7 (pdf)

Printed by Jyväskylä University Press
Jyväskylä 2005

Contents

ABSTRACT	7
PREFACE	11
I INTRODUCTION	
1 Quality and equity in education	15
1.1 The Finnish education policy and educational equality	15
1.2 The concept of educational equality	17
1.3 Equity of the European Educational Systems	18
1.4 The objectives of education and the principle of equity	22
1.5 Principle of equity at student and school level	23
1.6 Conclusions	24
2 School differences and inequities in educational outcomes in Finland	26
2.1 Finnish school differences and inequities in PISA 2000	26
2.2 Finnish school differences and inequities in national studies	32
2.3 Finnish school differences in international studies	35
2.4 Conclusions	36
3 Research problems	38
II DATA AND METHODS	
4 Data and statistical analyses	45
4.1 PISA 2000 data	45
4.2 Reading literacy performance score	46
4.3 Inequity factors	47
4.4 Engagement in reading	52
4.5 Process of data analyses	53

Contents

5	Multilevel modelling of school data	55
5.1	Hierarchical structure of school data	56
5.2	Intra-cluster correlation	58
5.3	Two-level linear model	60
5.4	Applications of two-level linear model	62
5.5	Explanatory power of the model	67
III	RESULTS	
6	School differences in reading literacy performance	73
6.1	Finnish school differences in reading literacy: Baseline model	73
6.2	Comparisons with Germany and Sweden	76
6.3	Conclusions	80
7	Two-level one-factor models	83
7.1	The fixed effects of two-level one-factor models	83
7.2	Explanatory power of the two-level one-factor models	88
7.3	Conclusions	95
8	Two-level multifactor model	97
8.1	Minority group factors: Language, grade and immigrant background	98
8.2	Location of the school: Regional unit and municipality type	100
8.3	Gender	101
8.4	Socio-economic background factors: Socio-economic status of the family, parents' educational level and wealth of the family	103
8.5	Explanatory power of multifactor model	106
8.6	Decomposition of variance components	109
8.7	Inequity model	113
8.8	Ranking of schools by controlling for the effects of the inequity factors	120
8.9	The effect of one single school	122
8.10	Comparisons with Germany and Sweden	125
8.11	Conclusions	133
9	Reducing inequities with reading activity	135
9.1	Effects of engagement in reading	135
9.2	Extended inequity model	137
9.3	Conclusions	142

IV DISCUSSION	
10 Conclusions and implications of the study	147
10.1 Conclusions	147
10.2 Striving for equity in reading literacy	151
10.3 Methodological reflections	154
10.4 Recommendations for further studies	155
10.5 Challenges for strengthening equity in comprehensive school	156
YHTEENVETO	159
REFERENCES	167
APPENDICES	
Appendix 1 Regional units in Finland, used in PISA 2000	173
Appendix 2 Means of students' socio-economic status index in regional units and municipality types, estimated by two-level model	174
Appendix 3 Fixed and random parameter estimates of the inequity model for the combined reading literacy in Finland, when the effect of one single school is controlled for	175
LIST OF FIGURES	
FIGURE 1 School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Finland	75
FIGURE 2 School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Germany	77
FIGURE 3 School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Germany, after controlling for the effect of school type	78
FIGURE 4 School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Sweden	79
FIGURE 5 Scatterplot of student scores in the combined reading literacy around the school means in Finland	81
FIGURE 6 Scatterplot of student scores in the combined reading literacy around the school means in Germany	81
FIGURE 7 Scatterplot of student scores in the combined reading literacy around the school means in Sweden	81
FIGURE 8 School residuals of the socio-economic status index with 95 % confidence intervals in rank order	94

Contents

FIGURE 9	The spread of school residuals of the combined reading literacy score in two-level models in model order	108
FIGURE 10	School residuals of the combined reading literacy score with 95 % confidence intervals in rank order, after controlling for the effects of inequity factors	109
FIGURE 11	Unadjusted and adjusted regional differences in the combined reading literacy score in Finland, with 95 % confidence intervals ...	115
FIGURE 12	The between-school and within-school regression lines of the combined reading literacy score on the standardized socio-economic index in Finland	117
FIGURE 13	School ranks in baseline model and model IV	121
FIGURE 14	Standardized school residuals of the combined reading literacy score by normal equivalent scores, after controlling for the effects of background variables	123
FIGURE 15	The between-school and within-school regression lines of the combined reading literacy score on the standardized socio-economic index in Germany	127
FIGURE 16	The between-school and within-school regression lines of the combined reading literacy score on the standardized socio-economic index in Sweden	131
FIGURE 17	School residuals of the combined reading literacy score with 95 % confidence intervals in rank order, after controlling the effects of inequity factors and engagement in reading	141

LIST OF TABLES

TABLE 1	Two-level variance component model for the combined reading literacy score in Finland	74
TABLE 2	The fixed effects of two-level one-factor models on the combined reading literacy score	84
TABLE 3	Variance components and related statistics of two-level one-factor models on the combined reading literacy score	90
TABLE 4	Change of the fixed parameter estimates of language, grade and immigrant background in constructing the two-level model explaining the combined reading literacy score	98
TABLE 5	Change of the fixed parameter estimates of regional units in constructing the two-level model explaining the combined reading literacy score	101

TABLE 6	Change of the fixed parameter estimates of municipality type in constructing the two-level model explaining the combined reading literacy score	101
TABLE 7	Change of the fixed parameter estimates of gender in constructing the two-level model explaining the combined reading literacy score	102
TABLE 8	Fixed parameter estimates of socio-economic status of the family in constructing the two-level model explaining the combined reading literacy score	104
TABLE 9	Fixed parameter estimates of parents' educational level in constructing the two-level model explaining the combined reading literacy score	105
TABLE 10	Fixed parameter estimates of family wealth in constructing the two-level model explaining the combined reading literacy score	105
TABLE 11	Random part estimates and related statistics of two-level models	107
TABLE 12	Proportional reductions in variance components and proportions of explained variances in two-level models	107
TABLE 13	Variance components in gender-only model and Model IV	110
TABLE 14	Variance components in the regional-units-only and full model	112
TABLE 15	Fixed parameter estimates of the inequity model for the combined reading literacy in Finland	114
TABLE 16	Random parameter estimates of the inequity model for the combined reading literacy in Finland	119
TABLE 17	Fixed parameter estimates of the inequity model for the combined reading literacy in Germany, after controlling for the differences between school types	126
TABLE 18	Random parameter estimates of the inequity model for the combined reading literacy in Germany when the differences between school types are controlled for	129
TABLE 19	Fixed parameter estimates of the inequity model for the combined reading literacy in Sweden	130
TABLE 20	Random parameter estimates of the inequity model for the combined reading literacy in Sweden	132
TABLE 21	One-factor two-level model of the combined reading literacy score on the index of engagement in reading in Finland	136
TABLE 22	Fixed parameter estimates of the extended inequity model for the combined reading literacy score in Finland	138
TABLE 23	Random parameter estimates of the extended inequity for the combined reading literacy score in Finland	140

Malin, A. 2005

**School Differences and Inequities in Educational Outcomes.
PISA 2000 Results of Reading Literacy in Finland**

University of Jyväskylä. Institute for Educational Research. Research Reports 15.
ISSN 1455-447X, ISBN 951-39-2247-2 (nid.) ISBN 978-951-39-3041-7 (pdf)

Abstract

The widely accepted educational objectives combine high quality and level of performance with equitable distribution of outcomes. The objectives of the Finnish education policy have been to raise the general standard of education and promote educational equality by providing all population groups and regions of the country with equal educational opportunities.

The purpose of this study was to investigate the level and distribution of reading literacy performance in Finland from the perspective of equity both at the student and school level, as they were observed in the international PISA 2000 reading literacy data. The Finnish PISA 2000 data consisted of a representative sample of 15-year-old students and schools, and it consisted of 4864 students in 155 schools. The aim was to find out the amount of the school differences in reading literacy performance in Finland, the force of the effects of selected background factors on reading literacy performance, and the extent to which the between-school as well as the total student variation were attributable to these factors.

Due to the hierarchical nature of the data, multilevel statistical modelling was used in the data analyses.

The unadjusted school differences in reading literacy performance were very small. There were some clear inequities in the student performances due to background fac-

tors that the individual students cannot escape or control, like gender, the socio-economic status of the family, parents' educational level, region of the country, language of the school and students' immigrant background. These inequity factors were associated with systematic variation between the students and schools in reading literacy performance. On average, higher socio-economic status and parents' higher education meant better results. Regional differences did exist, but they were associated with regional variation in the social background. Gender difference was notable in reading literacy, but it was not due to the low performance of boys rather than the extremely good performance of girls. The average results were lower in Swedish-speaking than in Finnish-speaking schools, and students with immigrant background did not do as well as native students. After controlling for the effects of the background factors, only few schools were clearly below or above the national mean in reading literacy performance. In the international comparison, the inequities were small but, however, existing. The results also indicated that the inequities in performances can be reduced with the effort from students, parents and schools.

The Finnish educational system and the comprehensive school have succeeded in keeping the school differences small. Finland has also succeeded reasonably well in the efforts to provide all population groups and regions of the country with equal educational opportunities in the comprehensive school. The biggest problem is not the variation between student groups or schools, but the variation between students within schools. Some schools lag behind others in their average performance, although there is no evidence that the same schools lag behind constantly year after year, and in every subject area. There are individual schools that need help, and especially the low achieving students in these schools are in need of cognitive and motivational support.

Key words: school differences, inequities in education, reading literacy, survey-based assessment, multilevel modelling, hierarchical data

Author's address Antero Malin
Institute for Educational Research
University of Jyväskylä
P.O. Box 35
FIN-40014 University of Jyväskylä
Finland
antero.malin@ktl.jyu.fi

Supervisor Professor Jorma Kuusinen
Department of Education
University of Jyväskylä, Finland

Reviewers Docent, PhD Ritva Jakku-Sihvonen
Faculty of Behavioural Sciences
University of Helsinki, Finland

Professor Risto Lehtonen
Department of Mathematics and Statistics
University of Jyväskylä, Finland

Opponent Docent, PhD Ritva Jakku-Sihvonen
Faculty of Behavioural Sciences
University of Helsinki, Finland

Preface

This study is the result of my long-standing work at the Institute for Educational Research. I have had the possibility to work in various research projects, which have mainly been concerned with school data. However, in the PISA project it became unavoidable to address the methodological challenges in the statistical analysis of school data. Since the data available in PISA was excellent material for applying and developing methodological knowledge and skills, the decision to start working with this thesis was easy to do.

I am grateful to numerous persons for helping me in completing my thesis. My deepest gratitude is addressed to my supervisor, Professor Jorma Kuusinen. He has guided my work with profound expertise, showing me the way to the essence of the topic and helping me to avoid all the sidetracks, however tempting they were. I want to show my gratitude to the reviewers of my thesis, Docent Ritva Jakku-Sihvonen and Professor Risto Lehtonen. Their valuable comments and suggestions helped me to improve the final version of my thesis.

In the Institute for Educational Research, I especially want to thank Professor Pirjo Linnakylä for her encouragement and inspiring feedback during the whole process of thesis writing. I have had the privilege of working with her for years, in research projects like PISA and SIALS among others. This collaboration has taught me a lot about research work. I am also indebted to the Institute for Educational Research and the Director of the Institute, Professor Jouni Välijärvi, for the possibility to complete my thesis. The inspiring discussions with Professor Rainer Lehmann, Humboldt-Universität zu Berlin, Abteilung Empirische Bildungsforschung und Methodenlehre, gave me valuable ideas on how to conduct my work.

Preface

I want to thank all my colleagues in the Institute for Educational Research, especially the members of the Methods and Data Management group and the head of the group, Doctor Kari Törmäkangas. Profound discussions with my best colleagues, not only of my thesis, but of whatever topic that has been in need of thorough discussion, have helped me to stand on steady ground and to avoid the desperate moments of lonely research work. I am also indebted to the PISA research team of IEA and all of its members for offering me such an encouraging and challenging working environment.

I also express my thanks to the Publication and information unit of IER for their contribution in bringing this book into its final format. Especially I want to thank Kaija Mannström for her work in editing this book. I am also grateful to MA Timo Törmäkangas for proofreading my English manuscript. My friend Vesa Huhtala has designed the cover of the book, for which I am indebted to him.

Finally I want to thank my family. My wife Paula has always encouraged me, when doubts about the importance of my thesis have sneaked into my mind, and my sons Aarno and Sampo have encouraged me by their own example.

This research has been supported by the Academy of Finland (grant no 206176).

Jyväskylä, August 2005

Antero Malin

I

INTRODUCTION

1

Quality and equity in education

1.1 The Finnish education policy and educational equality

The widely accepted educational objectives of knowledge societies combine high quality and level of performance with equitable distribution of outcomes. These are the guiding principles of the education policy in Finland as well. According to the Ministry of Education (2004), the traditional long-term objectives of the Finnish education policy have been to raise the general standard of education and to promote educational equality. Efforts have been made to provide all population groups and regions of the country with equal educational opportunities. These are the basic tenets of the educational reforms carried out over the last few decades. Special attention has been paid to the content of education and the methods of instruction, as well as to educational standards and equality. Increasing overall flexibility and opportunities for individual choice have also been considered important, and internationalization has emerged as a key objective.

The principle of equality in education is also stated in the Basic Education Act from 1998: "The objective of basic education is to support pupils' growth towards humanity and ethically responsible membership of the society, and to provide them with the knowledge and skills necessary in life. The instruction shall promote equality in the society and the pupils' abilities to participate in education and to otherwise develop themselves during their lives."

Human rights, equality and democracy are among the basic values on which the education is based in the New Framework Curriculum confirmed in the beginning of 2004 by The Finnish National Board of Education (2004). The basic education is intended to increase regional equality and equality between individuals, and to promote equality between the genders. Basic Education in Finland means the general education provided for each age group in its entirety. It is intended for children from seven to sixteen years of age, and its completion in the comprehensive school takes nine years. The task of the comprehensive school is to provide all children and young people with equal opportunities for learning regardless of their particular school, background or circumstances. Upon completing comprehensive school, pupils will have fulfilled their compulsory education.

The leading principles of the education policy in Finland are the principles of equality and social justice, and special attention has been paid to the goal of supporting the weakest performers. The principle of equality has a long tradition in the Finnish basic education (Ahonen 2003), although some researchers have argued that equality is no longer fulfilled in the Finnish comprehensive school. In a study based on students and schools of one town only, Olkinuora and Rinne (2001) argued that the schools within the town had divided into low achieving schools, where students had low level of social background, and to high achieving schools, where students came from families with high level of social background.

Rinne and Nuutero (2001) have claimed that the Finnish educational policy is changing and abandoning the principles of equality and social justice. According to them, in the new educational policy the schools are seen as business enterprises that should operate efficiently and successfully. At the same time, the old principles of regional, social, economic and gender equality are being abandoned.

However, equality is an aim highlighted in the education policies across the Nordic countries. Like in other Nordic countries, providing all students with equal access to education and removing obstacles to learning, especially among students from disadvantaged backgrounds, are the leading objectives for education in Finland. Attaining high overall performance and, at the same time, evening out disparities in learning outcomes are key aims not only in the Nordic countries (Husén 1989; OECD 2001; Lie, Linnakylä & Roe 2003). In comparison across the OECD countries, the Nordic comprehensive schools seem to function quite well in providing the children with equal opportunities for learning regardless of their particular school, background or circumstances (Väljjarvi & Malin 2003, 125).

1.2 The concept of educational equality

Three major stages have been distinguished in the development of the *concept of educational equality* (Husén 1974, 1975, 1986). These are (a) the conservative; (b) the liberal; and (c) the radical or “redemptive” conception of educational equality. The *conservative* conception of educational equality was prevalent in most industrialized countries until World War I. The conception was based on the idea that God had bestowed different amounts of talent upon each human being, and it was up to the individual to make the best possible use of that capacity. The *liberal* conception of educational equality puts emphasis on the equality of educational *opportunity*. The task is primarily to remove external barriers of economic and geographical nature so as to make the original capacity in each child develop. The success and failure in school primarily depends upon the individual student, and his natural intellectual and moral resources are the decisive factors. The liberal vision of equal opportunity prevailed in most Western industrial countries well until the late 1960s.

The question was asked whether one should not also consider what comes out of the system, that is to say whether equality of *results* is not more important than equality of initial opportunity. The *radical* conception of educational equality sees that the student’s success and failure must be ascribed mainly to the school situation. The basic problem turns on the extent to which the school has been able to provide the conditions conducive to satisfactory student development. The *more radical* conception of educational equality is that, in order to achieve the long-range objective of more equality in occupational career and standard of living, remedial action must be taken in the wider context within which the schools are operating – that is, the society at large. Educational reform cannot be a substitute for social reform. (Husen 1975.)

Husén (1974) defines provision of equal educational opportunities as a conservative interpretation of the principle of equity. A more liberal view places emphasis on the active removal of instructional and pedagogical obstacles for the most disadvantaged students and provision of special support for learners from weaker socio-economic backgrounds or with lower capabilities. A more radical interpretation adds to the previous ideas the aim of reducing inequality of learning achievement (Välijärvi 1994).

The principle of equity in education cannot mean only the equality of opportunities. The radical conception of educational equality, equality of results, can be expressed as a weaker principle of equitable distribution of results, in stead of equal results. Of the last two concepts of educational equality, the concept of equality of educational oppor-

tunities is more basic and it comes before the concept of equality of results. The equality of results makes sense only in the situation of equal opportunities. Equal results are worth of pursuit only when factors which are irrelevant with respect to educational opportunities, like gender, ethnicity or family background, have no influence on the educational opportunities and thereby to the results. And even then, the results cannot be the same for each individual student. There will still be variation in individual achievements, but not because of unjust circumstances or disadvantaged background, but due to talent, motivation, effort and other relevant factors influencing the results. In other words, the radical conception of educational equality can be understood as equitable distribution of results.

1.3 Equity of the European Educational Systems

The concept of equity in terms of education is far from being univocal. The meaning of equity is connected with the concept of fairness and justice. Inequalities in education have been discussed for a long time in the international publications, particularly in the OECD *Education at a Glance* (e.g. OECD 2003), but the treatment has been rather limited. In order to define the concepts of equality and equity and to measure and compare the equity of the education systems in the European Union Member States, the European Group of Research on Equity of the Educational Systems (2004) published the report *Equity of the European Educational Systems. A Set of Indicators*. The report is the result of collaboration between six European university teams which formed the project "Devising international indicators on equity of educational systems". In the first part of the report, the concepts of equality and equity are defined, and the framework of indicators and its guiding principles are presented.

The authors of the report admit that equity is a more difficult concept than equality, and that it allows, in its principle, inequalities. However, it is a concept that allows going beyond a purely formal examination to perform a multidimensional analysis, and we need to go beyond the concept of formal equality.

The discussion about equity begins when it is necessary to define the assets that should be equalized, or which principles of distribution equity it demands. A strictly egalitarian approach is impossible and we must take account of a multiplicity of principles of justice, assets connected with education or groups of individuals. To answer the question presented by Sen (1992): "Equality of what?", it must be specified which

equalities we are talking about and, by doing this, envisaging the discussion in terms of equity.

The report gave four major principles of equality in terms of education, adapted from Grisay (1984):

1. Equality of access or opportunities;
2. Equality of treatment;
3. Equality of achievement or academic success;
4. Equality of actualization or social output.

Whether we talk about equal opportunities, equal treatment, equal achievement or equal results, we stumble over the practical or theoretical limitations connected with the adoption of a particular theory. In the first case, the wish is that the social background does not influence success at school, but it is subject to criticism by those who claim that this leaves the possibility open to give better educational conditions to those with greater ability. In the second case, the same educational conditions are given to all, but this is open to criticism from those who think that some people, because they suffer from a handicap of one kind or another, need better educational conditions. In the third case, equality of results is desired, at least for a certain level of knowledge, but this comes up against those who claim that by pursuing this objective, the best pupils are deprived of the possibility of progressing as far as possible. (European Group of Research on Equity of the Educational Systems 2004.)

While each of the three principles has its own limits, the adoption of one principle or another poses an even more fundamental problem. They are "local" theories about justice, which consider education as a final asset. The last principle broadens the perspective by putting school back into its societal context. (European Group of Research on Equity of the Educational Systems 2004.)

The social consequences of failure at school are considered as increasingly important, and the idea that children and young people must also be treated justly is gaining ground. Consequently, the educational systems are being given increasingly precise targets, which are, in some cases, quantified, in terms of both efficiency and equity. People are also interested in the efficiency of their educational system, i.e. its cost, its internal and external efficiency, or its capacity to pass on skills that are useful to the society and to the nation. The emphasis on equity does not replace any other concern. On the contrary, citizens are demanding both efficiency and equity. (European Group of Research on Equity of the Educational Systems 2004.)

There are many principles of justice that can be invoked to justify actions, practices, or situations, but the existence of several principles does not lead them to invalidate each other. The European Group of Research on Equity of the Educational Systems (2004) developed eight guiding principles to structure their work. The second principle guided the team to consider justice in education from three angles: *The relevant educational inequalities for the majority of assets distributed in the context of educational systems may be grouped into three main families: the discrepancies between individuals, the inequalities between categories, and the proportion of individuals who find themselves below a minimum threshold.*

The first type of inequality, that of *inequalities between individuals* (skill differences between those who are most and least successful, for example), is justified by several reasons. Where they are too sizeable, they threaten social cohesion, maintaining a feeling of inferiority on the part of the least educated, which is contrary to an equitable sharing of self-respect. Furthermore, they threaten the feeling of equality which is necessary for the democratic principles of political society (Rawls 1971). In addition, they will convert, in the following generation, into inherited inequalities, which will be entirely unjust (Benadusi 2001). Thus, too large discrepancies between individuals and individual skill inequalities may jeopardize the social cohesion.

The second type of inequality is that of *inequalities between groups*, and it illustrates the principle of equal opportunities. The social membership of an individual must not handicap his or her success at school.

The third type of inequality is that of a *threshold* beneath which fairness dictates that nobody should find themselves. Being below certain skill thresholds is probably the educational situation that can have the most serious social consequences for the individual. Of course, the definition of the threshold and therefore of skills that are situated below and beyond it, may vary. It may refer to the "employability" of individuals, or to minimum skills for participating in democratic life, and to asserting one's rights (Gutmann 1999; Benadusi 2001). No student should leave the educative system being below the minimum skill threshold, in order to have a decent life in the modern society.

The third principle states: *Among relevant categories of individuals, the most important are those from which the individual cannot escape.* The concept of "what is important" is historically determined, and the establishment of relevant categories is more in the field of political or social movements than administrators or philosophers (Orfield 2001). The priority categories must be, according to the authors' opinion, those to which the individual belongs whether he or she wants to or not. For example, it is

impossible to change social origins, nationality, and gender or to escape a disability from birth.

Once we focus on developing the same basic skills among all individuals, there is a necessary link between effectiveness and equity, framed in terms of equality of results. If we consider matters from the viewpoint of equality of results, the idea of linking effectiveness and equity appears to be an integral part of the principal of equality of learning and its valuation by society. From this viewpoint, an education *system* will be considered effective *if, while raising the average level of knowledge, it reduces the overall variance of internal and external results of the education system*. In any case, this is the ideal advanced by Bloom (1976) since, according to him, effective education is characterized by three joint effects present at the end of each phase of learning:

- A rise in the average standard of results;
- A reduction in the variance of results;
- A reduction in the correlation between the social origin of the pupil (and in general, his/her initial characteristics) and achievement.

According to the report, Coleman (1966) writes on this subject: "Another way of putting this is to say that the schools are successful only insofar as they reduce the dependence of a child's opportunities upon his social origins. We can think of a set of conditional probabilities: the probability of being prepared for a given occupation or for a given college at the end of high school, conditional upon the child's origins. The effectiveness of the schools consists, in part, of making the conditional probabilities less conditional – that is, less dependent upon social origins."

The international comparison showed a widespread tendency towards inequity. However, the most important result was that of the pronounced differences between countries. Obviously, not all education systems are equivalent in their ability to treat students fairly.

School is not an island: it depends on the social and economic system in which it exists, but in return, the school can also modify the society by contributing to a greater or lesser extent in the reduction of inequalities that are considered unfair.

1.4 The objectives of education and the principle of equity

In large-scale survey-based assessments, the inference concerning the level of performance and the distribution of educational outcomes is usually based on quantitative data. From the quantitative methodological point of view, the level of performance and distribution of outcomes are not separate entities, but they are intertwined together. High level of performance can be defined as high average performance in the student population, and the equitable distribution of outcomes means that the performance variation in the population is acceptable. Thus, the distribution of outcomes is defined by statistics describing both the level and variation of performance in the population.

In statistical analyses of educational achievement data, more attention should be paid on the variation of the results. Indicators of central tendency have attracted more attention than they deserve, while it should also be paid on two additional elements: the overall dispersion of achievement around the central tendency, and the relationship of achievement to important social categories such as family background (Porter and Gamoran 2002, 15).

The distribution of educational outcome consists of systematic and non-systematic part of variation. Systematic variation is associated with factors describing the "source" of this variation, while the variation is called non-systematic (or random) when its "source" is not determined. Of course, some part of the non-systematic variation may be found to be systematic, if we later find out "sources" for this variation. The amount of the systematic variation which is attributable to known characteristics in the population may be estimated by methods which measure the proportion of the variation in the dependent variable that is explained by the independent variables, like the multiple coefficient of determination in the regression model.

At the level of the society, high level of performance cannot be achieved without equitable distribution of outcomes. Low level of performance may also be associated with acceptable and just distribution of outcomes, but this is not in accordance with the twin objectives of education. If the aim is to achieve high overall level with equitable distribution of outcomes, the general average has to be high and the variation around the average has to be small.

By uncovering the causes of systematic variation in the data one can judge, if the educational outcomes are equitably distributed in the population. This also exposes the population subgroups whose performance level is not high enough. If there is no

(known) systematic variation in the distribution and the non-systematic (random) variation is decent, in addition to high level of performance, one can be satisfied, from the point of view of the twin objectives of education.

When we discuss of the equitable distribution of educational outcomes, we have to tolerate the individual variation around the acceptable level of performance. Equality as equal outcomes in achievements is a group level expression, not individual level. As stated by Jakku-Sihvonen and Kuusela (2002, 7): "In terms of educational policy, we consider educational equality to refer to the inexistence of systematic differences between girls and boys, between different social groups and between geographical areas. Differences between individuals' results are natural." Equality as equal outcomes may not be possible to achieve at individual level, but it may be achievable on average, at the group level. However, at individual level the distribution of outcomes can be just or tolerable. Variation due to relevant factors with respect to educational outcomes, e.g. to talent, motivation and effort, is acceptable, and as long as human beings are concerned, individual variation will stay.

In conclusion, high overall level is easier to achieve, if there is no systematic variation in the distribution, and systematic variation may be interpreted as a sign of inequitable distribution of educational outcomes.

1.5 Principle of equity at student and school level

The school achievement data, like in PISA 2000 data, have two separate levels, the student level and the school level. The sample design includes both students and schools, and both the population of students and the population of schools are represented in the sample. Accordingly, the fulfilment of the principle of equity has to be investigated at both levels.

Traditionally, the student level has been the focus of interest. This is natural, since the educational outcomes are ultimately student performances. Equity at student level is reflected in the lack of the unacceptable systematic variation between students. The smaller this variation is, the more equal the students are with respect to their performance. If the average level of educational outcomes is high, and if the systematic variation is small or even lacking completely, and if the non-systematic variation is also tolerable, high general level and equitable distribution of educational outcomes are achieved at the student level.

The differences between schools are one aspect of the equitable distribution of educational outcomes. Students are grouped according to the schools they attend. In this sense, variation between schools is also systematic variation between students. However, this variation may have “sources” as well as the variation between students, and the proportion of the between-school variation due to different “sources” is systematic variation between schools. Actually, this variation is systematic variation between students as well. For example, a difference between urban and rural schools is the difference between urban and rural students.

The educational objectives of high level of performance and equitable distribution of outcomes are finally realized in single schools. In studying the fulfilment of the principle of equity in the comprehensive school, the between-school differences in the educational outcomes cannot be outside the study. By including the schools in the study as separate units and investigating the between-school variation simultaneously with between-student variation, we can achieve more detailed and reliable information of the realization of equity in education than studying the effects of inequity factors at student level only. If the quality of educational outcomes is considered high, and if the systematic variation between schools is small or even lacking totally, and if the non-systematic variation between schools is also acceptable, high general level and equitable distribution of educational outcomes are achieved at the school level. And if this is true, it also reduces the proportion of systematic variation between students.

1.6 Conclusions

The research areas in this study stem from the second guiding principle of the European Group of Research on Equity of the Educational Systems (2004) and they concentrate on the inequalities between individuals and the inequalities between groups of students.

The educational objectives of high level of performance and equitable distribution of outcomes cannot mean that every individual student actually reaches the very same level in educational outcomes, independently of his or her individual characteristics. But it can mean that systematic variation between individual students does not exist, if they are attributable to unacceptable factors with respect to educational outcomes, *like disadvantaged background, gender, location of residence and membership of a minority group, or schools which they attend.* If the disparities in educational performances asso-

ciated with these factors are eliminated, there will still exist variation between students which is then due to factors relevant to educational outcomes, and the diversity in outcomes is easier to consider acceptable and just. In this case, high level of performance is associated with high level of equality.

In the Nordic countries, education policy has been grounded in the principle of equity, and the countries have paid special attention to the goal of supporting the weakest performers (Linnakylä and Malin 2001, 68). According to the Finnish National Board of Education (2000), equal opportunity for personal development and self-enhancement according to individual skills is the central content of educational equality. One of the indicators of the quality of education policy is that students coming from different cultural, religious and ethnic background have equal opportunities to pursue even most demanding careers. As stated by Jakku-Sihvonen (2002, 24), in order to promote equality in education, it is crucial that during the basic education we strive to prevent the appearance of differences in learning results and attitudes. This is because a student's basic education opens up the student's path into higher education and forms a basis for his or her lifelong learning.

An important question is how much systematic variation is considered just and acceptable. However, it remains outside this study to define what the proportion of tolerable systematic variation is and when it cannot be tolerated. The aim is to reliably estimate the amount of systematic variation in one educational outcome, i.e. reading literacy, and the sources of variation, like gender, region and home background, and thus make at least a part of the systematic variation visible and object to judgement.

2

School differences and inequities in educational outcomes in Finland

In this chapter, the findings of previous studies concerning the school differences and the distribution of educational outcomes of the comprehensive school in Finland are described, with respect to the most important background factors. First, the PISA 2000 results are shortly described, then the results of the national studies are presented, and finally the findings of some international studies are described.

2.1 Finnish school differences and inequities in PISA 2000

The OECD Programme for International Student Assessment (PISA) is a collaborative effort among the Member countries of the OECD to measure, how well young adults, at age 15 and therefore approaching the end of compulsory schooling, are prepared to meet the challenges of today's knowledge societies. PISA is an internationally standardised assessment, and the outcomes of education systems are monitored in terms of student achievement in *reading literacy*, *mathematical literacy* and *scientific literacy*. The three domains represent knowledge and skills that are regarded relevant for adult life. Finland participated in the first round of the survey, which took place in 32 countries during 2000. In 2000 the primary focus was on reading literacy. The survey will be repeated every three years, with the primary focus shifting to mathematics in 2003, science in 2006 and back to reading in 2009. (OECD 2001.)

In PISA 2000, reading literacy was defined as the ability to understand, use and reflect on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate effectively in the society. Reading literacy was assessed using a series of texts, students being set a number of tasks on each text. PISA 2000 employed 141 items representing the kinds of reading literacy that a 15-year-old would require in the future (OECD 2001). A description of the conceptual framework underlying the PISA assessment of reading literacy is provided in *Measuring Student Knowledge and Skills – A New Framework for Assessment* (OECD 1999).

The first international report with the main initial results was published in 2001 (OECD 2001), and a thematic report on reading in 2002 (OECD 2002a). The main Finnish results are published in two national reports (Väljjarvi, Linnakylä & al. 2001; Väljjarvi & Linnakylä 2002) and in a thematic report on the secondary analyses of reading literacy (Linnakylä, Sulkunen & Arffman 2004).

In the light of PISA 2000, Finland seemed to have achieved high average level of performance and more equitable distribution of outcomes in reading literacy compared to other countries. On the combined reading literacy scale, the mean of Finnish students was 546 points and the standard deviation was 89, while the mean of OECD countries was 500 and the standard deviation was 100. The Finnish average was higher than the international average, and the distribution was narrower than the international distribution. This was reflected in the distribution of students in the five levels of reading proficiency. The highest proficiency level, Level 5 in the combined reading literacy, was attained by 18 per cent of students, which was the second highest percentage among the participating countries and significantly higher than the OECD average of 10 per cent. The lowest proficiency at Level 1 or below was attained by 7 per cent of Finnish students, as compared to the OECD average of 18 per cent. Finnish students seemed to do exceptionally well in retrieving information and in interpreting texts, where Finnish students had the highest mean. In the reflection and evaluation subscale, however, the mean performance was the third highest but still excellent. (OECD 2001; Väljjarvi & Linnakylä 2002.)

Even though the reading literacy skills of Finnish students proved high quality on average, inequities in the distribution of performance still existed. The results of PISA showed that gender differences in reading literacy existed in all participating countries, and girls reached higher level of performance than boys (OECD 2001, 122–123). In Finland, however, the gender difference on the combined reading literacy scale was the largest, 51 points on average, which is about 57 % of the Finnish standard deviation.

The difference, nevertheless, was not due to Finnish boys doing poorly, but rather to Finnish girls performing exceptionally well. After all, Finnish boys scored better than boys in any other OECD country and even better than girls in many of the participating countries (Väljærvi & al. 2002, 30). Reading literacy performance was strongly associated with reading interest and activity. Girls in Finland were more interested and engaged in reading than boys. Boys also reported spending much less time in reading for enjoyment than girls did, and girls were more active in reading fiction. The gender difference in reading literacy, after controlling for the effects of reading interest and activity factors, reduced by about 70 % to only 16 points, compared to the unadjusted gender difference (Linnakylä and Malin 2003).

In PISA, students were asked about their parents' occupations and the activities associated with those occupations. The socio-economic index of occupational status (Ganzeboom & al. 1992) was derived from students' responses on parental occupation. Values on the index ranged from 16 to 90, low values representing low socio-economic status and high values representing high socio-economic status. The index was based on either the father's or mother's occupation, whichever was higher.

The PISA results indicated that socio-economic background was strongly associated with reading literacy performance in all participating countries, but this effect also differed considerably between countries (OECD 2001, 2002a). In addition to the effect of students' socio-economic background on student performance, the schools' mean socio-economic background had an effect as well (OECD 2001). However, in Finland it was one of the smallest among the participating countries, and Väljærvi and Malin (2003) did not find a statistically significant effect of the schools' mean socio-economic background in Finland. Though the total effect of the socio-economic background was one of the smallest in Finland and the effect at school level was not self-evident, the association of the combined reading literacy score with socio-economic index at student level was obvious, and students with highest socio-economic background outperformed those with lower socio-economic status (Väljærvi & al. 2002, 33).

Students in PISA were also asked information about two other important family background characteristics: parents' education and family wealth. Due to the construction of the socio-economic index of parents' occupational status, it already covered part of the total variation in parents' education and family wealth. In Finland, the correlation between the socio-economic index and parents' education was 0.50, and the proportion of the variance in parents' education associated with the socio-economic index was about 25 %. The correlation between the socio-economic index and family wealth

was 0.25 and between parents' education and family wealth it was 0.22, and the overlap in variance was about 7 % and 5 %, respectively. Though these three variables were correlated and they had overlap in variance, there is still potential explanatory power left for the parents' education and family wealth, after controlling for the effect of the socio-economic index.

The parents' level of education was a significant source of disparities in student performance. The parents' education was defined in PISA on the basis of the International Standard Classification of Education (ISCED). According to PISA results, in all countries students whose mothers had completed upper secondary education achieved higher levels of performance in reading than students with less educated mothers, and in most countries the mother's completion of tertiary education gave a further advantage (OECD 2001).

The national analyses of PISA in Finland also showed that students whose mothers had second or third level education had better results in reading test compared to the students whose mothers had only basic education (Linnakylä 2002). Even though the parents' educational background was not as a significant factor in Finland as in some other countries, high parental level of education still had positive effect on reading performance. The Finnish results of the Second International Adult Literacy Survey (SIALS) even showed that the father's and mother's educational background had a strong correlation with their children's literacy level even as adults (Linnakylä et al. 2000).

Relative wealth of the family is generally an advantage, since wealthier people can afford, for example, access to more resources. To assess the relationship between family wealth and the performance in PISA, reports were obtained from students on the availability of various items in their homes (OECD 2001, 223). The items asked were a dishwasher, a room of their own, educational software, a connection to the Internet, and the number of cellular phones, television sets, computers, motor cars and bathrooms. A composite index of family wealth was derived from these reports. For the OECD countries, the average of this index was set to zero and the standard deviation to one.

The relationship between family wealth and performance in PISA was generally positive. Students from wealthier families typically did better than students from less wealthy families, although the pattern was less pronounced than in the case of parental occupational status. The relationship between family wealth and student performance was comparatively weak in the Nordic countries. In Finland, the average performance in the lowest national quarter of family wealth was clearly above the OECD average

performance, and the difference in performance averages between the lowest and highest national quarters was only 21 points (OECD 2001). Although family wealth seemed to have a small positive association with reading literacy performance in Finland, this association was, however, comparatively weak, and weaker than the association between reading performance and the socio-economic status of the family.

The students' immigration background was one source of systematic variation in reading literacy. The PISA results showed that in most countries with significant immigrant populations, first-generation students, i.e. students who were born in the country, where the assessment took place, but whose parents were born in another country, read clearly below the level of native students. The non-native students, i.e. students who were born outside the country where the assessment took place and whose parents were also born in another country, tended to lag even further behind native students than do first-generation students. The amount of non-native students in Finland was extremely small, only 1 % of the students participating in the assessment. Their mean score on the combined reading literacy score was 468 (OECD 2001), which is more than one reading literacy level below the national mean.

The national PISA findings indicated that some regional variation on the combined reading literacy score in Finland existed, but that it was small. For the sampling of PISA, Finland was divided in five regional units, according to The Nomenclature of Territorial Units for Statistics (Nomenclature des Unités Territoriales Statistiques, NUTS), established by Eurostat (see Appendix 1). The average performance scores of these regional units in the combined reading literacy score varied between 543 and 555 points, the difference being only 12 points. The highest mean values were in the Southern and Northern Finland and the differences between regional units were not statistically significant (Väljörvi, Linnakylä & al. 2001). Interesting, however, was, that after controlling for the effect of the socio-economic background at student and school level the Uusimaa Region along the southern coast around the capital city, Helsinki, had statistically significantly lower average than other parts of Finland, the difference with other regional units being between 18 and 26 points (Väljörvi and Malin 2002). The difference between the means of urban and rural municipalities were small, and it seemed that boys in rural areas especially in Northern Finland, but also in Mid-Finland and Eastern Finland, had lower average performance than in the Southern Finland and Uusimaa Region (Väljörvi, Linnakylä & al. 2001).

In most PISA countries, a considerable portion of the variation in student performance was attributable to *the differences between schools*. On average, across the OECD

countries, differences between schools accounted for 36 % of the OECD average total student variance on the combined reading literacy scale. However, in several countries, even over half of the variation in student performance was attributable to differences between schools. The intra-class correlation on the combined reading literacy scale was over 0.50 in nine PISA countries: Hungary (0.67), Poland (0.62), Austria (0.60), Belgium (0.60), Germany (0.59), Italy (0.55), Czech Republic (0.53), Mexico (0.53) and Greece (0.51) (OECD 2002a). In these countries, differences between schools accounted for more than 50 % of the total student variance on the combined reading literacy scale. (OECD 2001.)

Between-school variation was very small in the Nordic countries compared with the other OECD countries. In Denmark the variation was greater than in Finland, Iceland, Norway and Sweden, but even there differences between schools were remarkably smaller than in the OECD countries on average (Väljörvi & Malin 2003). The smallest intra-class correlations on the reading literacy scale were in Iceland (0.08), Sweden (0.09), Norway (0.10) and Finland (0.12). The highest ICC among the Nordic countries was in Denmark (0.19). (OECD 2002a, 246.)

Small school differences in Finland mean that students' performance was largely unrelated to the schools in which the students were enrolled. This is due to the fact that Finland has non-selective school system that provides all students with the same comprehensive schooling, usually in the nearest school. Even the least successful schools attained a relatively high level of reading literacy achievement, when compared to other countries (Väljörvi & al. 2002). Still, we cannot ignore the fact that there existed differences between individual schools.

From the equality point of view the variation between schools cannot be considered insignificant even in Finland. In OECD countries on average the difference between the best and poorest performing quarters of the schools was 146 points on the combined reading literacy scale, i.e. two proficiency levels. In Finland this difference was 67 points, which was close to one proficiency level (Väljörvi & Malin 2003). Thus, ensuring equal educational opportunities for all children and young people still seems to be a central challenge of education policy in Finland, though in most other PISA countries this challenge is even bigger.

From the PISA 2000 results, we can conclude, that the highest inequity in reading literacy in Finland was associated with gender. The inequities associated with socio-economic and educational background of the family were clear as well. The regional differences were small, and family wealth was only slightly associated with differences

in reading as well. School differences in performance were real, but small in comparison with other countries.

2.2 Finnish school differences and inequities in national studies

In recent years, differences between schools have become a focus of several studies also in Finland. The task of school achievement studies in general has usually been to investigate the student level associations between the achievements and the background factors, like gender, place of residence and family background. This means that in sampling the purpose has been to achieve a representative sample of students, not of schools. And when the between-school variation has been studied, the interest has focused on unadjusted school differences, without studying the between-school variation after controlling for the known student and school level effects. In some recent studies also the schools have been considered as units of statistical analyses and sampling has usually been designed to consist of a representative sample of schools as well, not only of students.

Kuusinen (1985, 1986, 1992) showed that the parent's socio-economic background, in addition to the student's ability measured at the beginning of school or just before it, had a strong influence on the student's school achievement in the comprehensive school. A good family background was a better guarantee for success at school than the ability. The results indicated that school favoured students with higher socio-economic background and discriminated students with lower socio-economic background. The object of equity in education was not fulfilled, especially, if the educational careers after comprehensive school were considered.

One of the first national efforts to estimate the between-school variation in educational achievements in the Finnish comprehensive school using developed statistical methods was done by Malin and Salmela (1993). They analyzed data sets from a national assessment project of the comprehensive school. The data was gathered in 1990–1991 and consisted of several subjects. The aim of the study was to find out, how large the school differences were in educational outcomes in Finland. The method used was a simple two-level model, which divides the total student variance of the response variable in two parts, describing the between-school and the within-school variation.

In 1994, the new curriculum was put to effect in Finland, and the schools were given more independence in deciding their own curriculum and instructional methods.

The study of Malin and Salmela (1993) was motivated by two reasons: to document the differences between Finnish schools in the beginning of the 1990s before important changes were made in the educational system, and to introduce the statistical methodology for estimating school differences in the Finnish context.

Malin and Salmela (1993) found, that the estimates of school differences in Finland varied according to the subject. Of the total student variance, 2–17 % was attributable to the differences between schools, depending on the subject. The smallest school differences were found in reading literacy. The between-school variance component was 6.0 % of the total student variance at grade 3 and 1.9 % at grade 8 in Finnish-speaking schools, and 4.7 % and 2.8 % in Swedish-speaking schools. The number of Finnish-speaking students who participated in the study was 1552 at grade 3 and 1279 at grade 8, representing 71 schools at both grades. In Swedish-speaking schools 1244 students from 73 schools at grade 3 and 3288 students from 40 schools at grade 8 participated.

According to Malin and Salmela (1993), the largest school differences were in English language and Swedish as a second language, 13.9 % and 17.3 %, respectively. In mathematics, the between-school variance component was 6.7–12.1 % of the total student variance, depending on the grade. In science the estimated between-school variance component varied between 2.8 % and 12.5 % depending on grade and precise subject, but the amount of students was much smaller than in other subjects and the estimates were not as reliable.

Although the samples were not designed to study school differences but to be representative student samples and the number of schools may be considered too small, the results of Malin and Salmela (1993) showed, however, that the school differences were quite small in Finland in the beginning of 1990s.

Saari and Linnakylä (1993) also concluded from the same national assessment data of the comprehensive school that differences between schools and teaching groups or classes within schools were minimal in Finland, and the students' opportunities to learn in different parts of the country were quite equal. In the international comparison, they saw the school differences to belong to the smallest in the world. This was not surprising, since in the beginning of the 1990s, the schools did not differ very much from each other with respect to the way the schooling was organized. They concluded that the Finnish comprehensive school was of high level and also fulfilling the principle of equity in the sense that the place of residence or the school itself did not affect the average level of educational achievement. There was also some indication that the com-

prehensive school had evened the differences between students in reading literacy (Linnakylä 1995).

Malin and Linnakylä (2001) used multilevel modelling to study the school satisfaction and changes in it between 1991 and 1995 among students at grade 8 in the comprehensive school. They found some between-school variation in school satisfaction at both time points, even after the effects of several factors associated with school satisfaction were controlled for. The school satisfaction was very stable at the school level, and real changes were observed only in few schools during the four year period.

The national analyses of the Finnish Third International Mathematics and Science Study (TIMSS) data showed no gender difference in mathematics, although in science boys were little better. No differences between regions were found, but in rural schools the average results were somewhat lower than in urban schools. Also the family background had an effect on the results: good educational resources at home were associated with better results. (Kupari & al. 2001.)

The Finnish National Board of Education has regularly arranged national assessments since 1998. The summary report of eight national assessments during the years 1998–2001 is presented by Jakku-Sihvonen and Kuusela (2002). The aim of these assessments has been to give information on how well the aims set in the national framework curriculum have been fulfilled. These assessments were based on national student and school samples, with typical student sample size around 4500 students with 5–7% of the students in the age group taking part in each assessment. Assessments were made on the final stage of basic education at grade 9, when students in the secondary schools are aged 15 and 16 and undergoing a transition phase leading to either vocational training or senior secondary schools. The subjects in the assessments were science, mathematics (2), English, first-language Finnish (2), and Swedish (2).

In the summary report, differences in the results between schools were identified. About ten per cent of the differences in students' results were interpreted as variation between schools (Kuusela 2002, 31). The proportions of total variation in the students' learning-based results attributable to schools were: mathematics 8–11 %, science 11%, Finnish as the first language 9–11 %, English 11 %, and Swedish as a second language 15 %. In all assessments except mathematics and science girls were performing better than boys. Equality between boys and girls seemed to be more real in the better performing schools of the capital city area than elsewhere. Regional differences were also found. School-specific learning results tended to vary, to some extent systematically between regions. (Kuusela 2002.)

The results showed that good results are linked to factors that define regional welfare. Such factors included the population's level of education, the numbers of people employed in management or working in business or finance. Also, high income levels and low rates of communal tax appeared to be related to good learning-based results. Good results were on average achieved more often in wealthy and densely populated areas. The results demonstrated an association between school results and variables that described the welfare of the region. (Kuusela 2002.)

In the national assessments reported by Lappalainen (2000; 2001) the learning performance of students in their mother tongue, Finnish, was evaluated at the end of their comprehensive education. Proportion of between-school variation of the total variation was 9 % in the year 1999 and 11 % in 2001. The general level of reading skills was described as relatively good. The difference in test results between boys and girls was substantial for the benefit of girls, and regional differences were also found.

The command of English among 9th grade comprehensive school students was evaluated in a nationwide study carried out in 1999 by Tuokko (2000) with respect to the objectives of the curriculum. In this assessment, girls were better than boys in all subsectors of language skills. Swedish-speaking students were also better than Finnish-speaking students. A relatively significant difference between provinces was identified, but the type of municipality did not seem to affect performance. One conclusion was that differences between schools seemed to have grown since the beginning of the last decade.

Also in a recent report about learning to learn in the lower secondary schools (Hautamäki & al. 2003), the between-school variation in students' skills seemed to be of the same size as in other national assessments. The proportion of the between-school variance component was 8 % of the total student variance, while the differences between schools explained 11 % of variance.

2.3 Finnish school differences in international studies

In international comparisons, the Finnish school differences in student achievements are found to be among the smallest in the world. The IEA Study of Reading Literacy was conducted in 32 educational systems in the period 1989 to 1992. In the Finnish population of 9-year-olds at grade 3, the intra-class correlation was 0.07. Thus, the approximate amount of differences in reading literacy achievement attributable to the between-school variation was only seven percent. The highest proportion of variation in

student achievements attributable to school differences in this study was 30 percent in Indonesia (Postlethwaite & Ross 1992; Linnakylä 1995). Also in the IEA study of science, Finland was one of the most homogenous countries in the achievements (Postlethwaite & Wiley 1992).

Lietz (1996) studied changes in reading comprehension across cultures and over time analyzing the data sets of two international reading literacy studies conducted by the International Association for the Evaluation of Educational Achievement (IEA). The Reading Comprehension Study was undertaken in 1970/71 and the Reading Literacy Study was administered in 1990/91. Eight educational systems participated in both studies at the secondary school level, Finland among them. The populations of both studies were 14-year-old students at the secondary school level.

Among various methods, Lietz (1996) applied multilevel modelling in the statistical analyses and estimated the proportion of between-school variance from the total student variance. In Reading Comprehension Study in 1970/71, the estimated proportion of the between-school variance component in Finland was as high as 35 % of the total variance. For Sweden this was only 5 %. In the Reading Literacy Study in 1990/91 the proportion of the between-school variance component was only 2 %, and in Sweden it was 8 %. During the time of the first study, the comprehensive school had not yet been established in Finland.

2.4 Conclusions

There is clear evidence that educational achievements in the Finnish comprehensive school depend on several factors describing the student's background, like gender, region and the family's socio-economic background, and also the schools themselves. Reading differences were found to be strongly related to gender. Even though Finnish boys achieved the highest literacy level, when compared to boys in the other PISA countries, their underachievement in comparison with Finnish girls was evident. In addition to the student level effects in background factors, also aggregated level effects of several regional factors seemed to exist. There are some obvious problems in students' equal opportunities for learning in the Finnish comprehensive school.

Even though the international comparison showed that the school differences seemed to be small in Finland, from the equality point of view the variation between schools cannot be considered insignificant. Detailed research is needed to find out possible factors which explain the school differences, since the individual schools have

the most essential role in striving for the educational objectives of high levels of performance with equitable distribution of outcomes. Ultimately, the success in fulfilling the principle of equity is realized in individual schools. Some worry is also raised by the results from the national assessments, where differences between schools are constantly larger than in the international comparisons.

The family background does not only influence students' school achievement, but it has consequences to the students' future. The link between family background and children's educational career still exists also in Finland (Isoaho & al. 1990; Kivinen & Rinne 1995; Laiho & Silvennoinen 1992; Stenström 1997). The influence of social background, measured by the level of educational attainment of the father, on the educational career of young people is clear. Father's educational level has a strong impact on the decision regarding the route of education for his children. The higher the father's educational attainment level is, the more likely his children are to continue in general upper secondary education, the less likely they are to choose a vocational education or no further education at all, the longer they will continue their studies, and the more likely they are to go to university (Haven 1999).

But there are also signs that the inequities may change during the adult life. According to the Finnish results of the Second International Adult Literacy Survey, the gender differences in literacy proved fairly small in Finland, although women had a slight advantage in prose literacy while the opposite was true for quantitative literacy. From the international perspective the situation in Finland speaks of great equality in this respect. On the other hand, among the various background factors explored in SIALS the extent of initial education proved to be the most important predictor of the level of adults' literacy skills in Finland. The longer the formal education attained, the higher was the performance level in the literacy tests. (Linnakylä & al. 2000.)

Literacy skills seemed to be related to employment as well. The literacy performances of the unemployed were clearly lower than those of the employed. The results also indicated that poor readers face a higher risk of unemployment than do other adults. Literacy skills were related to income levels, as well. (Linnakylä et al. 2000.)

Since literacy bears a relation to employment and a successful work career, it is not without meaning, what kind of systematic differences the students leaving the comprehensive school carry with themselves, when continuing their educational career and entering into work life. In this respect, it seems extremely important to promote the equity in the comprehensive education by evening out the obstacles of the educational career and future life.

3

Research problems

The aim of this study is to investigate the distribution and relative equality of educational outcomes in Finland, measured as reading literacy performance in the PISA 2000 data. The distribution of educational outcomes is examined both at the student and school level. The purpose is to find out, how large the school differences in reading literacy performance are in Finland, how strong the effects of selected background factors on reading literacy performance are, and how much of the between-school as well as the total student variation is attributable to these factors.

The background factors are divided into two groups. The first group consists of several factors, called the inequity factors, and the second group includes only one factor, which is the engagement in reading. The inequity factors describe different student and school characteristics which are known to introduce variation in the distribution of educational outcomes. The inequity factors are: language of the school, grade the student attends, student's immigrant background, location of the school as a regional unit and municipality type, gender of the student, socio-economic status of the family, parents' educational level, and wealth of the family. The engagement in reading score describes the students' own reading activities outside the school, and it is used to test, if the effects of the inequity factors can be reduced by students' own activities. Detailed description of these factors is presented in chapter 4. The important question is: how much of the distribution of reading literacy performance is attributable to the inequity factors, and how much engagement in reading modifies the effects of the inequity factors.

In estimating the school differences in educational outcomes, the estimation of only unadjusted between-school variation, without controlling for the effects of known sources of this variation, does not tell the whole truth of the differences. As important, or maybe even more important, is to distinguish the proportion of total variation attributable to the known sources of variation, both at the school and student level. If the differences between schools arise from the factors which are not under schools' or students' control, it is fair to adjust the school comparisons on these effects. The school differences which are cleaned from the effects known to increase inequities in the distribution of educational outcomes may better describe real school differences.

The problems addressed in this study are expressed in the following six questions:

1. How large is the between-school variation in reading literacy performance in Finland?

To answer the first question, the Finnish PISA 2000 data of reading literacy is examined to establish a proper and reliable estimate of the unadjusted between-school variation and school differences in reading literacy performance. This is done by means of the two-level variance component model which divides the total student variance to the between-school and within-school variance components. The intra-class correlation, which in this case could also be called the intra-school correlation, is used as an estimate of between-school variation. It estimates the amount of the total student variation which is attributable to the differences between schools in their average performance. The variation between schools in Finland is compared to the equivalent variation in Sweden and Germany. Sweden was chosen, since it has a very similar educational system and comprehensive school to Finland, while in Germany the students are divided in differentiated school types during the basic education.

2. How much of the between-school and total student variation in reading literacy performance is attributable to each of the inequity factors?

To answer the second question, reading literacy performance is modelled as a function of each inequity factor separately, by means of the two-level model. The effect of each factor on reading literacy performance is studied with emphasis on the amount of between-school and total student variation it explains. As a result, the unadjusted effect of

each factor is obtained. The amount of explained between-school and total student variance is considered to be an estimate of systematic variation that each inequity factor alone introduces to the variation of reading literacy performance. By comparing the variance components of two-level one-factor models to the baseline model, estimates of the between-school variation attributable to each inequity factor separately are derived.

3. How much of the between-school and total student variation in reading literacy performance is attributable to the inequity factors together?

After studying the effect of each inequity factor separately, they are all included in the same model to explain the variation in reading literacy performance. As a result, adjusted effects of the inequity factors are obtained. The amounts of the between-school and total student variation attributable to the inequity factors together are estimated from the reduction in the variance components, in comparison to the baseline model. An estimate of adjusted school differences, corrected by the effects of the inequity factors, is obtained. The result of this phase of the analysis is a statistical model called the "inequity model". Comparisons with equivalent Swedish and German models are carried out.

4. How large is the between-school and total student variation in reading literacy performance associated with engagement in reading?

The effect of engagement in reading on reading literacy performance is first studied separately with the two-level one-factor model, and the unadjusted effect of the factor is obtained. The reductions in the between-school and total student variation are estimated to find out how much of the variation is associated with engagement in reading. The school differences, adjusted for engagement in reading, are derived from the variance components of the one-factor model, again comparing them to the baseline model.

5. Does the effect of engagement in reading modify the effects of inequity factors on reading literacy performance?

To answer the question five, engagement in reading is included in the inequity model developed under question three. The effects in the model, now called the "extended

inequity model", are studied to find out, if controlling for the effect of engagement in reading modify, change or even reduce the effects of inequity factors. Thus, the effect of each inequity factor is adjusted to the effect of engagement in reading as well as other inequity factors. Comparing the extended inequity model to the inequity and baseline models is the basis in estimating, how much the engagement in reading reduces the between-school and within-school variation in reading literacy performance.

6. Is there still, and how much, between-school and total student variation in reading literacy performance, after controlling for the effects of the inequity factors and engagement in reading?

The final question can be answered on the basis of the extended inequity model. The unexplained between-school and total student variance components are used as estimates of the adjusted between-school and total student variation after controlling for the effects of inequity factors and engagement in reading. The amount of variation explained by all the factors is the part of the variation in reading literacy attributable to inequity factors and engagement in reading together. Finally we can conclude, if there is real between-school variation in the reading literacy performance in Finland, when the effects of important sources of variation, the inequity factors and engagement in reading, are controlled for.

II

DATA AND METHODS

4

Data and statistical analyses

This chapter describes the data and the variables used in the study, as well as the process of data analyses using statistical two-level models. The data consisted of selected variables from the PISA 2000 data set. The variables measure students' reading literacy performance and students' background. In the statistical analyses, variation in the students' reading literacy performance is explained by background factors connected with inequity, using two-level linear model.

4.1 PISA 2000 data

The data used in this study comes from the PISA 2000 reading literacy study, and the educational outcome is the students' reading literacy performance in the PISA test. The framework and design of the PISA study was reported in two international publications (OECD 1999, 2000).

PISA 2000 was a paper and pencil test of 7 hours, 2 hours for each student. The test items were a mixture of multiple-choice items and tasks requiring students to construct their own responses. The items were organised in units based on a text passage setting out a real-life situation. The study included 37 reading units with 141 items. The students also answered a 30-minute background questionnaire, with questions about their home background, reading interests and attitudes towards school. School principals were given a 30-minute questionnaire asking about their schools.

The Finnish PISA 2000 data consisted of a representative sample of 15-years old students and schools they attended, and it consisted of 4864 students and 155 schools. The size of the school population was 832 and the student population about 66 000. In Finland, like in most PISA countries, stratified two-stage cluster sampling was used. The strata were based on regional units, municipality type and language of the school. The first stage consisted of the sampling of individual schools in which 15-years old students were enrolled. Schools were sampled with systematic PPS sampling, the measure of size being a function of the number of 15-years old students enrolled. The population of schools was regionally stratified before sampling operations. All the sample schools and about 93 % of the sampled students participated in the study. (Malin & Puhakka 2002; OECD 2002b.)

4.2 Reading literacy performance score

The outcome variable in this study is the reading literacy performance score. There were two kind of reading test scores measuring students reading performance available in the PISA 2000 data: plausible values and weighted likelihood estimates.

As is the case with all item response scaling models, student proficiencies of reading literacy were not directly observed. Reading literacy performance scores are like missing data that must be inferred from the observed item responses. There are several possible alternative approaches for making this inference. PISA used two approaches: Maximum likelihood, using Warm's (1985) Weighted Likelihood Estimator (WLE), and plausible values (PVs). The WLE approach estimates the actual score that the student attained the most likely, while PVs are a selection of likely proficiencies for students that attained each score (OECD 2002b, 105). Using item parameters anchored at their estimated values from the international calibration, the plausible values are random draws from the marginal posterior of the latent distribution for each student. For details on the uses of plausible values, see Mislevy (1991) and Mislevy *et al.* (1992). (OECD 2002b, 106.)

It is very important to recognise that plausible values are not test scores and should not be treated as such. They are random numbers drawn from the distribution of scores that could be reasonably assigned to each individual – that is, the marginal posterior distribution. As such, plausible values contain random error variance components and are not optimal as scores for individuals. Plausible values as a set are better suited to describing the performance of the population. This approach, developed by Mislevy and

Sheehan (1987, 1989) and based on the imputation theory of Rubin (1987), produces consistent estimators of population parameters. (OECD 2002b, 107.)

Plausible values are intermediate values provided to obtain consistent estimates of population parameters using standard statistical analysis software such as SPSS and SAS. The PISA student file contained five plausible values for the combined reading literacy scale. It is recommended that the set of plausible values is used, when analysing and reporting statistics at the population level (OECD 2002b, 107). On the other hand, weighted likelihood estimates can be treated as (essentially) unbiased estimates of student *abilities*, and analysed using standard methods (OECD 2002b, 106).

In this study, the weighted likelihood estimates were used as the outcome variable measuring students reading literacy performance. The aim of this study was not to report statistics at the population level, but to find out associations and relationships between the outcome variable and the background factors using statistical modelling. The background factors were answers to direct questions from individual students (or derived from them) and hence individual student scores, not random numbers drawn from the distribution of scores that could reasonably be assigned to each student. Therefore, the response variable chosen to be used is the variable describing the performance of the same individual student, i.e. the weighted likelihood estimate of reading literacy performance, and not the random numbers drawn from the distribution of scores that could be reasonably assigned to the student. Since the test score used in this study was based on the whole range of reading literacy tasks, not on any subset of them, it is called the combined reading literacy score.

The weighted likelihood estimates were transformed to have a mean of 500 and a standard deviation of 100 by using the data of the participating OECD member countries only (OECD 2002b, 255).

4.3 Inequity factors

The background factors used in the study are of two kinds: a group of factors describing the possible sources of inequities in educational outcomes, called inequity factors, and a factor supposed to modify the effects of the inequity factors, describing the students' reading interest and activities.

There are nine inequity factors whose influence on reading literacy performance was studied. Three of them, language of the school, location of the school as regional unit and municipality type, are school level variables which have no within-school var-

iation. They have the same value for all the students in the school. Six of the inequity factors are student level variables: grade, immigrant background of the student, gender, socio-economic status of the family, parents' educational level, and wealth of the family. These factors vary also within each school.

Language of the school

There were 8 Swedish-speaking schools with 242 students in the Finnish PISA 2000 sample (Malin & Puhakka 2002). This was about 5 percent of the total amount of sampled schools and students, which is very close to the proportion of Swedish speaking population in Finland (5.5 %). This background factor was included in the study to test and estimate the difference between the two language groups. The difference was modelled using a dummy variable with Finnish speaking schools as the reference category with value 0, and Swedish-speaking schools having value 1. The language of school was also benefited in sampling stratification. However, the amount of Swedish speaking students in the sample was not large enough to make a very precise estimate of the population difference between the language groups, but the language of the school was included in the model to control for the effect of a possible difference in the student variation.

Grade

The sampling procedure in PISA was not based on grade but on age. In Finland, most of the 15-years old students attend grade 9. However, about 11 % of the sampled students were below this level, 520 students at grade 8 and 10 students even at grade 7. The grade difference was also modelled using a dummy variable, with students at grade 9 as a reference category with value 0. To test and estimate the difference between the students at grade 9 and below it, in practice at grade 8, the students below grade 9 were combined into one category with value 1.

Immigrant background of the student

In PISA 2000, students were grouped according to the immigrant background into three categories: native students, who were born in the country of assessment and who had at least one parent born in that country, first-generation students, who were born in the

country of assessment, but whose parents were born in another country, and non-native students who were born outside the country of assessment and whose parents were also born in another country (OECD 2001, 220–221).

The amount of students with immigrant background was extremely small in the Finnish sample. Only 46 students, about 1 % of the students in the Finnish sample, were non-native. The amount of first-generation students was even smaller, only 11 students. The first-generation students were combined with native students in the same category, called native students. According to the recoding, students were called native, if they were born in Finland and non-native, if they were born outside Finland. The effect of students' immigrant background was estimated and tested using native students as the reference category having value 0, and the non-native students were coded as 1.

Location of school: Regional unit

At the time of PISA 2000, Finland was divided in five regional units, according to The Nomenclature of Territorial Units for Statistics (Nomenclature des Unités Territoriales Statistiques, NUTS), established by Eurostat. The five regions were: Uusimaa Region, Southern Finland, Eastern Finland, Mid-Finland, and Northern Finland. This regional partition is presented in Appendix 1 and it was benefited in the sampling procedure. The regional variation was estimated and tested by four dummy variables. The Uusimaa Region is the most developed part of the country along with the southern coast around the capital city Helsinki, and it was used as the reference category with value 0. For each of the other four regional units, a dummy variable with value 1 was created. These variables were used to test, if the four regions differed from the Uusimaa Region in their average performance.

Most of the Swedish speaking schools (5 of 8) in the sample were located in Uusimaa Region. Accordingly, 64.5 % of Swedish speaking students in the sample were from the same part of the country. Rest of the Swedish speaking schools and students were from Southern Finland and Mid-Finland.

There were students at grade 8 in almost every school in the sample. Only in 8 schools (5 % of the sampled schools) there were students only at grade 9. Students with immigrant background came from 28 schools, and 61 % of these students (28 of 46) came from Uusimaa Region. There were only 8 schools with more than one student with immigrant background in the sample.

Location of school: Municipality type

Another variable describing the location of the school is the type of municipality, where the school was located. Finland is divided in 444 municipalities, forming local authority areas. Municipalities were divided into three categories: urban, densely populated and rural. The criteria used to classify municipalities are developed by Statistics Finland. The main criterion is the population density in the municipality. In this study, the urban and densely populated municipalities were combined into one category. If at least 60 % of the community population lives in densely populated parts of the community, the community is considered to be an urban or at least densely populated. Rural communities are more sparsely inhabited areas, where less than 60 % of the population lives in densely populated parts of the community.

According to the type of the municipality, the schools were divided into two groups: urban schools and rural schools. Hence, the rural schools were located in sparsely populated and urban schools in densely populated areas of the country.

All the Swedish speaking students and schools in the sample were from urban areas, as well as all the students with immigrant background. The same proportion of students, about 11 %, was at grade 8 both in urban and rural schools.

Gender

In gender comparisons, boys were used as the reference category and coded into 0. Accordingly, girls were given the value 1 in this dummy variable.

Socio-economic status

In PISA 2000, the students' socio-economic status was measured by the occupational status of their parents. Students were asked to report their mothers' and fathers' occupations, and to state whether each parent was in full-time paid work, part time paid work, not working but looking for a paid job, or "other". The open-ended responses were coded in accordance with the International Standard Classification of Occupations (ISCO 1988). The socio-economic index of occupational status (Ganzeboom & al. 1992) was derived from the students' responses on parental occupation. The index was based on either the father's or the mother's occupation, whichever was higher. (OECD 2002a, 187.)

Values on the index ranged from 16 to 90, low values representing low socio-economic status and high values representing high socio-economic status. To make the interpretations easier, the socio-economic status index was rescaled so that the average in OECD countries was 0 and the standard deviation was 1. On the original scale, the OECD average was 48.9 and standard deviation was 16.3, and in Finland they were 50.1 and 16.2, respectively. The empirical values of the rescaled socio-economic index in Finland varied between the values -2.01 and 2.52 with mean 0.08 and standard deviation 1.00 . School's socio-economic index was derived by averaging the student indices of the school. The empirical value of the school's socio-economic index in Finland varied between -0.71 and 1.28 with mean 0.09 and standard deviation 0.40 .

Parents' educational level

Students were asked to classify the highest level of education of their mother and father on the basis of national qualifications, which were then coded in accordance with the International Standard Classification of Education (ISCED 1997) in order to obtain internationally comparable categories of educational attainment. The resulting categories were (OECD 2001, 221):

1. Did not go to school.
2. Completed ISCED Level 1, primary education. In Finland this means 6 years of comprehensive school.
3. Completed ISCED Level 2, lower secondary education. In Finland this means 9 years of comprehensive school, which is the extent of compulsory schooling.
4. Completed ISCED Level 3B or 3C, upper secondary education, aimed in most countries at providing direct entry into the labour market. In Finland this means three years of vocational education after the compulsory schooling.
5. Completed ISCED Level 3A, upper secondary education, aimed in most countries at gaining entry into tertiary education. In Finland this means three years of education, after the compulsory schooling.
6. Completed ISCED Level 5A, 5B or 6, tertiary education. In Finland this means education in universities and polytechnics.

A short description of the Finnish education system is found in Välijärvi & al. (2002).

For the analyses of this study, parental level of education was based on either the father's or the mother's education, whichever was higher. The effect of parental education was again estimated and tested using dummy variables. The reference category was

the completed vocational education (ISCED Level 3B or 3C). Due to small amount of cases, category 1 was combined with category 2. In statistical modelling, categories from 1 to 3 were combined to form a group, where parents had completed at the most comprehensive school. Categories 5 and 6 were combined to form a group, where parents had completed at least academically oriented upper secondary level education. A variable describing the school's parental level of education was derived by averaging the original category values of the students in the school.

Wealth of the family

The PISA index of family wealth was derived from the students' reports on the availability of a dishwasher, a room of their own, educational software, a connection to the Internet, the number of cellular phones, television sets, computers, motor cars and bathrooms at home (OECD 2001, 223). Positive values indicated more wealth-related possessions and negative indicated fewer wealth-related possessions. This index was standardised so that the mean of the index value for the OECD student population was zero and the standard deviation was one, countries being given equal weight in the standardisation process (OECD 2002b, 218).

The empirical values of the PISA index of family wealth varied between -2.93 and 3.38 with mean 0.23 and standard deviation 0.72 in Finland. The school's index of family wealth was derived by averaging the student indices for each school. The empirical value of the school's index of family wealth varied between -0.20 and 1.03 with mean 0.23 and standard deviation 0.21 .

4.4 Engagement in reading

In addition to the inequity factors described above, one factor describing the students reading interest and activities outside school was added into the model. This factor, called engagement in reading, is supposed to modify the effects of the inequity factors (Guthrie & Wingfield 2000).

The PISA index of engagement in reading was derived from the students' level of agreement with the following statements: I read only, if I have to; reading is one of my favourite hobbies; I like talking about books with other people; I find it hard to finish books; I feel happy, if I receive a book as a present; for me, reading is a waste of time; I enjoy going to a bookstore or a library; I read only to get information that I need; and, I cannot sit still and

read for more than a few minutes. A four-point scale with the response categories 'Strongly disagree', 'Disagree', 'Agree' and 'Strongly agree' was used. The engagement in reading index was standardized so that the mean of OECD countries was 0 and standard deviation was 1. The Finnish mean was 0.20 and the Finnish standard deviation was 1.05.

4.5 Process of data analyses

Due to the choice of the response variable, one student out of 4864 was left out of the data, because of incomplete responses in the reading tasks and the impossibility of deriving the likelihood estimate of reading performance. One of the schools was a special education school and it represented a different student and school population than the rest of the schools, and this school was excluded from the baseline model and the rest of the statistical analyses.

The process of data analyses consists of four subsequent parts. It starts (Chapter 6) with the aim to estimate the amount of between-school variation in reading literacy in Finland and to establish the baseline model for later comparisons.

The second part (Chapter 7) of the statistical analyses consists of two-level one-factor models. The effect of each inequity factor on the reading literacy performance and its variation is studied in separate models to find out, how these factors relate separately to reading literacy performance and how much they explain of the between-school and total student variation. The estimated effect of each factor is thus unadjusted, since each factor is the single explanatory variable in the statistical model. These models also show, how much of the between-school and total student variation is attributable to each individual inequity factor.

However, the inequity factors do not operate in isolation from each other. In the third cycle (Chapter 8) of the statistical analyses, the effects and the explanatory power of the inequity factors are studied together after adjusting for their interdependence. The model is built by adding the factors to the model in four consequent blocks:

1. The first block consists of the three factors describing the minority group membership: language of the school, students' grade and immigrant background. Language is a school level variable, while grade and immigrant background are student level variables. Although the effects of these factors are of interest on their own, they are added to the baseline model mainly for controlling purposes. They are known to produce variation in the performances, and the aim is to control this variation to estimate the effects of other factors more accurately.

2. The second block consists of factors describing the location of school: regional unit and municipality type. These variables are school level variables and they are used to estimate the regional variation in the schools' performance, adjusted for language, grade and students' immigrant background.

3. The third block consists only of one student level individual characteristic, gender, which is included in the model, when the effects of minority group membership and location of the school are already controlled for.

4. The fourth block consists of three variables characterising the students' socio-economic background. These variables are: socio-economic status of the family, parents' educational level and wealth of the family. These variables are student level variables. Both the individual as well as the contextual effects of these variables are studied, after having controlled for the effects of the factors comprising the three previous blocks.

The result of the statistical analyses is a statistical model called the "inequity model". This model shows which of the factors are associated with inequities in reading literacy performances, after adjusting for their interdependence, and how strong these effects are. The model also enables the estimation of the amount of between-school and total student variation, attributable to all inequity factors together as well as the adjusted school differences in performance after controlling for the effects of the inequity factors.

When the assumptions of the model were studied, the effect of one single school on the results needed to be investigated in detail. The variance structure of the inequity model was also modelled further. At the end of the chapter, comparisons with equivalent Swedish and German models are presented.

Part four (Chapter 9) of statistical analyses consists of the efforts to find out if engagement in reading modifies the effects of the inequity factors. Engagement in reading is a student level variable. First, the contextual and individual effects of this factor are studied using separate two-level one-factor model. After that, all four factors are added into the inequity model, and the effects of the inequity factors and their changes are studied as well as the reductions in the variance components. This model is called the "extended inequity model". The effect of one single school on the results is again investigated in detail.

Multilevel modelling is applied in the data analyses, and the next chapter consists of a short description of this statistical methodology. The software used to estimate and test the models is MLwiN (Rasbash & al. 2000). For statistical analyses, the student weights were rescaled to have an average of 1.

5

Multilevel modelling of school data

In educational achievement research, designs and analyses that produce valid estimates of between-school variance are needed (Porter and Gamoran 2002). School data have a natural hierarchical structure. In the data, students are divided into the groups or clusters according to the schools they attend. Multilevel statistical models (Aitkin & Longford 1986; Bryk & Raudenbush 1992; Goldstein 1987, 1995, 2002; Hox 1994; Snijders & Bosker 2002) are a powerful technique to study hierarchically structured empirical data. On the other hand, if the hierarchical nature of the data is not taken into account, statistical analyses produce incorrect results. The statistical model is then misspecified, since the existence of an important source of variation, i.e. the data structure, is neglected.

In addition to the hierarchical data structure, a complex sampling design, such as multi-stage stratified cluster sampling, is often used in school studies. The structure of the data is benefited in the sampling design and data collection. Accounting for the sampling complexities is essential for reliable estimation and analysis, and design-based methods are useful for that purpose. The design-based methods regard the clustering effects as a disturbance to estimation and testing, and the aim is to obtain valid results by eliminating these effects. (Skinner, Holt & Smith 1989; Pahkinen & Lehtonen 1989; Pahkinen & Kupari 1991; Lehtonen, Nissinen & Pahkinen 1992; Lehtonen & Pahkinen 2004.)

Although design-based methods are a suitable alternative for analyzing school data, they do not consider the clustering effects as interesting but as a nuisance. Since the estimation and testing of school differences is one of the main interests in this study,

the multilevel modelling approach to analyse the complex school data is preferred. The use of multilevel modelling in the Finnish context is presented in several articles (Malin 1994, 1997; Malin & Linnakylä 2001; Malin, Lehtonen & Pahkinen 1995; Malin & Nissinen 1996; Malin & Salmela 1993; Linnakylä & Malin 2003; Välijärvi & Malin 2002, 2003).

In this chapter, the basic principles of multilevel modelling and the applications used in this study are briefly presented.

5.1 Hierarchical structure of school data

Hierarchical structure is a natural characteristic of school data: students are grouped into schools, and within schools into teaching groups or classes. Students, teaching groups and schools define the different levels in the data, and at all levels, different units of analyses exist. The data usually also include variables of all levels. The hierarchical structure is a characteristic of the population, not a result of the sampling method. Accordingly, a representative sample of the population reflects the structure of the population and represents the data at all levels. A representative sample is not a sample of students only, but it is a representative sample of schools as well.

In school research, it is natural to utilize the population structure in sampling, data collection and statistical analyses. In the PISA 2000 stratified two-stage cluster sampling was used in obtaining the Finnish sample. The first stage consisted of sampling of individual schools in which 15-year-old students were enrolled. Schools were sampled with systematic PPS sampling, the measure of size being a function of the estimated number of eligible 15-year-old students enrolled. The population of schools was regionally stratified before sampling operations. In the second stage, samples of students were selected within the sampled schools. Once the schools were selected, a sampling frame list of each sampled school's 15-year-old students was prepared. From this list, 35 students were then selected with equal probability. All 15-year-old students were selected, if fewer than 35 were enrolled.

In hierarchically structured data, level 1 units which are usually students in the school data are in clusters. The structure of the data is reflected in the intra-class correlation, often abbreviated as ICC and also called intra-unit, intra-cluster or intra-school correlation. Student level variables of school data are usually more or less intra-cluster correlated. In statistical analyses this means, that students are not statistically independ-

ent units, and the statistical methods requiring statistical independence of observations are not applicable.

The dependence of statistical units can be seen as a result of characteristics common to all the units in the same cluster, but varying between clusters. In school data this dependence, which is reflected in the ICC, may be result of the fact that the same teacher is teaching all the students in the teaching group, and the students are exposed to the same 'treatment'. Individual characteristics affected by this group 'treatment' tend to homogenize the individuals in the same group. The same effect may be a result of the school characteristics shared by all the students in the school.

Two types of group level variables may describe groups or clusters. The natural group level variables, like size of the school, are important and interesting variables. If they effect the individual level performance, changing these conditions may change the individual level performance of all the students in the same group. For example, if the teacher starts using a new instructional method, it may improve the individual performances of all the students in the teaching group and consequently raise the average level of the whole group. Differences between individual students may still stay the same.

In addition to natural group level variables, aggregated values can be derived from the group members' individual level variable values. The value of the aggregated variable is the same for all group members, like the school average of socio-economic background index. Raudenbush (1989) call these contextual variables. For example, if the average socio-economic level of the school has effect on individual performances, in addition to the student's own socio-economic status, this is called contextual effect. This group characteristic then affects all the individual performances, in addition to the individual level effect. These variables are also called compositional variables since they measure an aspect of the composition of the school to which the individual student belongs (Goldstein 1995).

Analyzing school data is subject to several demands. The units of analyses, i.e. the students are not statistically independent. There are at least two levels in the data, student and school level, and there are variables describing each level. Often the effects of contextual variables are interesting and they are used in statistical analyses. All these characteristics of the data have to be taken into consideration in the analyses, if we want to get reliable information of the school achievements and the student and school level factors associated with achievements. Multilevel modelling enables this, and, if we want to estimate the amount of between-school variation in addition to the effects of

the background factors, multilevel modelling is an appropriate technique for this purpose.

5.2 Intra-cluster correlation

The starting point in the statistical analyses of school differences is to divide the total student variance of the outcome variable into variance components. If the data has two levels, like in the PISA data, where students are level 1 units and schools are level 2 units, the total variance is divided into two components: the between-school and the within-school variance component. This is done by means of a simple variance component model:

$$\begin{aligned}
 y_{ij} &= \beta_0 + u_j + e_{ij} \\
 u_j &\sim N(0, \sigma_u^2) \\
 e_{ij} &\sim N(0, \sigma_e^2) \\
 \text{cov}(u_j, e_{ij}) &= 0
 \end{aligned} \tag{5.1}$$

where the index i denotes the student (level-1 unit) and j denotes the school (level-2 unit). The outcome variable is y_{ij} , for example the student's reading literacy score like in this study, and the parameter β_0 is the intercept. The school residual u_j is the random effect of school j having mean 0 and variance σ_u^2 , whereas e_{ij} is the student level residual having mean 0 and variance σ_e^2 . The random effects u_j and e_{ij} are assumed to be independent from each other. The normality assumption is usually made for u_j and e_{ij} .

The school residual u_j is the deviation of school j from the intercept β_0 , which can be interpreted as the average school mean. The expected mean of school j is the sum of the intercept and school residual: $\beta_{0j} = \beta_0 + u_{0j}$. The student residual e_{ij} is the deviation of student i from his or her school mean.

Since the residuals at different levels do not correlate, the total student variance σ^2 in formula (5.1) can be divided in two components:

$$(5.2)$$

where σ_u^2 is the between-school variance component and σ_e^2 is the between-student variance component within schools, i.e. within-school variance component. If there

are more than two levels in the data, the amount of variance components increases to correspond the number of levels.

The units in naturally existing clusters tend to be more similar or homogeneous than units selected at random from the population. This means that the level 1 units in the data are not statistically independent, which is reflected in the intra-cluster correlation. The intra-cluster correlation can be used to estimate the proportion of total student variance attributable to mean differences between schools. In two-level data, the ICC is estimated by dividing the between-school variance component by total variance and is expressed in the following form (Goldstein 1995; Snijders & Bosker 2002):

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} = \frac{\sigma_u^2}{\sigma^2} . \quad (5.3)$$

The ICC indicates, how much of the total student variance is attributable to the differences between schools. If there is no between school variation, $\sigma_u^2 = 0$, then the ICC is $\rho = 0$. In this case the school means are equal and they do not differ from the intercept value β_0 in formula (5.1), and all variation of the outcome variable is between-student variation. On the other hand, if there is no within-school variation, $\sigma_e^2 = 0$, then the ICC is $\rho = 1$. In this case, all the variation is between-school variation. This is not likely in the case of student level variable, but self-evident in the case of school level variable.

The ICC has a positive value, if there is variation between schools. The more the school means vary with respect to the total variance, the bigger is the between-school variance component. The statistical null hypothesis then states that $\sigma_u^2 = 0$, with the alternative hypothesis of $\sigma_u^2 > 0$. If the null hypothesis is rejected, the conclusion is that there is between-school variation in the variable, and the ICC estimates the proportion of it with respect to the total student variance.

It is important to notice that the ICC is a correlation between two level 1 units in the same cluster, not a correlation between two variables. It measures how homogeneous the cluster members are with respect to their individual characteristic. The units in naturally existing clusters tend to be more similar or homogeneous than units selected at random from the population, and this is reflected in the positive ICC of the variable.

The intra-cluster correlation has important consequences for the statistical analyses of the data. If the level 1 units are homogeneous and their ICC is positive, the effective sample size of level 1 units is smaller than the actual total sample size. Therefore,

we cannot use the actual sample size in estimating the standard errors. The effective sample size is the equivalent total sample size that we should use in estimating the standard errors (Snijders & Bosker 2002, 23).

In the case of equal cluster sizes of n with k cluster, the effect of the intra-cluster correlation on the sample size can be seen in the formula

$$n_{eff} = \frac{n \times k}{1 + (n-1)\rho}$$

where n_{eff} is the effective sample size of level 1 units and $n \times k$ is the actual or nominal sample size, consisting of n units in k clusters. If the intra-cluster correlation is $\rho = 0$, the effective sample size is equal to the nominal sample size. If the intra-cluster correlation is $\rho = 1$, the effective sample size is equal to the number of clusters k . The denominator in the formula is also known as the design effect. In the case of unequal group sizes, a good estimate is achieved using the average cluster size instead of n .

The positive ICC reduces the effective sample size of level 1 units. If this effect is not taken into account in the statistical modelling, the standard errors of the coefficient estimates will be too small. As a consequence, statistical test of coefficient estimates will not be reliable.

5.3 Two-level linear model

In school data, statistical methods based on the assumption of statistical independence of students are not applicable. The two-level statistical model (Goldstein 1987, 1995; Bryk & Raudenbush 1992; Snijders & Bosker 2002) incorporates the structure of the data into the model and takes into account the statistical dependences in the data. The general two-level model of school data with students as level 1 units and schools as level 2 units can be presented in the formula

$$(5.4)$$

where y_{ij} is the outcome variable, the index i denotes the student and j the school, x_{kij} is the student level explanatory variable, z_{kj} is the school level explanatory variable, β_{kij} is the coefficient of the student level explanatory variable, γ_l is the coefficient of the

school level explanatory variable, u_{kj} is the school level residual, and e_{kij} is the student level residual. The coefficients β_{kj} and γ_l are referred to as fixed parameters. For fixed parameter estimation, the normality assumption will be made for the residuals: $u_{kj} \sim N(0, \sigma_{uk}^2)$ and $e_{kij} \sim N(0, \sigma_{ek}^2)$. The random effects u_{kj} and e_{kij} are assumed to be independent from each other across the levels. The variances and covariances associated with the residuals are referred to as random parameters.

Each student level coefficient β_{kj} can be modelled at school level as one of three general forms:

1. A fixed student level coefficient; e.g.,

$$\beta_{kj} = \beta_k$$

2. A non-randomly varying student level coefficient; e.g.,

$$\beta_{kj} = \beta_k + \gamma_1 z_{1j}$$

3. A randomly varying student level coefficient; e.g.,

$$\beta_{kj} = \beta_k + u_{kj}$$

The student level coefficient can also have both non-random and random sources of variation,

$$\beta_{kj} = \beta_k + \sum_l \gamma_l z_{lj} + u_{kj} \quad .$$

In addition, the student level residuals e_{kij} can be used to model the student level variance structure, for example in case of heteroscedasticity.

Modelling of the coefficients β_{kj} makes the two-level model a very flexible way of modelling and analyzing the data. The general linear two-level model enables the inclusion of explanatory variables from both levels, interactions between them and modelling of complex variance structure.

In the next section, some specific two-level models and associated statistical hypotheses derived from the general two-level model (5.4) are presented. They enable the modelling and testing of the school differences in performances, of between-school variation in the effects of student level explanatory variables, school level variables affecting the student performances, interactions between student and school level variables, variance structure at different levels and contextual variables.

5.4 Applications of two-level linear model

In the next section, some applications of the two-level model derived from the general two-level model (5.4) are presented. These applications are used later in statistical modelling of the data.

1. Simple variance component model. This model tests and estimates the existence of any between-school variation in the outcome variable. This model is often called the null or empty model, and it is used as the reference or baseline model for comparisons in later model building. This model consists only of the intercept and the random part, and it is also called the random-intercept model. It estimates the intercept and divides the total student variance into two variance components:

$$\begin{aligned}
 y_{ij} &= \beta_{0j} & (5.5) \\
 \beta_{0j} &= \beta_0 + u_{0j} + e_{0ij} \\
 u_{0j} &\sim N(0, \sigma_{u0}^2) \\
 e_{0ij} &\sim N(0, \sigma_{e0}^2) \\
 \text{cov}(u_{0j}, e_{0ij}) &= 0 \quad .
 \end{aligned}$$

The model can be expressed in the following form:

$$\begin{aligned}
 y_{ij} &= \beta_0 + (u_{0j} + e_{0ij}) & (5.6) \\
 \text{var}(u_{0j} + e_{0ij}) &= \sigma_{u0}^2 + \sigma_{e0}^2 \quad .
 \end{aligned}$$

The interesting statistical hypothesis in simple variance component model tests, if there is between-school variation in the outcome variable. The statistical null hypothesis is $\sigma_{u0}^2 = 0$ with the alternative hypothesis stating that $\sigma_{u0}^2 > 0$.

2. Two-level model with one student level explanatory variable. One explanatory variable x_{1ij} at the student level is added to the null model. The coefficient of the explanatory variable may vary between schools. The model is

$$\begin{aligned}
 y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} \\
 \beta_{0j} &= \beta_0 + u_{0j} + e_{0ij} \\
 \beta_{1j} &= \beta_1 + u_{1j} \\
 u_{0j} &\sim N(0, \sigma_{u0}^2), u_{1j} \sim N(0, \sigma_{u1}^2) \\
 e_{0ij} &\sim N(0, \sigma_{e0}^2) \\
 \text{cov}(u_{0j}, e_{0ij}) &= \text{cov}(u_{1j}, e_{0ij}) = 0, \\
 \text{cov}(u_{0j}, u_{1j}) &= \sigma_{u01} \quad .
 \end{aligned}
 \tag{5.7}$$

The model can be expressed in the following form:

$$\tag{5.8}$$

Since the coefficient β_1 varies between schools, the complexity of the variance structure increases. In addition to the null model, there are two interesting statistical hypotheses:

(1) Does the effect of the student level explanatory variable x_{1ij} vary between schools?

~~$\sigma_{u1}^2 = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} + u_{0j} + u_{1j}x_{1ij} + \dots + u_{kj}x_{kij} + e_{0ij}$~~
 ~~$\text{Var}(\bar{u}_0 + \bar{u}_1 x_{1ij} + \dots + \bar{u}_k x_{kij}) = \sigma_{u0}^2 + 2\sigma_{u01}x_{1ij} + \dots + \sigma_{u1}^2 x_{1ij}^2 + \dots + \sigma_{e0}^2$~~
 ~~$\beta_{1j} = \beta_1 + u_{1j}$~~ β_1 varies between schools. If the hypothesis is rejected, the usual alternative hypothesis $\sigma_{u1}^2 > 0$ is accepted. The conclusion is that the coefficient β_1 varies between schools.

(2) Does the coefficient β_1 differ statistically significantly from 0? The statistical null hypothesis states: $\beta_1 = 0$, with the usual alternative hypothesis $\beta_1 \neq 0$.

3. Adding a school level explanatory variable into the model. In addition to the student level explanatory variable, one school level explanatory variable z_{1j} is added to the model, having an effect on the intercept of the model. The model is

$$\tag{5.9}$$

The model can be expressed in the following form:

$$\tag{5.10}$$

An interesting statistical hypothesis, in addition to the former ones, is now associated with the coefficient of the school level explanatory variable. Does the school level explanatory variable z_{1j} have an effect varying non-randomly on the student performances through the school means? The statistical null hypothesis states: $\gamma_0 = 0$, with the usual alternative hypothesis $\gamma_0 \neq 0$. If the null hypothesis is rejected, the conclusion is that the school level explanatory variable z_{1j} is associated with systematic variation between school means in the student performances, and hence, it explains the variation between students through the variation between school means.

4. Interaction between the student and school level explanatory variables. If, in addition to affecting the intercept, the school-level explanatory variable z_{1j} also affects the coefficient of the student-level variable x_{1ij} , the interaction between the student- and school-level variables should be modelled, the model is

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} & (5.11) \\ \beta_{0j} &= \beta_0 + \gamma_0z_{1j} + u_{0j} + e_{0ij} \\ \beta_{1j} &= \beta_1 + \gamma_1z_{1j} + u_{1j} \quad . \end{aligned}$$

This model can be expressed in the following form:

$$(5.12)$$

An interesting statistical hypothesis is now associated with the coefficient γ_1 of the interaction term. Do the student and school level explanatory variables x_{1ij} and z_{1j} have an interaction effect on the student performance? The statistical null hypothesis states: $\gamma_1 = 0$ with the usual alternative hypothesis $\gamma_1 \neq 0$. If the null hypothesis is rejected, the conclusion is that the effect of the student level explanatory variable depends on the value of the school level explanatory variable: $\beta_1x_{1ij} + \gamma_1z_{1j}x_{1ij} = (\beta_1 + \gamma_1z_{1j})x_{1ij}$.

5. Complex variance structure at the student level. Sometimes it is interesting, and even necessary for the reliable estimation of the fixed parameters, to model the variation explicitly as a function of explanatory variables. Different within-school variances for boys and girls can be modelled in the following way:

$$\begin{aligned}
 y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} \\
 \beta_{0j} &= \beta_0 + u_{0j} \\
 \beta_{1j} &= \beta_1 + e_{1ij} \\
 \beta_{2j} &= e_{2ij} \\
 \text{var}(u_{0j}) &= \sigma_{u0}^2 \\
 \text{var}(e_{1ij}) &= \sigma_{e1}^2, \text{var}(e_{2ij}) = \sigma_{e2}^2, \text{cov}(e_{1ij}, e_{2ij}) = 0
 \end{aligned}
 \tag{5.13}$$

$$x_{1ij} = \begin{cases} 1, & \text{if girl} \\ 0, & \text{if boy} \end{cases}, \quad x_{2ij} = \begin{cases} 1, & \text{if boy} \\ 0, & \text{if girl} \end{cases}.$$

The model consists of two explanatory variables, which are the dummy variables for boys (x_{1ij}) and girls (x_{2ij}). There is only one school level residual u_{0j} associated with the intercept estimating the between-school variation. Coefficient β_{0j} consists of the fixed parameter β_0 estimating the gender difference and of the residual e_{1ij} used to estimate the within-school variation of girls. Coefficient β_{2j} consists only of the residual e_{2ij} used to estimate the within-school variation of boys.

The model can be expressed in the following form:

$$\begin{aligned}
 \text{Total variance for girls} &: \sigma_{u0}^2 + \sigma_{e1}^2 \\
 \text{Total variance for boys} &: \sigma_{u0}^2 + \sigma_{e2}^2.
 \end{aligned}
 \tag{5.14}$$

An interesting statistical hypothesis, in addition to testing the mean gender difference with the statistical hypothesis of $\beta_1 = 0$, is now associated with the equality of the within-school variation between boys and girls. The statistical null hypothesis of equal within-school variance components for boys and girls states: $\sigma_{e1}^2 = \sigma_{e2}^2$ with the alternative hypothesis $\sigma_{e1}^2 \neq \sigma_{e2}^2$. If the null hypothesis is rejected, the conclusion is that the within-school variation is not equal for boys and girls. Whichever is larger is seen in the variance components. If there are considerable differences in the variance components, it may also be important to include the variance structure in the model to achieve

reliable estimates and standard errors for the fixed parameters.

The between-school variation may also be different between boys and girls:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} \quad (5.15)$$

$$\beta_{0j} = \beta_0$$

$$\beta_{1j} = \beta_1 + u_{1j} + e_{1ij}$$

$$\beta_{2j} = u_{2j} + e_{2ij}$$

$$\text{var}(e_{1ij}) = \sigma_{e1}^2, \text{var}(e_{2ij}) = \sigma_{e2}^2, \text{cov}(e_{1ij}, e_{2ij}) = 0$$

$$\text{var}(u_{1j}) = \sigma_{u1}^2, \text{var}(u_{2j}) = \sigma_{u2}^2, \text{cov}(u_{1j}, u_{2j}) = \sigma_{u12}$$

$$x_{1ij} = \begin{cases} 1, & \text{if girl} \\ 0, & \text{if boy} \end{cases}, \quad x_{2ij} = \begin{cases} 1, & \text{if boy} \\ 0, & \text{if girl} \end{cases}.$$

The only difference in this model, compared to the former one, is that the school residuals are separate for girls (u_{1j}) and boys (u_{2j}). The model can be expressed in the following form:

$$(5.16)$$

An interesting statistical hypothesis is now associated with the equality of the between-school variation for girls and boys. The statistical null hypothesis of equal between-school variance components for girls and boys states: with the alternative hypothesis $\sigma_{u1}^2 \neq \sigma_{u2}^2$. If the null hypothesis is rejected, the conclusion is that the between-school variation in the outcome variable is different for girls and boys.

6. Modelling and testing the contextual effect. The school level variable may be either a natural or contextual variable. In addition to the student level explanatory variable x_{ij} , a contextual variable $\bar{x}_{.j}$, the mean of students' values of the variable x_{ij} in each school j , is included in the model. The model is

$$\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_{1j}x_{ij} \\
\beta_{0j} &= \beta_0 + \gamma_0\bar{x}_{.j} + u_{0j} + e_{0ij} \\
\beta_{1j} &= \beta_1
\end{aligned}
\tag{5.17}$$

with usual assumptions about the residuals. The contextual variable is associated with the intercept, thus inducing the systematic variation between the school means in the outcome variable. The model can be expressed in the following form:

$$\begin{aligned}
y_{ij} &= \beta_0 + \gamma_0\bar{x}_{.j} + \beta_1x_{ij} + (u_{0j} + e_{0ij}) \\
\text{var}(u_{0j} + e_{0ij}) &= \sigma_{u0}^2 + \sigma_{e0}^2
\end{aligned}
\tag{5.18}$$

An interesting statistical hypothesis concerns now the contextual effect. Does the school level contextual variable $\bar{x}_{.j}$, which is the school mean of x_{ij} , have an effect on the student performances through school mean performances in addition to the individual effect? The statistical null hypothesis states: $\gamma_0 = 0$ with the alternative hypothesis $\gamma_0 \neq 0$.

If the statistical hypothesis is rejected and the contextual effect exists, the student level effect β_1 is the same in all schools, but the school intercept β_{0j} depends on the contextual effect. The systematic part of the school intercept is $\beta_{0j} = \beta_0 + \gamma_0\bar{x}_{.j}$. The within-school school regression lines are parallel, but the school intercepts depend on the contextual variable.

5.5 Explanatory power of the model

An interesting question in modelling school data with a two-level model is, how much of the student and school level variation does the model explain. In the linear regression model without a hierarchical data structure, the concept of 'explained variance' gives an answer to the question, how much of the variability of the dependent variable is accounted for by the linear regression on the explanatory variables. The usual measure for the explained proportion of variance is the squared multiple correlation coefficient, R^2 . For the hierarchical linear model, however, the concept of 'explained variance' is more problematic. (Snijders & Bosker 2002, 99.)

Proportional reduction in variance component

One way to approach this question is to transfer the customary treatment of R^2 straightforwardly to the hierarchical linear model and to treat *proportional reductions in the estimated variance components as analogous to R^2 values*. In this case, two estimates for the proportional reduction in residual variance components are available: one for the between-school variance component and another for the between-student within-school variance component. In addition, proportional reduction in the total variance is available.

If we write the estimated variance components for the null model as $\hat{\sigma}_{u(0)}^2$ and $\hat{\sigma}_{e(0)}^2$, and for the model with explanatory variables with $\hat{\sigma}_{u(x)}^2$ and $\hat{\sigma}_{e(x)}^2$, the analogue of R^2 value at the school level for the between-school variance component is

$$R_B^2 = \frac{\hat{\sigma}_{u(0)}^2 - \hat{\sigma}_{u(x)}^2}{\hat{\sigma}_{u(0)}^2} = 1 - \frac{\hat{\sigma}_{u(x)}^2}{\hat{\sigma}_{u(0)}^2} , \quad (5.19)$$

and at the student level for the within-school variance component it is

$$R_W^2 = \frac{\hat{\sigma}_{e(0)}^2 - \hat{\sigma}_{e(x)}^2}{\hat{\sigma}_{e(0)}^2} = 1 - \frac{\hat{\sigma}_{e(x)}^2}{\hat{\sigma}_{e(0)}^2} . \quad (5.20)$$

Since the total student variance is the sum of the two variance components, the proportional reduction in the total residual variance is

$$R_{Tot}^2 = \frac{(\hat{\sigma}_{u(0)}^2 + \hat{\sigma}_{e(0)}^2) - (\hat{\sigma}_{u(x)}^2 + \hat{\sigma}_{e(x)}^2)}{\hat{\sigma}_{u(0)}^2 + \hat{\sigma}_{e(0)}^2} = 1 - \frac{\hat{\sigma}_{u(x)}^2 + \hat{\sigma}_{e(x)}^2}{\hat{\sigma}_{u(0)}^2 + \hat{\sigma}_{e(0)}^2} = 1 - \frac{\hat{\sigma}_{tot(x)}^2}{\hat{\sigma}_{tot(0)}^2} . \quad (5.21)$$

It is convenient to express the proportional reduction in the variance components and total variance in percentages, i.e. to multiply it by 100.

However, the definitions of R^2 as the proportional reductions in the residual variance components are not completely unproblematic. It sometimes happens that the addition of explanatory variables increases rather than decreases some of the variance components, and in that case, even negative values of R^2 are possible (Snijders & Bosker 2002, 99). However, with R_{Tot}^2 this does not happen, since the total residual variance does not increase, when explanatory variables are added into the model. For the total

variance, the proportional reduction R_{Tot}^2 may be viewed as an equal of the squared multiple correlation coefficient R^2 in the linear regression model without hierarchical data structure.

Proportion of explained variance

Another way to approach the concept of 'explained variance' is to define measures of explained (or modelled) variation by the principle of *proportional reduction of prediction error* (Snijders & Bosker 1994; 2002). The difference between the observed value of the outcome variable and the predicted value of the statistical model is the prediction error. Two concepts of explained proportion of variance in a two-level model can be defined. The first one is the *proportional reduction of error for predicting an individual outcome*. The second one is the *proportional reduction of error for predicting a group mean*.

At the student level, the proportion of explained variance can be expressed by

$$R_1^2 = 1 - \frac{\text{var}(Y_{ij} - X_{ij}\beta)}{\text{var}(Y_{ij})} = 1 - \frac{\hat{\sigma}_{u(x)}^2 + \hat{\sigma}_{e(x)}^2}{\hat{\sigma}_u^2 + \hat{\sigma}_e^2} . \quad (5.22)$$

At the student level, R_1^2 is the proportional reduction in the value of $\hat{\sigma}_u^2 + \hat{\sigma}_e^2$ due to including the explanatory variables in the model. This is equal to R_{Tot}^2 , the proportional reduction in the total student variance.

At school level, the proportion of explained variance can be expressed by

$$R_2^2 = 1 - \frac{\text{var}(\bar{Y}_{.j} - \bar{X}_{.j}\beta)}{\text{var}(\bar{Y}_{.j})} = 1 - \frac{\hat{\sigma}_{u(x)}^2 + \frac{\hat{\sigma}_{e(x)}^2}{n}}{\hat{\sigma}_u^2 + \frac{\hat{\sigma}_e^2}{n}} . \quad (5.23)$$

This is different from R_B^2 presented earlier. The reason is that unexplained within-group variability is represented completely by $\hat{\sigma}_e^2$. Unexplained between-group variability, however, is constituted by variation ascribed to $\hat{\sigma}_u^2$ as well as variation ascribed to $\hat{\sigma}_e^2$, since $\text{var}(\bar{Y}_{.j} | X_{1j}, \dots, X_{nj}) = \hat{\sigma}_u^2 + \frac{\hat{\sigma}_e^2}{n}$. However, if the group size n is large, these two

measures are quite close. The advantage of R_2^2 , however, is that it will not have negative values as easily as R_B^2 , even though this may happen, as will be seen later.

The formula for R_2^2 is defined here for balanced data with equal group size n . In the case of varying group sizes, one possibility is to use the harmonic mean, defined by $N/\{\sum_j (1/n_j)\}$, as the group size (Snijders & Bosker 2002, 103).

III RESULTS

6

School differences in reading literacy performance

The statistical analysis of the data begins with a study into the Finnish school differences in reading literacy without controlling for the effects of the background factors. The result is an estimate of the amount of unadjusted between-school variation for Finnish schools. Comparisons are made with school differences in Germany and Sweden. The estimated variance component model is used as the reference model for later analyses, where background factors are added into the model as explanatory variables, in order to find out, what are their effects on the outcome variable and how much these variables explain the between-school as well as the total student variance in the outcome variable.

6.1 Finnish school differences in reading literacy: Baseline model

The Finnish data set includes one exceptional school. This school is a special education school for students with profound developmental disabilities. Because of the extremely low performance level of this school, it has usually been excluded from the national analyses. This is also the reason, why the estimate for Finnish school differences in the national reports (e.g. Välijärvi & al. 2001, 27) may slightly differ from the results described in the international reports (e.g. OECD 2001, 61).

The special school is so different from the other schools in every respect that it is not fair to the school to include it in the same comparison with the rest of the schools. An important principle of evaluation research is that it has to be fair towards the schools and students who participate in it, and the importance of fairness in testing and assessment is an issue that cannot be overridden (Standards for educational and psychological testing 1999, 73). This special education school is left out of the statistical analyses of this study, as well, since it does not represent the same population of schools and students as the others. In statistical terms, it is an outlier, which may produce bias in the results (Malin 2004).

Table 1. Two-level variance component model for the combined reading literacy score in Finland

Fixed Effect	
Intercept	545.0
S.e. of intercept	2.09
Random Effect	
Between-school variance component	435.0
Within-school variance component	7307.4
Total variance	7742.4
ICC	0.056
Number of schools	154
Number of students	4858

The baseline model for the combined reading literacy in Finland is presented in Table 1. The estimated intra-class correlation of the combined reading literacy score is 0.056. Only about 5.6 % of total student variance in the combined reading literacy performance was attributable to the differences between the schools. The between-school standard deviation of the combined reading literacy scale, estimated from the between-school variance component, was about 21 points. This means that about two-thirds of the estimated school means in the population of schools the 15-year old students attended were between 524 and 566 points and about 95 % between 503 and 587 points, both above the OECD average of 500 points.

In the international PISA comparisons, the Finnish ICC of the combined reading literacy score was reported to be 0.12 (OECD 2002a, 246). This estimate was calculated

using the five plausible values of the combined reading literacy and including the special education school in the analysis. Leaving the special school out of the analysis and calculating the ICC using the five PVs gave an estimate of 0.064, which is about half of the internationally reported value and just a little larger than the one achieved using likelihood estimates instead of plausible values. This estimate was also smaller than the lowest reported value among the participating countries, which was 0.08 for Iceland (OECD 2002a, 246). The school differences in student performances in reading literacy were the smallest in Finland among the 32 countries participating in PISA 2000.

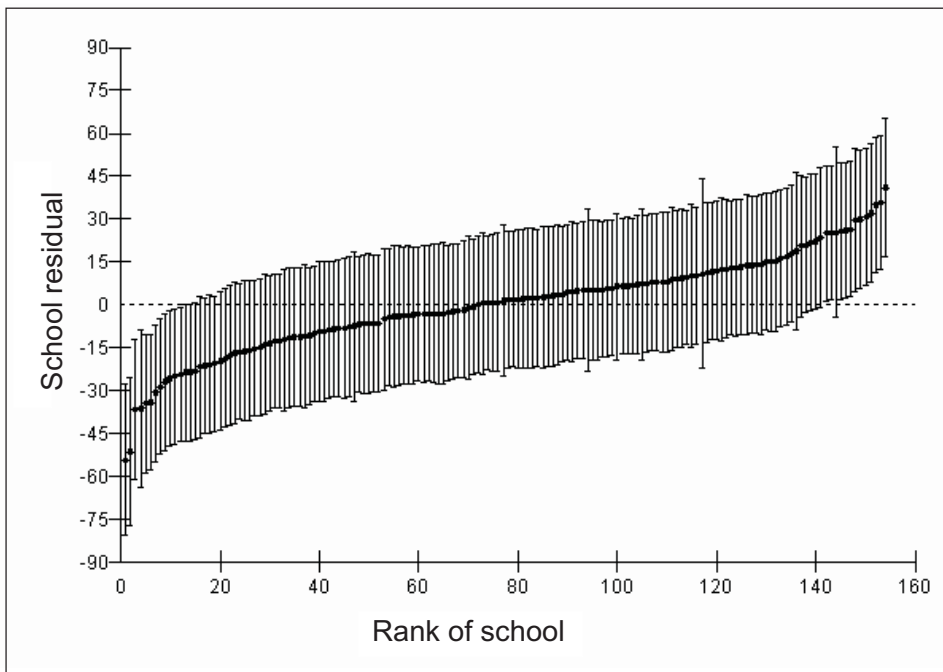


Figure 1. School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Finland

The Finnish school differences are shown graphically in Figure 1. The school residuals of the combined reading literacy score with 95 % confidence intervals are plotted in rank order. The school residuals are the schools' estimated deviations from the intercept in Table 1. For most of the schools, the 95 % confidence interval includes the ref-

reference line 0, which corresponds to the intercept of the baseline model. The means of these schools did not differ statistically significantly from the intercept estimate of 545 points, which may be interpreted as the average of the school means in Finland. However, there were schools clearly above (12 schools) and below (12 schools) the reference line. The 95 % confidence intervals of these schools do not include the reference line, and their means were statistically significantly different from the intercept. Since these 24 schools are 16 % of the sampled 154 schools, we can conclude that less than one fifth of the school means clearly differed from the average of the school means in the combined reading literacy scale in Finland.

6.2 Comparisons with Germany and Sweden

It is an interesting point of reference to compare the Finnish between-school variation with Germany. The German educational system with several school types is very different from the Finnish comprehensive school, and the between-school variation in Germany also seems to be much larger than in Finland (OECD 2002a). To ensure the comparability between Finland and Germany, the special education schools were left out of the German PISA 2000 data set as well, and the data then consisted of a sample of 215 schools and 5013 students. In all, 2.5 % of the 15-year-old students were enrolled in special education schools in Finland in 2000 (Linnakylä 2003, 162), while in Germany the proportion was little higher, between 3.3 and 5.5 % depending on the state (Döbert & al. 2003, 308). For statistical analyses, the student weights in German data were normalized to have an average of 1.

In Figure 2, the German school residuals of the two-level intercept-only model for the combined reading literacy score with 95 % confidence intervals are plotted in rank order. The amount of schools with the whole confidence interval above and below the reference line was much larger than in Finland. This is no wonder since the between-school variance component in the German PISA 2000 data was 4909 while the within-school variance component was 5353. The estimated intra-class correlation in Germany was 0.478. Almost half of the total student variance in the combined reading literacy performance is attributable to differences between schools. In the international PISA comparisons, the reported German ICC was even larger, 0.59, since it was estimated using the five plausible values and including the special schools in the analysis (OECD 2002a, 246).

According to the differentiated school system, the 15-year-old students in Germany attend several different school types, while in Finland the whole age cohort still attend the same comprehensive school. In addition to the special schools, six different school types were distinguished in the German data (Baumert & al. 2001, 38). The school type is an important source of variation in Germany, and the performance differences between the school types were clear (Baumert & al. 2003, 273). In the national study, the total student variance in the combined reading literacy was divided into three components: 56 % of the total variance was due to differences between school types, 5 % due to differences between schools and 39 % of the total variance was due to differences between students (Baumert & al. 2003, 268).

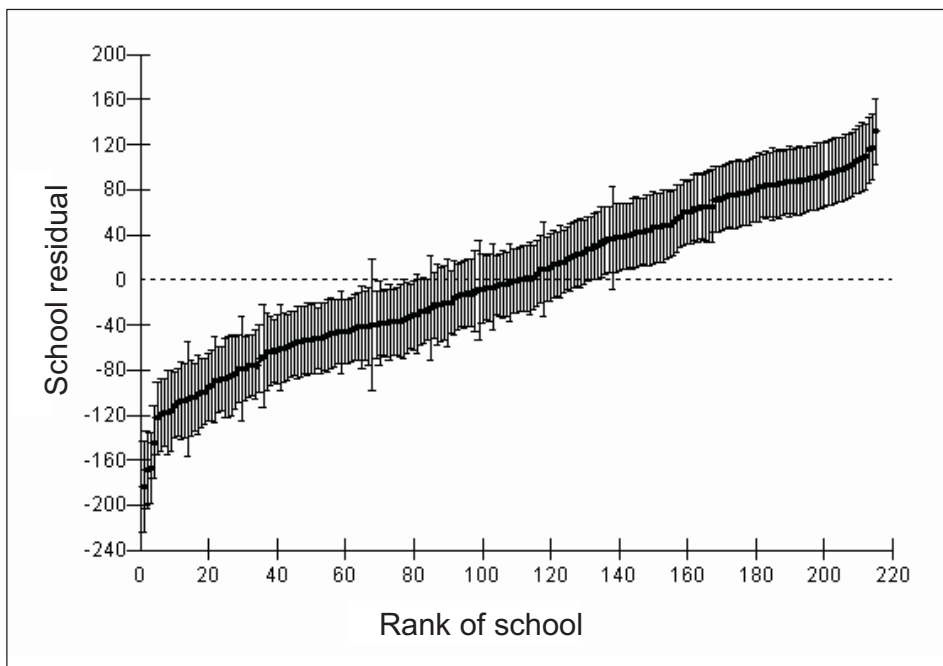


Figure 2. School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Germany

When the effect of the school type was controlled for, the German school differences looked quite different. The between-school variance component reduced to 806, while the within-school variance component was 5350; about the same as earlier. The Ger-

man between-school variance component was now about twice the Finnish one, and the within-school variance component was even smaller than for Finland. The ICC, adjusted for the school types, was only 0.131. The reduction in the between-school variance component was 84 %. Although the within-school variance component did not change, the huge reduction in the between-school variance component reduced the total student variance by 40 %. The school types explained this amount of the total student variance in the combined reading literacy for Germany.

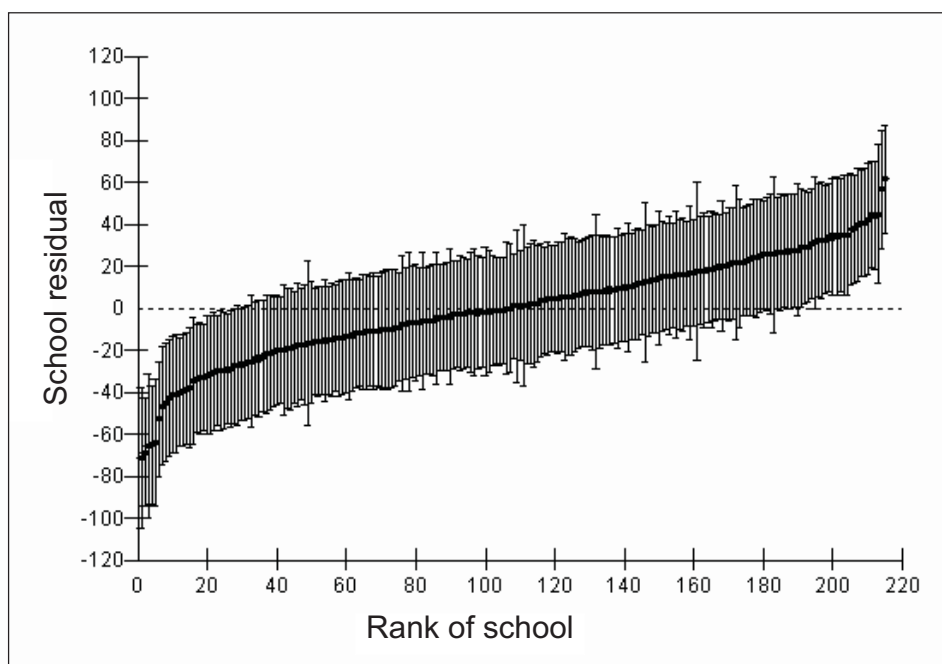


Figure 3. School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Germany, after controlling for the effect of school type

The German school residuals, adjusted for the school types, are presented in Figure 3. The reference line corresponds to the school type means, and the 95 % confidence interval shows, if the school's average performance is statistically significantly different from the mean of its own school type. By visual inspection, although the scales are somewhat different, the figure does not differ much from the equivalent Finnish figure (Figure 1), but is very different from the figure for the unadjusted German school resid-

uals. However, the conditional German ICC was still larger than the unconditional Finnish one, and this is reflected in the larger proportion of schools which clearly are above or below the reference line.

Most of the difference between Finland and Germany in the amount of unadjusted between-school variation is explained by the fact that the educational systems are different in these two countries. The variation in the average performance between different school types seems to explain most of the school differences in reading literacy performance in Germany.

Unlike Germany, Sweden has a very similar educational system to Finland, and the whole age cohort attends the same comprehensive school. The Swedish PISA 2000 data set consisted of a sample of 154 schools and 4415 students, and there were no special schools in the data. For statistical analyses, the student weights were normalized to have a mean of 1. The between-school variance component in Sweden was 706, clearly larger than in Finland, and the within-school variance component was 7942, just a little larger than in Finland.

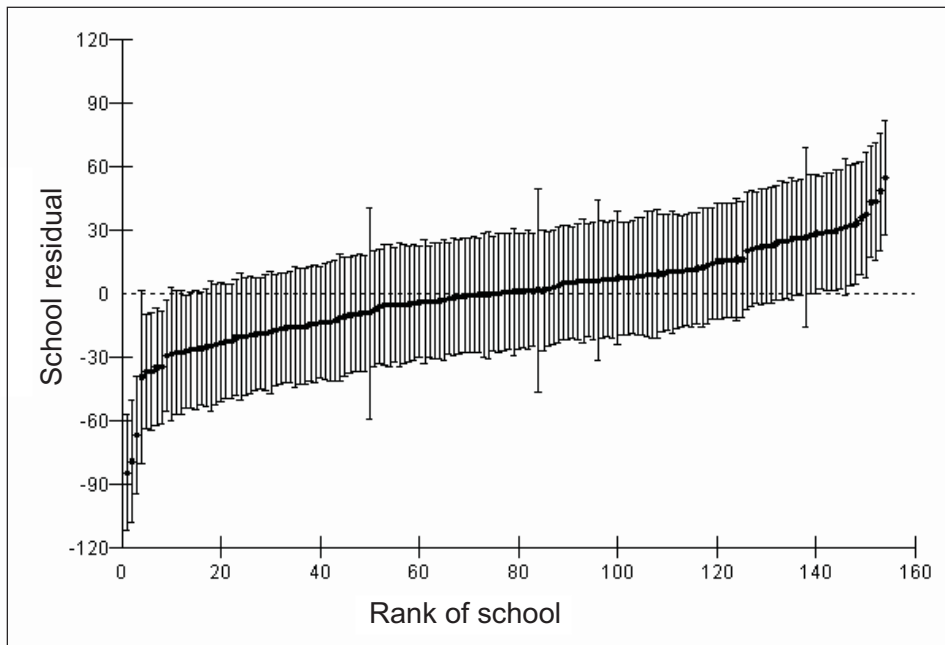


Figure 4. School residuals of the combined reading literacy score with 95 % confidence intervals in rank order in Sweden

The Swedish school residuals with 95 % confidence intervals are presented in Figure 4. The Swedish figure of unadjusted school residuals looks very similar to the Finnish one. However, the residuals of the first and last schools in rank order were little larger than in the Finnish data. This is reflected in the ICC, which in Sweden was 0.082, a little higher than in Finland. In the international PISA report, where the plausible values were used in estimation, the ICC was about the same, having a value of 0.09 (OECD 2002a, 246).

Another way to compare performance differences graphically between schools in separate countries is presented in Figures 5–7. In these figures, the schools are in rank order, but instead of ranks, the empirical school means of the combined reading literacy score are on the x-axis. The y-axis determines the spread of the empirical reading literacy scores of individual students around the school mean. The larger school mean variation in Germany, compared to Finland and Sweden, is easily seen, and the best and poorest performing schools of the three countries were in Germany. The Finnish and Swedish figures look very similar, but the best performing schools of these two countries were in Finland, and the poorest performing schools in Sweden. In Finland and Sweden, the students with the best test scores did not seem to attend the schools with the highest test means.

6.3 Conclusions

In the international PISA 2000 comparison, the unadjusted between-school variation in Finland was very small, less than 6 % of the total student variance, and it was actually the smallest among the participating countries. Small differences between the schools the 15-years old attend in Finland mean that the students' performance in reading literacy is largely unrelated to the schools in which they were enrolled. This may largely be due to the fact that Finland has a non-selective school system that provides all students with the same comprehensive schooling, and usually in the school that is located nearby. Even the least successful schools attained a relatively high level of reading literacy, when compared to other countries. Thus, the equality of access and opportunity seemed to be realized.

The Finnish results were compared with two reference countries, Germany and Sweden. The unadjusted between-school variation was a little larger in Sweden than in Finland, about 8 % of the total student variance. In Germany, the unadjusted between-school variation was the largest of the three countries, almost half of the total student

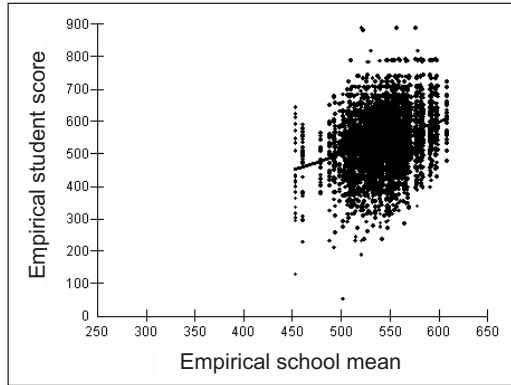


Figure 5. Scatterplot of student scores in the combined reading literacy around the school means in Finland

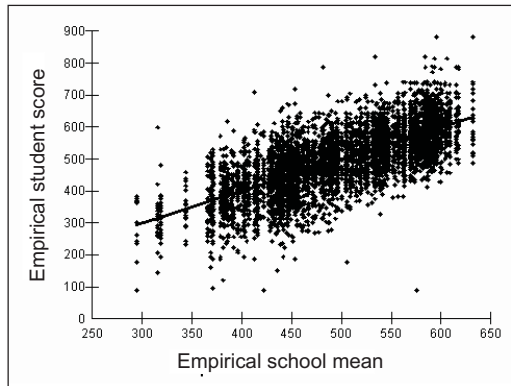


Figure 6. Scatterplot of student scores in the combined reading literacy around the school means in Germany

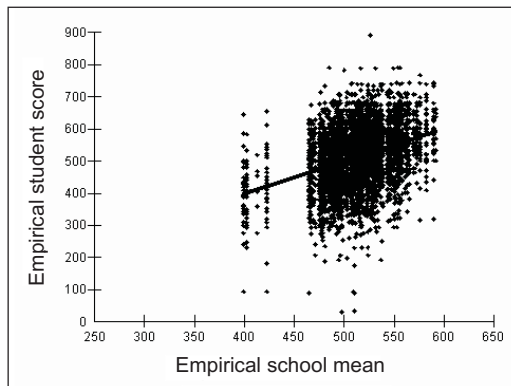


Figure 7. Scatterplot of student scores in the combined reading literacy around the school means in Sweden

variation. When the German results were adjusted for the different school types, the between-school variation in the average performance was still larger than the unadjusted results in Finland and Sweden, about 13 % of the total student variance.

Still, we cannot ignore the fact that there exist differences between individual schools in Finland. From this point of view, the variation between schools even in Finland cannot be considered insignificant. Finland has succeeded in keeping the between-school variation very small, but it is not in contradiction to the fact that there are individual schools that could do better.

The inequalities between individual students are reflected in the total student variance and standard deviation. In this sense, the equity between individual students seemed to be better realized in Finland than in the two reference countries, Germany and Sweden. The student standard deviation, calculated from the total student variance, which is the sum of the between-school and within-school variance components in the baseline model, was smaller in Finland (88) than in Sweden (93) or Germany (101). However, within-schools discrepancies between individual students were smallest in Germany, while in Sweden they were somewhat larger than in Finland.

In PISA 2000, the reading literacy scales were divided into five levels of knowledge and skills (OECD 2001, 2002a). Students who were at proficiency level 1, which is the lowest level or even below it in reading literacy, may be considered to lie beneath the threshold of equity. These students may not be acquiring the necessary literacy knowledge and skills to benefit sufficiently from educational opportunities, and there is evidence suggesting that in later life it is difficult to compensate for learning gaps present in initial education (OECD 2001, 48) and to be employed (Linnakylä & al. 2000). In Finland, 7 % of students were at or below the proficiency level 1, while the respective amount in Sweden was 12 % and in Germany 23 % (OECD 2001, 45). Also in this respect, Finland had managed better than the reference countries, although there is still work to be done to ensure that all students pass the threshold level of equity.

7

Two-level one-factor models

After constructing the baseline model and estimating the between-school and total student variance, the next step in the statistical analyses was to answer the question, how much of the variation between schools and students in reading literacy was systematic and due to each separate inequity factor. If systematic variation is associated with a factor, controlling its effect reduces the unexplained variance. The effect of each inequity factor on the reading literacy performance is studied first by separate two-level one-factor models. The effect of each background factor is thus unadjusted, i.e. each background factor is the only explanatory variable in the two-level statistical model. The statistical analyses of this chapter also show the preliminary results which guide the modelling of further analyses.

7.1 Fixed effects of two-level one-factor models

The two-level one factor models for the combined reading literacy score with only one inequity factor as the explanatory variable at a time are given in Table 2. The fixed effects, i.e. the estimates of the fixed parameters in each model, are presented in this table. Also later shown are the effect of each factor on the variance structure, i.e. the change in the variance components compared to the baseline model and related statistics.

The intercept of the two-level baseline model was 545 points. This is not exactly the same as the average of the students' combined reading literacy score, but it can be interpreted as the student mean in the average school, or school mean. To be more exact, it

Table 2. The fixed effects of two-level one-factor models on the combined reading literacy score

Variable	Coefficient				Sample size	
	Estim.	s.e.	t-value	p	Students	Schools
Intercept (baseline model)	545.0	2.09			4858	154
Language						
Finnish speaking schools (ref.)	546.9	2.04			4616	146
Swedish speaking schools	-36.7	8.99	-4.08	0.000	242	8
Grade						
Grade 9 (ref.)	550.8	2.13			4328	154
Grade 8	-53.0	3.89	-13.62	0.000	530	146
Immigrant background						
Native students (ref.)	545.8	2.09			4812	154
Non-native students	-79.7	12.8	-6.23	0.000	46	28
Regional unit						
Uusimaa Region (ref.)	538.5	4.25			1135	37
Southern Finland	13.2	5.55	2.38	0.017	1630	52
Eastern Finland	2.0	6.86	0.29	0.771	697	22
Mid-Finland	3.8	6.90	0.55	0.582	695	22
Northern Finland	9.4	7.02	1.34	0.181	701	21
Municipality type						
Urban (ref.)	547.2	2.38			3659	117
Rural	-9.1	4.82	-1.89	0.059	1199	37
Gender						
Male (ref.)	519.7	2.43			2354	154
Female	49.1	2.38	20.63	0.000	2504	154
Socio-economic status						
Intercept	544.4	1.98				
Contextual effect	3.6	5.05	0.71	0.476	4765	154
Individual effect	17.6	1.32	13.33	0.000	4765	154
Parents' educational level						
Vocational secondary (ref.)	537.4	3.56			1277	154
Contextual effect	6.8	4.35	1.56	0.118	4622	154
6-years basic education	-17.8	4.47	-3.98	0.000	500	141
9-years basic comprehensive	-19.8	4.44	-4.46	0.000	513	143
General secondary	14.8	4.08	3.63	0.000	665	152
Academic/Tertiary	19.5	3.27	5.96	0.000	1667	154
Wealth of the family						
Intercept	541.8	3.11				
Contextual effect	6.4	10.19	0.63	0.530	4845	154
Individual effect	7.7	1.78	4.33	0.000	4845	154

is the combined reading literacy score of the student, whose school and student level residual is zero.

The estimate of the intercept depends on the explanatory variables in the model, and when explanatory variables are added to the model, its value changes. This means that the interpretation of the intercept depends on the explanatory variables of the model. The change in the intercept during model building is studied in more detail in the next chapter. In the one-factor models of Table 2, the intercept is the expected value of the reference group in the case of a categorical background factor. If the background factor is a continuous variable, the intercept is the expected value of the students for whom the background factor value is zero.

There were 8 Swedish-speaking schools with 242 students in the Finnish sample. This was about 5 percent of the total amount of sampled schools and students, which is very close to the proportion of Swedish speaking population in Finland (5.5 %). The average in the Swedish speaking schools was 37 points lower than in the Finnish speaking schools. This was slightly more than the difference between Sweden and Finland, which was 30 points (OECD 2001). The difference was statistically significant, although the standard error was quite large and due to the small amount of schools and students in the sample. For more reliable comparisons between the two language groups, the Swedish speaking schools and students should be oversampled. However, the language group was a clear source of variation in the sample, and this background factor is included in later analyses to control for the variation due to the difference between the two official language groups.

In Finland, most of the 15-years old students attend grade 9. However, 530 students, about 11 % of the student sample size, were at grade 8. The definition of the population in PISA is based on the age of the student, not on the grade. The target population in Finland consisted of all the students who were born between February 1984 and January 1985 (Malin & Puhakka 2002). About 8 % of the students in the sample were born in 1985. Since students usually begin school in the year they turn seven, practically all students born in 1984 (96.6 %) were at grade 9 and all born in 1985 (95.4 %) were at grade 8. However, about 29 % of all grade 8 students (154 students of 530) were born in 1984. These students had either started their school one year later than usual or they were repeating the grade. Only a few of grade 9 students (18 of 4329) were born in 1985.

The difference in the mean scores between students at grades 8 and 9 was 53 points to the benefit of the grade 9 students. Since grade 8 students were not a representative sample of all the students at that grade, this is not a reliable estimate of the grade effect

in reading literacy between grades 8 and 9. In addition, the students at grade 9 who were born in January 1984 were not included in the sample.

However, the difference between the two grades in the average performance is essential for later statistical analyses. The proportion of students attending grade 8 varied between schools, ranging from 0 to 31 %. In general, the more there were grade 8 students in the school, the lower was the average of that school, and this affected the estimate of the between-school variance component. The between-school variation in the proportion of students at grade 8 is a source of variation in estimating the school differences at the end of the basic education at grade 9, and its influence on the between-school variance component has to be controlled for.

The amount of students with immigrant background in the Finnish sample, as in the population, was extremely small. Only 46 students, 1 % in the Finnish sample, were non-native students, i.e. both the parents and the student were born outside Finland. Students who had been in the country less than a year were excluded from the study. The reading performance of the non-native students was much lower than that of native students. Students with immigrant background were about 80 points below the native Finnish students on average. The difference was almost one student standard deviation. The small amount of the non-native students was reflected in the large standard error of the estimate. These students were a very heterogeneous group; some of them were very good and some extremely weak in reading performance. Their reading score ranged from 261 to 639 points with mean score 466. As a factor contributing to a large difference and an important and interesting source of variation, this variable was also included in the subsequent analyses.

There were two variables describing the location of the school, namely regional unit and municipality type. The unadjusted differences between the five regional units were small. Uusimaa Region was the reference category, and when the differences were compared to it, only one statistically significant deviation was found. The results in the schools of Southern Finland were 13 points better than in the Uusimaa Region. The second largest difference was between Northern Finland and the Uusimaa Region, 9 points for the benefit of Northern Finland, but this was not statistically significant. When the four regional estimates were compared with each other, there were no statistically significant differences between them. Only schools in Southern Finland had clearly achieved better results than schools in the Uusimaa Region, and this difference was quite small, only 15 % of the student standard deviation of the combined reading literacy score.

There was no statistically significant difference in the average reading performance between the two municipality types, urban and rural, though the urban schools were 9 points better on average.

The gender difference was large in Finland. Girls were 49 points better than boys, on average. This difference is equal to 56 % of the student standard deviation of the combined reading literacy score. However, it has to be remembered that the Finnish boys were not weak in reading, since they were the best among boys in the comparison of the 32 countries (OECD 2001, 276). The large gender difference was a result of the fact that Finnish girls were extremely good readers.

The socio-economic status of the family was associated with variation in the students' reading performance. The effect of this factor was divided in two parts, contextual school level and individual student level effect. At the student level, the coefficient estimate was almost 18, which is 20 % of the student standard deviation of the reading literacy score. Since the socio-economic index was normalized with a standard deviation of one, this was equal to the average difference in reading between two students who went to the same school, but whose families had a difference of one standard deviation in the socio-economic index. With four standard deviations, the difference was almost 70 points, which was approximately as large as the range of one proficiency level. At the school level, there was no contextual effect of the socio-economic status in addition to the individual effect. The contextual coefficient estimate was less than 4 and it was equivalent to 2.5 between-school standard deviations in the socio-economic index. The socio-economic composition of the school did not seem to affect the results in reading performance, at least according to the one-factor model. The index of the socio-economic status was missing from 93 students in the sample, and they come from 68 schools.

The measure of the educational background of the students' families was based on either the father's or mother's level of education, whichever was higher. In the one-factor model, the parents' education was divided in five categories: (1) completed 6-years basic education, (2) completed 9-years basic education of comprehensive school, (3) completed vocational secondary level education, which is the reference category, (4) completed upper secondary education, aimed at gaining entry into tertiary education and called general secondary education, (5) completed academic or tertiary education. The contextual value was the school average of parents' educational level values. There were 236 students with missing information on their parents' educational level, and they came from 116 different schools.

Parental level of education was also associated with the reading literacy performance. In general, the higher the parents' education, the better were the students' average results. The students whose parents had completed only the 6-year basic schooling had 18 points and students whose parents had completed the 9-year comprehensive schooling had 20 points lower average reading score, compared to students whose parents had completed vocationally oriented secondary level education. On the other hand, if at least one of the parents had completed academically oriented upper general secondary education, the students performed 15 points better compared to students, whose parents had vocationally oriented education. Parents' completion of academic or tertiary education still slightly increased the difference. The average result was 20 points better than in the reference group. However, the difference between 6-year and 9-year basic education was not statistically significant, nor the difference between upper secondary general or tertiary education. In later analyses, these groups were combined to form only one category below and one category above the reference category. The parents' education had no statistically significant contextual effect.

Family wealth had a small effect on reading performance at the student level. The individual coefficient estimate was 8 points, which was equivalent to an increase of one international student level standard deviation in the index of family wealth. Since the Finnish standard deviation was narrower, this was equivalent to 1.4 national standard deviations. Family wealth had no statistically significant contextual effect at the school level. Although the estimate of the contextual coefficient was almost as much as on the individual level, it was equivalent to almost five national between-school standard deviations. The information of family wealth was missing from 13 students in 8 schools.

7.2 Explanatory power of two-level one-factor models

The explanatory power of each of the two-level one-factor models reveals the magnitude of the systematic variation between schools and students in reading literacy, associated with each of the inequity factors separately, or, in other words, how much are explained by the two-level one-factor models of the between-school and total student variation in reading literacy performance.

Variance components and related statistics for two-level one-factor models of the combined reading literacy score are presented in Table 3. In the table, the total student

variance in the combined reading literacy score was divided in two components, after fitting each one-factor model in the data. They are the between-school and within-school variance components, which were used as estimates of the between-school and within-school variation. The intra-class correlations, ICCs, were estimated by dividing the unexplained between-school variance component by the unexplained total variance. The ICCs estimates the proportion of the total unexplained variance attributable to the differences between schools after controlling for the effect of each inequity factor. In this sense, they are now conditional ICCs.

Two estimates of the explained between-school variation are presented in the Table 3. The first one is the reduction in variance component, describing how much the variance component reduces in percents after controlling for the effect of the explanatory variable. The second one is the proportion of explained variance, estimating the reduction of the between-school variance. Reduction in variance components is also presented for the within-school variance component, as well as for the total variance. The proportion of explained variance is presented for the total variance, although it is and it should be the same as the reduction in total variance. The between-school, within-school and total standard deviations, derived as square roots of the respective variance components and total student variance, are presented in the last three columns.

The baseline model in Table 3 is the intercept-only model, which divides the total student variance in the between-school and within-school variance components. This is used as the reference model in estimating the amount of the systematic variation between schools and students, associated with each of the inequity factors separately. This model is the same as was presented already in the previous chapter.

Controlling for the effect of language, which is a school level variable, reduced the between-school variance component by 14 %, compared to the variance component of the baseline model. On the other hand, the proportion of explained variance was 0.092 (or about 9 %). Since this variable did not reduce the within-school variance component, it explained only about 1 % of the total student variance. The proportion of explained total variance was small, since in the baseline model the between-school variance component was only 5.6 % of the total student variance. After controlling for the difference between Finnish-speaking and Swedish-speaking schools, the conditional ICC was 0.049. The between-school variance component was less than 5 % of the total unexplained variance.

Table 3. Variance components and related statistics of two-level one-factor models on the combined reading literacy score

Variable	Variance components			ICC	Reduction in variance components (%)			Proportion of explained variance			Standard deviations		
	Between School	Within School	Total variance		Between School	Within School	Total variance	Between School	Total	Between School	Within School	Total	
Intercept (baseline model)	435.0	7307.4	7742.4	0.056	--	--	--	--	--	20.9	85.5	88.0	
Language	372.8	7307.4	7680.2	0.049	14.3	0.0	0.8	0.092	0.008	19.3	85.5	87.6	
Grade	441.6	7032.0	7473.6	0.059	-1.5	3.8	3.5	0.003	0.035	21.0	83.9	86.4	
Immigrant background	437.1	7248.8	7685.9	0.057	-0.5	0.8	0.7	0.000	0.007	20.9	85.1	87.7	
Regional unit	422.7	7307.1	7729.8	0.055	2.8	0.0	0.2	0.018	0.002	20.6	85.5	87.9	
Municipality type	424.0	7307.4	7731.4	0.055	2.5	0.0	0.1	0.016	0.001	20.6	85.5	87.9	
Gender	464.3	6697.3	7161.6	0.065	-6.7	8.3	7.5	-0.014	0.075	21.5	81.8	84.6	
Socio-economic status	353.6	6905.9	7259.5	0.049	14.4	3.7	4.3	0.105	0.043	18.8	83.1	85.2	
Parents' educational level	376.8	7024.5	7401.3	0.051	12.9	2.7	3.3	0.092	0.033	19.4	83.8	86.0	
Wealth of the family	421.4	7233.6	7655.0	0.055	0.8	0.4	0.4	0.007	0.004	20.5	85.1	87.5	

Grade is a student level variable, and it explained less than 4 % of the total student variance. Now it is interesting to notice that the reduction in the between-school variance component had a negative sign, indicating that the between-school variance component had actually increased 1.5 %, when compared to the baseline model. The within-school variance component and the total variance decreased, when grade alone was the explanatory variable in the one-factor model, but the between-school variance component increased.

The between-school and within-school variance components are derived from the unexplained total variance. The between-school variance component may sometimes increase, when adding a student level variable into the model, but the total unexplained variance is always smaller than the total variance in the baseline model. Although the total unexplained variance reduced, as should happen, when new variables are added in the model, the relationship between the variance components changed in a way that increased the between-school variance component in comparison to the baseline model. This is a phenomenon that sometimes happens with multilevel data (Snijders & Bosker 2002). This happens more seldom to the proportion of explained between-school variance, which in the case of grade was positive in sign but practically zero.

This phenomenon was reflected in the small increase of the ICC from 0.056 to 0.059. If all the 530 students at grade 8 were actually excluded from the data, the between-school variance component would increase even more, to 490.6, and the within-school variance component to 6785.5, and total variance to 7276.1. The between-school variance component would increase and the within-school variance component decrease, and as a result, the ICC would increase. Calculated from these values, the ICC would be 0.067. In other words, without grade 8 students, 6.7 % of the variation in student performances in the combined reading literacy was attributable to the school differences at grade 9.

The effect of students' immigrant background was very small to the total student variance. Only less than 1 % of the total variance was explained. The reduction of the between-school variance component was slightly negative again. The proportion of immigrant students varied between schools from 0 to 0.17, and the effect of this variable on the school mean varied, depending on the amount of immigrant students in the school sample. However, the proportion of explained between-school variance was zero, indicating that the students' immigrant background did not explain school differences in the combined reading literacy.

The variables indicating the location of the school, i.e. the regional unit and municipality type explained almost nothing of the variation in student performances, since the regional differences were quite small. Both of these variables reduced the between-school variance component only by 3 % and the proportion of explained between-school variance was less than 0.02. The reduction in the within-school variance component was zero, since the variables were at school level, and as a result, the proportion of the explained total student variance was practically zero.

The effect of gender on the between-school variance component was even more interesting than the effect of grade. The difference between boys and girls explained almost 8 % of the total student variance in the combined reading literacy. However, in this one-factor model the between-school variance component increased from 435 in the baseline model to 464. The within-school variance component decreased from 7307 to 6697 and the total variance from 7742 to 7162. As a result, when the effect of gender difference was controlled for, decrease in the total variance and increase in the between-school variance component increased the ICC to 0.065.

When only all the three variables, grade, immigrant status and gender, were modelled as explanatory variables in the same model, the total unexplained variance was 6894, the between-school variance component was 477 and the within-school variance component was 6418. The three background factors together explained 11 % of the total student variance. The reduction in the within-school variance component was 12.2 %, but an increase of 9.5 % was observed in the between-school variance component. However, the proportion of explained between-school variance was only slightly negative, -0.019. To avoid confusing conclusion, it may be more reliable to use the latter statistic as the estimate of the explained between-school variance. The change in variance components meant a change in the ICC from 0.056 to 0.069. In other words, controlling the effects of grade, immigrant status and gender, 6.9 % of the remaining variance in student performances in the combined reading literacy was attributable to the school differences.

Socio-economic status of the family was the most powerful single variable explaining the between-school variation of the combined reading literacy. Now the reference values used to estimate the reductions in variance components and the proportions of explained variances were obtained from the intercept-only model which included exactly the same students as the one-factor model. The students with a missing value in the explanatory variable were excluded from the baseline model. For this reason, com-

parison of the results of the one-factor model directly with the baseline model in Table 3 gives slightly different results.

The between-school variance component was reduced by 14 % and the within-school variance component by 4 %, and together they explained little over 4 % of the total student variance. The proportion of the explained between-school variance was also high, 0.105. Although this is a student level variable and the contextual effect was not statistically significant, it was the best of the nine inequity factors in explaining the school differences. This implies that the index of socio-economic status was also intra-school correlated, i.e. there were differences between schools in their socio-economic intake. Using the two-level model with the socio-economic index as the response variable, the ICC of the socio-economic index was estimated to be 0.129. This is more than twice as high as the ICC of the combined reading literacy score. In Figure 8, school residuals of the socio-economic status index with 95 % confidence intervals in rank order of schools are presented. Comparing this figure with Figure 3, an equivalent figure for the combined reading literacy score, also shows that the between-school variation of the socio-economic index was larger than that of the reading literacy score. School differences in reading literacy were smaller than school differences in the social intake, and in this respect, the schools were able to even out the inequities.

Parents' educational level alone explained the between-school variation also quite well. Now again, the reference values were estimated without students with a missing value in parents' educational level. Parents' education reduced the between-school variance component by 13 %, but the within-school variance component was only reduced 3 %. Combining these two effects, reduction in the total student variance was, however, only little more than 3 %. The proportion of explained between-school variance was also quite high, 0.092. Like the socio-economic status of the family, parents' educational level was an intra-school correlated student level variable. The ICC was 0.090, calculated from the original values from the educational levels ranging from 1 (did not go to school) to 6 (completed tertiary education). Like the ICC of the socio-economic index, this was also higher than the ICC of the combined reading literacy score. This again indicates that the between-school variation in reading literacy was smaller than the between-school variation in the social intake.

The effect of family wealth on the variation of reading literacy seemed to be quite small. The reference values were again calculated from students with family wealth known. The reduction in the total student variance was less than 1 %, and the propor-

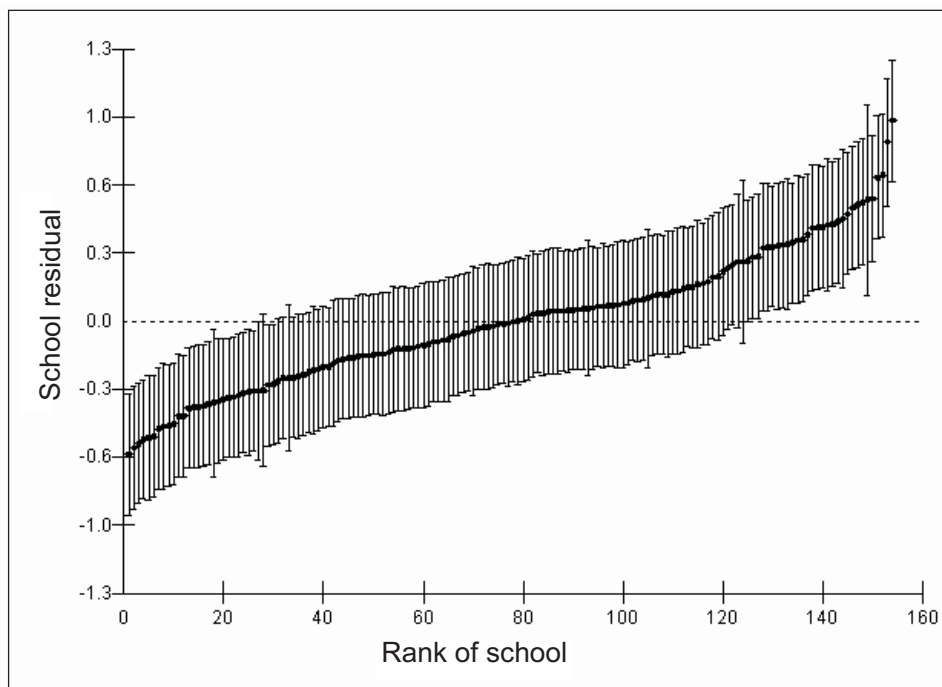


Figure 8. School residuals of the socio-economic status index with 95 % confidence intervals in rank order

tion of explained between-school variance was practically 0. The ICC of the family wealth index was 0.051. The wealth of the family did not explain much of the variation in reading literacy in Finland, since variation in this index was small. Its variance in the Finnish sample was only 0.513, half of the variance in the total OECD sample.

In Table 3, estimates of the between-school, within-school and total student standard deviations are also presented, for an insight in the variation between schools and students. These statistics were calculated as square roots of the variance components and the total student variance. The school level standard deviation was about 21 in the baseline model, and its average was close to 20 in the one-factor models. Smallest school level standard deviation was in the model with the socio-economic status of the family as an explanatory variable, being about 19, since this factor was the best one in explaining the between-school variation. According to the baseline model, two-thirds of the population school means were within the range of 42 points and 95 % within the range of 84 points, and there clearly were some differences between schools in the

average combined reading literacy score. After controlling for the effect of the socio-economic status of the students' family, two-thirds of the school means were within the range of 38 points and 95 % within the range of 76 points. Controlling for the effect of the socio-economic status of the students' family decreased the school differences. Or, to put it in other way, the variation in the schools' social intake at the student and school level increased the school differences in the combined reading literacy.

The ICC of the baseline model was 0.056. The average conditional ICC for the one-factor models was about the same, 0.055. After controlling for the effects of the each explanatory variable separately, almost 6 % of the unexplained total variance in student performances in the combined reading literacy was still attributable to the school differences. On average, the proportion of the between-school variance component of the total unexplained variance did not change practically at all.

7.3 Conclusions

The conclusions to be drawn from the statistical analyses of the two-level one-factor models are that the Swedish-speaking schools did not achieve as good results as the Finnish-speaking, and the non-native students were below native students in their performance. Regional differences in Finland were small, but the gender difference for the benefit of girls was large. Socio-economic background was associated with reading performance. The higher the socio-economic status of the family, the better educated were the parents and the more wealth there was in the family, the better were the student performances in the reading literacy. However, these effects are unadjusted, and the effects of other background factors are not yet taken into account in the estimation of the model coefficients.

Most of the systematic variation between schools was associated with the socio-economic status of the family, parents' educational level and language used in the school. School differences in reading literacy were smaller than school differences in their social intake, and in this respect, the schools were able to even out the inequities in the students' reading literacy performance. Most of the systematic variation within schools was associated with gender, which is the best explanatory factor of total student variation.

In this chapter, the effect of each inequity factor on reading literacy performance was studied in separate models. However, this kind of examination only gives information of the pairwise associations between reading performance and each background

factor at a time. To get a more informative view of the simultaneous associations between all important background factors and reading literacy, we have to use more developed statistical methods. In the next chapter, all the inequity factors are combined in the same statistical model, to find out what is the adjusted effect of each factor while controlling for the effects of the others, and how much of the between-school and total student variation in student performances can be ascribed to these factors in all.

8

Two-level multifactor model

In further analyses, the two-level multifactor model consisting of all the inequity factors simultaneously as explanatory variables is constructed. The aim was to find out, how large is the effect of each inequity factor on the reading literacy performance, when the effects of the other factors are controlled for in the model. After this we can evaluate, how much of the between-school and total student variation in reading literacy is attributable to all the inequity factors together.

The multifactor model is constructed by adding inequity factors in a groupwise manner into the model. The factors are divided in four groups, and the model is expanded by adding these groups one after another into the statistical model. This procedure yields four consecutive models, in addition to the baseline model, and it enables tracing the change in the coefficient estimates of the background factors, when more explanatory variables are added and their effects are controlled for. Finally, the model with all the inequity factors included is presented and studied in detail, and comparisons with equivalent German and Swedish models are carried out.

The four groups of the inequity factors are: I. Minority group factors, consisting of the language of the school, student's grade and immigrant background. II. Factors describing the location of the school, consisting of the regional unit and the municipality type. III. Gender. IV. Socio-economic background factors describing the student's family: socio-economic status index of the family, parents' educational level and wealth of the family. The groups of factors are added in this sequence into the model. However, the final model would be the same, regardless of the sequence in which they are added.

8.1 Minority group factors: Language, grade and immigrant background

In addition to the baseline model, four models are presented in Table 4 describing the change in the intercept and the coefficient estimates of three variables indicating the membership of minority group, i.e. language of the school, student's grade and immigrant background. The models were constructed in sequence by adding the four factor groups into the model. The table describes the change in the estimates while controlling for the effects of the factors already in the model.

Table 4. Change of the fixed parameter estimates of language, grade and immigrant background in constructing the two-level model explaining the combined reading literacy score

Model	Intercept		Swedish language		Grade 8		Non-native students	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
0: Baseline model	545.0	2.09	-	-	-	-	-	-
I: 0 + minority groups	553.2	2.08	-38.3	8.95	-51.2	3.90	-64.0	12.59
II: I + location of the school	551.7	4.21	-38.9	9.06	-51.3	3.90	-64.5	12.61
III: II + gender	530.4	4.41	-41.6	9.07	-47.4	3.73	-68.8	12.08
IV: III + socio-economic background	510.0	6.53	-46.7	8.38	-42.6	3.78	-60.9	12.17

Reference categories are: Finnish language, grade 9 and native students. Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

The intercept of the baseline model was 545. The estimate of the intercept changed, when explanatory variables were added into the model, since the intercept is the expected value for those students whose all background factors are equal to 0. In model I, the estimate increased slightly to 553 points, compared to the baseline model, since it is now the average score of the native students in the Finnish speaking schools at grade 9.

In model II, two variables describing the location of the school were added into the model. The intercept is now the average of urban schools in the Uusimaa Region for native Finnish speaking students at grade 9. In model III, the gender was added. The intercept went down, since it is the mean score for boys with the same characteristics as in model II.

In model IV, three variables describing the socio-economic background of student's family were added into the model. The intercept is the expected value for the students, whose mother or father, whoever had the higher value, has vocational secondary education and both the value of the socio-economic status and the wealth of the family is the OECD-average, i.e. they are both zeros. In addition, these students have the same characteristics as in the previous model.

The unadjusted difference between the Finnish and Swedish speaking schools in the combined reading literacy score was 37 points in the one-factor model shown in Table 2, for the benefit of the Finnish speaking schools. By adding the other inequity factors into the model, the difference increased. In model I, controlling for the effects of grade and immigration background, the difference was 38 points, and the difference increased to 47 points, when all the background variables were added into the model. The main reason is that the average socio-economic status index of the Swedish speaking students was higher than that of the Finnish speaking students. The difference in the index was about 28 % of its standard deviation, which was a statistically significant difference. With the same socio-economic background and other characteristics being equal, students attending Finnish speaking schools were 47 points better than in Swedish speaking schools. This was ten points more than in the one-factor model.

In Finland, most of the 15-years old students were at grade 9, but about 11 % of students in the sample were at grade 8. The unadjusted grade difference in the one-factor model (Table 2) was 53 points, for the benefit of grade 9 students. In model I, the difference between grades was 51, and it decreased by 10 points to 43, when all the inequity factors were included in the model. The change in the grade difference was also associated with the socio-economic background, since the average socio-economic status index of students at grade 8 was a little lower than that of students at grade 9. The difference was about 15 % of the standard deviation of the index, and this difference was statistically significant.

Only 1 % of the students in the Finnish sample were non-native. The reading performance of these students was lower than that of native students, the difference being 80 points in the one-factor model (Table 2). Controlling for the effects of the inequity factors, especially when the socio-economic background factors were added, reduced the difference, but non-native students were still on average about 61 points below the native Finnish speaking students.

8.2 Location of the school: Regional unit and municipality type

The change in the estimated differences between regional units in the sequential two-level models is presented in the Table 5. In the one-factor model, only Southern Finland differed statistically significantly from the reference category of the Uusimaa Region by 13 points for the benefit of Southern Finland (Table 2). When the effects of language, grade and immigration background were controlled for, even this difference reduced and there was no statistically significant variation between regional units. On the other hand, as can be seen in Table 6, the difference between urban and rural areas was now statistically significant, 13 points for the benefit of urban municipalities. This difference was not statistically significant in the one-factor model.

When gender was included in the model, differences between regional units remained practically the same. However, the difference between urban and rural areas increased to 23 points. This may imply an interaction between gender and municipality type, which is studied in the next step of model building.

Adding the socio-economic background factors into the model changed the results. The average performances in all four regional units were statistically significantly higher than in the reference category of the Uusimaa Region. The difference varied between 15 and 19 points. On the other hand, the difference between urban and rural municipalities reduced to 8 points, and it was not statistically significant any more. These changes in the estimates are due to the effects of the socio-economic background factors. The average value of the socio-economic background index was higher in urban than in rural municipalities, and it was higher in the Uusimaa Region than in other regions of the country (Appendix 2). Thus, with equal individual and contextual socio-economic background, students in other parts of Finland achieved 15–19 points better results in the combined reading literacy test on average, compared to the Uusimaa Region. This difference was equal to 17–22 % of the total student standard deviation in the combined reading literacy score.

Table 5. Change of the fixed parameter estimates of regional units in constructing the two-level model explaining the combined reading literacy score

Model	Southern Finland		Eastern Finland		Mid-Finland		Northern Finland	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
0: Baseline model	-	-	-	-	-	-	-	-
I: 0 + minority groups	-	-	-	-	-	-	-	-
II: I + location of the school	10.1	5.37	0.3	6.89	4.0	6.62	5.0	6.83
III: II + gender	9.0	5.34	-0.7	6.91	2.2	6.63	2.1	6.84
IV: III + socio-economic background	18.8	5.19	15.3	6.82	15.0	6.39	15.3	6.58

Reference category: Uusimaa Region

Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

Table 6. Change of the fixed parameter estimates of municipality type in constructing the two-level model explaining the combined reading literacy score

Model	Rural municipalities	
	Coeff.	s.e.
0: Baseline model	-	-
I: 0 + minority groups	-	-
II: I + location of the school	-12.7	4.83
III: II + gender	-22.6	5.57
IV: III + socio-economic background	-7.8	5.69

Reference category: Urban municipalities

Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

8.3 Gender

The gender difference in the one-factor model was 49 points for the benefit of girls (Table 2). When the effects of the minority group factors and the location of the school were controlled for, the difference was about 43 points (Table 7). However, a statistically significant interaction effect of gender and municipality type was now included in the model. The effect of this interaction was about 21 points. In other words, the gender difference was not constant, as assumed in the one-factor model, but it depends on the municipality type where the school is located.

Table 7. *Change of the fixed parameter estimates of gender in constructing the two-level model explaining the combined reading literacy score*

Model	Main effect: Female		Interaction: Female X Rural	
	Coeff.	s.e.	Coeff.	s.e.
0: Baseline model	-	-	-	-
I: 0 + minority groups	-	-	-	-
II: I + location of the school	-	-	-	-
III: II + gender	42.8	2.68	20.5	5.38
IV: III + socio-economic background	43.3	2.68	20.0	5.35

Reference categories: Male and Urban areas

Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

To interpret the results, we have to take into consideration the main effect of rural areas in model III (Table 6). The difference between urban and rural areas was -22.6 points, for the benefit of urban areas. When gender and the interaction of gender and municipality type were included in the model, this difference between urban and rural municipalities is the difference for boys only. The boys in urban areas were about 23 points better than boys in rural areas. The main effect of gender (42.8) in model III is the difference between boys and girls in urban municipalities, for the benefit of girls. The gender difference in rural areas is the main effect of gender (42.8) plus the interaction effect of gender and municipality type (20.5), i.e. 63.3 points. The gender difference in rural areas was 20.5 points larger than in urban areas. However, there was no difference between the girls in urban and rural areas, since the main effect of rural areas (-22.6) was about equal to the interaction effect of gender and municipality type (20.5), but negative in sign. Thus, the interaction effect evens out the main effect of rural areas for girls. Girls in urban and rural areas were equally good in reading literacy, when controlling for the effects of school's location and the differences between minority groups, but rural boys were performing worse in reading than urban boys.

However, the picture of gender differences changes, when the socio-economic background factors are included in the model (model IV in Table 7). This leads only to minor changes in the estimates of the main effect of gender and the interaction effect of gender and municipality type. The estimate of the main effect of municipality type changed, however, as seen earlier, and it was now only -7.8 points and not statistically

significant (Table 6). This means that the difference between boys in urban and rural areas was only about eight points for the benefit of urban boys, and this difference was also not statistically significant. The difference between rural and urban girls was now $20.0 - 7.8 = 12.2$ points for the benefit of rural girls, which was a statistically significant difference ($p = .032$).

When the effects of the socio-economic background factors in addition to the minority group differences and the location of school were controlled for, there was no average difference between boys in urban and rural schools, and girls in schools located in rural municipalities were better than girls in urban municipalities. The gender difference was larger in rural municipalities, about 63 points or 72 % of the student standard deviation of the combined reading literacy score, while in the urban municipalities it was about 43 points or 49 % of the student standard deviation.

8.4 Socio-economic background factors: Socio-economic status of the family, parents' educational level and wealth of the family

The fourth block of the inequity factors added into the model consists of the socio-economic status index of the family, parents' educational level and wealth of the family. In Table 8, the estimates of the socio-economic status index are presented. The effect of socio-economic status was divided into two parts, individual and contextual effect. The individual within-school effect was statistically significant, when controlling for the effects of all the other factors in the model. The higher the socio-economic index of the family was, the better was the student's average reading performance in the test. The individual coefficient estimate was about 12, which was the average change within the school in the student's reading literacy score equal to one standard deviation change in the socio-economic index. In the multifactor model, the coefficient estimate of the individual effect was smaller than in the one-factor model, where it was about 18 (Table 2).

Although the coefficient estimate of the contextual effect of the socio-economic index was higher in model IV (Table 8) than in the one-factor model, it was not statistically significant. The standard deviation of the contextual school level socio-economic status index was 0.40. Though the coefficient of 11.4 as such looks high, it was equal to the average difference in the reading literacy scores between two schools which were

2.5 standard deviations apart in their average socio-economic index. One school level standard deviation in the socio-economic index was equal only to 4.6 points in the average reading literacy score.

The effect of parents' educational level on the student performances, when the effects of other factors were controlled for, is presented in Table 9. Students whose parents had only basic education had 15 points lower and students whose parents had completed at least general secondary education had 12 points better results than students in the reference category, i.e. the students whose parents had received vocational secondary education.

The contextual variable of the parents' educational level was derived by averaging the student level values of parents' education, ranging from 1 (did not go to school) to 6 (completed tertiary education), in each school. As in the one-factor model, the coefficient estimate of the contextual variable was not statistically significant, and there was no contextual effect of parents' educational level on the reading literacy performance.

Table 8. Fixed parameter estimates of socio-economic status of the family in constructing the two-level model explaining the combined reading literacy score

Model	Socio-economic status			
	Contextual		Individual	
	Coeff.	s.e.	Coeff.	s.e.
0: Baseline model	-	-	-	-
I: 0 + minority groups	-	-	-	-
II: I + location of the school	-	-	-	-
III: II + gender	-	-	-	-
IV: III + socio-economic background	11.4	8.90	12.3	1.42

Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

The effect of family wealth on the reading literacy performance is presented in Table 10. Unlike in the one-factor model (Table 2), the wealth of the family had neither contextual nor individual effect on the student performances in reading literacy, when the effects of the other factors were controlled for. Though the estimate of the contextual coefficient was 9 and larger than in the one-factor model, the coefficient was not statistically significant. The between-school standard deviation of the family wealth index

was 0.21, and the coefficient of the contextual effect was equal to almost five standard deviations. The coefficient of the individual effect, which was statistically significant in the one-factor model, was now practically zero. The individual effect of the family wealth in the multifactor model is already explained by other socio-economic background factors, i.e. the socio-economic status of the family and parents' educational level.

Table 9. Fixed parameter estimates of parents' educational level in constructing the two-level model explaining the combined reading literacy score

Model	Parents' educational level					
	Contextual		Basic education		General secondary or tertiary	
	Coeff.	s.e.	Coeff.	s.e.	Coeff.	s.e.
0: Baseline model	-	-	-	-	-	-
I: 0 + minority groups	-	-	-	-	-	-
II: I + location of the school	-	-	-	-	-	-
III: II + gender	-	-	-	-	-	-
IV: III + socio-economic background	2.4	7.20	-15.3	3.34	11.7	3.04

Reference category: Vocational secondary education

Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

Table 10. Fixed parameter estimates of family wealth in constructing the two-level model explaining the combined reading literacy score

Model	Wealth of the family			
	Contextual		Individual	
	Coeff.	s.e.	Coeff.	s.e.
0: Baseline model	-	-	-	-
I: 0 + minority groups	-	-	-	-
II: I + location of the school	-	-	-	-
III: II + gender	-	-	-	-
IV: III + socio-economic background	9.4	10.36	0.8	1.75

Statistically significant coefficients at $p \leq 0.05$ are indicated in bold typeface.

8.5 Explanatory power of multifactor model

Until now, the fixed part of the two-level multifactor model has been explored by going through a sequence of models which were formed by adding inequity factors in a group-wise manner, one after another, into the model. Now the focus is on the random part of these models, in order to find out, how much of the between-school and total student variation in reading literacy is attributable to all of the inequity factors together.

The random part estimates and related statistics of the successive two-level models are presented in Tables 11 and 12. In Table 11, the variance components and statistics derived from them, and in Table 12, the proportional reductions in variance components and proportions of explained variances of the models are presented.

The between-school variance component reduced from 435 in the baseline model to 261 in the model including all the inequity factors. The proportional reduction in the between-school variance component was 38 %, and the proportion of explained between-school variance was 0.300. Adding each group of explanatory factors into the model decreased the between-school variance component, except for gender. Adding gender into the model increased the between-school variance component slightly, reflecting the phenomenon discussed in the previous chapter. The within-school variance component reduced from 7307 to 5990 after adding all the factors into the model, which is equal to a proportional reduction of almost 17 %. Although quite a large amount of the between-school variation was explained by the background factors, the proportion of explained total student variance was only 17.8 %, since the between-school variance component constitutes only a small amount of the total student variance. This is the amount of the total student variance in the combined reading literacy score that was attributable to all the inequity factors together.

The intra-class correlation reduced from 0.056 in the baseline model to 0.042 in model IV. There still was between-school variation after controlling for the effects of background factors, but only about 4 % of the unexplained total student variance was attributable to the differences between schools.

Table 11. *Random part estimates and related statistics of two-level models*

Model	Variance components			ICC	Standard deviations		
	Between-School	Within-School	Total variance		Between-School	Within-School	Total
0: Baseline model	435.0	7307.4	7742.4	0.056	20.9	85.5	88.0
I: 0 + minority groups	377.0	6994.6	7371.6	0.051	19.4	83.6	85.9
II: I + location of the school	349.0	6994.2	7343.2	0.048	18.7	83.6	85.7
III: II + gender	369.8	6400.4	6770.2	0.055	19.2	80.0	82.3
IV: III + socio-economic background	261.4	5990.3	6251.7	0.042	16.2	77.4	79.1

Table 12. *Proportional reductions in variance components and proportions of explained variances in two-level models*

Model	Proportional reduction in variance components (%)			Proportion of explained variance	
	Between-School	Within-School	Total variance	Between-School	Total variance
0: Baseline model	-	-	-	-	-
I: 0 + minority groups	13.3	4.3	4.8	0.101	0.048
II: I + location of the school	19.8	4.3	5.2	0.143	0.052
III: II + gender	15.0	12.4	12.6	0.141	0.126
IV: III + socio-economic background	37.9	16.6	17.8	0.300	0.178

Note: In model IV, the reference variance components are estimated with a sample of 4577 students since 281 students have missing values in socio-economic background factors.

In Figure 9, the spread of the school residuals of the combined reading literacy score in two-level models is presented in model order. The school residuals were estimated as school deviations from the intercept of the model. A slight decrease in the spread of the residuals can be seen in the figure, reflecting the decrease in the between-school variance component after controlling for the effects of the inequity factors, even though the variation does not seem to change very much. However, in model IV, the spread of the residuals around the intercept is clearly smaller than in the baseline model. One school is an exception, since it seems to have quite large residual in every model, even in the one including all the inequity factors. A closer look at this school was needed and will be shown below.

About 30 % of the between-school variance and 38 % of the between-school variance component of the two-level model were attributable to all the inequity factors together. The school differences in reading literacy, after controlling for the effects of the background factors are shown in Figure 10. The school residuals of the combined reading literacy score with 95 % confidence intervals are presented in rank order, after controlling for the effects of inequity factors. Only two schools are clearly above and four schools are clearly below the reference line of 0, denoting the intercept. The 95 % confidence intervals of the rest of the school residuals include the reference line and there is not enough evidence to claim that they are different from each other in their average performance, when the effects of the background factors are controlled for. In the baseline model, the total amount of the schools clearly above or below the intercept was 24. The school furthest below the intercept is the same as noted already in Figure 9.

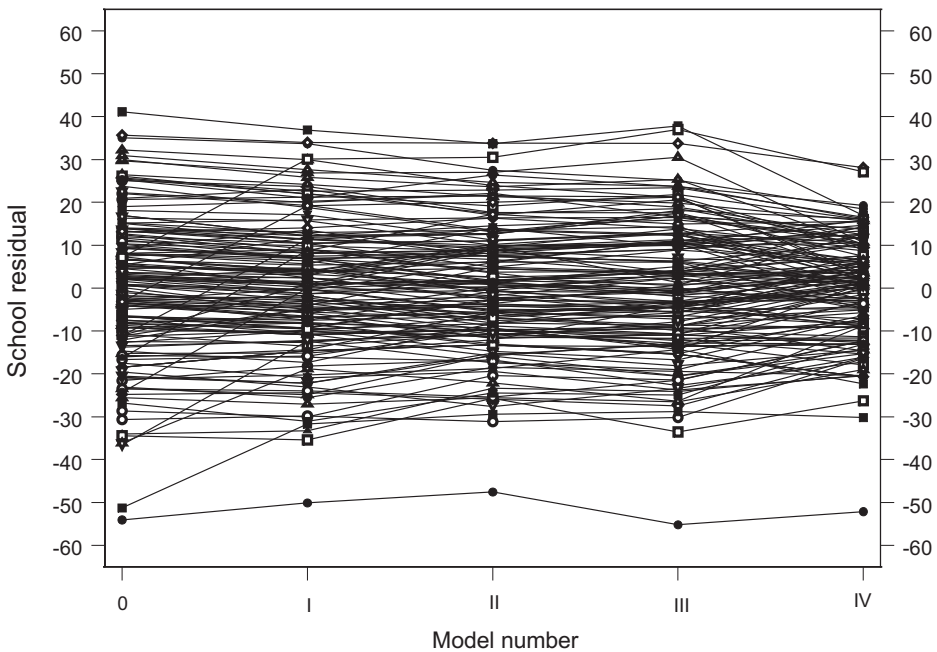


Figure 9. The spread of school residuals of the combined reading literacy score in two-level models in model order

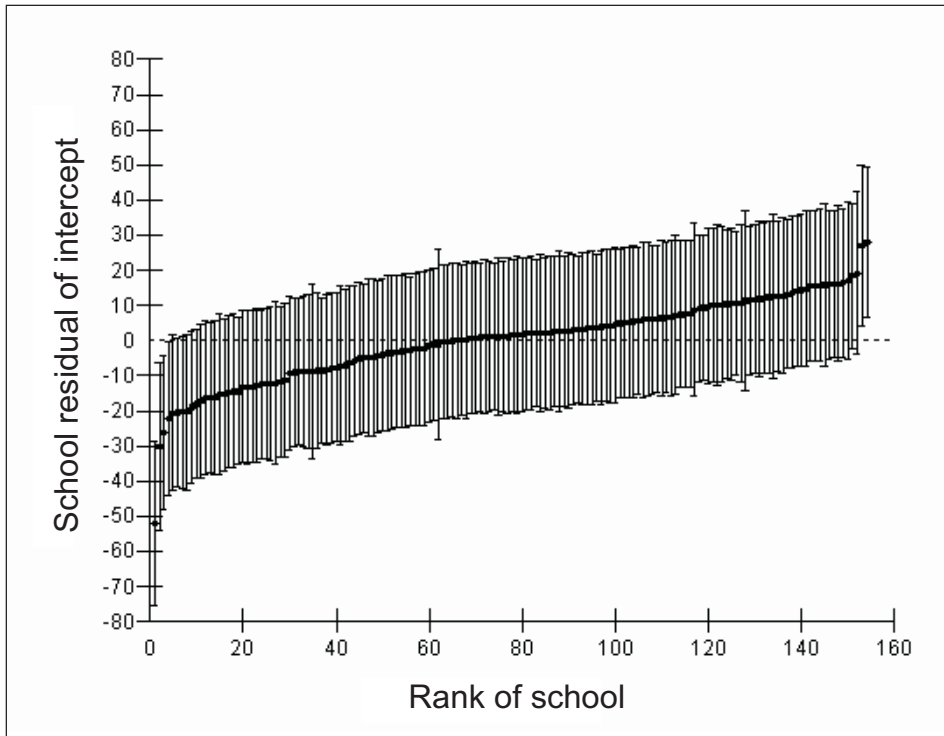


Figure 10. School residuals of the combined reading literacy score with 95 % confidence intervals in rank order, after controlling for the effects of inequity factors

8.6 Decomposition of variance components

In developing the two-level model, the assumption has been that a single variance component at both levels of the data describes the random variation, and the variance is constant across the range of the explanatory variables. However, it may be possible that the variance depends on the explanatory variable, and the variation has to be modelled explicitly as a function of it. For example, the variances may be unequal in different subpopulations. In this case, the variance structure should be modelled as a function of the variable defining the subgroups, to estimate different variance components for these subgroups and to test their equality.

The change in the between-school variance component after adding gender into the model (Table 12) may indicate that the variance components are not equal for boys and girls. Initially, the variance components are decomposed by modelling the variance as a function of gender, and investigating if the between-school, within-school and total student variation depend on the gender.

Table 13. Variance components in gender-only model and Model IV

	Gender-only model			Model IV			Reduction in variance component (%)
	Variance component	s.e.	p	Variance component	s.e.	p	
Between-school							
Females	279.5	77.1	0.000	121.1	54.8	0.027	56.7
Males	745.4	143.5	0.000	513.2	111.4	0.000	31.2
Females/Males	452.2	82.7	0.000	240.8	58.5	0.000	
Within-school							
Females	6002.2	180.0	0.000	5454.0	163.6	0.000	9.1
Males	7149.6	223.2	0.000	6480.9	202.3	0.000	9.4
Total variance							
Females	6281.7			5575.1			11.2
Males	7895.0			6994.1			11.4

Note: Gender-only model is estimated with the same cases as full model (n of students = 4577).

The variance structure was modelled as a function of gender by estimating separate between-school and within-school variance components for boys and girls. In Table 13, the variance components for two models are presented. The first one is the one-factor model with only gender as explanatory variable. This model is used as a reference model for later comparisons. The estimated gender difference in the average performance in this model was about 48 points for the benefit of girls, which is practically the same as in the one-factor model in Table 2. The second model in the table is Model IV including all the inequity factors as explanatory variables.

In the gender-only model, the between-school variance component was much larger for boys than for girls, 745 and 280 respectively. They were both statistically significantly different from zero, and the statistical hypothesis of their equality was rejected ($p=0.001$). Likewise, the within-school variance component was larger for boys (7150)

than for girls (6002), and the statistical hypothesis of their equality was rejected again ($p < 0.001$). As a result, the total student variance was also larger for boys (7895) than for girls (6282). The conclusion is that the between-school variation for boys is larger than for girls, as is the total student variation. The total standard deviation for boys was 89, ten points larger than the standard deviation for girls, which was 79.

In Model IV, after controlling for the effects of the inequity factors, the between-school variance component for girls was only 121 while for boys it was much larger, 513. They were both statistically significantly different from zero, although for girls the p -value of the approximate test was 0.027. The statistical hypothesis of the equality of the between-school variance components was rejected ($p = 0.001$). Likewise, the within-school variance component was larger for boys (6481) than for girls (5454), and the statistical hypothesis of their equality was again rejected ($p < 0.001$). After controlling for the effects of the inequity factors, the unexplained between-school variation for boys was still larger than for girls, as was the total unexplained student variance. The total standard deviation for boys reduced to 81, while for girls it was 74.

Modelling the variance as a function of gender indicates that the between-school variation in the reading literacy performance was larger for boys than girls, as well as the within-school variation. As a result, the total student variance was larger for boys than girls. The result was the same both in the gender-only model and Model IV, where the variation attributable to the inequity factors has been controlled for.

When the variance was modelled as a function of the municipality type, i.e. separate variance components were estimated for rural and urban schools and students, no statistically significant unadjusted or adjusted differences could be found in variance components between the municipality types. The between-school and within school variation in reading literacy was equal in urban and rural communities.

It is interesting to take a look at the differences of the variance components between the regional units. However, the amount of schools and students were quite low in some regions for this kind of comparison, and the results may not be completely reliable, but more tentative in nature. To avoid computational problems due to small amount of units in subgroups, the within-school variance component was constrained equal in all five regional units and estimated using all the students in the sample.

Table 14. Variance components in the regional-units-only and full model

	Regional-units-only model			Model IV			Number of schools
	Variance component	s.e.	p	Variance component	s.e.	p	
Between-school							
Uusimaa Region	732.5	226.7	0.001	666.5	204.5	0.001	37
Southern Finland	196.1	85.0	0.021	32.9	46.2	0.477	52
Eastern Finland	372.5	176.8	0.035	288.0	143.3	0.044	22
Mid-Finland	129.7	107.3	0.227	160.0	106.9	0.134	22
Northern Finland	759.4	305.2	0.013	219.2	131.3	0.095	21
Within-school							
Total	7310.2	150.7	0.000	5989.7	127.3	0.000	154

In Table 14, two models are presented with between-school variance component modelled as a function of the regional units. The first one is the one-factor model with only regional unit as an explanatory variable. This model is used as the reference model for later comparison. The second model is Model IV including all the inequity factors.

The simultaneous test of the statistical hypothesis that all five between-school variance components in the regional-units-only model were equal to zero was rejected ($p < 0.001$), indicating that at least one of the variance components differs statistically significantly from zero. The separate tests in Table 14 indicated that only one of the five between-school variance components did not differ from zero, namely that of Mid-Finland.

In Model IV, the overall test showed again that at least one of the variance components was larger than zero ($p < 0.001$), after controlling for the effects of the inequity factors. The separate tests showed that statistically significant between-school differences existed in the Uusimaa Region and also in Eastern Finland, but not in the other three regions. The between-school variation within regions seemed to be somewhat smaller in other parts of Finland than in the Uusimaa Region.

The student level variance may also depend on the other explanatory variables in the model. However, modelling the variance as a function of the socio-economic status index, parents' educational level or wealth of the family led to the conclusion that the variance was constant with regard to these variables. Likewise, no statistically significant random coefficients of the student level inequity factors were found, indicating that the effects of these factors did not vary between schools.

8.7 Inequity model

The results from the process of model building described above were combined in the final step in constructing the model of inequities in the reading literacy performances in Finland. In the final model, called the inequity model, the statistically significant effects of the inequity factors were included in the same two-level model. Correspondingly, the effects which were not statistically significant were fixed to zeros and left out of the model. Also, the variance was modelled as function of gender by estimating different variance components for boys and girls. These specifications of the two-level model lead to some changes in the estimates of the coefficients and variance components, when compared to the previous results.

The fixed parameter estimates of the inequity model for the combined reading literacy are presented in Table 15 and the random parameter estimates later in Table 16. One of the variables describing the socio-economic background, wealth of the family, was left out of the inequity model. Although the wealth of the family had a statistically significant individual effect in the one-factor model, it had no statistically significant effect on the reading literacy at the student or school level, when the effects of all the other background factors were controlled for. The effect of this variable was already explained by other socio-economic factors.

In the inequity model (Table 15), all three minority group factors had negative effect on the combined reading literacy. The difference in the average performance between Swedish and Finnish speaking schools was 40 points, 45 % of the total student standard deviation, for the benefit of the Finnish speaking schools. Students at grade 9 were 42 points better than students at grade 8, and native students were 61 points better than non-native, when the effects of the other background factors were controlled for.

Table 15. Fixed parameter estimates of the inequity model for the combined reading literacy in Finland

Fixed Parameter	Coeff.	s.e.	t-value	p
Intercept	511.1	5.07	100.81	0.000
1. Minority groups				
Swedish language (ref. Finnish)	-39.9	7.62	-5.24	0.000
Grade 8 (ref. grade 9)	-42.0	3.77	-11.14	0.000
Non-native students (ref. native)	-61.0	11.96	-5.10	0.000
2. Location of school				
Regional unit:				
Uusimaa Region (ref.)				
Southern Finland	19.1	5.00	3.82	0.000
Eastern Finland	15.4	6.34	2.43	0.015
Mid-Finland	15.5	6.09	2.55	0.011
Northern Finland	15.8	6.21	2.54	0.011
3. Gender				
Males (ref.)				
Females, urban schools	44.6	2.74	16.28	0.000
Females, rural schools	59.7	4.11	14.53	0.000
4. Socio-economic background				
Socio-economic status				
Contextual effect	16.5	5.40	3.06	0.002
Individual effect	12.2	1.39	8.78	0.000
Parents' educational level				
Basic education	-15.8	3.29	-4.80	0.000
Vocational secondary (ref.)				
General secondary or tertiary	12.6	3.01	4.19	0.000

Number of schools: 154

Number of students: 4579

Differences between the regional units were clear. In the Uusimaa Region, the average school performance was 19 points lower than in Southern Finland and 15–16 points lower than in other parts of Finland. The differences between regional units were statistically significant only, when the Uusimaa Region was compared to the other regions, but there were no statistically significant differences between the other regions. The average school performance in the combined reading literacy was 17–22 % total student standard deviations better in other parts of Finland, compared to the Uusimaa Region, when the effects of the other background factors were controlled for.

The regional differences, both unadjusted and adjusted for the effects of the background factors, are shown in Figure 11. The Uusimaa Region is the reference category, and the bars for the other regions represent the regional differences in comparison to Uusimaa. The light grey bars are the unadjusted and the dark grey bars are the adjusted differences. The thin lines are the 95 % confidence intervals for the estimates. Among the unadjusted differences, the average of Southern Finland was clearly above that of the Uusimaa Region, but in other regions, the confidence intervals include the average of Uusimaa. Among the adjusted means, all four regions differed from the Uusimaa Region, but not from each other.

As noted earlier, the gender effect had an interaction with the municipality type. According to the municipality type, the schools were called either urban or rural schools. There was no difference in the average performance between boys in urban and rural schools. Girls in urban schools were 45 points better than boys on average, a difference of 51 % of the total student standard deviation in the combined reading literacy score. In rural schools girls were even better, and the difference between boys and girls was 60

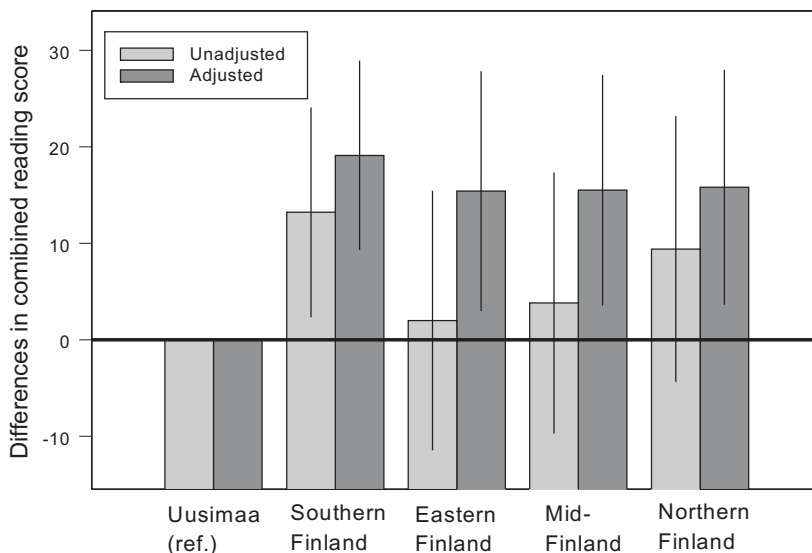


Figure 11. Unadjusted and adjusted regional differences in the combined reading literacy score in Finland, with 95 % confidence intervals

points, which is equal to 68 % of the standard deviation. The difference between girls in urban and rural schools was statistically significant amounting to 15 points, equal to 17 % of the standard deviation.

Only two of the three factors used to describe student's socio-economic background had statistically significant effects on the reading literacy, when the effects of all the other inequity factors were controlled for. Unlike in the one-factor model and the model in Table 8, the socio-economic status of the family was included both as a contextual and as an individual effect in the inequity model. Leaving out the contextual effect of the parents' educational level and the contextual and individual effects of family wealth, since they were not statistically significant, increased the contextual coefficient estimate of the socio-economic status index, and the estimate became statistically significant. On the other hand, if the regional units were not included in the model, the estimate decreased to about half of the one in Table 15 and it was not statistically significant anymore. Also, if the socio-economic status index was completely excluded from the model, the estimates of the regional differences became smaller. Like in the one-factor model, the difference between the Uusimaa Region and Southern Finland was then the only statistically significant difference, the estimate of the difference being about 11 points. The reason for the interdependence between the coefficients of regional units and socio-economic status index on reading literacy performance was the regional variation in the socio-economic status index (Appendix 1).

In the inequity model, the contextual between-school effect of the socio-economic status of the family was 17 points, and the individual within-school effect was 12 points (Table 15). Better socio-economic status was associated with better results within the school, but in addition to that, better socio-economic status of the school increased the average performance of the school.

The empirical values of the rescaled socio-economic index in the Finnish sample varied between the values -2.01 and 2.52 with the mean at 0.08 and a standard deviation of 1.00 . The coefficient of the individual effect of socio-economic status index, 12.2 points, is the increase in the reading literacy score, which is associated with the increase of one standard deviation in the student's socio-economic index within school.

The empirical values of the school's socio-economic index in the Finnish sample varied between -0.71 and 1.28 with mean 0.09 and standard deviation 0.40 . The contextual coefficient of 16.5 is equal to 2.5 standard deviations in the schools' socio-economic index. One school level standard deviation in the socio-economic index is equal to 6.6 points in the schools' average reading literacy score.

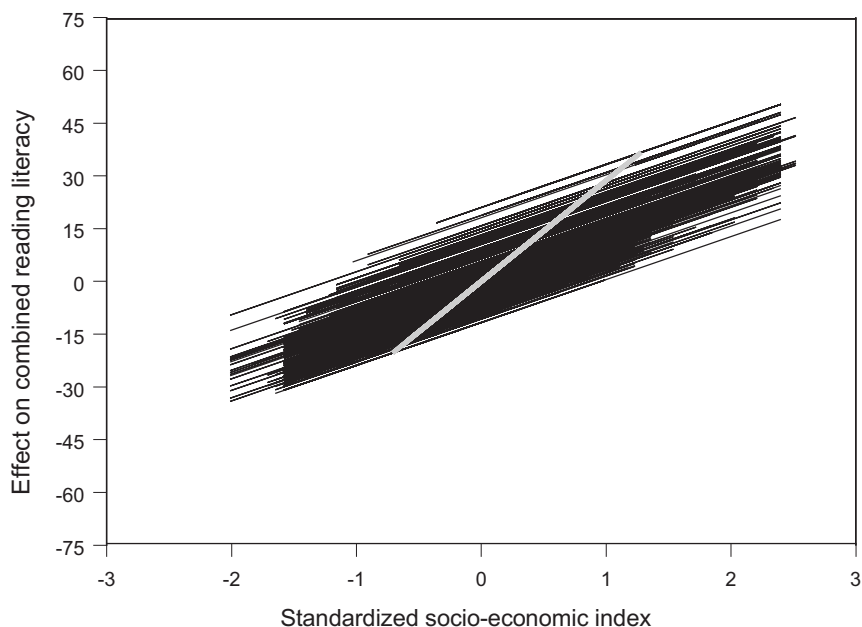


Figure 12. The between-school and within-school regression lines of the combined reading literacy score on the standardized socio-economic index in Finland

The effect of the socio-economic status on reading literacy in the sample schools is shown in Figure 12. The thin black lines are the within-school regression lines of students' reading literacy score on the socio-economic index, and the thick grey line is the between-school regression line of the schools' average reading literacy score on the schools' average socio-economic index.

The student level within-school regression lines are parallel. These regression lines are moving up or down, depending on the average socio-economic index of the school. The within-school effect was clearly positive: on average the better the student's socio-economic index was, the better was his or her result in the reading literacy.

The between-school regression line links the school's average socio-economic index to its average performance. It shows, how the within-school regression lines are moving up or down, depending on the schools' average socio-economic index. The between-school regression line is steeper than the student level regression lines or the regression line based on the contextual coefficient alone, since the coefficient of the

between-school regression line is the sum of the contextual and individual coefficient, an estimate of 28.7 in this model. If there is no contextual effect, the within-school and between-school regression lines collapse to one regression line with an equal coefficient.

In addition to the socio-economic status of the family, the parents' education had an additional effect on the reading literacy. Students whose parents had only basic education achieved 16 points lower results on average, and students whose parents had completed general secondary or tertiary education achieved about 13 points better results, on average, compared to students in the reference category. The difference between these two categories is equal to 33 % of the total student standard deviation in reading.

The factors in the inequity model in Table 15 did not explain all of the between-school or total student variation. The random effects, i.e. the variance components and the total student variance unexplained by the model, are presented in Table 16.

Earlier in Model IV in Table 12, the variance structure was modelled as a function of the intercept. In that model, the proportion of the explained between-school variance was 0.300 and the proportion of the explained total student variance was 0.178. The proportional reductions in the variance components were 37.9 % in the between-school and 16.6 % in the within-school variance component, and the proportional reduction in the total variance was the same as the explained total student variance, 17.8 %. This means that there is variation in the average reading literacy performance between schools unexplained by the model. The between-school variance component is the estimate of this variation. In model IV, the between-school variance component was 261, which gives the estimated between-school standard deviation of 16. In other words, 95 % of the school intercepts vary within the range of 64 points, assuming that the school-level residuals are normally distributed.

In association with Model IV, the variance structure was also modelled as a function of gender, and the reductions in the variance components by gender were considered in Table 13. But Model IV included coefficients that were not statistically significant, and this model was not the same as presented in Table 15. For this reason, the variance components of the inequity model in Table 16 and the proportional reductions in them are slightly different.

As seen before, the between-school and within-school variance components were larger for boys than for girls, as well as the total unexplained student variance. The inequity model seemed to reduce the between-school variance component more for girls than for boys, 54 % and 32 % respectively. The reduction in the within-school variance

Table 16. Random parameter estimates of the inequity model for the combined reading literacy in Finland

Random Parameter	Variance component	s.e.	Reduction in variance component (%)
Between-school			
Females	129.6	55.8	53.6
Males	506.3	110.6	31.8
Females/Males	242.1	58.8	
Within-school			
Females	5451.4	163.5	9.2
Males	6476.4	202.0	9.4
Total variance			
Females	5581.0		11.2
Males	6982.7		11.5
Conditional ICC			
Females	0.023		
Males	0.073		

Number of schools: 154

Number of students: 4579

component was about the same for boys and girls, little over 9 %, as was the case for the reduction in the total student variance, little over 11 %. The reduction in the within-school variance component and total student variance was larger in Model IV, when the variance structure was modelled as a function of the intercept. The reason for this is that gender alone explained 7.5 % of the total student variance and reduced the within-school variance component by 8.3 % (Table 3). Now gender does not explain the within-school variation, since it is estimated separately for boys and girls.

According to the inequity model, the best 15-years old students in reading literacy in Finland had the following combination of background characteristics: They were girls going to school in rural areas in Southern Finland, or at least outside the Uusimaa Region. They came from families where parents had high socio-economic status and were well educated, and they attended schools with high average socio-economic status. In addition, the best 15-years old readers were native Finnish speaking students at grade 9. The poorest 15-years old readers were a combination of the opposite characteristics: They were boys going to school in the Uusimaa Region. They came from families, where parents had low socio-economic status and short education, and they attended

schools with low socio-economic status. However, one has to remember that this is a rough generalization. Most of the individual variation still remained unexplained by the inequity factors of the model.

8.8 Ranking of schools by controlling for the effects of the inequity factors

An attempt to compare schools is to put them in rank order according to the average performance and straightforwardly conclude that the higher the rank is, the better is the school. However, control over the effects of background factors may change the rank order.

An example of the change in rank order is presented in Figure 13. In this figure, the schools in the Finnish PISA 2000 sample are ranked according to the results of the two models considered earlier. The first ranking is based on the baseline model and the second ranking on the results of Model IV. The rank order in both cases is based on the school residuals, i.e. to the estimates of the extent the school means deviate from the intercept of the respective model. These deviations are recoded into ranks, with value 1 for the best performing school and value 154 to the poorest performing school.

The ranking seems to exaggerate school differences, at least in the case of Finland. Ranking introduces differences between schools even in cases, where there are none. According to Figure 1 and the intra-class correlation of 0.056 in the baseline model, the unadjusted school differences were quite small. On the other hand, the impression of school differences based on ranks looks much bigger. Comparing Figure 10 and the rank order based on Model IV in Figure 13 makes the difference between the two approaches even clearer. By visual inspection, the school differences look as large as in the baseline model, even though the variance component, on which the school residuals and ranks are based, has reduced about 38 % and there are only few schools which clearly deviate from the adjusted average performance.

Most interesting in Figure 13, however, is the change in the rank order of the schools, when moving from the baseline model to Model IV. The schools with the largest change in the rank order are presented with thicker lines. These schools consist of ten schools which improved their rank the most and ten schools which worsened their rank the most. One can clearly see that in some cases, the changes were great. There are some schools which were among the best performing schools according to the baseline model, but which were among the poorest performing schools according to Model

IV. Same changes happened to the other direction. Some of the poorest performing schools improved essentially their relative position in the rank order.

The school differences in the performances are small in Finland. Small changes in the school residuals, when shifting from one model to another, may introduce large changes in the rank order. Even though the correlation of school residuals between the two models was as high as 0.76, the rankings of the two models are for some schools dramatically different. In other words, the intensity of school differences is lost, when changing from residuals to ranks.

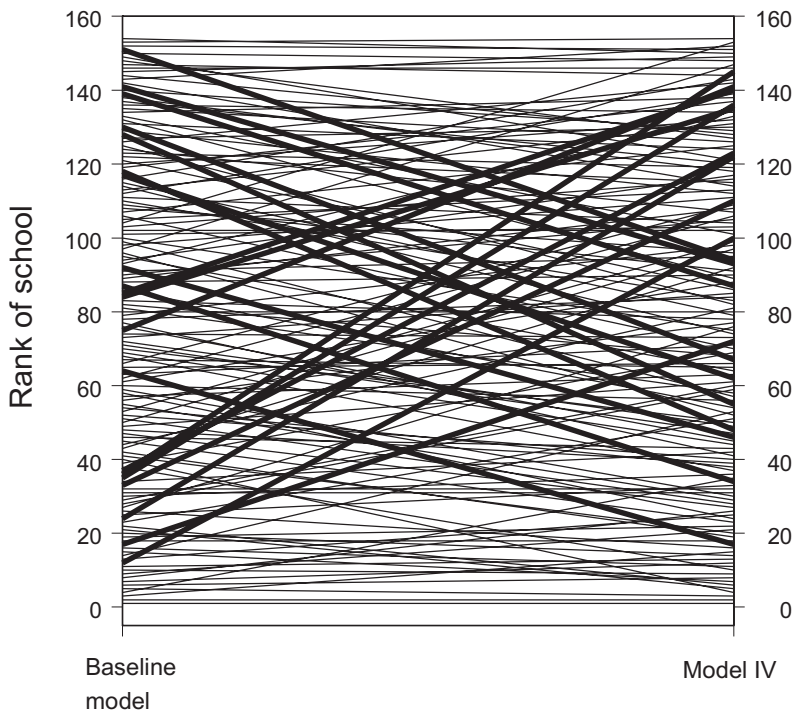


Figure 13. School ranks in baseline model and model IV

8.9 The effect of one single school

Already in Figures 9 and 10 one could see that in the Finnish data set, there was one school with extremely low average performance and it was not in agreement with the other schools. Figure 14 presents the standardized school residuals of the combined reading literacy score by Normal equivalent scores, after controlling for the effects of the inequity factors. The standardized school residual for the school was almost -5 , with normal equivalent score less than -3 . This is the same school that was detected in Figure 9 where school residuals of the combined reading literacy score were presented in the order of model building. This school may violate the model assumption of normally distributed residuals, and the effect of this school on the model estimates has to be studied in detail. One way to test and estimate its effects on the results is to include it in the inequity model using a dummy variable and compare the results with the previous model.

There were 24 students in this school. It was a Finnish speaking school located in the Uusimaa Region. The average socio-economic index of that school was above the average, 0.772 in the standardized scale and 61.5 in the original scale.

The fixed and random parameter estimates of the inequity model, when the school is parametrized using dummy variable, are presented in Appendix 3. The average of this school in the combined reading literacy scale was 117 points below the intercept, which is equal to 1.3 times the total student standard deviation. Controlling for the effect of this exceedingly low-performing Finnish speaking school increased the difference between language groups by 2 points. The effects of grade and immigrant background did not change.

Since this school was located in the Uusimaa Region, controlling for its effect increased the average of the Uusimaa Region and decreased the regional differences by 2–3 points. The differences between the Uusimaa Region and other parts of Finland looked now a little smaller. However, the regional differences between the Uusimaa Region and other parts of Finland were still statistically significant. Regional units are school level variables, and the great deviation of this school from the other schools increased the standard errors of the estimates of regional differences in the inequity model. Accordingly, the standard errors decreased, when the effect of the school was controlled for. The same thing happened to the standard error of the language group estimate.

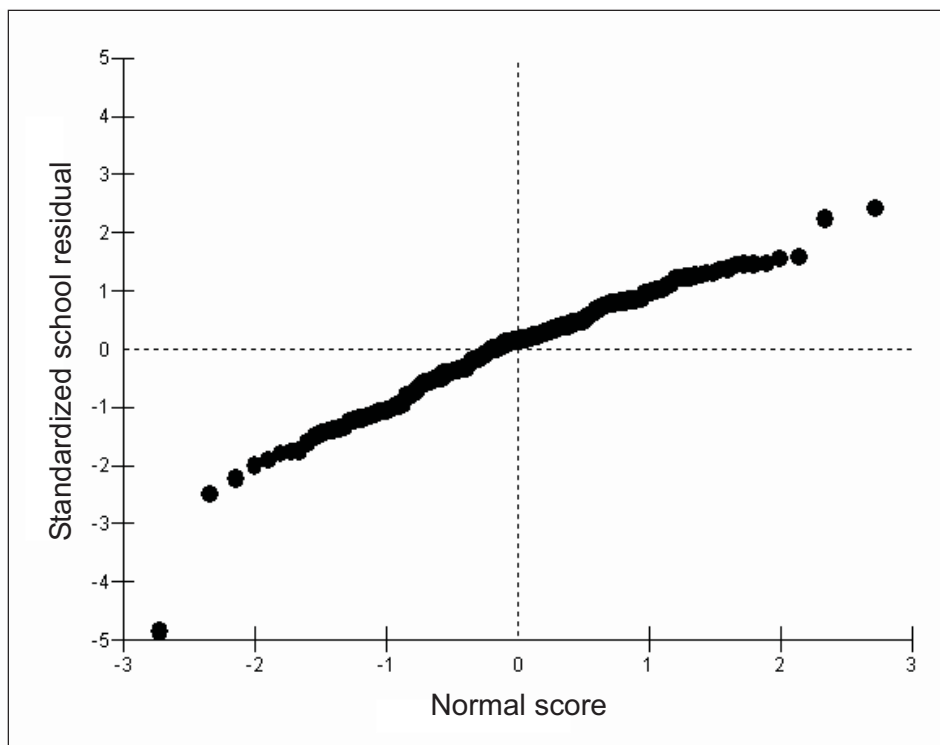


Figure 14. Standardized school residuals of the combined reading literacy score by Normal equivalent scores, after controlling for the effects of background variables

The gender difference did not change, nor the individual effect of the socio-economic status. The coefficient estimate for the contextual effect of the socio-economic status increased slightly, since this school was a school with high socio-economic status. With exceptionally low average performance, it decreased the contextual coefficient estimate in the inequity model. Controlling for this contradiction between the low performance and the high socio-economic status of this school increased the association between them at the school level, and the contextual coefficient became larger for the rest of the schools. The standard error of the coefficient decreased as well and the estimate of the coefficient became more precise. The estimated effect of the parents' education remained practically the same.

This school alone seems to have a big effect on the between-school variance component. Controlling for the effect of this single school decreased the between-school var-

iance component from 261 in Model IV (Table 11) to 188, when the variance structure was modelled as a function of the intercept only. This is equivalent to a 28 % reduction in the estimate of the between-school variance component, and the reduction was 56 %, when compared to the baseline model. The conditional ICC was now only 0.030; about 3 % of the total unexplained student variance is attributable to the differences between schools. At the same time, the proportion of the explained between-school variance increased from 0.300 to 0.411, a change due to a single school alone. However, the proportion of explained total student variance increased only slightly, from 0.178 to 0.188.

When the variance was modelled as a function of gender, the reductions in the between-school variance components looked even more dramatic. The between-school variance component for girls was only 26, which did not differ statistically significantly from 0. The reduction in the variance component, compared to the baseline model, was 91 %. There were practically no between-school differences in the girls' average performances any more. The boys' between-school variance component reduced as well but not as much. The reduction in the variance component was about 40 %. The within-school variance components did not change, since the dummy variable used in the model had no within-school effect, but there was a small increase in the reduction of total variances.

As for the between-school variance components of the regional units presented in Table 14, controlling for the unexpectedly low performance of one single school by using a dummy variable reduced the between-school variance component of the Uusimaa Region both in the regional-units-only model and in the full model. In the regional-units-only model, the variance component reduced from 733 by 25 % into 553, and in the full model from 667 by as much as 48 % into 347. This single school seemed to have a strong effect on the estimated between-school variation in its region. However, the between-school variation was still largest in this region in the inequity model and second largest in the regional-units-only model.

It seems that this single school alone had an effect on the fixed parameter estimates and certainly to the random parameter estimates. It may be the case that this school introduced some bias to the estimation results of the inequity model. However, since this school came from the same school population as the other schools, it may be preferable to include it in the analyses without controlling for its effect.

This analysis also shows how much one deviant school may influence the estimated between-school variation, if the school differences are actually small. The behaviour

of this school was not as expected, nor in agreement with the rest of the data, and the results raise the question, if the students in this school did their best in the PISA 2000 reading literacy test? Or were there some disturbances in the organisation of the reading test which were reflected on the student performances?

8.10 Comparisons with Germany and Sweden

In this section, the Finnish results are compared with the equivalent German and Swedish results. These results were obtained by analyzing the German and Swedish PISA 2000 data sets using two-level models that are as close to the Finnish inequity model as possible. In German and Swedish data sets, variables describing the location of the schools by regional units or municipality type were not available, although at least in Germany, regional variation existed between states (Baumert & al. 2002, 2003). Unlike in Finland, these countries have only one official language, and language comparisons are not included in the models. As minority group variables, the models include student's grade and immigration background. Like in the Finnish model, the only individual characteristic is gender, and variables describing the socio-economic background are the socio-economic status index of the family and the parent's educational level.

The fixed parameter estimates of the German model are presented in Table 17 and the random parameter estimates in Table 18. In the German model, the estimates were adjusted for the differences in the average performances between the school types; i.e. the systematic variation associated with school types was controlled for. Also, the special education schools were left out of the model.

In Germany, the difference between grades 8 and 9 was practically the same as in Finland, 41 points for the benefit of grade 9 students, after controlling for the effects of the other inequity factors. The difference between grades 9 and 10 was smaller, 33 points for the benefit of grade 10 students. The adjusted difference between the native and non-native students was less than a half of the Finnish one, about 29 points for the benefit of the native students. The adjusted gender difference was only 15 points, which was one third of the gender difference in urban schools and one fourth of the gender difference in rural schools in Finland.

Table 17. Fixed parameter estimates of the inequity model for the combined reading literacy in Germany, after controlling for the differences between school types

Fixed Parameter	Coeff.	s.e.	t-value	p
Intercept	496.4	2.73	181.83	0.000
Minority groups				
Grade 8 (ref. grade 9)	-41.0	3.24	-12.65	0.000
Grade 10 (ref. grade 9)	33.1	2.69	12.30	0.000
Non-native students (ref. native)	-28.7	4.25	-6.75	0.000
Gender				
Males (ref.)				
Females	14.5	2.11	6.87	0.000
Socio-economic background				
Socio-economic status				
Contextual effect	26.8	7.61	3.52	0.000
Individual effect	6.4	1.30	4.92	0.000
Parents' educational level				
Primary education	-42.0	6.93	-6.06	0.000
Lower secondary education	-22.9	4.11	-5.57	0.000
Vocational secondary (ref.)				
General upper secondary education	-2.5	3.63	-0.69	0.491
Tertiary education	1.9	2.54	0.75	0.454

Number of schools: 215

Number of students: 4416

When compared to Finland, the effect of the socio-economic status of the family was different in Germany after controlling for the effect of school type. The contextual effect was somewhat stronger than in Finland and the individual effect within schools, accordingly, somewhat weaker. The coefficient estimate of the contextual effect was about 27, and the coefficient estimate of the individual effect within schools was about 6. Within the different school types, the socio-economic status had not as much effect within schools as in Finland, but between schools the effect was still stronger.

The empirical values of the rescaled socio-economic index in Germany varied between the values -2.01 and 2.52 with mean 0.06 and standard deviation 0.95 . The coefficient of the individual effect of socio-economic status index, 6.4 points, is the increase in the reading literacy score which is associated with the increase of about one standard deviation in the student's socio-economic index within school. It was about half of the effect in Finland.

The empirical values of the school's socio-economic index in Germany, without the special schools, varied between -1.20 and 1.37 with mean 0.02 and standard deviation 0.47 . The contextual coefficient of 26.8 is equal to a difference of 2.1 standard deviations in the schools' socio-economic index. One school-level standard deviation in the socio-economic index is equal to 12.6 points in the schools' average reading literacy score. This was the increase in the average reading literacy score associated with the increase of one standard deviation in the school's socio-economic index within school type in Germany. It was almost twice that of Finland.

The effect of the socio-economic status on reading literacy in the German sample schools is shown in Figure 15. The thin black lines are the within-school regression lines of students' reading literacy score on the socio-economic index, and the thick grey line is between-school regression line of schools' average reading literacy score on the

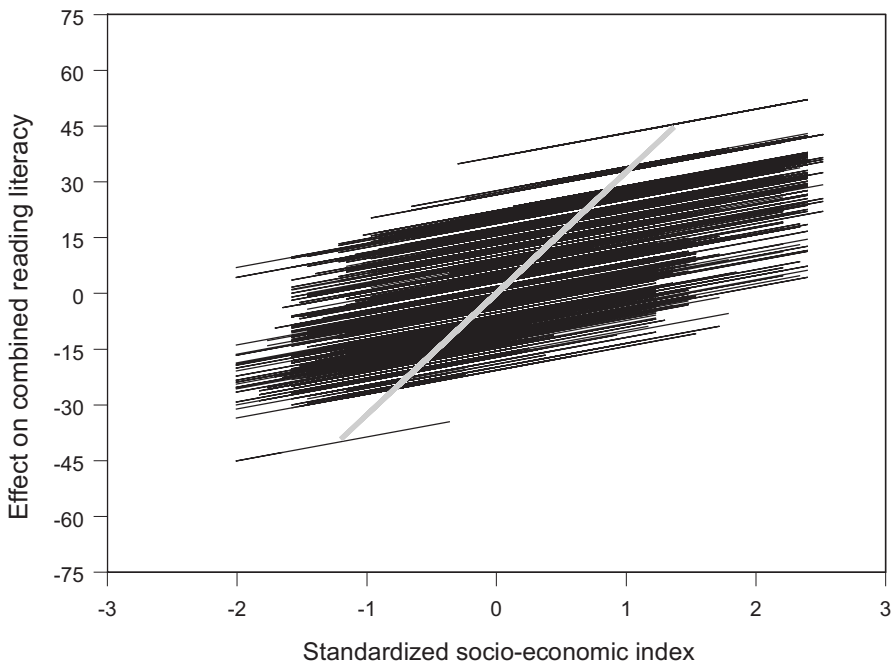


Figure 15. The between-school and within-school regression lines of the combined reading literacy score on the standardized socio-economic index in Germany

schools' average socio-economic index. The figure describes systematic variation between schools within school types. The results in schools with high average socio-economic status were clearly better than in schools with low average socio-economic status. The within-school regression lines are not as steep as in Finland. Within schools, students with high socio-economic status in schools of low average socio-economic status did not reach as good results as students with low socio-economic status in schools of high socio-economic status.

The effect of the parents' educational level was also associated with the students' reading performances, after controlling for the effects of other inequity factors. However, the effect was somewhat different from that in Finland. The parents' low educational level lowered the average student performance in reading, but the parents' high educational level did not raise the average performance, compared to the reference category with vocationally educated parents. Also unlike in Finland, there was a difference between primary education and lower secondary education. The average performance was 42 points below the average of the reference category, if the parents had only primary education, and 23 points below, if the parents had completed lower secondary education. General upper secondary or tertiary education increased the average performance above that of the reference group.

The random parameter estimates of the German model are presented in Table 18. The between-school variance component was of a decent size, and the conditional intra-class correlation was only 0.117. About 12 % of the variance in student performances was attributable to the between-school differences, after controlling for the effects of the inequity factors and the differences between school types. The model reduced the between-school variance component by 26 % and the within-school component by 14 %, compared to the baseline model with the school type. However, one has to remember that controlling for the effect of school types has already reduced the unconditional between-school variance component by 84 % and the total student variance by 40 %. Within school types, the model explained about 15 % of the total student variance, and the proportion of the explained between-school variance was 0.230. There was also some evidence that the within-school variation for boys was larger than for girls in Germany, but the detailed analyses of the variance structure of the German data were not included in this study.

Table 18. *Random parameter estimates of the inequity model for the combined reading literacy in Germany when the differences between school types are controlled for*

Random Parameter	Variance component	s.e.	Reduction in variance component (%)	Proportion of explained variance
Between-School	558.8	75.9	26.2	0.230
Within-School	4203.8	91.7	13.7	
Total Variance	4762.6		15.4	0.154
Conditional ICC	0.117			

Number of schools: 215

Number of students: 4416

The fixed parameter estimates of the Swedish model are presented in Table 19 and the random parameter estimates in Table 20. In Sweden, the adjusted difference between grades 8 and 9 was twice the difference of Finland, 86 points for the benefit of grade 9 students. However, there were only 87 students at grade 8 in Sweden, which was only 2 % of the student sample size. The difference between the native and non-native students was less than in Finland, about 38 points for the benefit of native students, after controlling for the effects of the inequity factors. The adjusted gender difference was only 34 points, which was close to a half of the gender difference in rural schools and 10 points less than the gender difference in Finnish urban schools.

The effect of the socio-economic status of the family was somewhat stronger both at the student and school level in Sweden than in Finland. The coefficient estimate of the contextual effect was about 23, and the coefficient estimate of the individual effect within schools was about 22.

The empirical values of the rescaled socio-economic index in Sweden varied between the values -2.01 and 2.52 with mean 0.11 and standard deviation 0.99 . The coefficient estimate for the individual effect of the socio-economic status index, 21.9 points, is the increase in the reading literacy score which is associated with the increase of one standard deviation in the student's socio-economic index within school. It was almost twice the effect in Finland.

The empirical values of the school's socio-economic index in Sweden varied between -1.16 and 1.28 with mean 0.09 and standard deviation 0.40 . The contextual coefficient of 23.2 was equal to a difference of 2.5 standard deviations in the schools' so-

Table 19. Fixed parameter estimates of the inequity model for the combined reading literacy in Sweden

Fixed Parameter	Coeff.	s.e.	t-value	p
Intercept	507.3	4.79	105.91	0.000
Minority groups				
Grade 8 (ref. grade 9)	-86.3	9.36	-9.22	0.000
Non-native students (ref. native)	-37.9	5.90	-6.42	0.000
Gender				
Males (ref.)				
Females	33.9	2.55	13.29	0.000
Socio-economic background				
Socio-economic status				
Contextual effect	23.2	4.38	5.30	0.000
Individual effect	21.9	1.48	14.80	0.000
Parents' educational level				
Primary education	-27.9	10.65	-2.62	0.009
Lower secondary education	-10.6	6.57	-1.61	0.107
Vocational secondary (ref.)				
General upper secondary education	-0.5	5.14	-0.10	0.923
Tertiary education	-8.5	4.86	-1.75	0.080

Number of schools: 154

Number of students: 4137

cio-economic index scale. One school-level standard deviation of the socio-economic index is equal to 9.3 points in the schools' average reading literacy score. This was the increase in the average reading literacy score associated with an increase of one standard deviation in the school's socio-economic index in Sweden. It was almost one and a half times the effect in Finland.

The effect of the socio-economic status on reading literacy in the Swedish sample schools is shown in Figure 16. The thin black lines are again the within-school regression lines of students' reading literacy score on the socio-economic index, and the thick grey line is the between-school regression line of schools' average reading literacy score on schools' average socio-economic index. The figure describes the systematic variation between schools.

Also in Sweden the results in schools with high average socio-economic status were clearly better than in schools with low average socio-economic status. At the student level, students with high socio-economic status in schools of low average socio-economic

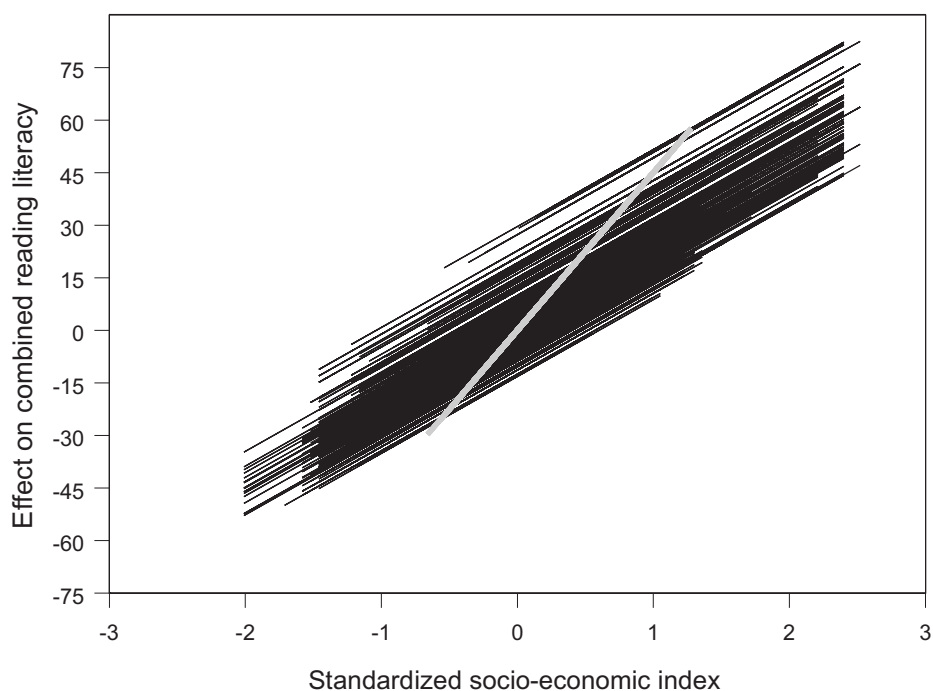


Figure 16. The between-school and within-school regression lines of the combined reading literacy score on the standardized socio-economic index in Sweden

economic status were able to reach better results than students with low socio-economic status in schools of high socio-economic status. Within schools as well as between schools, the socio-economic status had a greater impact on the reading literacy performance in Sweden than in Finland.

The effect of the parents' educational level also had an association with the students' reading performance in Sweden, after controlling for the effects of other inequity factors. Like in Germany, low educational level lowered the student performances on average, but high educational level did not raise the performances, compared to the reference category, where parents had vocational education. Unlike in Finland again, there was a difference between groups of primary education and lower secondary education. The average performance was 28 points below the average of the reference group, if the parents had only primary education, but only 11 points below, if the parents had com-

pleted lower secondary education. However, the latter difference was not statistically significant. The students, whose parents had completed the general upper secondary education, or even tertiary education, did not differ from students with vocationally educated parents in average performance.

The random parameter estimates for the Swedish inequity model are presented in Table 20. The between-school variance component was very small, and the conditional intra-class correlation was only 0.023. Only about 2 % of the variance in student performances was attributable to the between-school differences, after controlling for the effects of the inequity factors in the model. The model reduced the between-school variance component even by 73 %, but the within-school component only by 12 %, when compared to the baseline model. The model explained almost 17 % of the total student variance. In stead of the large reduction in the between-school variance component, the proportion of the explained between-school variance was 0.497, about half of the unconditional between-school variance. Like in Germany, there was also some evidence that the within-school variation for boys was larger than for girls in Sweden, but the detailed analyses of the variance structure were not included in this study.

Table 20. Random parameter estimates of the inequity model for the combined reading literacy in Sweden

Random Parameter	Variance component	s.e.	Reduction in variance component (%)	Proportion of explained variance
Between-School	151.7	45.5	73.4	0.497
Within-School	6566.7	147.1	12.1	
Total Variance	6718.4		16.5	0.165
Conditional ICC	0.023			

Number of schools: 154

Number of students: 4137

8.11 Conclusions

In Finland, there are background factors that introduce systematic variation and hence inequities in the student performances in reading literacy, in the light of the PISA 2000 data.

The gender difference was large for the benefit of girls. However, the gender difference was not constant, but it depended on the municipality type, where the school was located. Girls in rural schools were better than girls in urban schools, but no difference between boys in urban and rural schools could be found. Girls' between-school and within-school variance components and the total variance as well, were smaller than boys'.

Regional differences in student performances existed, but the regional differences were due to differences in the socio-economic background between the regions. When the effect of the students' and schools' socio-economic status were controlled for, the reading literacy performances were lower in the Uusimaa Region around the capital Helsinki than in other parts of Finland. There is also indication that the between-school variation was larger in the Uusimaa Region as well as in Northern Finland than in other parts of Finland.

The students' socio-economic status had both contextual and individual influence on the performances. Higher socio-economic status of the school meant higher school average in reading literacy. In addition, the effect was positive at the student level within schools: higher socio-economic status of the student's family was associated with better results. After controlling for the effect of the socio-economic status, the parents' educational level still had an influence on the performances. Better performances were associated with parents' higher educational level.

In comparing the results of Finland, German and Sweden, the difference between grades 8 and 9 was the largest in Sweden, and it was twice the size of Finland and Germany. The difference between native and non-native students was the largest in Finland and the smallest in Germany, after controlling for the effects of school type and other background factors. Girls were better than boys in each country. The gender difference was the largest in Finland and smallest in Germany.

The effect of the socio-economic status differed between the countries. The contextual effect was the largest in Germany, even after controlling for the effect of the school type. In Sweden the contextual effect was almost as high as in Germany, and clearly higher than in Finland. In addition to the contextual effect, the coefficient of the indi-

vidual effect was statistically significant in every country. This effect was the strongest in Sweden and weakest in Germany. However, it was important to note that the German results were adjusted for the school type.

The effects of the parents' educational level were different in the three countries. A low parental educational level had a stronger worsening effect on the student performances in Germany and Sweden than in Finland. Parents' general upper secondary education or tertiary education did not improve the results in Germany and Sweden, compared to the reference group, where the parents had vocational education, but in Finland it did.

The inequity factors introduced systematic variation between schools and students in all three countries, although the amounts of between-school and total student variation the inequity factors explained were different. In Sweden, the between-school variation was reduced more than in Finland and Germany. The between-school variation was the largest in Germany, even after controlling for the effects of the inequity factors and school type, but the within-school variation was the smallest there. Of the total student variance, about the same amount, 15–18 % was explained in each country. However, the German results were again adjusted for school types.

In Finland, about one third of the between-school variation and little less than one fifth of the total student variance were explained by the inequity model. In Germany, about a quarter of the between-school variation and fifteen percent of the total student variance was explained by the model, after controlling for the variation between school types. In Sweden, about half of the between-school variation and seventeen percent of the total student variance was explained by the model.

In the Finnish PISA 2000 data, the effect of one single school was very strong, especially on the estimate of the between-school variation. Controlling for this effect reduced the unexplained between-school variance component of the inequity model further by one quarter.

Ranking of schools by controlling for the inequity effects changed the unadjusted rank order of schools in Finland. The intensity of school differences is lost in ranking, and small changes in the residuals introduced large changes in the rank order, since the school differences in Finland in reading literacy were small.

9

Reducing inequities with reading activity

In this chapter, the possibility to reduce or moderate the impact of inequity factors is studied. In order to find out, if the effects of inequity factors change while reading activities are controlled for, the inequity model is extended by adding a variable describing the student's engagement and interest in reading into the model. This variable is a student level variable, called Engagement in reading. Unlike inequity factors, the individual values of this factor depend very much on the activities and interests of the students themselves and their families, and they can be influenced by schools or families or students themselves.

The effect of this factor on reading literacy was first studied using a two-level one-factor model, where engagement in reading was the only explanatory factor. After that, the factor was included into the inequity model presented in the previous chapter. The aim was to find out, how much this factor changed the effects of the inequity factors and reduced the between-school and total student variation still present in the inequity model.

9.1 Effects of engagement in reading

The engagement in reading index is a measure that is standardized to have mean 0 and standard deviation 1 in the crossnational data. The Finnish mean was 0.20 and the Finnish standard deviation was 1.05. This index was associated with most of the inequity factors in Finland. Girls were more engaged in reading than boys. The difference in

the index between boys and girls was 0.92 for the benefit of girls, a difference which is almost one student standard deviation. The differences between regional units were not statistically significant, nor between municipality types. The correlation with socio-economic status index was 0.092, a small but statistically significant correlation in a large sample. There were also differences associated with the parents' education. If the parents had basic education only, the difference to the group, where parents had general secondary or tertiary education was 0.23, which was also a statistically significant difference. There was very small, but statistically significant between-school variation in the engagement in reading score but the intra-class correlation was only 0.020.

In Table 21, the parameter estimates of the two-level one-factor model are presented. The combined reading literacy score is regressed on the index of engagement in reading. The effect of the index was divided into two components, contextual and individual. The coefficient estimate of the contextual effect was quite close to zero and not statistically significant. On the other hand, the coefficient estimate of the individual effect was large, about 37. On average, a shift of one international standard deviation in the scale of the index is equal to a shift of 37 points in the combined reading literacy scale, and a shift of one Finnish standard deviation is equal to a shift of 39 points.

Table 21. One-factor two-level model of the combined reading literacy score on the index of engagement in reading in Finland

Fixed Effect	Coeff.	s.e.	t-value	p
Intercept	537.8	2.52	213.41	0.000
Engagement in reading				
Contextual effect	1.7	8.03	0.21	0.832
Individual effect	37.2	1.07	34.77	0.000
Random Effect	Variance Component	s.e.	Reduction in variance component (%)	Proportion of explained variance
Between-school	391.4	65.7	8.0	0.125
Within-school	5721.9	118.8	20.6	
Total variance	6113.3		19.9	0.199
ICC	0.064			

Number of schools: 154

Number of students: 4796

Random part estimates of the model, i.e. between-school and within-school variance components and total residual variances of the combined reading literacy score, and related statistics are also presented in Table 21. When estimating the reductions in the variance components and the proportions of explained variances, only the students with a known value in the engagement in reading factor are included in the baseline model and students with missing values are excluded.

Engagement in reading had a strong effect in reducing the total student variance of the combined reading literacy score. Engagement in reading explained as much as 20 % of the total student variance. The reduction in the between-school variance component was 8 %, but the proportion of explained between-school variance was 0.125. The effects of engagement in reading introduced quite a lot of systematic variation into the reading literacy performance, especially between students, but also between schools.

9.2 Extended inequity model

Next, the inequity model was extended by adding engagement in reading into the inequity model as an explanatory variable. The aim was to find out, whether the effect of this factor would change the effects of the inequity factors, and to what extent the between-school and total student variation in reading literacy performance was attributable to the inequity factors and engagement in reading together?

The fixed parameter estimates of the extended inequity model are presented in Table 22. The coefficient estimate of the individual effect of engagement in reading was 30.2, meaning that one international standard deviation change in the student's engagement in reading was equal to 30 points change in the reading literacy score. This equals to 34 % of the Finnish student standard deviation in reading literacy score. Students, who were four standard deviations apart in their engagement in reading, had a difference of 120 points in their reading literacy performance on average, all other factors in the model being equal.

All the coefficient estimates of the inequity factors were smaller in the extended inequity model, compared to the inequity model of Table 15, except the difference between native and non-native students. This means in general, that the inequity factors did not influence the reading literacy performance as much as earlier, when the effect of the engagement in reading was controlled for. In other words, the variation associated with the inequity factors was smaller among students who were equal with respect to the engagement in reading factor than among all the students.

Table 22. Fixed parameter estimates of the extended inequity model for the combined reading literacy score in Finland

Fixed Effect	Coeff.	s.e.	t-value	p	Reduction in estimate (%)
Intercept	520.8	4.87	106.94	0.000	
1. Minority groups					
Swedish Language (ref. Finnish)	-35.5	7.26	-4.89	0.000	11.0
Grade 8 (ref. grade 9)	-39.6	3.55	-11.15	0.000	15.0
Non-native students (ref. native)	-63.9	11.42	-5.60	0.000	-4.8
2. Location of school					
Regional unit					
Uusimaa Region (ref.)					
Southern Finland	17.5	4.79	3.65	0.000	8.4
Eastern Finland	11.0	6.08	1.81	0.070	28.6
Mid-Finland	11.2	5.83	1.92	0.055	27.7
Northern Finland	12.3	5.95	2.07	0.039	22.2
3. Gender					
Males (ref.)					
Females, urban schools	18.5	2.72	6.80	0.000	58.5
Females, rural schools	27.6	3.97	6.95	0.000	53.8
4. Socio-economic background					
Socio-economic status					
Contextual effect	14.4	5.17	2.79	0.005	12.7
Individual effect	10.4	1.30	8.00	0.000	14.8
Parents' educational level					
Basic education	-11.5	3.08	-3.73	0.000	27.2
Vocational secondary (ref.)					
General secondary or tertiary	9.8	2.82	3.48	0.001	22.2
5. Engagement in reading					
Individual effect	30.2	1.16	26.03	0.000	

N of schools: 154

N of students: 4497

Of the inequity factors, the change in the estimates of gender difference was the largest. Between boys and girls in urban schools, the difference reduced by 26 points, being now only 19 points. Between boys and girls in rural schools, the difference reduced as much as 32 points, being now 28 points. The difference between urban boys and girls reduced by 59 % and between rural boys and girls by 54 %. The rural girls were still better than urban girls by 9 points, which is a statistically significant difference ($p=0.020$), while the difference was 15 points in the inequity model.

Controlling for the effect of engagement in reading reduced the average difference between the Finnish and Swedish schools by 11 % and by 15 % between grade 8 and 9 students. The difference between the non-native and native students increased slightly, by 5 %.

The differences between regional units also became smaller. The largest difference was still between Southern Finland and the Uusimaa Region, almost 18 points. The reduction in this regional difference was the smallest, only 8 %. Differences between other regional units and the Uusimaa Region reduced more, by 22–29 %, being now about 11–12 points. In addition to Southern Finland, only schools in Northern Finland were statistically significantly better, about 12 points on average, compared to schools in the Uusimaa Region. There were no statistically significant differences between the other parts of Finland, only when they were compared to the Uusimaa Region.

The effects of the socio-economic background factors also decreased when engagement in reading was included in the model. The individual effect of the socio-economic status index decreased by 15 % from 12.2 to 10.4 and the contextual effect decreased by 13 % from 16.5 to 14.4. The effect of the parents' educational level decreased as well. The difference between students, whose parents had basic education, changed by 27 % from 15.8 to 11.5, when compared to the reference category of students whose parents had vocational education. The difference was 9.8 points, corresponding to a reduction of 22 % from 12.6 in the inequity model, when students whose parents had general secondary or tertiary education were compared to the reference category.

When the variance of the extended inequity model was modelled as a function of the intercept only, not as a function of gender, the model explained 28.2 % of the total student variance. This was more than in the inequity model, where it was about 18 %. The reduction in the between-school variance component was 39.7 % and in the within-school variance component it was 27.5 %, when compared to the baseline model. There was still variation unexplained by the model in the average performance in reading literacy between schools. In the extended inequity model, the between-school variance component was 258 while the within-school variance component was 5194. Based on these estimates, the conditional ICC was 0.047.

The random parameter estimates, i.e. the variance components and total student variance unexplained by the model, now as a function of gender, are presented in Table 23. As before, the between-school and within-school variance components in the extended inequity model were larger for boys than for girls, as was the case also for the total unexplained student variance. The model reduced the between-school variance

Table 23. *Random parameter estimates of the extended inequity for the combined reading literacy score in Finland*

Random Parameter	Variance component	s.e.	Reduction in variance component (%) Baseline model	Reduction in variance component (%) Inequity model
Between-school				
Females	110.7	48.6	59.4	0.9
Males	510.0	104.2	34.2	3.3
Females/Males	254.0	54.2		
Within-school				
Females	4761.1	143.2	20.2	12.5
Males	5592.4	175.7	20.6	12.9
Total variance				
Females	4871.8		21.9	12.3
Males	6102.4		22.0	12.1
Conditional ICC				
Females	0.023			
Males	0.084			

component for girls more than for boys, 59 % and 34 % respectively, when compared to the baseline model. The reduction in the within-school variance component was about the same for boys and girls, about 20 %, as well as the reduction in the total student variance, about 22 %. The separate variance components for boys and girls seemed to explain less of the within-school variance component and of the total student variance than the model with variance as a function of the intercept only. The reason for this was again that gender in this model did not explain the within-school variation since it was estimated separately for both genders.

When the reductions in variance components were compared to the inequity model, the model reduced the between-school variance component slightly more for boys than for girls, with only about 3 % and 1 %, respectively. The reduction in the within-school variance component was again about the same for boys and girls (13 %), as well as the reduction in the total student variance (12 %). Engagement in reading reduced the student level variation in reading performance more than the inequity factors (Table 16), while the reduction in school-level variation was smaller.

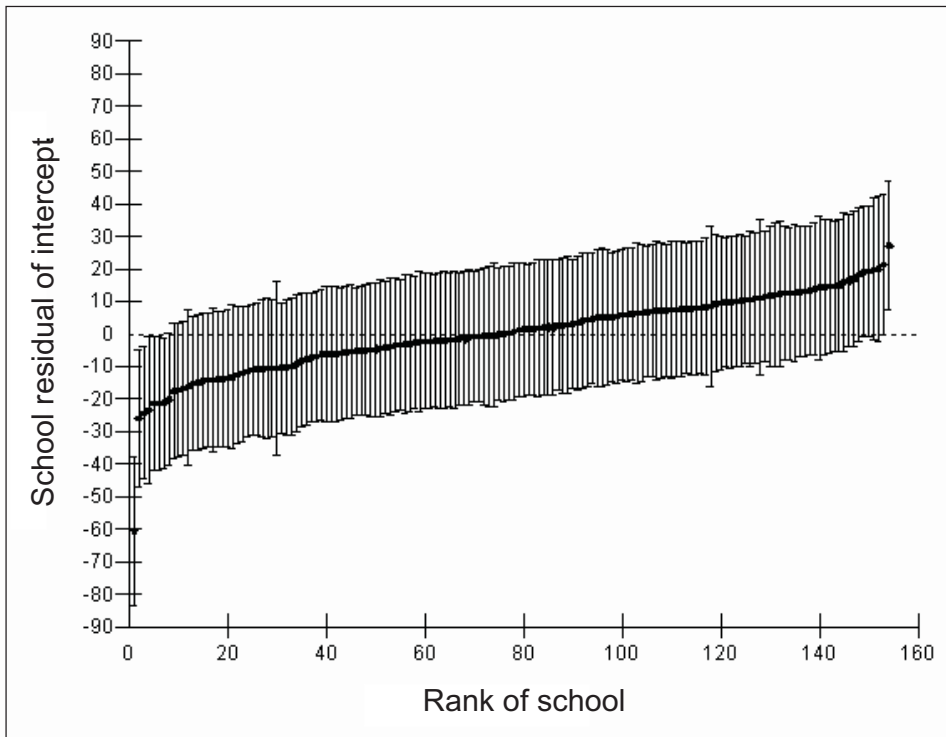


Figure 17. School residuals of the combined reading literacy score with 95 % confidence intervals in rank order, after controlling the effects of inequity factors and engagement in reading

The school differences in the reading literacy performance, after controlling for the effects of the inequity factors and engagement in reading, are shown in Figure 17, where the school residuals of the combined reading literacy score with 95 % confidence intervals are presented in rank order. The school residuals were estimated from the model with the variance as a function of the intercept only. In the figure, only one school is clearly above the intercept and three schools are clearly below the intercept. The 95 % confidence intervals of these schools do not include the intercept, denoted by zero on the y-axis. The confidence intervals of the rest of the school residuals include the intercept and there is not enough evidence to claim that these schools are different from each other in their average performance, when the effects of the inequity factors and

engagement in reading are controlled for. Practically all the schools were so close to each other that any reasonable difference between them cannot be claimed to exist. However, the residual of one school was quite large and negative in sign, and the standardized residual was -5.7. This is the same school already noted, and its effect on the results demands a closer look.

Controlling for the effect of the single school decreased the between-school variance component from 258 to 172, when the variance was modelled as a function of the intercept only. This is equivalent to a 33 % reduction in the estimate of the between-school variance component. The within-school variance component was 5190, and the conditional ICC was 0.032. Only about 3 % of the total unexplained student variance was attributable to the differences between schools, when the effect of this school was controlled for, in addition to the effects of inequity factors and engagement in reading.

When the variance structure was modelled as a function of gender, the between-school variance component for girls was only 10 and not statistically significantly different from zero. This single school alone was responsible for the between-school variance component for girls. The between-school variance component for boys reduced as well, by 18 % to 416, when compared to the previous model, but the variation between the boys' school means in reading performance was still statistically significant.

9.3 Conclusions

Engagement in reading had only a small influence on the between-school variation, and not as much as the inequity factors. At the student level this relationship was reversed. The engagement in reading induced more within-school and total student variation into the data than the inequity factors. However, in addition to the inequity factors, the between-school variation was still reduced to some extent, when the effect of engagement in reading was controlled for.

The effect of engagement in reading indicated that it is possible to compensate for the effects of the inequity factors. When the effect of engagement in reading was controlled for in the extended inequity model, the effects of the inequity factors became smaller among students with equal engagement in reading.

In the extended inequity model, one standard deviation increase in the student's engagement in reading was almost twice the difference between the Uusimaa Region and Southern Finland and more than twice the difference between the Uusimaa Region and other parts of Finland in the average performance. It was also three times as

much as the effect of one between-student standard deviation change within schools in the socio-economic status index of the family, and equal to the effect of five between-school standard deviations' change in the school's socio-economic index, as well. In a hypothetical situation, where the boys' average engagement in reading would be one standard deviation higher than the girls', the extended inequity model would predict that the boys would be better than girls in their reading performance. Actually, the gender difference in urban schools was equal to the effect of 0.61 standard deviations in the engagement in reading scale, and in rural schools it was equal to the effect of 0.91 standard deviations.

Even though we do not actually know, what our model would look like, if boys were reading as much as girls, the model nevertheless indicates that the effects of the inequity factors on reading performance are not so substantial that they could not be overcome by the students' own efforts with support from their families and schools. It is possible to compensate and even out, and maybe to some extent even prevent the effects of inequity factors. Nevertheless, the effect of engagement in reading cannot completely prevent the effects of the inequity factors. They still influence on the performances, both at the school and student level.

Controlling for the effect of one single school reduced the girls' between-school variation practically to zero. This again shows that one school may have a great impact on the estimates of the between-school variation, when the differences between schools are small.

IV

DISCUSSION

10

Conclusions and implications of the study

Combining the high quality of performance with equitable distribution of outcomes are widely accepted educational objectives. The aim of this study was to investigate, from the perspective of equity, the distribution of students' reading literacy performances in Finland both at the student and school level, as they were observed in the international PISA 2000 reading literacy data. The Finnish PISA 2000 data consisted of a representative sample of 15-years old students and schools they attended, and it consisted of 4864 students in 155 schools. The data were analyzed using two-level models in order to find out, how large the between-school variation in student performances was, how strong were the effects of the inequity factors on performances, and how much of the variation at the student and school level was systematic and attributable to the inequity factors. The question was also asked, if the effects of these factors could be reduced. The inequity factors used were the following: language of the school, grade the student attended, student's immigrant background, location of the school as a regional unit and municipality type, gender of the student, socio-economic status of the family, parents' educational level, and wealth of the family. Students' own engagement in reading was used to test, if the effects of the inequity factors could be reduced.

10.1 Conclusions

In the international comparison, the quality of performance in reading literacy was high in Finland. The national average was highest among the PISA 2000 participating countries, and it was associated with a relatively small student variation. However, there

were some clear, although mainly quite small, inequities in student performances with respect to the background factors that the individual students could not escape or control.

The between-school variation in reading literacy performance was small in Finland, particularly in the international comparison. The unadjusted between-school variation was less than 6 % of the total student variance, which was the smallest amount among the PISA 2000 participating countries. Only about 8 % of the sampled schools were clearly below the national average and the same amount of schools were clearly above it. The majority of the schools did not differ statistically significantly from the national mean and there was no clear evidence of the differences between these schools. Even the least successful Finnish schools attained a relatively high average level. The mean performance score of only 5 % of the sampled schools was below the OECD mean.

In this study, the Finnish results were compared to two reference countries, Sweden and Germany. Sweden was chosen, since it has a very similar educational system and comprehensive school to Finland, while in Germany the students are divided in differentiated school types during the basic education. In Sweden, the unadjusted between-school variation was a little larger than in Finland, about 8 % of the total student variance. In Germany, the unadjusted between-school variation was largest of the three countries, 48 % of the total student variance. When the German results were adjusted for the different school types, the between-school variation was still larger in Germany than the unadjusted results in Finland and Sweden, being about 13 %, although controlling for the differences between school types in their average performance reduced the between-school variance component by 84 %. Within-schools discrepancies between individual students were the smallest in Germany, while in Sweden they were somewhat larger than in Finland. About 7 % of the Finnish students did not surpass the minimum level of equity, beneath which nobody should be situated since these students may not be acquiring the necessary literacy knowledge and skills to benefit sufficiently from their educational opportunities. In Sweden this amount was almost twice as large, and in Germany even three times larger than in Finland.

In Finland, about one third of the between-school variance and little less than one fifth of the total student variance were explained by the inequity model, i.e. by the inequity factors together. Most of the systematic variation between schools was associated with the socio-economic status, which was based on the parents' occupations, and parents' educational level, and most of the systematic variation within schools was due to the gender difference.

The gender difference in reading literacy was large, for the benefit of the girls. However, this difference was not constant, but depended on the municipality type, where the school was located. Girls in rural schools were better than girls in urban schools, but no difference between boys in urban and rural schools could be found. Girls' between-school and within-school variance components and also the total variance were smaller than boys'.

Regional differences in student performances existed, but they were largely due to differences in the students' socio-economic background between the regions. When the effect of the students' and schools' socio-economic status was controlled for, the reading literacy performances were lower in the Uusimaa Region around the capital Helsinki than in other parts of Finland. There was also some indication that the between-school variation was somewhat larger in the Uusimaa Region as well as in Northern Finland than in other parts of Finland.

Students' socio-economic status had both contextual and individual influence on performances. Higher socio-economic status of the school meant higher school average in reading literacy. In addition, the effect was positive at the student level within schools: Higher socio-economic status of the student's family was associated with better results. After controlling for the effect of the socio-economic status, the parents' educational level still had an influence on the performances. Better performances were associated with the parents' higher educational attainment. The wealth of the family, instead, had no influence on the performance, when the effects of the other two components of the social background, the parents' socio-economic status and educational attainment, were controlled for.

Inequity was also associated with the students' immigrant and language background. The Swedish-speaking schools did not achieve as good results as the Finnish-speaking ones and the non-native students were below the native students in their performance. However, the numbers of students in these minority groups were small, while larger samples of these students are required for precise estimates of the differences. Students at grade 8 were not as good as students at grade 9.

The effect of one single school was found to be very strong on the estimate of the between-school variation. Controlling for this single school effect reduced the unexplained between-school variance component by one quarter. This finding emphasizes the importance of investigating the existence of influential observations in school data.

Ranking of schools by controlling for the inequity effects changed the unadjusted rank order of schools. Small changes from unadjusted to adjusted school means intro-

duced large changes in the rank order. Some schools were among the best performing schools in the unadjusted ranking, but among the poorest performing schools, when the effects of inequity factors were controlled for. Also some of the poorest performing schools improved essentially their relative position in the rank order.

In comparisons between the results of Finland, Sweden and Germany, we observed that girls were better than boys in each country. The gender difference was the largest in Finland and smallest in Germany. The difference between native and non-native students was the largest in Finland and smallest in Germany, after controlling for the effects of the school type and other background factors. The difference between grades 8 and 9 was the largest in Sweden, where it was twice the size of Finland and Germany.

The contextual effect of the socio-economic status was the largest in Germany, even after controlling for the effect of the school type. In Sweden, the contextual effect was almost as high as in Germany and clearly higher than in Finland. In addition to the contextual effect, the individual effect was the strongest in Sweden and weakest in Germany. However, the German results were adjusted for the school type.

The effect of the parents' educational level was different in the three countries. The parents' low educational level had a stronger worsening effect on the student performances in Germany and Sweden than in Finland. The parents' general upper secondary education or tertiary education did not improve the students' results in Germany and Sweden, when compared to the reference group, where the parents had vocational education, while in Finland it did.

The inequity factors induced systematic variation between schools and students in all three countries, although the amounts of between-school and total student variation explained by the inequity factors were different. By controlling for the effects of the inequity factors, the between-school variation was reduced more in Sweden than in Finland and Germany. The between-school variation was the largest in Germany, even after controlling for the effects of the inequity factors and school type, while the within-school variation was the smallest. Of the total student variance, about the same amount, 15–18 % was explained in each country, after the German results were adjusted for school types.

In the extended inequity model, the between-school variation in Finland was still reduced slightly by controlling for the effect of students' own engagement in reading in addition to the effects of the inequity factors. At the student level, engagement in reading alone explained more of the within-school and total student variation in the data than the inequity factors together. When the effect of engagement in reading was controlled for, the effects of the inequity factors decreased in size.

After controlling for the effects of the inequity factors as well as the students' own engagement in reading, there was still variation between schools in reading literacy performances, though the total reduction in the between-school variance component was 40 %. Of the total student variance, 28 % was explained. The unexplained between-school and within-school variation was larger for boys than for girls and, as a consequence, the total student variance was larger as well. When the effect of one single school was controlled for, the girls' between-school variation was reduced practically to zero. This again emphasizes the importance of locating the influential observations in the data.

10.2 Striving for equity in reading literacy

The aim of the comprehensive school is to provide all children and young people with equal opportunities for learning regardless of their particular school, socio-economic background or the region where they live. The objective of equity in education means providing students with skills and knowledge that ensures the realisation of equity. The associations between students' school achievements and background characteristics are good indicators of the realisation of this objective. School achievements in basic education are the foundation for the students' access into further studies. The associations between the students' school achievements and background characteristics or the schools they attend will be reflected in the limited possibilities and choices in their adult life. Therefore, "in order to promote equality in education, it is crucial that during basic education we strive to prevent the appearance of differences in learning results and attitudes", as stated by Jakku-Sihvonen (2002, 24).

If the differences between schools in educational outcomes are used as indicators of equity in education, the distinction has to be made between school differences before and after controlling for the effects of known inequity factors, such as the socio-economic background. The school differences are partly results of the effects of background factors. School differences may indicate that schools are working with different student compositions, but they may also indicate that the schools have not been able to provide equal opportunities for all children for succeeding and developing in their studies.

In Finland, the majority of total student variance in the reading literacy performances, about 94 % of the total student variance, was due to variation in the students' performances within schools. Thus, only a very small amount of the total student var-

iation was attributable to between-school variation, and therefore the differences between schools cannot influence the students' further education to a high degree. The students in PISA 2000 were aged 15 years and they will be moving from the lower secondary school to either vocational training or upper secondary school. School differences may, however, for some students have an impact on their opportunities for taking the next step in their educational career.

In the international context, Finland has succeeded in keeping the between-school variation very small. However, we cannot ignore the fact that some school performances were clearly below the national mean. Although 84 % of the schools were so close to the Finnish average performance that there were no clear evidence of differences between these schools, this does not contradict the fact that there were individual schools which could do better. If we aim to reduce still the differences in average school achievements, special attention should be paid directly to the schools where the average performance level is the lowest in Finland.

Most of the explained between-school variation was induced by two inequity factors. These factors were the socio-economic status of the family, based on parents' occupations, and parents' educational attainment. Other factors had practically no influence at all on the between-school variation, although the average difference between Finnish and Swedish speaking schools was also reflected in the between-school variation. Probably the schools, where most of the students come from families with low socio-economic background and where the parents' educational level is low in general, are in need of special support.

Gender was the most influential single factor on the total student variation in reading literacy. It was responsible for almost 8 % of the total student variance, while both the socio-economic status and the parents' educational level explained only 3–4 % of the total student variance. Boys seem to need special motivation and help in reading. Instead of concentrating on all boys, the extra help should be directed to the low-achieving boys. These boys need help in becoming familiar with different types of reading materials, and they have to be encouraged to find their own personal area of interest in reading. Also, the between-school as well as the within-school variation was larger for boys than for girls, which may also indicate that low-achieving boys in low-achieving schools are the target group that is mostly in need of support.

When the effect of the students' own engagement in reading was controlled for, the effects of the inequity factors decreased in size. This indicates that it is possible to compensate the effects of the inequity factors (cf. Malin & Linnakylä 2003; Guthrie & Wieg-

field 2000). The effects of the inequity factors on the reading performance were not so large that they could not be overcome by students' own efforts and support from their families and schools. Active engagement in reading may, to some extent, prevent the effects of inequity factors on reading literacy. The schools were able to even out the inequities between students since the between-school variation in the social intake was about twice as large as the unadjusted between-school variation in reading literacy performance.

Regional differences were small and they were associated with regional variation in the students' social background. There was some indication that the school differences were larger in the Uusimaa Region than in other parts of Finland. It is to be hoped that this is not an indication of a process towards increasing school differences and division of the schools into low and high achieving schools (cf. Olkinuora & Rinne 2001) or abandoning the principles of equality and social justice (Rinne and Nuutero 2001).

In a recent study (Jakku-Sihvonen 2004), where the data of several national assessment studies were combined, a stronger concern about the school differences and inequities in Finland was expressed. In that study, the school results were lowest in Northern Finland and highest in the capital area, and better in urban than rural schools. The gender difference was for the benefit of the girls in the learning of languages. The school differences were considered too large, from the point of view of equality in education. The proportion of variation explained by the unadjusted school mean deviations from the grand mean was 10.3 %. This was somewhat larger than the estimate of the between-school variation found in this study, where the Restricted Iterative Generalized Least Squares method was used in estimation (Goldstein 1995, 21–23). When the equal ANOVA-type method was applied instead of multilevel modelling, the proportion of the between-school variation was 8.3 % of the total student variance in the Finnish PISA 2003 reading literacy. In addition to the methodological choices, one reason that explains the differences in the results of these two studies is the focus of the assessments. In national assessments the focus was on curriculum contents, and in PISA it was on knowledge and skills that are learnt both at school and outside it and which are regarded relevant for adult life. It is very likely that the results are not exactly the same in studies where the object of measurement and estimation method is different.

10.3 Methodological reflections

The Finnish PISA 2000 data was a representative sample of Finnish 15-years old students and the schools they attended. The response rate was as high as 93 % of the sampled students and 100 % of the sampled schools participated in the study (Malin & Puhakka 2002). The sampling procedure was stratified according to the regions, which guaranteed regional representativeness in the sample. The combined reading literacy scale was based on 141 items which were a representative sample of a wide variety of various reading tasks. Based on the good representativeness of the data with respect to the student and school populations and the coverage of the reading tasks, the reliability of the results achieved by multi-level modelling is high.

In addition to the results concerning the quality and distribution of reading literacy performances, the study proposes an analysis method for future use in large scale assessment studies. This is a statistical model that estimates simultaneously both the effects of the inequity factors and the between-school variation, after controlling for the effects of the background factors known to induce differences between schools and students. The model takes into account the interdependence of the inequity factors and adjusts the results accordingly. Since the school data have a hierarchical structure, to achieve reliable estimates of the effects of inequity factors and the variation they induce between schools and students, the proper statistical technique is based on multi-level modelling. This method also allows the analyses of longitudinal data as well as data with more than two levels, e.g. the municipalities as level three units.

One important methodological implication of the study is that the effect of one school on the estimate of the between-school variation may be large, especially when the school differences are small. This emphasizes the importance of studying the school effects in detail to avoid overestimating the real differences, since in survey assessments there may always be one or two schools where, for different reasons, the results of the test are lower than the students' knowledge and skills are in reality. If the effect of this kind of influential school is overlooked, the conclusions drawn from the results may be misleading, especially when the real school differences are small.

Some methodological questions would have deserved more attention in the study. In using statistical methods, the researcher has to decide between a number of different methodological alternatives. The details in the results may depend on these choices, and, accordingly, different choices may lead to somewhat different results. Some of the methodological solutions in this study were employed without explicit methodologi-

cal argumentation, following the *PISA 2000 Technical Report* (OECD 2002b). However, the development of methodology would require a thorough discussion and study of the effects of these methodological solutions on the results, although they may interest mainly the statistically oriented readers.

In PISA, appropriate sampling weights were internationally constructed for each national sample, and these weights, after rescaling to sum up to the sample size, were used in this study. However, a more detailed description of the construction of weights and the effects of weighting on the results would be illustrative. More should have been said about the construction of the two different performance scores, plausible values and weighted likelihood estimates, and of the methods used to estimate them, and also of the possible differences in the results between them. A closely argued description of the estimation method and its effects on the results, compared to other estimation methods and results obtained by them, would help to convince methodologically oriented reader of the results. The discussion of the effects of missing data on the results and comparisons with results obtained by using imputed data would help to develop the quality of the assessment survey data.

In studying school differences, the statistical methodology is advanced and complicated. This implies that further education in the application of advanced statistical methodology during research training is needed. In the implementation of assessment surveys, the collaboration between experts covering different areas of research is essential.

10.4 Recommendations for further studies

In this study, the combined reading literacy score of PISA 2000 was used as the response variable in statistical modelling. The first results of PISA 2003, where the primary focus was on the mathematical literacy, are already available (OECD 2004; Kupari & al. 2004), and it will be interesting to test and estimate the models of this study using mathematical literacy data. In future, it will be as interesting to apply the models to the data of PISA 2006, where the primary focus is on the scientific literacy, or to the interdisciplinary subject areas such as problem solving, which was included in PISA 2003.

There are also important student-level factors associated with educational outcomes, like student's motivation, self-regulated learning, ICT competences, talent and effort, which should be included into the statistical models. This would reveal to what extent the effects of the inequity factors could be modified and the between-school variation

reduced, when the effects of these factors are controlled for. In addition, further study is needed about the effects of real school level factors, like size of the school, which were not included in this study. Studies are also needed on the effects of learning outside of the school.

The development of regional differences in performances is an important topic to follow. It may be that the division of Finland in regional units used in the assessment surveys does not reflect the regional variation in the best possible way, and it should be redefined.

One important question is how constant the school differences with respect to yearly changes are. New students enter school every year, and there may be fluctuation between age cohorts in the performances. This means that the average performance of the school may change along with different cohorts, and the rank order of the school is not constant from year to year, especially when the school differences are small. It is likely that in future the amount of students with immigrant background increases at least in schools of the big cities and parents may increasingly use the opportunity to choose a school for their children. This development may also change the rank order of the schools.

Another important question is, how constant are the school differences with respect to the variation between different subject areas. A school's average may be better in one subject than in another. When this is combined with yearly fluctuation, the accurate ranking of schools is largely a delusion until we know how large the variation between student cohorts and subject areas are and can take it into account in addition to the effects of the inequity factors.

10.5 Challenges for strengthening equity in comprehensive school

There were some clear inequities in the student performances in Finland with respect to the background factors that the individual students cannot escape. These inequity factors induced systematic variation between students and schools in the performances. On average, higher socio-economic background and parents' higher education meant better results. Regional differences existed, but they were associated with regional variation in the social background. Gender difference was notable in reading literacy, but not because of the low performance of boys but rather of the extremely good performance of girls. Some systematic variation was observed in the schools' average

performance, but the majority of the schools were within a decent distance around the national mean. In the international comparison, these inequities were small but, however, existing. The results also indicated that the inequities in performances can be reduced with the efforts of the students, parents and schools.

In general, it seems that the Finnish educational system and the comprehensive school have succeeded in keeping the school differences small. Finland also seems to have succeeded reasonably well in the efforts to provide all population groups and regions of the country with equal educational opportunities in the comprehensive school. The biggest problem is not the variation between student groups or schools, but the variation between students within schools. At the school level, the problem is that there are some schools that lag behind others in their achievement, although there is no evidence that the same schools lag behind constantly year after year, and in every subject area. This means that there are individual schools that need help, and especially the low achieving students in these schools are in need of cognitive and motivational support. If we want to help the schools to achieve better average results in the first place, we have to support low-achieving students in these schools.

At the same time we also have to admit that at least some of the best-performing students could also do even better. However, this cannot very much raise the national average performance which is already very high, since a small amount of best-performing students doing even better does not have a great impact on the average performance. But from the individual point of view the best-performing students are entitled to obtain support as well. All students are entitled to personal and individual support in their studies, and more attention should be paid to the development and implementation of individual educational plans and to personalize education for every student.

Yhteenveto

Malin, A. 2005

Koulujen väliset erot ja koulusuoritusten
tasa-arvoinen jakautuminen.

PISA 2000 -lukutaitotutkimuksen tuloksia

Jyväskylän yliopisto. Koulutuksen tutkimuslaitos. Tutkimuksia 15.
ISSN 1455-447X, ISBN 951-39-2247-2 (nid.) ISBN 978-951-39-3041-7 (pdf)

Johdanto

Koulutuksellisen tasa-arvon edistäminen on suomalaisen koulutuspolitiikan keskeinen tavoite. Kaikille pyritään tarjoamaan tasa-arvoiset mahdollisuudet saada koulutusta asuinpaikasta, varallisuudesta, sukupuolesta tai äidinkielestä riippumatta. Peruskoulu pyrkii tarjoamaan kaikille nuorille yhtäläiset oppimismahdollisuudet riippumatta siitä, millainen on heidän kotitaustansa ja millaisissa olosuhteissa tai mitä yksittäistä koulua he käyvät. Peruskoulutuksen tavoitteena on edistää yksilöiden välistä tasa-arvoa niin, että heidän mahdollisuutensa ja valmiutensa kouluttautua ja kehittää itseään aikuisina olisivat yhtäläiset. Oikeudenmukaisuus ja tasa-arvo ovat koulutuspolitiikan johtavia periaatteita, ja koulutuksen tavoitteena pidetään korkeatasoisten kouluosaavutusten yhdistymistä niiden oikeudenmukaiseen ja tasa-arvoiseen jakautumiseen. Tällaisen koulutuksen tuloksia kuvaavat seuraavat seikat: suoritusten keskimääräinen taso on korkea, tulosten vaihtelu on pientä ja suoritusten yhteydet sellaisiin taustamuuttujiin, joihin oppilaat eivät itse voi vaikuttaa ja joita siksi voidaan nimittää tasa-arvon toteutumista ehkäiseviksi tekijöiksi, ovat heikot.

Koulusuorituksia tutkittaessa huomiota kiinnitetään liian paljon vain oppilaiden ja koulujen keskimääräistä tasoa kuvaaviin tunnuslukuihin. Entistä enemmän huomiota

pitäisi kiinnittää tulosten vaihtelun kuvaamiseen ja analysointiin. Osa oppilaiden ja koulujen välisestä suoritusten vaihtelusta on systemaattista, oppilaita ja kouluja kuvaaviin taustatekijöihin liittyvää vaihtelua. Selvittämällä systemaattisen vaihtelun lähteet voidaan arvioida, kuinka tasa-arvoisesti koulusuoritukset jakautuvat oppilaiden ja koulujen perusjoukoissa kokonaisuudessaan. Myös koulujen välinen vaihtelu kertoo koulusaavutusten tasa-arvoisesta jakautumisesta.

Korkean keskimääräisen suoritustason yhdistyminen tulosten tasa-arvoiseen jakautumiseen ei tarkoita, että jokainen oppilas saavuttaisi saman tuloksen, riippumatta hänen yksilöllisistä piirteistään. Se tarkoittaa, että yksilöiden välinen systemaattinen vaihtelu ei liity sellaisiin oppilaita ja kouluja kuvaaviin taustatekijöihin, joiden vaikutusta koulusaavutuksiin ei voida hyväksyä, kuten puutteellinen kotitausta, sukupuoli, asuinpaikka ja vähemmistöryhmään kuuluminen, tai koulu jota oppilas käy. Jos näihin tekijöihin liittyvät erot koulusuorituksissa voidaan minimoida tai poistaa kokonaan, jää yksilöllisiin suorituksiin vielä vaihtelua. Jos tämä vaihtelu on yhteydessä koulussa menestymisen kannalta tärkeinä pidettyihin tekijöihin, niin eroja suorituksissa voidaan pitää hyväksyttävänä. Koulutuksellisen tasa-arvon edistämisen kannalta on olennaisen tärkeää, että peruskoulutuksen aikana pyritään ehkäisemään koulusaavutuksissa ilmenevää ei-toivottua systemaattista vaihtelua. Peruskoulutuksen tulokset ovat yksilön kannalta olennaisen tärkeitä hänen jatko-opintojensa ja elinikäisen oppimisen kannalta.

Tämän tutkimuksen tarkoituksena oli tarkastella koulusuoritusten tasoa ja jakautumista sekä oppilaiden että koulujen välillä Suomessa, sekä suoritusten yhteyttä niihin taustamuuttujiin, joihin oppilaat eivät itse voi vaikuttaa, kuten sukupuoli, kotitausta ja asuinpaikka. Tutkimusaineistona käytettiin PISA 2000 –lukutaitotutkimuksen aineistoa. Aineistoa analysoitiin kaksitasoisella lineaarisella mallilla sen selvittämiseksi, kuinka laajaa oli koulujen välinen vaihtelu lukutaidossa ja kuinka paljon koulujen ja oppilaiden välisestä vaihtelusta oli taustamuuttujiin liittyvää systemaattista vaihtelua. Tutkimuksessa pyrittiin myös selvittämään, voidaanko näiden tekijöiden vaikutuksia vähentää.

Tutkimusongelmat, aineisto ja analyysimenetelmät

Tutkimuksen ongelmat voidaan tiivistää seuraavasti: Ensinnäkin haluttiin selvittää, kuinka suurta oli koulujen välinen lukutaidon vaihtelu. Seuraavaksi tutkittiin, kuinka tasa-arvon toteutumista ehkäisevät tekijät olivat yhteydessä koulujen ja oppilaiden suo-

ritusten vaihteluun sekä erikseen että yhdessä. Sen jälkeen tutkittiin, muuttuvatko näiden tekijöiden vaikutukset, kun lukuharrastukseen sitoutumisen vaikutus kontrolloidaan. Lopuksi haluttiin selvittää, kuinka paljon koulujen ja oppilaiden suoritukset vielä vaihtelivat, kun tasa-arvon toteutumista ehkäisevien tekijöiden ja lukuharrastukseen sitoutumisen vaikutus kontrolloitiin.

Tutkimusaineistona käytettiin PISA 2000 –tutkimuksen Suomen aineistoa, joka on edustava otos 15-vuotiaista koululaisista ja heidän käymistään kouluista vuonna 2000. Otoksessa oli 4864 oppilasta ja 155 koulua, ja oppilasperusjoukon koko oli noin 66 000 ja kouluperusjoukon koko 832. Otantamenetelmänä käytettiin kaksiasasteista ositettua ryväsotantaa. Perusjoukon osittamisperusteina olivat EU:n ja Tilastokeskuksen käyttämät suuralueet, koulun opetuskieli ja kuntamuoto. Kaikki otoskoulut ja oppilaista 93 % osallistuivat tutkimukseen.

Selittävänä muuttujana oli oppilaiden lukutaitoa kuvaava pistemäärä, joka muodostettiin 37 lukutehtävän ja 141 osion perusteella. Lukutehtävistä ja niihin liittyvistä osioista muodostettiin 9 tehtävävihkoa, jolloin yksittäinen oppilas vastasi vain 32–52 lukutaitoa mittaavaan osioon. Lukutaidon pistemäärä estimoitiin IRT-menetelmällä.

Tutkimuksessa oli yhdeksän taustamuuttujaa, joihin oppilas itse ei voi vaikuttaa. Kolme muuttujista oli koulutason muuttujia, nimittäin koulun kieli (suomi ja ruotsi), koulun sijainti sekä alueellisesti (5 suuraluetta) että kuntamuotona (maaseutu ja taajama). Kuusi tekijää oli oppilastason muuttujia: luokka-aste (8 ja 9), oppilaan maahanmuuttajatausta, sukupuoli, perheen sosioekonominen asema, vanhempien koulutustaso ja perheen varallisuus.

Lisäksi tutkimukseen valittiin yksi muuttuja, johon oppilas voi itse vaikuttaa, nimittäin lukuharrastukseen sitoutuminen. Se kuvasi oppilaan lukemista kohtaan tuntemaa kiinnostusta ja hänen lukemisaktiivisuuttaan koulun ulkopuolella. Tämän tekijän oletettiin muokkaavan ja tasoittavan tasa-arvon toteutumista ehkäisevien muuttujien vaikutuksia.

Koska tutkimusaineisto sisälsi kaksitasoisen rakenteen, siis sekä oppilas- että koulutason, ja koska koulujen välisten erojen estimointi ja testaus oli keskeinen tutkimuksen tavoite, tilastollisiin analyyseihin käytettiin kaksitasoista lineaarista mallia. Malli jakaa oppilaiden lukutaitopistemäärän kokonaisvarianssin kahteen osaan, joista toinen kuvaa koulujen välistä lukutaidon tason vaihtelua ja toinen oppilaiden lukutaidon vaihtelua koulujen sisällä. Menetelmä ottaa huomioon havaintoyksiköiden keskinäiset riippuvuudet, mahdollistaa eri tasoja kuvaavien muuttujien samanaikaisen käytön sekä kontekstuaalisten muuttujien vaikutusten luotettavan estimoinnin.

Tulokset

Suomalaisten nuorten keskimääräinen suoritustaso oli korkea, olihan Suomen lukutaidon keskiarvo PISA 2000 –tutkimukseen osallistuneiden maiden paras. Korkeaan tasoon yhdistyi kansainvälisesti verrattaessa suhteellisen pieni oppilaiden välinen vaihtelu. Kuitenkin oppilaiden ja koulujen suorituksissa havaittiin taustatekijöihin liittyvää systemaattista vaihtelua.

Koulujen välinen lukutaidon keskimääräisen tason vaihtelu oli Suomessa hyvin pientä. Koulujen välisen vaihtelun osuus oli alle 6 % oppilaiden kokonaisvarianssista, mikä oli osallistujamaiden pienin arvo. Otokouluista vain 8 % oli selvästi kansallisen keskiarvon alapuolella ja yhtä paljon kouluja oli selvästi sen yläpuolella. Koulujen valtaosa ei eronnut tilastollisesti merkitsevästi kansallisesta keskiarvosta eikä niiden välillä havaittu selviä suoritustasoeroja. Myös huonoiten menestyneet suomalaiset koulut saavuttivat suhteellisen korkean tason, sillä vain 5 % otokouluista oli OECD-maiden yhteisen keskiarvon alapuolella.

Suomen tuloksia verrattiin Ruotsin ja Saksan vastaaviin tuloksiin. Ruotsissa koulujen välinen lukutaidon keskimääräisen tason vaihtelu oli vähän suurempaa kuin Suomessa, noin 8 % oppilaiden kokonaisvarianssista. Saksassa koulujen välinen vaihtelu oli suurinta, peräti 48 % oppilaiden kokonaisvarianssista. Kun eri koulumuotojen välisen vaihtelun vaikutus kontrolloitiin, koulujen välinen vaihtelu oli Saksassa 13 % oppilaiden kokonaisvarianssista. Yksityisten oppilaiden väliset lukutaidon suorituserot koulujen sisällä olivat pienimpiä Saksassa, ja Ruotsissa ne olivat hieman Suomea suurempia. Noin 7 % suomalaisista nuorista jäi lukutaidoltaan alle sen tason, jota koulutusmahdollisuuksien kaikinpuolisen hyödyntämisen katsotaan edellyttävän. Ruotsissa tämä oppilasjoukko oli suunnilleen kaksinkertainen ja Saksassa jopa kolminkertainen Suomeen verrattuna.

Suomessa noin kolmasosa koulujen välisestä lukutaidon vaihtelusta ja lähes viidesosa oppilaiden kokonaisvarianssista selittyi tutkimukseen mukaan otetuilla tasa-arvon toteutumista ehkäisevillä tekijöillä, kun ne kaikki olivat tilastollisen mallin selittäjinä. Suurin osa koulujen välisestä systemaattisesta vaihtelusta liittyi vanhempien sosioekonomiseen asemaan ja koulutukseen. Suurin osa oppilaiden välisestä systemaattisesta vaihtelusta koulujen sisällä johtui tyttöjen ja poikien välisestä erosta.

Sukupuolten välinen ero lukutaidossa oli erittäin suuri tyttöjen hyväksi. Tämä ero oli kuitenkin erilainen maaseudulla ja kaupungissa. Tytöt olivat maaseudun kouluissa parempia kuin kaupunkikouluissa, mutta poikien kesken ei tällaista eroa ollut. Siksi myös tyttöjen ja poikien välinen ero oli maaseutukouluissa kaupunkikouluja suurem-

pi. Sekä koulujen välinen että koulujen sisäinen vaihtelu, kuten myös lukutaitopistemäärän kokonaisvarianssi, olivat tytöillä pienempiä kuin pojilla.

Oppilaiden suorituksissa oli jonkin verran alueellista vaihtelua, joka oli yhteydessä sosioekonomisen taustan alueelliseen vaihteluun. Kun oppilaiden ja koulujen sosioekonomisen taustan vaikutus kontrolloitiin, lukutaidon keskimääräinen taso oli heikoin Uudellamaalla. Viitteitä ilmeni myös siitä, että koulujen välinen vaihtelu oli Uudellamaalla ja Pohjois-Suomessa hieman suurempaa kuin muualla Suomessa.

Oppilaiden sosioekonominen tausta vaikutti sekä koulun että oppilaan tasolla. Koulujen korkeampaan sosioekonomiseen tasoon liittyi keskimäärin oppilaiden parempi lukutaito. Sen lisäksi oppilaan korkea sosioekonominen asema koulun sisällä paransi vielä lukutaitoa. Oppilaan tuloksia paransi myös vanhempien korkea koulutustaso. Perheen varallisuudella ei ollut yhteyttä suorituksiin, kun sosioekonomisen aseman ja vanhempien koulutuksen vaikutus kontrolloitiin.

Myös oppilaiden maahanmuuttajatausta ja kielitausta olivat yhteydessä suorituksiin. Ruotsinkielisten koulujen tulokset eivät olleet yhtä hyviä kuin suomenkielisten, ja ulkomaalaistaustaisten oppilaiden suoritukset olivat heikompia kuin Suomessa syntyneiden. Nämä vähemmistöryhmät olivat kuitenkin kooltaan varsin pieniä, ja erojen luotettavampi estimointi edellyttäisi suurempia oppilasmääriä näistä ryhmistä. Kahdeksaluokkalaisten suoritukset olivat selvästi heikompia kuin yhdeksaluokkalaisten.

Yhdellä koululla oli suuri vaikutus koulujen välistä vaihtelua kuvaavaan estimaattiin. Tämän koulun vaikutuksen kontrollointi vähensi koulujen välistä vaihtelua neljänneksellä. Yksittäisen muista poikkeavan koulun vaikutus analyysituloksiin voi olla suuri, erityisesti kun todelliset koulujen väliset erot ovat pieniä.

Tasa-arvon toteutumista ehkäisevien tekijöiden vaikutusten kontrollointi muutti koulujen välistä alkuperäistä paremmuusjärjestystä. Jotkut alkuperäisen järjestyksen mukaan parhaiten sijoittuvista kouluista olivat huonoimpien joukossa, kun näiden tekijöiden vaikutukset kontrolloitiin. Samoin jotkut huonot koulut paransivat asemaansa olennaisesti.

Sukupuolten välinen ero lukutaidossa oli Suomessa suurempi kuin Ruotsissa, ja Saksassa se oli pienin. Ulkomaalaistaustaisten ja kotimaassa syntyneiden välinen ero oli myös Suomessa suurin ja Saksassa pienin, kun taustamuuttujien vaikutukset kontrolloitiin. Sosioekonomisen aseman kontekstuaalinen vaikutus oli suurin Saksassa, vaikka koulumuodon vaikutus kontrolloitiin. Ruotsissa kontekstuaalinen vaikutus oli melkein yhtä suuri kuin Saksassa ja selvästi suurempi kuin Suomessa. Sosioekonomisen aseman yhteys lukutaitoon koulun sisällä oli vahvin Ruotsissa ja heikoin Saksassa.

Vanhempien koulutuksen vaikutus vaihteli hieman maiden kesken. Vanhempien matala koulutustaso heikensi oppilaiden suorituksia enemmän Saksassa ja Ruotsissa kuin Suomessa. Vastaavasti vanhempien korkea koulutustaso ei parantanut oppilaiden tuloksia Saksassa ja Ruotsissa kuten Suomessa.

Tasa-arvon toteutumista ehkäiseviin tekijöihin liittyi systemaattista koulujen ja oppilaiden välistä lukutaidon vaihtelua kaikissa kolmessa maassa, vaikkakin niiden määrä vaihteli. Nämä tekijät selittivät Suomessa noin kolmanneksen koulujen välisestä vaihtelusta ja hieman alle viidenneksen oppilaiden välisestä kokonaisvarianssista. Saksassa koulujen välisestä vaihtelusta selittyi noin neljännes ja oppilaiden välisestä kokonaisvarianssista noin viisitoista prosenttia, kun koulumuodon vaikutus kontrolloitiin. Ruotsissa koulujen välisestä vaihtelusta selittyi noin puolet ja oppilaiden välisestä kokonaisvarianssista noin seitsemäntoista prosenttia.

Kun tasa-arvon toteutumista ehkäisevien tekijöiden lisäksi kontrolloitiin oppilaiden lukuharrastukseen sitoutumisen vaikutus, koulujen välinen lukutaidon keskimääräisen tason vaihtelu Suomessa pieneni vielä hieman. Sitoutuminen lukuharrastukseen yksinään selitti enemmän oppilaiden välistä lukutaidon vaihtelua koulujen sisällä ja oppilaiden välistä kokonaisvaihtelua kuin tasa-arvon toteutumista ehkäisevät tekijät yhdessä. Kun lukuharrastukseen sitoutumisen vaikutus kontrolloitiin, tasa-arvon toteutumista ehkäisevien tekijöiden vaikutukset heikkenivät. Erityisesti tyttöjen ja poikien välinen ero pieneni.

Kun sekä tasa-arvon toteutumista ehkäisevien tekijöiden että lukuharrastukseen sitoutumisen vaikutukset kontrolloitiin, koulujen välillä oli yhä jonkin verran eroja lukutaidon keskimääräisen tason vaihtelussa. Alkuperäinen koulujen välinen vaihtelu pieneni 40 % ja oppilaiden välisestä kokonaisvarianssista selittyi 28 %. Kun sen lisäksi vielä yhden koulun vaikutus kontrolloitiin, tyttöjen koulujen välinen keskimääräisen lukutaidon vaihtelu hävisi kokonaan.

Pohdinta

Tutkimuksessa löytyi tasa-arvon toteutumista ehkäiseviä oppilaita ja kouluja kuvaavia taustatekijöitä, joihin liittyi systemaattista oppilaiden välistä lukutaidon vaihtelua. Jonkun verran vaihtelua oli myös koulujen välillä, mutta valtaosa kouluista sijoittui kohdullisen lähelle kansallista keskiarvoa. Kansainvälisesti verrattuna koulujen väliset erot olivat pieniä. Oppilaiden ja koulujen välisiä eroja voitiin myös tasoittaa oppilaiden omin, vanhempien ja koulujen ponnistuksin.

Yleisesti ottaen suomalainen koulujärjestelmä ja peruskoulu ovat onnistuneet pitämään koulujen väliset erot pieninä. Suomi näyttää myös selvinneen kohtalaisen hyvin pyrkimyksessään tarjota kaikille väestöryhmille ja eri alueilla asuville oppilaille yhtäläiset mahdollisuudet menestyä peruskoulussa. Suurin ongelma ei olekaan systemaattinen vaihtelu oppilasryhmien tai koulujen välillä, vaan oppilaiden suoritusten vaihtelu koulujen sisällä. Koulutasolla ongelmana on, että jotkut koulut näyttävät menestyvän selvästi muita jonkun verran huonommin. Tietoa ei kuitenkaan ole siitä, ovatko nämä koulut samoja vuodesta toiseen ja oppiaineesta toiseen. On kuitenkin olemassa kouluja, jotka tarvitsevat ylimääräistä tukea. Jos näiden koulujen keskimääräisiä suorituksia halutaan parantaa, onkin tuettava heikoimmin menestyviä oppilaita näissä kouluissa.

On myös todettava, että parhaiten menestyvilläkin oppilailla on vielä mahdollisuuksia parantaa suorituksiaan. Kaikilla oppilailla on oikeus henkilökohtaiseen ja yksilölliseen tukeen opinnoissaan, ja enemmän huomiota tulisi kiinnittää henkilökohtaisten opetussuunnitelmien ja yksilöllisen opetuksen kehittämiseen kaikille oppilaille. Parhaiten menestyvien oppilaiden ja koulujen tukeminen ei kuitenkaan kohota kansallista keskimääräistä suoritustasoa kovinkaan paljon, ellei samaan aikaan pidetä huolta myös heikkojen oppilaiden ja muita heikommin menestyvien koulujen kehityksestä.

References

- Ahonen, S. 2003. *Yhteinen koulu – tasa-arvoa vai tasapäisyyttä.* [Common school – equality or evenness.] Tampere: Vastapaino.
- Aitkin, M. & Longford, N. 1986. Statistical Modelling Issue in School Effectiveness Studies. *Journal of the Royal Statistical Society, Ser. A*, 1–43.
- Basic Education Act 628/1998. Online <<http://www.oph.fi/english/>>.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (Hrsg.) 2002. *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich.* Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Tillmann, K.-J. & Weiß, M. (Hrsg.) 2003. *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland.* Opladen: Leske + Budrich.
- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (Hrsg.) 2001. *PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich.* Opladen: Leske + Budrich.
- Benadusi, L. 2001. Equity and Education: A critical review of sociological research and thought. In W. Hutmacher, D. Cochrane & N. Bottani (Eds.) *In Pursuit of Equity in Education.* Dordrecht: Kluwer Academic Publishers.
- Birzea, C. 1982. *La pédagogie du succès.* Paris: PUF.
- Bloom, B. S. 1976. *Human Characteristics and School Learning.* New York: McGraw-Hill.
- Bryk, A., & Raudenbush, S. 1992. *Hierarchical Linear Models: Application and Data Analysis Methods.* Newbury Park: Sage Publications.
- Coleman, J. S. 1966. Equal schools or equal students? *The Public Interest*, 4, 70–75.
- Döbert, H., Klieme, E. & Sroka, W. (Eds.) 2003. *Conditions of School Performance in Seven Countries. A Quest for Understanding the International Variation of PISA Results.* New York: Waxmann Münster.
- European Group of Research on Equity of the Educational Systems 2004. *Equity of the European Educational Systems. A Set of Indicators.* Second edition. Draft version.
- Ganzeboom, H. B. G., De Graaf, P. & Treiman, D. J. 1992. A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1, 1–56.
- Goldstein, H. 1987. *Multilevel Models in Educational and Social Research.* London: Griffin.
- Goldstein, H. 1995. *Multilevel Statistical Models.* London: Edward Arnold.
- Goldstein, H. 2002. *Multilevel Statistical Models.* Third Edition. London: Edward Arnold.
- Grisay, A. 1984. Les mirages de l'évaluation scolaire. *Revue de la Direction Générale de l'Organisation des Études*, XIX, 8.

References

- Guthrie, J. T. & Wigfield, A. 2000. Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of reading research*. Vol. III LEA, 403–425.
- Gutmann, A. 1999. *Democratic Education*. Princeton: Princeton University Press.
- Hautamäki, J., Arinen, P., Hautamäki, A., Kupiainen, S., Lindblom, B., Mehtäläinen, J., Niemivirta, M., Rantanen, P., Ruuth, M. & Scheinin, P. 2003. *Oppimaan oppiminen yläasteella 2. Tilanne vuonna 2001 ja muutokset vuodesta 1997*. [Learning to learn at lower secondary schools 2. Situation in 2001 and changes from 1997.] *Oppimistulosten arviointi* 6/2003. Helsinki: Opetushallitus.
- Haven, H. (Ed.) 1999. *Education in Finland 1999: Statistics and indicators*. SVT Education 1999:4. Helsinki: Statistics Finland.
- Hox, J. J. 1994. *Applied Multilevel Analysis*. Amsterdam: TT-Publikaties.
- Husén, T. 1974. *Learning society*. London: Methuen.
- Husén, T. 1975. *Social influences on educational attainment*. Centre fo Educational Research and Innovation (CERI), Organisation for Economic Co-operation and Development.
- Husén, T. 1986. *The learning society revisited*. Oxford: Pergamon.
- Husén, T. 1989. *Skolan I fokus*. Stockholm: Almqvist & Wiksell.
- Isoaho, H., Kivinen, O. & Rinne, R. 1990. *Nuorten koulutus ja kotitausta*. [Education and family background of young people.] *Tutkimuksia* 171. Helsinki: Tilastokeskus.
- Jakku-Sihvonen, R. 2002. Differences between school-specific results in terms of equality in education. In R. Jakku-Sihvonen & J. Kuusela (Eds.) *Evaluation of the Equal Opportunities in the Finnish Comprehensive Schools 1998–2001*. *Evaluation* 11/2002. Helsinki: National Board of Education, 9–26.
- Jakku-Sihvonen, R. & Komulainen, E. 2004. *Perusopetuksen oppimistulosten meta-arviointi*. [Meta analysis of learning outcomes in basic education.] *Arviointi* 1/2004. Helsinki: Opetushallitus.
- Jakku-Sihvonen, R. & Kuusela, J. 2002. *Evaluation of the Equal Opportunities in the Finnish Comprehensive Schools 1998–2001*. *Evaluation* 11/2002. Helsinki: National Board of Education.
- Kivinen, O. & Rinne, R. 1995. *Koulutuksen periytyvyys. Nuorten koulutus ja tasa-arvo Suomessa*. [Heritability of education. Education and equality among young people in Finland.] SVT, Koulutus 1995:4. Helsinki: Tilastokeskus.
- Kupari, P., Reinikainen, P., Nevanpää, T., & Törnroos, J. 2001. *Miten matematiikkaa ja luonnontieteitä osataan suomalaisessa peruskoulussa? Kolmas kansainvälinen matematiikka- ja luonnontiedetutkimus Suomessa*. [How well are mathematics and science known in the Finnish comprehensive school? The Third International Mathematics and Science Study in Finland.] Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Kupari, P., Välijärvi, J., Linnakylä, P., Reinikainen, P., Brunell, V., Leino, K., Sulkunen, S., Törnroos, J., Malin, A. & Puhakka, E. 2004. *Nuoret osaaajat. PISA 2003 -tutkimuksen ensituloksia*. [The skilled young people. First results of the PISA 2003 study.] Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Kuusela, J. 2002. Links between school results and demographic factors. In R. Jakku-Sihvonen & J. Kuusela (Eds.) *Evaluation of the Equal Opportunities in the Finnish Comprehensive Schools 1998–2001*. *Evaluation* 11/2002. Helsinki: National Board of Education, 27–49.
- Kuusinen, J. 1985. Lukioon siirtyneiden lahjakkuus ja sosiaalinen tausta. [Talent and social background of students in the general upper secondary school.] *Sosiologia*, 22 (3), 191–196.

- Kuusinen, J. 1986. Koulumenestys, lahjakkuus ja sosiaalinen tausta. [Success in school, talent and social background.] *Kasvatus*, 17 (3), 192–197.
- Kuusinen, J. 1992. Hyvät, huonot ja keskinkertaiset. [The good, the bad and the average.] *Kasvatus*, 23 (1), 47–56.
- Laiho, I. & Silvennoinen, H. 1992. Tutkijankoulutetut yhä valikoitunutta joukkoa. [Graduate students are even more selected group.] *Korkeakoulutieto*, 19 (1), 26–32.
- Lappalainen, H.-P. 2000. *Peruskoulun äidinkielen oppimistulosten kansallinen arviointi 9. vuosiluokalla 1999*. [National evaluation of learning outcomes in the mother tongue at grade 9 in the comprehensive school in 1999.] Oppimistulosten arviointi 1/2000. Helsinki: Opetushallitus.
- Lappalainen, H.-P. 2001. *Perusopetuksen äidinkielen ja kirjallisuuden oppimistulosten kansallinen arviointi 9. vuosiluokalla 2001*. [National evaluation of learning outcomes in the mother tongue and literature at grade 9 in the comprehensive school in 2001.] Oppimistulosten arviointi 6/2001. Helsinki: Opetushallitus.
- Lehtonen, R., Nissinen, K. & Pahkinen, E. 1992. *Tilastollinen analyysi ryväsotantaan perustuvissa frekvenssiaineistoissa*. [Statistical analysis of frequency data based on cluster sampling.] Jyväskylän yliopisto. Tilastotieteen laitoksen julkaisuja 16.
- Lehtonen, R. & Pahkinen, E. J. 2004. *Practical Methods for Design and Analysis of Complex Surveys*. Chichester: Wiley. 2nd ed.
- Lie, S., Linnakylä, P. & Roe, A. 2003. Northern lights on PISA. In S. Lie, P. Linnakylä & A. Roe (Eds.) *Northern Lights on PISA. Unity and Diversity in the Nordic Countries in PISA 2000*. Oslo: University of Oslo, Department of Teacher Education and School Development, 7–20.
- Lie, S., Linnakylä, P. & Roe, A. (Eds.) 2003. *Northern Lights on PISA. Unity and Diversity in the Nordic Countries in PISA 2000*. Oslo: University of Oslo, Department of Teacher Education and School Development.
- Lietz, P. 1996. *Changes in reading comprehension across cultures and over time*. New York: Waxmann Münster.
- Linnakylä, P. 1995. *Lukutaidolla maailmankartalle. Kansainvälinen lukutaitotutkimus Suomessa*. [The IEA Study of Reading Literacy in Finland.] Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos.
- Linnakylä, P. 2002. Miten kotitausta näkyy oppimistuloksissa? [The effect of home background on learning outcomes.] In Välijärvi, J. & Linnakylä, P. (Eds.) *Tulevaisuuden osaajat. PISA 2000 Suomessa*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 89–107.
- Linnakylä, P. 2003. Finland. In Döbert, H., Klieme, E. & Sroka, W. (Eds.) *Conditions of School Performance in Seven Countries. A Quest for Understanding the International Variation of PISA Results*. New York: Waxmann Münster, 150–218.
- Linnakylä, P. & Malin, A. 2001. Demand for adult education and determinants of participation. In A. Tuijnman & Z. Hellström (Eds.) *Curious Minds. Nordic Adult Education Compared*. Copenhagen: Nordic Council of Ministers, 68–103.
- Linnakylä, P. & Malin, A. 2003. How to reduce the gender gap in reading literacy. In S. Lie, P. Linnakylä & A. Roe (Eds.) *Northern Lights on PISA. Unity and Diversity in the Nordic Countries in PISA 2000*. University of Oslo, Department of Teacher Education and School Development, 39–53.
- Linnakylä, P., Malin A., Blomqvist, I. & Sulkunen, S. 2000. *Lukutaito työssä ja arjessa. Aikuisten kansainvälinen lukutaitotutkimus Suomessa*. [Literacy at and beyond work. The International Adult Literacy Survey in Finland.] University of Jyväskylä, Institute for Educational Research.

References

- Linnakylä, P., Sulkunen, S. & Arffman, I. (Eds.) 2004. *Tulevaisuuden lukijat. Suomalaisnuorten lukijaprofiileja. PISA 2000*. [Readers of the future. Reading profiles of the Finnish young people. PISA 2000.] Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Malin, A. 1994. Monitasomallien estimointi ryväsotantaan perustuvissa koulututkimuksissa ML3-ohjelmalla. [The estimation of multilevel models in school data based on cluster sampling using ML3 software.] Helsinki. *Suomen Tilastoseuran vuosikirja 1993*, 235–261.
- Malin, A. 1997. Monitasomallit koulututkimuksessa. [Multilevel models in school studies.] Helsinki. *Suomen Tilastoseuran vuosikirja 1996*, 193–207.
- Malin, A. 2004. Examining for outliers in the Finnish PISA 2000 data. Unpublished manuscript. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Malin, A., Lehtonen, R. & Pahkinen, E. 1995. Koulusaavutusaineiston analyysi asetelmaperusteisilla ja kaksitasoisilla logit-regressiomalleilla. [The analysis of school achievement data using design based and two-level logit models.] *Kasvatus*, 26 (2), 101–115.
- Malin, A. & Linnakylä, P. 2001. Multilevel Modelling in Repeated Measures of the Quality of Finnish School Life. *Scandinavian Journal of Educational Research*, 45 (2), 145–166.
- Malin, A. & Nissinen, K. 1996. *Analysis of school achievement data with multilevel models: Likelihood and Bayesian approaches*. University of Jyväskylä. Preprints from the Department of Statistics, 18.
- Malin, A. & Puhakka, E. 2002. PISAn otannasta ja kvantitatiivisista menetelmistä. [Sampling and quantitative methods in the PISA study.] In J. Välijärvi & P. Linnakylä (Eds.) *Tulevaisuuden osaajat. PISA 2000 Suomessa*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 219–230.
- Malin, A. & Salmela, T. 1993. Millaisia saavutuseroja on koulujen välillä? [School differences in learning outcomes.] In V. Brunell & P. Kupari (Eds.) *Peruskoulu oppimisympäristönä. Peruskoulun arviointi 90 -tutkimuksen tuloksia*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos, 169–182.
- Ministry of Education 2004. *The Finnish education policy*. Online <<http://www.minedu.fi/minedu/education/index.html>> .
- Mislevy, R. J. 1991. Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56, 177–196.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. 1992. Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, 29, 133–161.
- Mislevy, R. J. & Sheehan, K. M. 1987. Marginal estimation procedures. In A. E. Beaton (Ed.) *The NAEP 1983–1984 Technical Report*. Report No. 15-TR-20. Princeton: Educational Testing Service, 293–360.
- Mislevy, R. J. & Sheehan, K. M. 1989. Information matrices in latent-variable models. *Journal of Educational Statistics*, 14(4), 335–350.
- National Board of Education 2000. *More quality to the quality policy of education*. Helsinki: National Board of Education.
- National Board of Education 2004. *Perusopetuksen opetussuunnitelman perusteet 2004*. [The foundations of curriculum for the basic education in 2004.] Online <http://www.oph.fi/info/ops/pops_web.pdf>.
- OECD 1999. *Measuring Student Knowledge and Skills. A New Framework for Assessment*. Paris: OECD Publications.
- OECD 2000. *Measuring Student Knowledge and Skills. The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. Paris: OECD Publications

- OECD 2001. *Knowledge and Skills for Life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD Publications.
- OECD 2002a. *Reading for Change – Performance and Engagement across Countries. Results from PISA 2000*. Paris: OECD Publications.
- OECD 2002b. *PISA 2000 Technical Report*. Paris: OECD Publications.
<http://www.pisa.oecd.org/Docs/Download/PISA_TechnicalReport_H.pdf>
- OECD 2003. *Education at a glance: OECD indicators 2003*. Paris: OECD.
- OECD 2004. *Learning for Tomorrow's World. First results from PISA 2003*. Paris: OECD Publications.
- Olkinuora, E. & Rinne, R. 2001. Miten menee peruskoulussa? [What is the situation in the comprehensive school?] Teoksessa E. Olkinuora & E. Mattila (toim.) *Miten menee peruskoulussa? Kasvatuksen ja oppimisen edellytysten tarkastelua Turun kouluissa*. Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja A: 195, 197–224.
- Orfield, G. 2001. Why data collections matter: the role of race and poverty indicators in American education. In W. Hutmacher, D. Cochrane & N. Bottani (Eds.) *In Pursuit of Equity in Education*. Dordrecht: Kluwer Academic Publishers.
- Pahkinen, E. & Kupari, P. 1991. Analysis of Educational Frequency Data from a Complex Sample Survey. *Scandinavian Journal of Educational Research*, 35(3), 213–225.
- Pahkinen, E. & Lehtonen, R. 1989. *Otanta-asetelmat ja tilastollinen analyysi*. [Sampling designs and statistical analysis.] Helsinki: Gaudeamus.
- Porter, A. C. & Gamoran, A. 2002. Progress and Challenges for Large-Scale Studies. In A. C. Porter & A. Gamoran (Eds.) *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington: National Academy Press.
- Postlethwaite, T. N. & Ross, K. 1992. *Effective schools in reading*. The Hague: IEA.
- Postlethwaite, T. N. & Wiley, D. 1992. *The IEA study of science II: Science achievement in twenty-three countries*. Oxford: Pergamon.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I. & Lewis, T. 2000. *A user's guide to MLwiN. Version 2.1b*. Multilevel Models Project. University of London. Institute of Education.
- Raudenbush, S. 1989. "Centering" predictors in multilevel analysis: Choices and consequences. *Multilevel Modelling Newsletter*, 1 (2), 10–12.
- Rawls, J. 1971. *A theory of justice*. Oxford: University Press.
- Rinne, R. & Nuutero, A.-K. 2001. Turkulainen peruskoulu, koulutuspolitiikan muutos ja vanhempien näkökannat. [Comprehensive school in Turku, change of educational policy and parent views.] Teoksessa E. Olkinuora & E. Mattila (toim.) *Miten menee peruskoulussa? Kasvatuksen ja oppimisen edellytysten tarkastelua Turun kouluissa*. Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja A: 195, 87–126.
- Rubin, D. B. 1987. *Multiple Imputations for Non-Response in Surveys*. New York: Wiley.
- Saari, H. & Linnakylä, P. 1993. Oppimisen tasosta peruskoulussa. [The level of learning outcomes in the comprehensive school.] In H. Saari and P. Linnakylä (Eds.) *Oppiiko oppilas peruskoulussa. Peruskoulun arviointi 90 -tutkimuksen tuloksia*. Jyväskylän yliopisto: Kasvatustieteiden tutkimuslaitos.
- Sen, A. 1992. *Inequality re-examined*. Oxford: Oxford University Press.
- Skinner, C.J., Holt D. & Smith T.M.F. (Eds.) 1989. *Analysis of Complex Surveys*. New York: John Wiley & Sons.
- Snijders, T. & Bosker, R. 1994. Modeled variance in two-level models. *Sociological Methods & Research*, 22, 342–363.

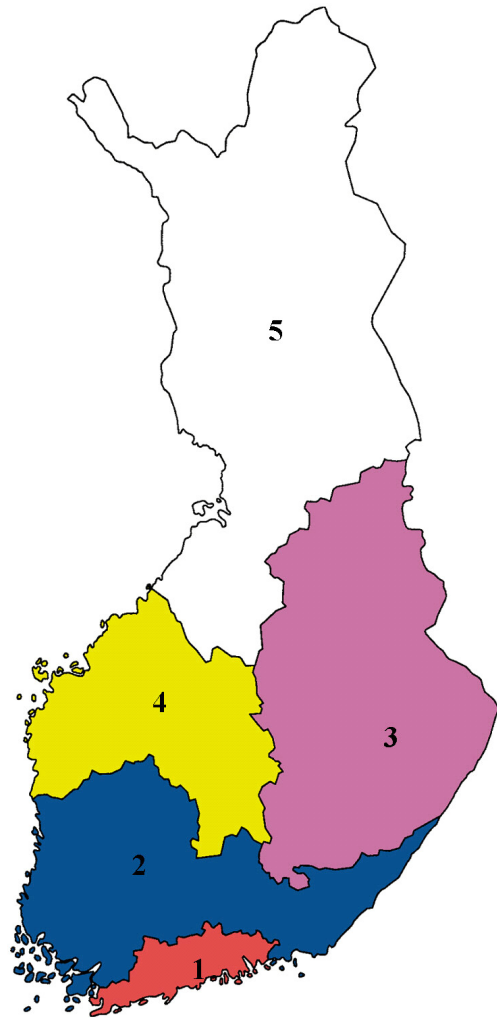
References

- Snijders, T. & Bosker, R. 2002. *Multilevel Analysis. An Introduction to Basic and Advanced Multi-level Modeling*. London: Sage Publications.
- Standards for educational and psychological testing 1999. Prepared by the Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. Washington, DC: AERA.
- Stenström, M.-L. 1997. *Educational and Gender Equality in Vocational Education. The case of Commercial Education in Finland*. Research Reports 1. University of Jyväskylä: Institute for Educational Research.
- Tuokko, E. 2000. *Peruskoulun 9. vuosiluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi 1999*. [National evaluation of learning outcomes in the English language at grade 9 in 1999.] Oppimistulosten arviointi 3/2000. Helsinki: Opetushallitus.
- Väljjarvi, J. 1994. Kaventuva tasa-arvo lisääntyvän vapauden hinta [Diminishing equality as a price for increased freedom.] *Kasvatus*, 25 (2), 219–221.
- Väljjarvi, J. & Linnakylä, P. (Eds.) 2002. *Tulevaisuuden osaajat. PISA 2000 Suomessa*. [Skillful actors of future. Future talents. PISA 2000 in Finland.] Jyväskylän yliopisto: Koulutuksen tutkimuslaitos.
- Väljjarvi, J., Linnakylä, P., Kupari, P., Reinikainen, P., Malin, A. & Puhakka, E. 2001. *Suomen tulevaisuuden osaajat. 15-vuotiaiden nuorten lukutaito sekä matematiikan ja luonnontieteiden osaaminen kansainvälisessä vertailussa*. [Skillful actors of future in Finland. Reading, mathematical and scientific literacy of the 15-year olds in an international comparison.] Jyväskylän yliopisto. Koulutuksen tutkimuslaitos.
- Väljjarvi, J., Linnakylä, P., Kupari, P., Reinikainen, P. & Arffman, I. 2002. *The Finnish success in PISA – and some reasons behind it*. University of Jyväskylä: Institute for Educational Research.
- Väljjarvi, J. & Malin, A. 2002. Onko koulutuspalvelujen laadussa eroja Suomessa. [Are there differences in educational services in Finland?] In Väljjarvi, J. & Linnakylä, P. (Eds.) *Tulevaisuuden osaajat. PISA 2000 Suomessa*. Jyväskylän yliopisto: Koulutuksen tutkimuslaitos, 109–120.
- Väljjarvi, J. & Malin, A. 2003. The two-level effects of socio-economic background. In S. Lie, P. Linnakylä & A. Roe (Eds.) *Northern Lights on PISA. Unity and Diversity in the Nordic Countries in PISA 2000*. University of Oslo, Department of Teacher Education and School Development, 123–131.
- Warm, T. A. 1985. Weighted maximum likelihood estimation of ability in Item Response Theory with tests of finite length. *Technical Report CGI-TR-85-08*. U. S. Coast Guard Institute, Oklahoma City.

Appendix 1

Regional units in Finland, used in PISA 2000

1. Uusimaa Region
2. Southern Finland
3. Eastern Finland
4. Mid-Finland
5. Northern Finland



Appendix 2

Means of students' socio-economic status index in regional units and municipality types, estimated by two-level model

Regional unit and municipality type	Socio-economic status index		Standardized socio-economic status index	
	Mean	s.e.	Mean	s.e.
Uusimaa Region	56.0	0.90	0.438	0.057
Southern Finland	49.6	0.67	0.045	0.048
Southern Finland, Urban areas	50.5	0.78	0.099	0.045
Southern Finland, Rural areas	46.4	0.70	-0.153	0.085
Eastern Finland	46.4	1.10	-0.152	0.073
Eastern Finland, Urban areas	50.3	0.81	0.090	0.066
Eastern Finland, Rural areas	42.5	1.17	-0.390	0.065
Mid-Finland	47.8	1.46	-0.065	0.073
Mid-Finland, Urban areas	50.2	1.80	0.089	0.099
Mid-Finland, Rural areas	42.6	0.80	-0.383	0.143
Northern Finland	48.4	1.29	-0.028	0.075
Northern Finland, Urban areas	51.0	1.39	0.130	0.080
Northern Finland, Rural areas	43.2	1.12	-0.347	0.114
Total, Urban areas	52.3	0.54	0.209	0.031
Total, Rural areas	43.8	0.57	-0.308	0.054
Total	50.2	0.52	0.085	0.032

Appendix 3

Fixed and random parameter estimates of the inequity model for the combined reading literacy in Finland, when the effect of one single school is controlled for

Table 1. Fixed parameter estimates of the inequity model for the combined reading literacy in Finland, when the effect of one single school is controlled for

Fixed Parameter	Coeff.	s.e.	t-value	p
Intercept	512.3	4.72	108.54	0.000
One single school	-116.9	17.98	-6.50	0.000
1. Minority groups				
Swedish language (ref. Finnish)	-41.9	4.72	-8.88	0.000
Grade 8 (ref. grade 9)	-41.4	3.76	-11.01	0.000
Non-native students (ref. native)	-60.9	11.89	-5.12	0.000
2. Location of school				
Regional unit				
Uusimaa Region (ref.)				
Southern Finland	16.4	4.47	3.67	0.000
Eastern Finland	13.2	5.63	2.34	0.019
Mid-Finland	13.2	5.41	2.44	0.015
Northern Finland	13.5	5.53	2.44	0.015
3. Gender				
Males (ref.)				
Females, urban schools	44.6	2.74	16.28	0.000
Females, rural schools	59.7	4.11	14.53	0.000
4. Socio-economic background				
Socio-economic status				
Contextual effect	18.2	4.88	3.73	0.000
Individual effect	12.2	1.38	8.84	0.000
Parents' educational level				
Basic education	-15.5	3.27	-4.74	0.000
Vocational secondary (ref.)				
General secondary or tertiary	13.0	3.00	4.33	0.000

Number of schools: 154

Number of students: 4579

Antero Malin

School Differences and Inequities in Educational Outcomes PISA 2000 Results of Reading Literacy in Finland

This study investigates the level and distribution of reading literacy performance in Finland both at the student and school level from the perspective of equity, using the international PISA 2000 reading literacy data. The main purposes were to find out the amount of the school differences in reading literacy performance, the force of the effects of selected background factors, and the extent to which the between-school as well as the total student variation were attributable to these factors.

There were some clear inequities in the student performances in Finland with respect to background factors that the individual students cannot escape or control. These inequity factors were associated with systematic variation between students and schools in reading literacy performance. The results also indicated that the inequities in performances can be reduced with the effort from students, parents and schools.

For interested readers, this study provides detailed analyses of the effects of selected background factors on reading literacy performance at the student and school level in Finland, based on two-level linear modelling. In addition, methodologically oriented researchers may benefit from reading this book.

Research Reports 15
ISSN 1455-447X
ISBN 978-951-39-3041-7

Cover: Vesa Huhtala



INSTITUTE FOR EDUCATIONAL RESEARCH
UNIVERSITY OF JYVÄSKYLÄ