Roma Education Fund

A Successful School Integration Program

Working paper №2

Gábor Kézdi – Éva Surányi

Gábor Kézdi – Éva Surányi

A Successful School Integration Program

An Evaluation of the Hungarian National Government's School Integration Program

2005-2007

ROMA EDUCATION FUND

ISBN- 978-963-9832-07-7

Cover design and layout: Fo-Szer Graphic Design studio

Contents

Foreword

Integrated education of students of differing family background is a hot policy issue in Central and Eastern Europe in general and in Hungary in particular. The most controversial problem is the integrated education of Roma and non-Roma students or, more generally, disadvantaged and non-disadvantaged students. The main questions are whether children of various backgrounds should be educated together or separately; what methods shall be used; and the effects of integrated versus separated education on the development of children of differing backgrounds. The debate is characterized by heated emotions and little evidence. The need for an open discussion is often expressed. Unfortunately, the need for appropriate evidence is less often heard. Our intention is to show evidence, because we believe that they are should form the basis of policy decisions.

Integrated education is a topic much broader than integrating Roma and non-Roma students. Nevertheless, the focus on ethnicity in Hungarian debates is inevitable. The labor market situation of the Roma minority is disastrous, with severe negative consequences both for the Roma minority and the entire society. One of the most important causes, albeit not the only one, is the lack of skills demanded in a modern economy. And one of the most important institutions where these skills can be developed is the school. Therefore the conclusion that education may be a key element of the future success of the Roma minority. Education, in Hungary, is also mostly a state-run activity and is thus an evident subject of public policy.

Supporters of integrated education believe that integration itself is crucial in the successful education of minorities. This view finds support in human rights arguments and international experience, primarily the example of the African American minority in the United States. Separation is viewed to be discriminatory in itself, regardless its consequences. A more outcome-oriented argument uses the fact that segregated education of disadvantaged minorities almost always results in low-quality education for them, which undermines the common goal of improving the labor market perspectives of the minority. Yet another argument emphasizes inter-group contact: Integrated education creates an opportunity for everyday interactions and can reduce prejudice. This is important because prejudice, often in very subtle forms, may be in part responsible for the lower success of Romanies both in education and the labor market.

Supporters of integrated education see two main tasks, one technical and one political. The technical task is to find appropriate institutional arrangements and educational methods in order to make sure that all students, minority and majority alike, progress better in an integrated environment. The political task is to convince the society of the mutual gains of integrated education.

Skeptics, on the other hand, are not convinced. They reckon that integrated education of Romanies may not bring the benefits projected by its supporters, and it may have negative

effects on some of the students. Skeptics with an eye on empirical evidence can point out that the ethnic composition of the student group is correlated with students' development: student in classes with more Roma and/or disadvantaged students perform worse in Hungary. A seemingly inevitable conclusion is that the higher proportion of Romani and/or disadvantaged students may hold everyone back. They conclude that integrated education may harm all students, especially those belonging to the majority. Some of our results show that such a conclusion is not necessary true.

One version of the skeptical reasoning holds that while integrated education may indeed reduce prejudice it does so at the expense of the skill development of some (the "better") students. There may therefore be a tradeoff between tolerance and academic success, which implies difficult choices for policy. How much shortfall in cognitive development should we allow for a certain increase in tolerance? What is the tradeoff between skill developments of majority versus minority students? Indeed, if these tradeoffs existed, one would not need to be racist to see more harm in integration than benefit.

Appropriate evidence would help a lot in this debate. The evidence could either support or refute the tradeoff between cognitive development and tolerance, or the tradeoff between the developments of minority students versus majority students. However, there is very little evidence in the context of Roma education in general, and in Hungary in particular. At most we have what some call "anecdotal evidence:" pieces of knowledge based on the selective use of few examples. There is a pressing need for real evidence.

This research report intends to provide such real evidence. We examine the impact of integrated education of Roma and/or disadvantaged students in elementary school, by looking at a program in which integrated education is accompanied with teacher training and other support. We examine the effects on the development of Roma and disadvantaged, as well as of non-Roma and non-disadvantaged students. The basis for the analysis is the complex model program of the National Educational Integration Network (OOIH) that focused initially on 45 schools. The most important goal of the program is providing quality education for all students in an integrated environment. The analysis of the program's impact, therefore, provides an excellent opportunity to demonstrate the impact of integrated education when it is accompanied by adequate pedagogical support, in Hungary, in the early years of the 21st century.

Skill development, which we interpret in a broad sense, is at the core of our study. The most recent labor market research demonstrates that, besides cognitive skills and competences (e.g. literacy and numeracy), other dimensions are of comparable importance. Positive self-esteem, the sense of controlling one's own destiny, and the ability to cope with difficult situations are found to be important determinants of labor market success. We therefore examine these skills besides cognitive development. Naturally, we also look at prejudice and social distance between Roma and non-Roma students.

This study presents the most important results of the research. The presentation is kept simple but with the aim of satisfying the appropriate scientific standards. Conclusions are drawn on the basis of appropriate methodology in terms of data collection, processing

and analysis. In order for the readers to be able to check out methods, we included some technical details in this report. Nevertheless, we attempt at making these details as clear as possible. At the end of the day, this study is about numbers. As a result, tables and graphs, not words dominate many pages of the report.

One of our goals is to show that the tools of modern social science can be used to examine the impact of programs financed from public funds in general and within the framework of educational policy in particular. We believe that such studies are necessary for open and rational public discussion. Naturally, the conclusions are subject to debate. Such debates stay within a rational framework if they are centered on methodology and focus on whether the conclusions are supported by the evidence. We would consider it to be a great success if we helped leading the debate on integrated education in this direction.

Throughout the three years of the research, we received valuable help from the program organizers, our colleagues, the teachers, students and parents of the schools in the study, and numerous research assistants and interviewers who took part in the data collection and organization.

Our first thanks go to the program organizers for providing resources and administrative assistance without any interference in the research. Their support was forthcoming regardless of our results, led only by a sincere need to get to know the real impact (and the belief that it would be positive). Among the organizers our special thanks go to Judit Szőke, Mariann Szemenyei, Tamás Varga and Orsolya Szendrey. The research would have been impossible without Gábor Kertesi's initiative and ongoing support. Our colleagues helped by valuable advice. In addition to Gábor Kertesi, we express our thanks to Gábor Havas, Vera Messing and András Semjén for their helpful comments. Benő Csapó and Mrs. Mária Bogár Németh provided significant assistance by making appropriate tests available to us.

Invaluable help was provided by the management, teachers, students and their parents from the participating schools, the control sample and the national standardization sample. The research is anonymous with respect to schools; therefore, we cannot express our gratitude to each of them. Some of the schools provided valuable help in the questionnaire design in the form of open discussions and pilot interviews. Of them we offer our special thanks to Mrs. Józsefné Zágoni-Szabó and to the teachers and students of the school managed by her. The research would have been impossible without the help of our friends, colleagues and assistants who helped in organizing and implementing the data collection. Therefore our final thanks go to Rami Malik in the first place, as well as Zsófia Szeremi, Imola Kincses, Balázs Tóth, Léna Szilvási and Katalin Medvecki as well as to nearly a hundred interviewers and assistants.

Gábor Kézdi
Assistant professor of economics at Central European University (CEU) and research fellow of the Institute of Economics of the Hungarian Academy of Sciences (MTA)

Éva Surányi
Research fellow of the Institute of Economics of the Hungarian Academy of Sciences (MTA) and PhD student of economics at Central European University (CEU)

Introduction

The National Educational Integration Network (Hungarian acronym: OOIH)[1] was established in 2002. Its primary goal has been the promotion of quality education of disadvantaged and Roma[2] children in elementary schools, in an integrated environment. The program examined in this study was launched in autumn 2003 with the participation of 45 schools. The program required education in integrated groups and provided complex educational, managerial and financial support.

The substantive elements of the program are summarized in the Integrated Pedagogical System (Hungarian acronym: IPR[3]). It is based on the idea that school and management is the key to successful integrated education, and the autonomy of the schools is crucial. The IPR defines several elements, some of which are compulsory, while others are optional. Some elements are specific to integration (compulsory integration within schools,[4] emphasizing multicultural content, etc.), but most elements aim at providing quality education for all children. Among other things, the IPR focuses on modern, competence-oriented and student-centered educational methods, effective classroom management, effective organization of schools, and improving the relationship of the school and its environment.

The subject of the research presented here is the OOIH integration program focusing on 45 featured schools. The scope of activities pursued by OOIH is much wider but we do not look at anything else. Our objective has been to evaluate the impact of the program. Students have been at the core of the impact assessment, that is, the issue was the impact on their development.

The regulatory background and financing of the program have undergone many changes. We have focused on schools that participated in the program from the very beginning, and

1 Országos Oktatási Integrációs Hálózat.

2 There is some controversy about the name of the Romani ethnic group. In Central and Eastern Europe the name Roma is used, as a noun (Roma plural) and also as an adjective. It is also used by some international organizations and initiatives, such as the Roma Education Fund or the Decade of Roma Inclusion. The United Nations, the U.S. Library of Congress and other international associations use the Romani name for an adjective and a noun as well (Romanies plural). The name Gypsy is used by many non-Roma but not by the Roma: It is a name created by outsiders and is derived from the misconception of Egyptian origin. Similarly to the alternative local names such as Tsigane, Cigany, Gitane or Gitano, the name Gypsy brings negative associations about lifestyle or project images that are inaccurate for many Roma (e.g. the romantic image of travelers). In this study we use Roma and Romani interchangeably.

3 Integrált Pedagógiai Rendszer.

4 About half of Hungarian elementary schools have more than one group within grade. Group assignment is usually fixed for the students' entire career within the school, and the groups are often different in terms of academic performance and social background.

we call them program schools. Within the program schools, we concentrate on children in those grades that were in the focus of the program itself.

Some studies discuss problems with the implementation of the program. Frequent changes in the organizational framework resulted in significant confusion from the schools' viewpoint, and many elements of the program were implemented in ways that are far from the ideal.[5] Our question is not relative to some ideal but relative to what would have happened without the program. In other words, we evaluate the effect of the program as it was implemented.

The main focus of this study is the students: We would like to know whether the program has a positive impact on the students' development, and if yes, in what way. We are especially interested in the potentially different impact on Roma and non-Roma students as well as disadvantaged and non-disadvantaged students. (Two thirds of the Roma students in the examined schools are disadvantaged, and two thirds of disadvantaged students were Roma; see later for the definitions.) In addition to cognitive achievement (reading comprehension, direction of further education[6]), we also examine broader personality traits (self-esteem, locus of control i.e. the belief of the extent of one's control of her/his own destiny, the ability to cope with difficult situations) as well as inter-ethnic relationships. Where possible, we concentrate on the students' development and compare test scores to earlier test scores.

Chapter 2 of the research report gives a summary of the program and the evaluation method. Our aim is to make clear what we think can learn about the impact of the program from the data. We discuss the choice of the control group, details of the data collection, measurement of ethnicity and disadvantaged status, as well as our handling of the potential distortion of questionnaire answers by social desirability.

A program's impact is defined as the results of the participants compared to what they would have achieved had they not taken part in the program (compared to counterfactual results). Our research was carried out in 60 schools. From the 45 program schools we selected 30 and matched each to a control school. The results of the control pairs are supposed to measure the counterfactual results, i.e. the results that the students of the program schools would have achieved in the absence of the program. Of course, one can never be absolutely sure that the program schools would really have achieved similar results to those of the control schools.

[5] Szilvia Németh and Attila Papp Z. "És mi adjuk az integráció vezérfonalát…" Dokumentumelemzés és a kvalitatív vizsgálat eredményei (*And it is us who give the guiding thread to integration…" Document analysis and the results of a qualitative examination)* in Szilvia Németh [edit.]: Integráció a gyakorlatban. A roma tanulók együttnevelésének iskolai modelljei (*Integration in practice. School models of co-educating Roma children*). National Public Education Institute, Budapest, 2006. http://www.oki.hu/oldal.php?tipus=cikk&hod=integracio_gyakorlatban-02_vezerfonal).

Other chapters of the same volume describe a far-from-ideal situation of the integrated education policy and educational practice in the program schools. These studies are based on document analysis and interviews.

[6] In Hungary, students of grades 9 to 12 can attend three types of schools:, one being vocational training schools without the possibility of further studies. The labor market value of the vocational training degree is low (see, for example, G. Kézdi, "Education and Earnings." in: K. Fazekas and J. Varga (eds.), The Hungarian Labour Market 2005. In Focus: Education and the Labour Market, Institute of Economics HAS, 2005. http://econ.core.hu/doc/mt/2005/en/infocus.pdf).

A randomized experiment could have lead to such conclusions with high certainty. Unfortunately, the design of the OOIH program does not allow for randomized evaluation. However, based on the results of econometrical models, we believe that, by comparing the program schools to the control schools in the data, it is mostly the impact of the program that we capture. Since the student composition in program and control schools is identical, we are convinced that students in program schools achieve better results because of their schools. Additional analysis makes us think that those schools are better at least in part as a result of the program.

The remainder of this report is organized the following way. In Chapter 3 we show that program schools achieved ethnic integration across parallel classes (permanent student groups within the same grade). Some of the control schools achieved similar degree of integration but many maintained significant differences in the ethnic composition of their groups. The level of integration is lower in some of the other dimensions, but in those dimensions, too, program schools are at least as well-integrated as control schools.

Based on hundreds of classroom observations, Chapter 4 looks at teachers, teaching methods, classroom management and interactions. No significant differences are found with respect to the teachers' general competence. At the same time, program schools are more likely to engage in student-centered education. Student autonomy was found higher in the program schools, both in the lower and the upper grades. Group work receives significantly more emphasis in the program schools and, when applied, it is used much more in accordance with the principles of cooperative learning. Probably as a result, cooperative and collaborative relationship among students is more widespread in program schools. Differentiated education in the control schools is barely detectable, and, although more frequent, it is rare in the program schools as well.

The following three chapters examine the differences in the students' achievements. Chapter 5 focuses on school grades, cognitive achievements and the admission into secondary schools. We demonstrate that students of the program schools achieve somewhat better grades and better results on standardized reading comprehension test. The advantage of program schools is generally small, which, in light of the children's skills development, is not surprising. Most importantly, however, we find no group with lower results in program schools. This is a powerful result: Program schools seem to achieve integrated education without hurting non-Roma and non-disadvantaged students' skills development.

Chapter 6 examines the development of some personality traits that are called often non-cognitive skills in the economics and program evaluation literature. Our estimates show that students in program schools, especially Roma students, have a more interior locus of control, i.e. they are more likely to think that they themselves are responsible for their destiny as opposed to outside forces. Students of the program schools have more positive self-esteem, regardless of their family background. The results are similarly more favorable with respect to the ability to cope with difficult situations. It seems that all these results can be attributed to the program.

Perhaps surprisingly, those non-cognitive skills are as important in a successful life as cognitive skills. Their labor market value is comparable, and they are important determinants

of success, or of avoiding failure, in many other areas of life. We also know that, to a much stronger extent than cognitive skills, these personality traits are malleable in the elementary school period. Therefore, it is an important result that the OOIH program has a positive impact on the development of non-cognitive skills, both among Roma and non-Roma, disadvantaged and non-disadvantaged.

In Chapter 7, we look at inter-ethnic relations: prejudice, social distance, as well as value judgments and anxiety that may operate in the background. We show that non-Roma students of program schools hold a smaller social distance from the Roma than non-Roma students in control schools. Roma students are much more accepting towards non-Roma both in program and control schools. Roma and non-Roma students show less social anxiety in program schools and are less likely to believe in social dominance.

In Chapter 8 we try to uncover the mechanisms behind the success of the program schools. Not surprisingly, identifying the mechanisms proves even harder than estimating the impact of the program. On the one hand, the complex nature of the program makes it practically impossible to separate the effects of different elements. On the other hand, the freedom of schools in choosing many elements of the program and the teachers' inevitable freedom in choosing teaching and classroom management methods the possibility that prior student achievements lead teachers to adopt certain teaching methods. This phenomenon is an example of reverse causality, i.e. a relationship that runs the opposite direction than the mechanism we are after.

With all those caveats, our results indicate that cooperative techniques and observed cooperation help Roma and disadvantaged students, and when applied, differentiated education seem to help non-disadvantaged students. The most important result is that, regardless of the students' family background, student-centered education and strong student autonomy seem to make a significant contribution to better results across all dimensions.

In the last chapter before the conclusions we look at the correlation between student achievements and the composition of the class. For several reasons, we expect that in general, a larger proportion of Roma and disadvantaged children is related to lower results in all dimensions, with the exception of inter-ethnic relations (where more contact is likely to decrease prejudice and social distance). Our question is whether the program broke this correlation when harmful and reinforced when helpful.

The results give clear support to the latter: the social distance kept by non-Roma students from the Roma continuously decreases as the proportion of Roma students increases, and the relationship is stronger in program schools. The picture is less clear in the other dimensions. In some cases the negative correlation between the fraction of Roma and student achievements (by both Roma and non-Roma) decreased, while in other cases, it did not. In those cases where the program did not break this correlation, program schools seem to have achieved significant results only in classes where Roma and disadvantaged students were a minority (below 25 per cent).

What do all these results imply? We believe that, despite its organizational problems, the program is successful. Students of program schools performed better in all dimensions. The improvement is not spectacular – one can say it is rather small in many dimensions – but it is present in all dimensions analyzed here. Even more importantly, we did not find any dimension in which the impact would be negative for any group, Roma or non-Roma, disadvantaged or non-disadvantaged.

The results support the arguments for integrated education. They imply that it is possible to promote the skills development of Roma and non-Roma students and reduce the social distance between them at the same time, in the framework of integrated education.

Naturally, all this does not happen on its own. Elimination of majority-Roma classes is probably an important part of the success. But the key is likely to be in improved educational methods and classroom management. In short, they imply that integrated education can be successful if complemented with quality education.

Background and Methodology

2.1 The OOIH Program

The National Educational Integration Network (Hungarian acronym OOIH for Országos Oktatási Integrációs Hálózat) was founded by the Ministry of Education of Hungary in 2002, with the aim to promote integrated and quality education of disadvantaged and Roma children in elementary schools in Hungary. One of the major activities of OOIH was the design and management of a small and intensive program. This study focuses on that program.

Schools applied by responding to an open call in 2003, and, of the applicants, 45 schools were selected for the program. Application was clearly a result of self-selection. Participants were chosen by the OOIH itself based on potential success (not on a one-dimensional score). The program was launched in the fall of 2003, in the first and the fifth grades, with expansion to all grades one through eight in a grandfathered way. The 45 program schools we supposed to serve as examples for neighboring schools in order to make them join the network. For this reason, the Hungarian label of the program schools is "bazisiskola" (basis school). The program was designed by a small group of experts led by Judit Szőke, the first director of OOIH.

The substantive elements of the program are summarized in the Integrated Educational System ("Integrált Pedagógai Rendszer," IPR henceforth).[7] The design is based on the premise that integrated education is an issue of school management and is inseparable from quality education for all children. Therefore, IPR goes beyond promoting educational methods; it also contains school management and classroom management elements. For program schools, some elements of the IPR have been compulsory, while other elements have been optional.

The IPR has become important outside the program itself. Starting 2005, each Hungarian elementary school has been eligible for additional support for each disadvantaged student if it adopted the IPR (again with some elements compulsory and others optional). The activities of the OOIH are thus not limited to the management of the examined program. However, the subject of this study is the impact assessment of the narrowly interpreted OOIH integration program covering the 45 program ("basis") schools. This is what we call "the program" throughout the study.

[7] See, e.g. http://www.sulinovadatbank.hu/index.php?akt_menu=1028, in Hungarian.

The program underwent major organizational changes. Originally financed by the Hungarian government, the program became part of those supported by the EU Structural Funds (in the Hungarian framework HEFOP 2.1) in its third year. According to anecdotal evidence, the transition was far from being smooth. Among other changes, program application was renewed, open to all elementary schools in Hungary. The set of successful applications in this second round overlapped with the original schools in large part but not entirely. Organization and management of the program also changed a few times. Even after those changes, the program proceeded to function unchanged in terms of its fundamental values, objectives and content-related elements.

For the evaluation study we choose a sample of 30 schools from the 35 (of the original 45) that had continuously participated in the program. This study focuses on children who were in their first and fifth grade in 2003-4. Due to administrative reasons, data collection could not start before the spring of 2005, more than one year after the program started. The first measurement took place when children in the first cohorts were in their second and sixth grade, respectively. Altogether, three waves of data collection took place, in the springs of 2005, 2006 and 2007. Data collection focused on the same students, following them to fourth and eighth grades, respectively. Hungarian elementary schools cover the first eight grades. From ninth grade on, students continue their studies in different types of secondary schools. The direction of further studies is an outcome variable we analyze for graduating eight-graders (see later for more details).

2.2 The Control Group

The primary focus of this evaluation is the effect of the OOIH program as it has been implemented. The effect of a program is captured by the results of its participants, compared to what they would have achieved had they not participated in the program. The results of the participants in the absence of the program are called counterfactual results. The main methodological problem of all evaluation studies lies in the fact that, of course, counterfactual results are not observed. All evaluations have to estimate the counterfactual results from somewhere else.

Evaluation studies measure the counterfactual results by observing the results of a control group. The effects of the program are then measured by some comparison of the results of the participants (also known as the treatment group) and the results of the control group. The assumption of such comparisons is that results of the control group are similar to the counterfactual results of the treatment group, at least on average. Obviously, selection of the control group is of key importance. The classic scientific method for selecting control groups is the randomized experiment. In a randomized experiment, treated and control status of the units (here schools) would be determined by a random rule. (Practically, any rule that is known to be unrelated to program results may qualify.) Unfortunately, randomized experiments are rare among social programs. The OOIH program was no exception: selection of treatment schools was not random at all. Instead, selection into the treatment group was based on self-selection into

applications and qualitative judgments by the administrators. As a result, the control group had to be chosen in some other way.

The treatment group of the evaluation study is a sample of the program schools that participated continuously in the program. The treatment sample consists of 30 schools. To each of these program schools we selected one control school that did not participate in the program but whose characteristics matched those of the program school in question. The entire sample, therefore, consists of 60 schools.

Control schools were selected by propensity score matching. The basic idea of propensity score matching is to find a control pair to each treated unit so that the two are similar in as many dimensions as possible that are relevant to the program under investigation. We used a set of pre-program variables, including information on ethnic composition from survey data,[8] official school-level data on size, location, and ethnic minority education,[9] and family background and achievement score results of the spring 2003 national competence measurement covering sixth graders.[10]

All information used for the pairing reflects the situation and results before the program. The average values of the variables used for the propensity score are shown in table 2.1. Besides the averages for the treatment and the control group, table 2.1 shows national average values as well.[11]

[8] Ilona Liskó–Gábor Havas Segregation in the Elementary School Education of Roma Students (Szegregáció a roma tanulók általános iskolai oktatásában). Research Institute of Higher Education, Budapest, 2005. Online: http://www.hier.iif.hu/hu/kutatas_kozben.php.

[9] Yearly data from KIR-STAT (http://www.kir.hu) and the number of students eligible for the Roma minority educational support in year 1999, the only year when it was available.

[10] http://www.sulinova.hu/cikk.php?sess=&alsite=27&rovat=31&alrovat=41&cid=103.

[11] Besides the variables listed in the table, the geographical location of the schools was also taken into account. The control schools had to be located in the same region as their base pair, and had to belong to the same type of settlement. From among the possible control schools that could be found in the same regions and of the same settlement type, we selected the most similar schools by the propensity score matching procedure. Following all this, we verified the selection with Gábor Havas, a sociologist with extensive local knowledge of Hungarian elementary schools with disadvantaged students. We asked his opinion about the closest school, and, when he raised serious objections (e.g. by pointing to inflated or suppressed reports on minority students), we opted for the second or third closest control school (again taking Havas's suggestions into account). It is the noisy and incomplete nature of our most important data sources that made this expert opinion valuable. We express our gratitude to Gábor Havas here for his extremely valuable help.

2.1 Table – Composition of program and control schools on the basis of the data used for selection (1999, 2003)

	Program schools	Control schools	National average
Population in town/village of school	87,363	83,399	225,992
School size (number of students)	320	311	286
Fraction eligible for Roma minority support in 1999 (%)	31	26	5
Fraction of students at risk (%)	16	15	9
Fraction of students (%) with			
Mother's education less than 8 grades	9	7	2
Mother's education exactly 8 grades	35	33	21
Father's education less than 8 grades	3	5	1
Father's education exactly 8 grades	26	25	14
No working parent	29	30	20
Number of books at home 0 to 50	29	26	12
Number of books at home approximately 50	16	15	12
Competence scores at school level, 6th grade, spring 2003			
Mathematics: school average	456	446	500
Reading: school average	449	436	500
Mathematics: school standard deviation	88	88	87
Reading: school standard deviation	92	95	89

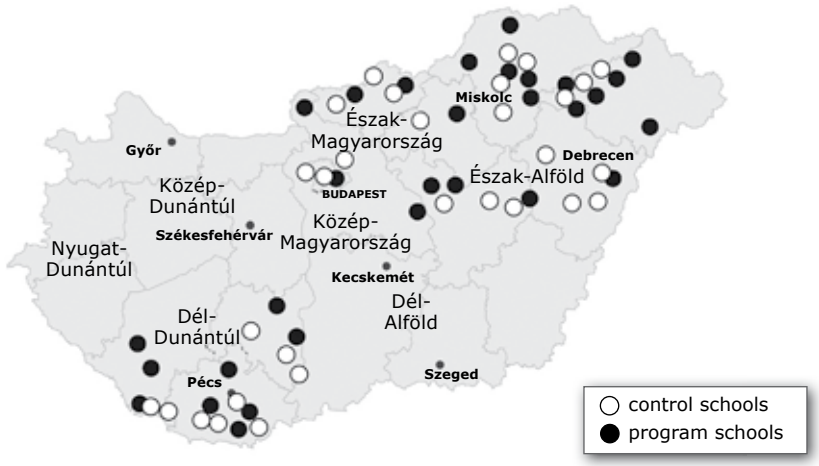
Program and control schools are very similar based on the size of the settlement, the size of the school, and also by the students' family background. The fraction of Roma students, those with low-educated mothers and with few books at home is slightly higher in program schools than in control schools. At the same time, pre-program competence scores in mathematics and reading comprehension are slightly higher in program schools than in control schools, with very similar within-school inequality. The differences are small and statistically not significant, which validates the selection method (see more validation later).

While the matching is of good quality, the data of table 2.1, even if insignificantly in a statistical sense, raise the potential problem that the program schools, on average, may have achieved somewhat better results from their somewhat more disadvantaged students already before the program. If that is true, then, by making a simple comparison of the program schools and control schools, we may attribute to the program's impact results that are in fact achieved by program schools but are not necessarily due to the program. When comparing students' achievements, we shall address this potential problem.

Note that our data are not representative from a national point of view. Compared to the national average, the program and control schools in our sample are located in smaller settlements; they scored lower on average on the national competence tests both in mathematics and reading comprehension; and, perhaps most importantly, their students come are significantly more likely to be Roma and come from disadvantaged families. This is a natural consequence of the program's focus. Where possible, we compare all results to national averages.

The geographic location of the schools is not representative either. Program and control schools are concentrated on the South-West, North-East and Mid-East of Hungary, as demonstrated in figure 2.1. These are the Hungarian regions where the Roma are overrepresented.

Figure 2.1 – Geographical location of the program and control schools



Within each school, we followed two cohorts of students: those in their second grade and those in their sixth grade in spring 2005, the time of the first wave of data collection. Half

of our sample consists of schools with one class per grade (classes are student groups of approximately 20 to 25 students). Students in larger elementary schools in Hungary are organized into rigid and permanent student classes: students assigned to a particular class stay with that class for the entire day, and class affiliation hardly changes through grade eight. In the large schools of our sample, we sampled two classes per grade. In the few schools with more than two classes per grade we selected the best and of the least good classes (on the basis of preliminary consultations with school headmasters). We followed all students within the selected classes.

2.3 Data Collection

The program was launched in the 2003-2004 school year with a focus on the student cohorts then in their first and fifth grades. For administrative reasons, our first data collection took place in the spring of 2005, while the last data collection had to be closed in 2007. We could follow students from second grade and sixth grade to the end of the fourth and eighth grade, respectively. The data collection took place in three waves. The first one in 2005 measured students' skills and some other characteristics. The third data collection, in spring 2007, measured students' results again (school achievements, cognitive and non-cognitive skills, inter-ethnic relations; see later for details). The two data points enable us to analyze year 2007 results by controlling for year 2005 results.

In addition to measuring the students' results and views, we carried out several complementary surveys. We conducted interviews with the school headmasters in 2005 about participation in other programs and, in program schools, whether the school had operated in an integrated way before the program (half did, half did not). In spring 2006, students' family background was measured by questionnaires to be completed at home. Also in the spring of 2006, we carried out standardized classroom observations in each school in each group in our sample, in two subjects (math and literature). Finally, after the 2007 skills measurement, we registered the students' ethnic background (in full compliance with Hungarian data protection regulations, see later), and we collected information on the secondary school admissions of graduating eight-graders.

The analysis does not focus on students with special educational needs (SEN), for two practical reasons and a conceptual one. First, the tests we used were not designed for children with special educational needs; therefore, their use may have been problematic for SEN students. Second, there were very few SEN students in our sample (only 167 of the more than 4000, in comparison to more than one thousand Roma students). The conceptual reason is that the issues of integrated education of SEN students do not overlap entirely with the issues of integrated education of non-SEN students of ethnic minorities and disadvantages. The OOIH program did not put an explicit focus on SEN students.

The details of the three waves of data collection are outlined next.

2.3.1 First wave

The first wave of data collection took place in the spring of 2005. We may call it input measurement of skills and other characteristics of the students. Ideally, input measurement would take place before or at the very beginning the start of the program. As we noted earlier, for administrative reasons, this particular measurement took place one year and a half into the program (the program started in the fall of 2003). As a result, results measured in 2005 may be in part attributable to the program itself. By looking at changes between 2007 and 2005, we restrict the focus to a two-year development. While many things may change in two years, we cannot expect a significant impact in such a short time period.

In the first wave of data collection, each student (then second and sixth graders) had to fill in two questionnaires. One measured cognitive skills. Sixth-graders completed an inductive thinking test, second-graders a mathematical thinking test, both designed by Benő Csapó and his team from Szeged, Hungary.[12] The other questionnaire contained some basic demographic information, self-esteem, coping with difficult situations and some sociometrics. Besides these self-administered tests, schoolteachers (the "master teacher" of each class) filled out a short datasheet for each student.

The composition of the sample in the program and control schools is reported in table 2.2. The data presented here refer to measurements after the sampling procedure and thus provide an ex-post validation of the control group. The table shows variables that the program was unlikely to change. National comparisons are not available for the data presented here.

[12] http://www.edu.u-szeged.hu/mtakcs/.

Table 2.2 – Composition of program and control schools based on the data of the first data collection of the evaluation study

	2ND GRADE		*6TH GRADE*	
	Program	**Control**	**Program**	**Control**
INFORMATION ON STUDENTS				
Number of surveyed students	1032	983	1099	1081
Fraction of girls (%)	47	46	47	50
Fraction of overage students (%)	6	8	8	11
Fraction of disadvantaged students (assessed by teacher, %)	51	46	44	42
Fraction of students at risk (assessed by teacher, %)	10	13	11	14
Fraction with behavioral problems (assessed by teacher, %)	18	15	15	16
Fraction with studying problems (assessed by teacher, %)	21	16	14	17
Fraction of SEN students (%)	7	7	5	4
Fraction of students with diagnosed deficits (e.g. dyslexia, %)	16	6	5	7
CLASS-LEVEL INFORMATION				
Number of classes	49	48	50	48
Class size	20.5	20.2	21.7	21.4
Fraction of Roma students (unweighted, %)	43	46	38	39
Fraction of Roma students (weighted by class size, %)	40	44	35	37
INFORMATION ON SCHOOLTEACHERS	**Program**	**Control**	**Program**	**Control**
Fraction female (%)	98	94	85	84
Average age	44	42	44	42
Average tenure (years since in present school)	21	19	20	17
Education: teachers' college for lower grades ("tanitokepzo," %)	77	79	4	4
Education: teachers' college for upper grades ("tanarkepzo," %)	17	12	71	87
Education: teachers' college MA degree ("egyetem," %)	2	0	23	4
INFORMATION ON SCHOOLS				
Number of students in school	344	348		
Number of teachers in school	34	34		
Fraction of Roma students in school (%)	40	41		
Fraction of Roma students in the neighborhood (%)	19	22		
Fraction of students in school from outside district (%)	12	12		
Fraction of students in district going to other school (%)	12	10		

According to the evidence provided in the table, student composition in program and control schools is practically identical, both in terms of individual and class-level measures. This provides important ex-post validation of the sampling procedure. Similarity in student composition means that differences in the results of program versus control schools are very unlikely to be caused by differences in family background.

Teacher composition is also similar, but small differences emerge. Program school teachers are somewhat older and more experienced. With respect to the educational background, no significant differences are found among the teachers of lower grades, but upper grade teachers in the program schools are more educated. Almost one fourth of program school teachers in upper grades had a five-year university degree (sort of a master's degree), while practically no teacher had such an education in the control schools. The rest of the school characteristics are very similar. These results imply that there may have been some quality differences between program and control schools even before the program. Such quality differences would imply that differences in the results of program versus control schools may have been present without the program. Self-selection of schools into the program also points to such an interpretation. When evaluating the differences in students' results, therefore, we address the problem in more detail.

2.3.2 Second wave

The second stage of the data collection took place in the spring of 2006. Two measurements were taken: one of family background, and one of teaching practices and classroom dynamics.

Family background information was collected by questionnaires filled out by the student and her/his family at home. The questionnaire was based on the family background questionnaire national competence measurement in 2006, with some extra questions added. Table 2.3 shows summary statistics for some variables for both the program and the control schools, along with the corresponding national averages. Note that the national competence measurement of 2006 took place in spring 2006 as well, around the time of our measures. While our measures covered grade 3 and 7 students, the national competence measurement covered grade 4 and 8 students. Comparison of program and control schools serves as our second ex-post validation of the matched sample. Comparisons to national averages reveal the differences of students in the focus of the program to the average Hungarian student.

Table 2.3 – Students' family background in the second wave; data from the national competence measurement of 2006 in comparison

	Grade 3			Grade 7		
	Program	Control	National	Program	Control	National
Number of students with information	810	909	91,349	757	897	92,588
Fraction girls (%)	47	48	50	49	50	49
Welfare eligible (%)	37	37	19	31	35	20
Eligible for free meal in school (%)	59	60	28	10	8	4
Eligible for free school books (%)	73	83	57	66	78	52
Average family size	4.9	5.0	4.5	4.7	4.8	4.4
Fraction with father in family (%)	78	76	83	76	73	79
Mother uneducated (8 grades or less, %)	43	43	21	40	40	19
Father uneducated (8 grades or less, %)	35	35	16	32	29	13
Fraction with working mother (%)	41	42	66	52	51	73
Fraction with working father (%)	64	64	84	69	65	80
Avg. size of apartment (sqmeter/capita)	19.0	19.0	n.a.	20.8	20.5	n.a.
Average monthly spending (HUF '000)	109	110	n.a.	120	115	n.a.
Fraction who think poor (%)	38	38	17	29	29	14
Fraction who were not on holiday (%)	41	41	21	28	31	18
Fraction without a car (%)	46	48	27	44	43	27
0 to 50 books at home (%)	35	35	16	29	27	12

Note. National results refer to grade4 and grade 8.

Response rate to the family background questionnaire is 85 per cent in grade 3 and 75 per cent in grade 7. Response rates are somewhat lower in program schools than in control schools. Sample size in the program and the control schools is less than a thousand students per grade each, while the national competence measurement data refer to over 90 thousand students per grade. Comparing students of program schools to students of control schools reinforces our previous findings: the composition of students by family background is very

similar. Also in accordance with our previous findings, we see that families of students in program and control schools are significantly poorer than the national average, and their parents are significantly less educated and are significantly less likely to be employed.

Teaching practices and classroom dynamics was measured with the help of structured classroom observations. The questionnaire focused on objective measures (physical environment, structure of class, types of activities etc.) as well as more subjective assessments (student cooperation, autonomy, respect, etc.). The observations were taken in two classes in each student group, one in mathematics and one in Hungarian literature. Each observation was carried out by two observers. Typical classroom observers were upper year psychology majors or students of teacher training colleges. They participated in intensive two-day training sessions including video training and trial observations.[13]

We considered classroom observations to be collaborative processes in which the observed teachers and the observers both had their role before, during and after the observation. Therefore, the observations were preceded and also followed by discussions with the teachers. The primary goal of the first discussion before was clarifying the roles and reducing tensions. The observers informed the teachers about the goals and aspects of the monitoring, answered the teachers' questions, and received some information about the "history" of the class. The goal of the second interviews was to provide feedback. The observers, where possible, took a back or side seat at a distance from the students and the teacher as well.

Observed classes are not representative of average classes. The presence of the observers results in inevitable changes. Teachers were notified about the observation well in advance.

They had time for preparing for the observed classes more thoroughly than usual, and they were likely to have done so. Aside from being better prepared, teachers are likely to have tried to make their and the students' best. Even with these biases, though, classroom observations are suitable for our purposes. Our main question is not about the level of the observed phenomena but the differences between program and control schools. While all measures may be biased, the comparisons of major interest may not be biased much if the behavioral distortions are similar in program and control schools. Moreover, observed classes were better (or simply different) in ways that were important for the teachers. Therefore, classroom observations can provide a good picture of the teacher's educational principles and the methods that she considered to be good.

More importantly, many things are difficult to change (e.g. contact with students, speed of speech, spontaneous reactions), in which distortions are probably small. The students' behavior is also less likely to be affected by classroom observations, partly due to the lower awareness (because of the positioning of the observers), and partly due to the lower incentives to change their behavior. Students' behavior, therefore, can be closer to

[13] Léna Szilvási and Katalin Medvecki video trainers provided excellent assistance both in the design of the classroom observations and the training sessions. We express our special thanks for them.

the representative classroom behavior, perhaps with the exception of extreme forms of behavior. When assessing the results of classroom observations, we take all these aspects into consideration.

From among the 197 classes participating in the research, classroom observation took place in 189, from 201 student groups (a few schools divided classes into smaller groups). Altogether 366 classes were evaluated, most but not all by two observers, resulting in 684 questionnaires filled out properly. Results of the classroom observations are described in Chapter 4.

2.3.3 Third wave

The third wave of data collection took place in the spring of 2007. This was the outcome measurement of the evaluation study. We assessed the cognitive and non-cognitive skills of children who participated in the research. Measurement was carried out with the help of two questionnaires: a cognitive and a non-cognitive test. The tests will be discussed in detail in the following chapters together with the results. Cognitive skills were again measured using tests developed by Benő Csapó and his colleagues (see above), this time with a focus on reading comprehension. The second questionnaire contained the rest of the outcome variables: self esteem, locus of control, coping, ethnic stereotypes, social distance, etc.

Besides the two questionnaires, we collected information on the children's ethnic background (using the parental assessments, see later), and, on the eight-graders, we collected data about thir admissions to secondary schools.

2.3.4 Connecting individual data

During the three years of research, we carried out several separate surveys on the same students. In order to comply with requirements of data anonymity, students' names were not mentioned on any of the questionnaires. Instead, the students were identified by their number in the class roster, which ensured a smooth matching of questionnaires within the same year. Matching questionnaires across years was made possible by matching the class rosters across grades.

2.4 Validation and Standardization of the Tests

Students' results were measured by self-administered questionnaires (under appropriate supervision). Whether those tests measure what they are supposed to measure is an important and non-trivial question. Another question is the interpretation of the results: what to make of the magnitude of the estimated differences. The first question is one of validity, while the second question is one of standardization.

As we mentioned before, our cognitive tests (inductive thinking in sixth grade, mathematical thinking in second grade, and reading comprehension test in fourth and eighth grades) were developed, validated and standardized by the research group led by Benő Csapó.[14] The other tests on non-cognitive skills and inter-ethnic relations were developed and standardized by ourselves. All were adaptations of standard international tests. Measurement of so many dimensions was possible by using short versions of the tests. Reduction was guided by theory and by maximizing internal validity. We discuss each test when we present their results, in the subsequent chapters.

We standardized all non-cognitive tests on national representative samples of elementary school students for both the fourth and the eighth grades. Besides means, standard deviations were standardized in order to achieve interpretable units of measurement. One unit of difference between two students (or two groups of students) will mean one national standard deviation, i. e. the expected difference between two randomly selected students.

2.5 Disadvantaged and Roma Students

Disadvantaged students were the program's explicit target group. In theory, disadvantage covers family circumstances which in themselves seriously impede children's development. Chronic poverty, social exclusion, unstable family structure as well as deviant parental behavior may all contribute to disadvantage. Disadvantage is therefore a complex notion.

Social sciences and social policy attempt to define disadvantage with proxy variables that are relatively easy to measure. The proxies in use often deviate from each other. At the time of the evaluation study, Hungarian educational policy defined disadvantage by welfare eligibility as assessed by the municipality, complemented with some family characteristics. Severe disadvantage was defined as welfare eligibility coupled with uneducated parents (eight grades or less). The definitions changed many times in the past. Our data allows for simultaneous measurement of various definitions. It also contains the schoolteachers' direct assessment of disadvantaged status. Perhaps not surprisingly, the groups of students identified by the various measures are quite different.

In this study we define disadvantage by the parents' education. Parental education is a strong proxy of permanent income in modern economies, with very low education predicting long-term poverty and poor home environment. At the same time, it is a relatively simple and clear measure. This is the reason why parents' (primarily mothers') educational level is often used in international comparisons.[15]

[14] http://www.edu.u-szeged.hu/mtakcs/.

[15] See, e.g. Micklewright, J.–Schnepf, S. V. (2004): Educational achievement in English-speaking countries. Do different surveys tell the same stories? IZA *Discussion Paper*, No. 1186. http://ftp.iza.org/dp1186.pdf.

We define disadvantaged children to be those both of whose parents have 8 grades of education or less. There is a bit of a controversy in Hungary as to whether three-year vocational training schools (of grades 9 to 11) qualify for secondary education.[16] Although there are powerful arguments against vocational education to be considered a "real" secondary a degree (those with vocational training degree are in many respects more similar 8-grade-educated),[17] we restricted disadvantage status to parents without even such a degree.

According to this definition, 35 per cent of the students in our sample are disadvantaged. The schools considered two thirds of these to be disadvantaged when asked directly, while half of the students they labeled as disadvantaged had at least one parent more educated than eight grades. If vocational training degrees were included into the disadvantaged definition, the extent of the mismatch remained similar with opposite magnitudes. Even the more complex definitions, including those closer to regulations, did not result in a larger overlap with schools' direct assessment of disadvantage.

Roma students are not explicitly targeted by the program. It may seem therefore as if the program took a position with respect to ethnic targeting and opted for a color-blind policy. At the same time, many of the program's elements explicitly target ethnicity (minority culture) or ethnic conflicts and prejudices. This provides a reason for measuring the impact of the program also separately for Roma and non-Roma students.

Two other reasons led us to look at the program's impact by ethnicity. One reason was that in Hungary integrated education is usually understood as integration by ethnicity. Therefore, independent of the program's objectives, it is of great political interest to assess the development and beliefs of Roma and non-Roma students separately. The second reason is based on a more general principle. The dismal labor market situation of the Roma minority is one of the most serious problems in contemporary Hungary. There are therefore powerful arguments in favor of measuring the impact of practically all policy interventions on the Roma minority, whether the policy itself is color-blind or not.[18]

However important, measuring ethnicity is somewhat difficult in Hungary. Similarly to most countries with a sizeable Roma minority, Hungarian regulation prohibits the collection of information on ethnicity without the knowledge and documented consent of the individuals involved. As a result, ethnicity is not part of any administrative record. In principle, aggregated data can be collected along ethnic lines, but data providers (including governmental statistical agencies) opt for the safer choice and collect no data by ethnicity at all.

[16] The Hungarian Census of 2001 does not consider vocational training schools as secondary schools, see http://www.nepszamlalas.hu/hun/egyeb/eurostat/tables/tabhun/load1_10.html.

[17] See, e.g. Gábor Kertesi and Júlia Varga, "Employment and Educational Attainment in Hungary", and Gábor Kezdi, "Education and Earnings." Both in: K. Fazekas and J. Varga (eds.), The Hungarian Labour Market 2005. In Focus: Education and the Labour Market, Institute of Economics, Budapest, 2005. http://econ.core.hu/doc/mt/2005/en/infocus.pdf.

[18] Glenn Loury puts forward a similar argument in his book (Glenn C. Loury: The Anatomy of Racial Inequalities, Harvard U.P. 2002.).

Our research took the measurement of ethnicity seriously. Fully complying with the regulations (Act LXIII of 1992 on personal data protection plus the recommendations of the data protection ombudsperson), we were able to identify the ethnicity of schoolchildren in a satisfactory way. We sent out a letter to the parents of all students in the sample. The letter asked them to provide a statement declaring the national and ethnic belonging of the student, and we explicitly asked for their consent for the information to be used in the research in an anonymous way. Importantly, we allowed for multiple ethnicity. The parents had an opportunity to provide the following answers:

My child is

- Hungarian, non-Roma
- Hungarian, Roma
- Hungarian, partially Roma
- not Hungarian, Roma
- Romanian
- Slovak
- Other

Based on the statements that were sent back to us, we considered all students to be Roma if the parents selected from the answers b, c or d. The first row of table 2.4 contains the proportion of Roma students defined this way from among those who submitted valid statements. The next row in the table contains teachers' estimates as to the proportion of Roma students. This latter information was taken from the interviews made before the classroom observations (the estimate is the average two estimates by class if the assessments of the two teachers were different, and the numbers shown in the table are weighted estimates of the class-level figures). The last row of table 2.4 shows the proportion of students whose parents did not submit statements.

Table 2.4 – Fraction of Roma students. Individual statements by the parents versus the class-level estimates by the teachers

	Grade 4		Grade 8	
	Program	**Control**	**Program**	**Control**
Fraction of Roma students (%)				
From individual statements by the parents	35	35	27	29
From class-level estimates of teachers (weighted by class size)	34	36	27	33
Missing individual statements from parents (%)	25	28	27	32

Twenty-seven per cent of the statements were not returned. But those who did return the statements gave the same ethnic identity as the schoolteachers. The proportion of Roma students measured from the parents' statements and the teachers' estimates are virtually identical except for a small discrepancy in control schools in grade 8. This means two things. On the one hand, the proportion of Roma and non-Roma parents among those who did not respond was the same. On the other hand, and this is the more important conclusion, there is no significant difference between ethnicity as stated by the parents and by the teachers. The figures calculated in the two different ways are very close to each other not only in total, but also by class. All this means that parents' statements measure the same Roma identity that would have been measured on the basis of the teachers' estimates. Recall that the parents' statements are in full compliance to the regulations.[19]

Not surprisingly, the joint distribution disadvantage (defined by low-educated parents) and ethnicity (defined by the parents' statements) show a close relationship. Two-thirds of the Roma students in the examined schools are disadvantaged, while somewhat less than two-thirds of disadvantaged students are Roma in the sample.

2.6 Identifying the Impact of the Program

As defined above, a program's impacts are the results of the program school students compared to what their results they would have been if they had not participated in the program. In the matched sample of this study, a control school is assigned to each program school in order to measure those counterfactual results. When measuring the program's impact, the starting point is the average difference of the results of program students compared the results of the control students. These simple comparisons are unbiased estimates of the impacts of the program if the matching is perfect in the sense that the results of control students, on average, equal to the results program students would achieved without the program.

Because of the nonexperimental design of this evaluation study, nothing guarantees that the control group is appropriate. Tables 2.1 to 2.3 demonstrate that the student composition is virtually the same in program and control schools but program schools may have better

[19] Census data in Hungary and neighboring countries systematically underestimate the proportion of the Roma population. According to the 2001 census, the number of the Roma population in Hungary is 190,046 persons (1.86 per cent of the total population). Based on the results of the most recent sociological research, the population of individuals identified as Roma by their environment is estimated at 600 thousand, 6 per cent of the total population (see, e.g., István Kemény –Béla Janky–Gabriella Lengyel: "The Gypsy Population of Hungary" (A magyarországi cigányság), 1971–2003. Gondolat Publisher, Budapest, 2004). The contradiction is usually explained by the assumption that a large number of Roma "deny" or "hide" their Roma ethnicity in order to avoid stigma. Our results, on the other hand, indicate that the problem of the census is due in large part to double identity. See also Béla Janky–István Kemény: On the Gypsy Ethnicity Related Data (A cigány nemzetiségi adatokról). Minority research, 2003/2, 309–315. Online: http://www.hhrf.org/kisebbsegkutatas/kk_2003_02/cikk.php?id=747.

prepared teachers and may have been doing slightly better even before the program. Although these latter differences are small, they point to the possibility of other differences as well in terms of unobserved school characteristics. If, due to these unobservable differences, the program schools would have produced better results than the control schools in the absence of the program, the simple comparisons of the program and control results can be misleading. This problem is shared by all nonexperimental impact assessments. While nothing fully makes up for the lack of randomization, econometric methods can provide useful evidence and more convincing conclusions on the program's real impacts.

Besides simple program-control comparisons, we estimated the program's impacts in two alternative ways. The first such measurement is based on multiple regressions, controlling for students' results from the first wave (2005) and a number of other covariates. Among those covariates we entered all the observable school characteristics that showed a difference. It is important to see that by controlling for 2005 results, we exclude all the students who repeated a grade or moved from or to the school.[20]

The second alternative measurement restricted the sample to those program schools (and their control pairs) that implemented integrated education only after the launch of the program and not before. In an interview in 2005, the headmasters of the program schools reported whether their school had been teaching in an integrated way (based on an approach close to that of IPR) even before the program. Half of the program schools indicated that they did, while the other half told that they started integrated education with the program. If the program-control differences (controlled for 2005 results and other covariates) indeed measure the program's impact, these differences must be the same in the restricted sample as well (provided the program's impact was not extremely heterogeneous). Therefore, repeating the measurement in the restricted sample provides an additional robustness check.

2.7 Controlling for Social Desirability

Measuring the attitudes towards oneself and others by self-administered questionnaires may be problematic if students project attitudes that are positively biased in order to make a good impression. Again, the question here is not the extent of the bias but whether it is different in program versus control schools. The tendency of people to present themselves in a more positive way in their statements than in reality is often considered the result of social desirability. Survey researchers have addressed such biases since the 1950s. The problem is more serious in surveys with children as the aspiration to provide answers that meet social expectations is part of normal development, although this tendency decreases

[20] The regressions could control for students' family background. However, the nearly identical student composition of program and control schools makes those variables unnecessary to include. In order to check robustness of the results, we estimated such richer models as well with virtually no changes in the results.

with age. Therefore, most of the tests that measure children's personality contain questions that aim at mitigating or eliminating the bias due to social desirability.[21]

For our purposes, we have adopted the Children's Social Desirability Scale (CSDS).[22] CSDS describes children's everyday actions (e.g. "Sometimes I don't like to share things with my friends", "I never shout when I am angry"). It has been used in a number of studies that targeted elementary and secondary schoolchildren. The original test contains 48 true-false statements of which in 26 questions "true" means an answer that matches social desirability while the rest of the answers are of reverse indication. We used a reduced, five question version of the CSDS test.[23] Among the questions there are statements such as: "I never say things that may hurt someone's feelings." and "I always behave respectfully with older people."

The list of questions on making a good impression were included and mixed with the non-cognitive tests in 2007. The average of the base and control school students and the differences between them are shown in table 2.5.

Table 2.5 – The result of the test on making a good impression (nationally standardized results)

	Program Schools	Control Schools	Difference
Grade 8	0.37	0.16	+0.21**
Grade 4	0.83	0.64	+0.19**
Grade 8, Roma	0.53	0.46	+0.07
Grade 8, Non-Roma	0.29	0.03	+0.26**
Grade 8, Disadvantaged	0.48	0.38	+0.10
Grade 8, Non-Disadvantaged	0.34	0.10	+0.24**

\+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

[21] See, e.g. Reynolds, C. R.–Paget, K. D.: National normative and reliability data for the revised children's manifest anxiety scale. School Psych. Rev. 1983, 12, 324-36.

[22] Crandall, V. C.–Crandall, V. J.–Katkovsky, W.: A children's social desirability questionnaire. J. Consult. Psychol., 1965/29, 27-36.

[23] The short version was developed by recording of the original 48 question test jointly with the Rosenberg test on positive self-esteem (see later) and the questions of the RCMAS (Revised Children's Manifest Anxiety Scale) lie detector test (the latter is a 37 item self-filled questionnaire that measures the children's manifest anxiety and is used primarily for clinical and diagnostic purposes). Analyzing the joint factor structure of the three tests, we selected those five questions from the CSDS test which explain a common factor with the items of the RCMAS lie detector test, but do not correlate with the items of the positive self-image test and provide the highest inter-item correlation. The selected items have been named as a „making a good impression" test.

Students of the program schools attempted to make a good impression to a larger extent than the students of the control schools. The difference is significant – one fifth of a standard deviation both in grade 4 and 8. The difference is strong among non-Roma and non-disadvantaged students and negligible among Roma and disadvantaged students. Table 2.6 demonstrates that the differences in attempts to make good impression are present among schools that started integrated education only with the program itself.

Table 2.6 – The result of the test on making a good impression in the narrowed sample (base-control pairs in which the program schools had taught with a non-integrated approach before joining the program), nationally standardized results

	Program Schools	Control Schools	Difference
Grade 8	0.30	0.02	+0.28**
Grade 4	0.78	0.40	+0.38**
Grade 8, Roma	0.53	0.38	+0.15
Grade 8, Non-Roma	0.20	-0.07	+0.27**
Grade 8, Disadvantaged	0.42	0.14	+0.28+
Grade 8, Non-Disadvantaged	0.26	0.00	+0.26**

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

The reasons for the significant differences are not obvious. It is possible that some program school teachers tried to encourage their students to provide more positive answers before recording the test. But it is also possible that the desire to make good impression is indeed stronger among program school students; perhaps it is stronger outside the survey situation as well. Indeed, it would not be surprising that the impact of the program on students' attitudes were realized through the students' social desirability. Greater demand for making a good impression may easily lead to a stronger conformity to social norms, which may in the long run result in more tolerant attitudes (provided the social norm is that of tolerance). In this sense, the desire to making a good impression may not be a biasing factor but part of the intended impact of the program.

But the possibility of survey bias cannot be excluded. That is the reason for our inclusion of the social desirability results among the control variables in the regressions. As we shall see, while measured social desirability does differ between program and control schools, controlling for social desirability does not change the estimated impacts in most dimensions.

Chapter 3

Integration of Roma and Disadvantaged Students

The OOIH program supports the education of Roma and disadvantaged students in an integrated environment. The starting point is integration. Therefore, we begin the examination of the program's impact by exploring the extent to which integration was realized in the program schools.

Integrated education of minority students means that they are educated together with the majority, within one school, within one class, not separated within the classroom, and they are treated equally with the majority students. The larger the differences in the ethnic composition of classes, the smaller the level of integration. Such differences may be caused by differences between schools and differences within schools. The program analyzed here focuses on individual schools. As a result, it cannot have immediate consequences on differences between schools.

Between-school differences in the ethnic composition are important in Hungary. In part, it reflects the residential distribution of minorities. In regions, towns, or neighborhoods with smaller (larger) fraction of minority residents, the proportion of minority students in schools will also be lower (higher). But between-school differences are larger in Hungary than residential patterns would imply. In the Hungarian system of free school choice, families may send their children not only to the district elementary school but to any other school as well (as long as that other takes the student). Some elementary schools (those characterized by extra applicants) can afford to select among the applicants. Such schools are, on average, less likely admit disadvantaged or Roma students. This practice was widespread in the 1990s, and, although in theory prohibited, continues to exist.

3.1 Integration Within Schools

The OOIH program takes the student composition of the school as given, and it focuses on integration within schools. To be eligible for program participation, a school had to have a moderate proportion of disadvantaged and Roma students, somewhere between 20 and 40 per cent. An important element of the program was that, in those schools with more than one group (class) per grade, there should be no differences in the composition of classes within a grade. On the one hand, the requirement aimed at the school assignment mechanism: heterogeneity had to be aimed for. On the other hand, the result itself was also a criterion: there could be no significant differences in the proportion of disadvantaged and Roma children. Table 3.1 demonstrates the results of the interviews conducted with headmasters, while table 3.2 examines the compliance with the requirements on the basis of our own data recordings.

Recall that in schools with more than two classes per grade, we focused on two classes. The two classes were the ones characterized by the highest and the lowest achievement levels (as classified by the school headmaster). Where there was only one class in the grade, between-class differences are of course not defined. In this section we exclude those from the analysis. Of the 30 schools, 18 are included in this analysis.

In the interviews made with the headmasters, we asked about the criteria they use when assigning would-be first graders to classes (recall that class assignment is permanent for the rest of the elementary school). The first line of table 3.1 shows the proportion of those schools where the headmasters made a separate mention that they had aimed at setting up heterogeneous classes or classes of similar composition. The second line shows the frequency of mentioning specialized classes (intensive foreign language education, music education, etc.). Such classes generate within-school differences in themselves: usually, schools select would-be first graders into such classes by their expected success. They are often thought as signals for middle-class parents so they can cluster in a "more decent" class.

Table 3.1 – Criteria used when assigning would-be first graders to classes

	Program	Control	Difference
Heterogeneous classes are a priority (%)	80	21	+59**
School has specialized classes (%)	14	41	-27

+ Statistically significant at the 10% level.
*** Statistically significant at the 5% level.**
**** Statistically significant at the 1% level.**

In 80 per cent of the program schools the headmasters mentioned the importance of heterogeneous classes and that they aimed at setting up classes that were similar in their composition. In control schools, this proportion barely exceeds 20 per cent. The reason for the large difference may be higher levels of integration as well as better knowledge of the "right answer" to such questions regardless of actual practice. Specialized classes are more widespread in control schools.

Table 3.2 shows between-class differences in the fraction of Roma students, disadvantaged students (measured in various ways), average years spent in kindergarten[24] (free of charge in Hungary, compulsory for one year, median is three years), and average cognitive test scores. The cognitive test scores were taken in the first wave of data collection, at the end of grade 2 or grade 6. The differences we look at in Table 3.2 are not normalized. To fix ideas, consider the fraction of Roma students. Between-class differences can be interpreted the following

[24] In Hungary, kindergarten is free of charge and compulsory for one year, and the median length of participation is three years.

way. If the fraction of Roma students is the same, the difference is zero per cent. If the fraction is 20 per cent in one class and 40 per cent in the other one, the difference is 20 per cent. Note that if all Roma students are concentrated in one class, the difference is likely to be large but it's not necessary 100 per cent. For example, if one class is 50 per cent Roma, the other class is zero per cent Roma, then the difference is 50 per cent.

Table 3.2 – Average differences between parallel classes (in absolute value) in program and control schools

	Program	Control	Difference
Between-class differences in the			
Fraction of Roma students (%)	13	28	-15**
Fraction of disadvantaged students (%)	18	24	-6
Faction of students who received child protection benefit (%)	17	26	-9+
Faction of students entitled to free textbooks (%)	20	22	-2
Years spent in kindergarten	0.25	0.59	-0.34**
Standardized cognitive test results for the year 2005	0.57	0.62	-0.04

+ Statistically significant at the 10% level.

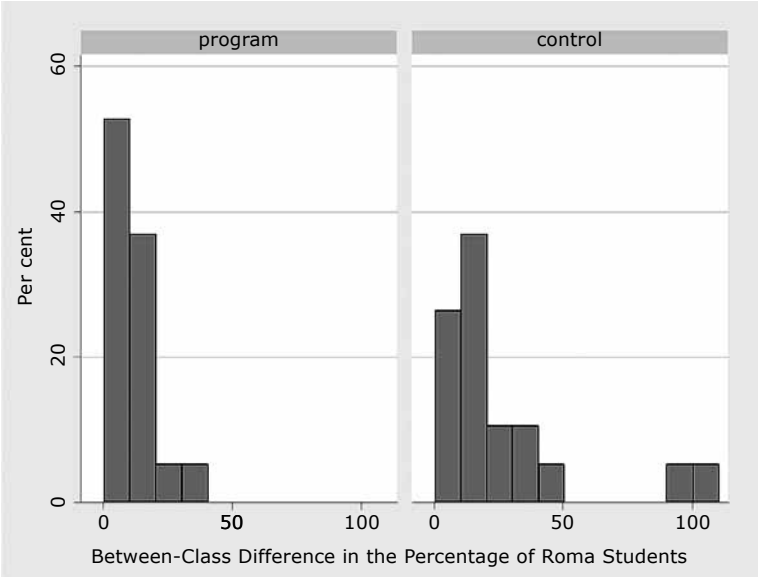
*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Table 3.2 shows that the between-class difference in the fraction of Roma students is 13 per cent on average, while it is 28 per cent in the control schools, more than twice as large. In a typical program school, one class is 25 per cent Roma, while the other class is 35 per cent. In a typical control school, one class is 15 per cent Roma, while the other class is 45 per cent.

The averages hide significant variation across schools. The distribution of the differences between classes is shown in Figure 3.1.

Figure 3.1 – Difference in the proportion of Roma students (in absolute value) in parallel classes



The figure shows that in more than half of the examined program schools, the fraction of Roma students is virtually identical across classes (the difference is below 10 per cent). There is no program school with a between-class difference of more than 40 per cent. On the other hand, in less than 25 per cent of the control schools do we see identical ethnic composition. In some control schools entirely Roma classes are observed along with entirely non-Roma classes. It should be noted that those few base schools where the differences were above 30 per cent fail to meet the program's expectations.

The remaining rows of table 3.2 show that between-class differences are more similar in program schools and control schools with respect to the other measures. Program schools seem more integrated along those lines as well, but program versus control differences are not statistically significant. The exception is years spent in kindergarten, which is likely to reflect ethnic composition (Roma children are known to spend fewer years in kindergarten).

The small program versus control differences in those other dimensions are the result of somewhat larger between-class differences in program schools than in the case of the ethnic composition, and somewhat smaller differences in control schools. Program schools are relatively less integrated in terms of various measures of student disadvantage than in terms of ethnicity. They are also less integrated in terms of cognitive test results of their students. The latter is especially striking: the average between-class difference is 0.6 standard deviations in both program and control schools. Given that one standard deviation means the average difference at the individual level, average differences at the individual level are surely less than one. The average number of students in the classes

was 20. Random assignment of students to classes thus would have resulted in average between-class differences less than 0.22 (one over square root of twenty). Measured between-class differences are almost three times larger.

Large cognitive between-class differences in program (as well as control) schools can mean two things. On the one hand, student assignment to classes may be very far from random even in control schools in terms of cognitive ability (even though ethnic composition was equalized). On the other hand, it is possible that different classes received different educational quality. Significant differences in the quality of education would result in differences in test scores measured at the end of second or sixth grade. Whether it is the selection or significant variation in educational quality, it seems integration in that program school is realized with deficiencies.

Finally, let us examine between-class differences in the narrow sample of program school that were not integrating before the program and their control pairs. The most important results are shown in table 3.3.

The results shown in table 3.3 are very similar to those in table 3.2, suggesting that the program versus control differences shown there are likely to be caused by the program itself.

Table 3.3 – Average differences between parallel classes (in absolute value) in base and control schools. Narrowed sample: those base school-control school pairs where the base schools had not educated in the spirit of integration before joining the program

	Program	Control	Difference
Heterogeneous classes are a priority (%)	82	18	+64**
Between-class differences in the			
Fraction of Roma students (%)	12	26	-14*
Faction of disadvantaged students (%)	14	23	-9
Faction of students who received child protection benefit (%)	17	28	-11+
Faction of students entitled to free textbooks (%)	20	23	-3
Years spent in kindergarten	0.22	0.53	-0.31*
Standardized cognitive test results for the year 2005	0.62	0.60	+0.02

+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

The main results of this section can be summarized the following way. Program schools achieved a relatively high level of integration by ethnicity, as shown by small between-class differences within schools. Control schools are, on average, significantly less integrated. It seems that the higher level of integration of program schools is the effect of the program itself. Integration in terms of other measures of family background (disadvantage) is somewhat weaker, and there program versus control differences are also smaller. With respect to the students' cognitive test results, significant between-class differences were identified both in program and control schools.

Teaching Methods, Classroom Management and Student Behavior

The most important elements of the program were teachers' training in modern, student-centered teaching methods and classroom management that deal with students' heterogeneity and reinforce cooperation. Training was continuous throughout the program, with follow-ups and elements that helped teachers' co-operation within schools.

We carried out structured classroom observations in order to measure teaching methods and classroom management in practice. Two courses were observed in each class, on in mathematics and one in Hungarian literature. Observations were carried out with the help of standardized questionnaires, developed by ourselves with the help of experts in teachers' training. Each classroom observation was carried out by two observers, specifically trained for this task. Classroom observers were selected from among university students in psychology and education. Each observer took part in a two-day training that included video training and trial recordings.

The concept of the questionnaire was to merge objective description of the methods and events observed in the classroom with subjective judgment of interactions, student autonomy and the teacher's work. Some of the questions in the questionnaire were related to the entire period, while others were related to each of the observed activities.

With respect to the whole of the period, we focused on the following objective and subjective questions.

- General physical conditions (light, furniture, decoration etc.).
- Seating order (clustering of students by ability or social background, flexibility in accommodating students' needs, etc.).
- Homework (whether there was any, how it was checked, etc.).
- Teacher's personality and motivation (subjective judgments).
- Use of educational tools.
- Teacher-student and student-student interactions, and general atmosphere in the class.
- Didactic and student-centered elements in the structure of the class.
- General subjective evaluation on various dimensions.

With respect to individual activities, we focused on work forms and general aspects.

- General characteristics of the activity (cognitive level, other characteristics).
- Work form.
- If group work, its characteristics (cooperative nature, the role of interactions, etc.).
- if individual work, its characteristics (role of self-assessment, how teacher helps, etc.).
- whether the activity entailed differentiated education, and if yes, its characteristics.

From among the 197 classes we followed in the evaluation study, we have valid classroom observations in 189, with results from 201 student groups (some classes were split into more groups), and a total of 366 periods that we could evaluate. The total number of valid classroom observations suitable for analysis is 684 (some observations were not valid and were thus dropped from the analysis).

4.1 General Physical Conditions

With respect to the general physical conditions, the circumstances were not necessarily better in the program schools (table 4.1). Twenty per cent of the classrooms in the grade 3 were found dark by our classroom observers, while in the control schools this proportion is below 10 percent. The furniture in the program school classrooms, however, significantly differs from that in the control schools. While classrooms in control schools are almost exclusively furnished with traditional desks suitable for frontal teaching, classrooms in program schools are dominantly furnished with desks that are movable and are suitable for group work. This, in principle, could be realized in the large majority of control schools as well, since the furniture is fixed only in few of the schools. Nevertheless, control schools seem to choose frontal arrangement. In one third of classrooms of the grade 3 classes in program schools there was a chat corner, while the proportion of these in the control schools was below 15.

Table 4.1 – Physical conditions and student arrangement in program and control schools

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Lighting OK (%)	78	91	-13**	79	79	0
Rich decorations (%)	92	94	-2	54	53	+1
Furniture suitable for frontal work (%)	28	82	-54**	45	83	-38**
Furniture suitable for group work (%)	67	15	+52**	49	12	+37**
Furniture fixed (%)	14	29	-15**	14	13	+1
Chat corner available (%)	30	14	+16**	1	1	0
Advanced and less advanced students seated separately (%)	14	29	-15**	14	13	+1
Roma and non-Roma students sat separately (%)	1	6	-5*	2	2	0

+ Statistically significant at the 10% level.

* Statistically significant at the 5% level.

** Statistically significant at the 1% level.

4.2 Classroom Seating

Students are clustered by ability in 14 per dent of program school classrooms in both grades 3 and 7, and in grade seven in control school classrooms as well. In grade 3, however, ability clustering is observed in almost 30 per cent of control school classrooms. At the same time, within-classroom segregation of Roma students is very rare (being found only in 6 per cent of control school classrooms).

4.3 Homework

Homework was assigned to children in 45 per cent of the observed program school classes and in 55 per cent of the control school classes (table 4.2). The difference is not significant, but it is of similar magnitude in the lower and upper grades. There are differences in homework evaluation as well. Rewarding is of similar proportion, but control schools are more likely to apply punishment in cases of poorly executed or missing homework.

Table 4.2 – Homework. Incidence in observed classes (per centage).

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Homework was assigned (%)	44	54	-10+	45	55	
Punishment was applied in connection with homework (%)	5	23	-18**	5	28	
Rewarding was applied in connection with homework (%)	30	22	+8	12	17	

+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

4.4 Personality and Motivation of the Teacher

From among the general teacher competences, we paid attention to three factors: patience, ability to share attention, and motivation (enthusiasm) (table 4.3).

Table 4.3 – Teacher's personality and motivation. Incidence in observed classes (percentage)

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Patience (%)	86	86	0	86	85	+1
Ability to share attention (%)	86	87	-1	84	73	+11+
Enthusiasm (%)	73	73	0	61	50	+11+

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

With respect to the examined characteristics, we found no differences between the program and control schools in the lower grades, and small and not very significant differences in the upper grades. Therefore, if the students of the program schools were more effective, the reasons are unlikely to be found in the teachers' general competence.

It is also worth noticing that, compared to the teachers of lower grades, enthusiasm of the teachers of upper grades is low. In 50 per cent in the control schools and in 40 per cent in the program schools our classroom observers found that the teachers in the upper grades were unmotivated ("lacking enthusiasm"). This proportion is especially high if we take into account the fact that observation itself was likely to increase teachers' motivation. In normal circumstances, teachers are likely to be even less motivated or enthusiastic.

4.5 Use of Educational Tools

Table 4.4 provides information about the use of educational tools. The results demonstrate that the teachers of the program schools use worksheets more often and the blackboard slightly less often. There is no significant difference in terms of the usage of other tools of demonstration and manipulation.

Table 4.4 – Teachers' use of resources. Incidence in per cent of observed classes

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Worksheets (%)	65	41	+24**	48	33	+15**
Blackboard (%)	70	83	-13**	77	84	-7
Resources for demonstration and manipulation (%)	52	52	0	27	21	+6

+ **Statistically significant at the 10% level.**
* **Statistically significant at the 5% level.**
** **Statistically significant at the 1% level.**

4.6 Interpersonal Relationships in the Classroom

One of the most striking differences between observed classes of program versus control schools is in terms of interactions. Table 4.5 demonstrates that the teachers of the program schools, both in the lower and upper grades, were more likely to establish personal relationships with the students.

Table 4.5 – Student-teacher and student-student relationships; student behavior. Incidence in per cent of observed classes.

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
The teacher established a personal relationship with the student (%)	54	35	+19**	39	26	+13**
Medium or high level of student-student cooperation (%)	95	46	+49**	82	34	+48**
Medium or high level of attention (%)	97	93	+4	78	72	+6
Medium or high level of activity (%)	81	88	-7	63	53	+10*
Medium or high level of student autonomy (%)	54	40	+14*	47	33	+14*

+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

Teachers establish personal relationship with the students in program schools significantly more frequently than in control schools, in both grade 3 and grade 7. Such relationships are very important elements of student-centered education. Almost all program school classes are characterized by high or moderate levels of student cooperation, while the same is true for less than half of control school classes. As we shall see, this is a direct consequence of program school teachers using cooperative educational methods. The students' level of attention was high both in the program and control schools, perhaps in part due to the classroom observation. The students' level of activity (remarks, engaging in discussing the topic, raising questions) is also high, with non-significant but negative difference in lower grades, and slightly significant but similarly small positive difference in upper grades.

Importantly, student autonomy is somewhat higher in program schools. It seems that in the program schools, teaching methods, classroom management techniques, and perhaps teachers' attitudes, lead to high or moderate levels of student autonomy in a somewhat larger fraction of program schools than in controls schools. Examples include students' decision in the pace of learning, self-assessments, etc. Even in program schools, though, almost half of the observed classes were characterized by low levels of student autonomy.

4.7 Didactic and Student-centered Elements

Table 4.6 shows some important structural characteristics of the observed classes. In the table, we group these characteristics by whether they belong to didactic or student-centered educational methods.

Table 4.6 – Didactic and student-centered elements in the structure of the period. Incidence in per cent of observed classes

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Didactic elements						
At the beginning of the class, the teacher shares the objectives of the period (%)	57	57	0	76	71	+5*
At the beginning of the class, the teacher shares the structure of the period (%)	18	10	+8+	19	15	+4
The structure of the class is logical (%)	88	95	-7*	81	96	15**
Summarizing of the content at the end of the period (%)	23	49	-26**	33	27	-6
Student-centered elements						
Warm up activity at the beginning of the class (%)	65	60	+5	39	32	+7
Closing up activity (%)	16	23	-7	13	3	10**

\+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

The picture is mixed in terms of how well teachers do by the standards of didactic education. In terms of some elements, program school teachers seem to work more in accordance with the objectives of didactic education, while in terms of other elements, control school teachers seem to do better. On balance, teachers in control schools seem to place a larger emphasis on the elements of didactic organization of the class (logical structuring, summarizing). The use of non-cognitive warm up and closing activities is somewhat more widespread in program schools, another sign of more room for student-centered education there.

Keep in mind that because of classroom observations, the teachers are likely to have concentrated on the elements that they themselves considered to be important. In light of that, the results of table 4 suggest that teachers of the control schools consider didactic elements to be more important, while there are more teachers in the program schools for whom student-centered elements are more important.

Observed characteristics of teachers' instruction, questioning and feedback practice reinforce the notion that teachers in the control schools are better rather in their didactic skills, while teachers in the program schools are somewhat better in terms of the affective elements, especially in lower grades (table 4.7). In observed control school classes, teachers ask more questions of more students. In case of wrong answers, however, it is rather the teachers themselves who provide the answers. At the same time, at least in lower grades, teachers of the program schools are better at giving constructive feedback and assessment are somewhat more likely to be concentrated on the development of the student.

Table 4.7 – Instruction, questioning techniques, feedback. Incidence in per cent of observed classes

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Teacher questioning techniques						
The teacher asked many questions (%)	81	86	-5	72	84	-12*
The answers given to the class were emphatic (%)	40	39	+1	40	39	+1
Everybody took their turn (%)	53	75	-22**	58	62	-4
In case of a wrong answer, the teacher provided the answer (%)	15	22	-7*	26	33	-7+
Feedback						
Frequent constructive feedback (%)	40	25	15**	25	23	+2
Frequent positive emotional feedback (%)	43	41	+2	22	21	+1
Frequent feedback concentrating on the development (%)	40	31	+9+	30	28	+2

+ **Statistically significant at the 10% level.**
* **Statistically significant at the 5% level.**
** **Statistically significant at the 1% level.**

4.8 Classroom Activities

In addition to the evaluation of the entire class period, we also examined the individual activities separately. The number of activities in the program and control schools is practically identical. In the lower grades the average is somewhat above 4, in the upper grades of around 3.5. The average covers few classes with many activities, and many one or two-activity classes.

Program and control schools show differences in the characteristics of the activities they chose for the classroom observations. These differences are typical primarily in four dimensions: the cognitive level of tasks, the nature of the tasks, the presence of group work and cooperative organization, and the presence of differentiated education.

Table 4.8 demonstrates that the cognitive level of the task instructed by the teacher is somewhat higher in the program schools. In these schools, slightly fewer tasks are reduced to passive reception from the children, and slightly more tasks are encouraging the students to a higher level of thinking (representation and construction).[25]

Table 4.8 – Task structure and cognitive level of tasks. In per centage of the observed classes

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Reception (%)	14	29	-15**	21	36	-15**
Application (%)	70	59	+11**	56	56	0
Representation or construction (%)	14	12	+2	22	6	+16**

+ Statistically significant at the 10% level.

* Statistically significant at the 5% level.

** Statistically significant at the 1% level.

Table 4.9 demonstrates that a larger fraction of the observed activities in the program schools are motivating, manual, playful, involving active motions, or creative. Again, keeping in mind that teachers were surely better prepared for the observed classes than usual, the higher proportion of the more creative, more playful etc. activities, indicates a shift in the direction of more student-centered educational methods in the program schools.

[25] If the observed activity contained tasks with varying cognitive levels, the monitors recorded the highest level.

Table 4.9 – Task structure. In proportion of observed classes

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Strongly motivating (%)	32	22	+10**	17	10	+7*
Manual (%)	20	12	+8**	13	6	+7*
Playful (%)	42	32	+10**	23	9	+12**
Creative (%)	32	16	+16**	23	8	+15**
Kinesthetic (%)	14	7	+7**	8	3	+5*

+ **Statistically significant at the 10% level.**
* **Statistically significant at the 5% level.**
** **Statistically significant at the 1% level.**

4.9 Work Form

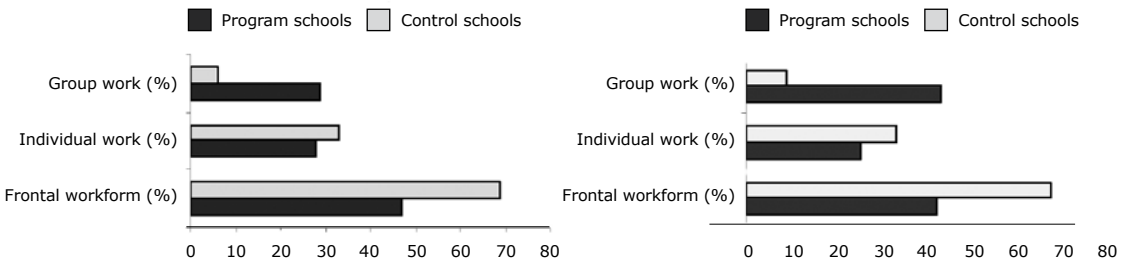
The work form of the observed activities shows significant differences between program and control schools. We differentiated among three work forms: frontal (the teacher works with the children who are seated in front of him or her), individual (the students work individually), and group work (the students work with each other in groups). The proportions of the incidence of work forms are shown in table 4.10 and the related figure 4.1. When interpreting the numbers, it should be taken into account that there may be related to activities classified simultaneously in two work forms (e.g. frontal activity combined with individual work).

Table 4.10 – Work forms of activity. In per centage of observed activities

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Frontal work form (%)	47	69	-22**	42	67	-27**
Individual work (%)	28	33	-5	25	33	-8**
Group work (%)	29	6	+23**	43	9	+32**

+ Statistically significant at the 10% level.
* Statistically significant at the 5% level.
** Statistically significant at the 1% level.

Figure 4.1 – Work forms of activity. In percentage of observed activities.



Grade 3 Grade 7

The dominant work form in the control schools is traditional, frontal teaching: in more than two thirds of the activities this is what we observed. The role of frontal teaching in the program schools is much smaller: less than half of the activities could be listed there. The difference is "filled" by group work: while in the control schools less than one tenth of the activities are in the form of group work, in the program schools it is 30 per cent in the lower grades and 40 per cent in the upper grades. The proportion of individual work is around 30 per cent, with somewhat smaller prevalence in program schools.

4.10 Group Work and Cooperative Learning

Traditional classroom education has been based on alternating phases of frontal teaching and individual work. Cooperative learning techniques are outside this traditional scheme: they foster cooperative, collaborative student work by making students work in groups in the classroom. Group work and cooperative learning techniques have become very popular in the past decades. Being an active learning form, it promises to increase students' attention and involvement and thus helps their cognitive development. By making students work in groups for common goals, it is also likely to provide an opportunity to develop students' social skills as well.

The impact of cooperative learning techniques on the student development has been examined in the literature. Results from 675 different studies were examined in a large-scale meta-analysis.[26] Overall, those studies show that cooperative learning techniques fulfill their promises, both on the cognitive and social development of children.

Group work does not automatically lead to such outcomes. Appropriate group incentives and the cooperative nature of the tasks are shown to be very important.[27] Competition among groups is shown not to provide positive effects. At the same time, it turns out to be important whether the group's results depend on each individual's contribution or not. These suggest that the success of cooperative techniques is in large part due to better motivation of students. This motivation is aimed not at defeating other groups but at encouraging the members of their own group.

The impact of cooperative techniques on non-cognitive dimensions is even more powerful. They are shown to promote friendships and positive interactions among children of different levels of achievement, family background and ethnicity. In addition, cooperative techniques often have positives impact on emotional components such as self-esteem, self-confidence or the feeling of being accepted.[28]

Placing students into groups does not necessarily mean that they will work together in a cooperative way. It is important for the students to understand the elements of group work that are indispensable for the success of the method. The meta-analysis referred to above identified the following criteria:

[26] Johnson, D. W.–Johnson, R. T.: Cooperative Learning and Social Interdependence Theory. In R. Scott (ed): Theory and Research on Small Groups. Plenum Press, Klatt, Bruce, New York, 1999.

[27] Slavin, R. E.: *Cooperative learning.* Longman, New York, 1983.

[28] Good, T. L.–Brophy, J. E.: *Looking in classrooms* (9th edition). Allyn and Bacon-Longman, Boston, 2003. In Hungarian: Good, T. L.–Brophy, J. E.: *Nyissunk be a tanterembe!* Translated by Katalin Abrudán, edited by Zsuzsa Kereszty. The basis of the translated version was the 10th edition. Educatio, social service provider public benefit organisation, Budapest, being published.

- » **positive interdependence** – the development of individual students or student groups depends on other students, and the success of the group depends on common effort;
- » **cooperative behavior and parallel interactions between group members** –there are simultaneous interactions among students in the learning process;
- » **clear individual responsibility** – each child in the group is responsible for the success of the group, everybody has to make a contribution to the group's goals;
- » **equal participation** – each child joins the work in an identical way (for instance, work is not voluntary).

Earlier we demonstrated that group work in program schools is significantly more widespread than in control schools. In addition, the classroom observations revealed that the characteristics of group work are also different in program and control schools. The most important aspects are shown in table 4.11. According to the table, not only do program schools apply group work more often, but they also they do it better, i.e. more in line with the principles of cooperative learning techniques.

Table 4.11 – Elements of cooperative learning in program and control schools

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Positive interdependence (%)	35	18	+17*	32	8	+24**
Peer cooperation (%)	47	45	+2	43	32	+11*
Clear individual roles	42	24	+18*	31	30	+1
Equal participation	95	100	-5	95	85	+10

+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

Since one of the most important goals of cooperative learning techniques is the development of social skills, we examined the extent to which teachers pay attention to such development. The results are summarized in table 4.12. They show that teachers of the program schools are more likely to formulate expectations towards children with respect to their social behavior, and they are also more likely to evaluate students' social behavior after the activity.

Table 4.12 – Teacher's focus on social behavior in the course of group work

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Teacher's explicit expectation regarding social behavior (%)	57	33	+24**	52	22	+30**
Teacher's evaluation of social competences (%)	49	22	+27**	27	12	+15**

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Summarizing the evidence, group work is substantially more frequent in the program schools (it is a rare exception in control schools), and to a large extent, it is in line with the criteria for successful cooperative learning, both in the lower and upper grades.

4.11 Individual Work

Approximately 30 per cent of the activities involved individual work in both the program and the control schools. If we look into the details of individual work (see table 4.13), we find some interesting differences, especially among third-grade classes. Students in the program are significantly more likely to ask for help or confirmation from the schoolteacher, and they are more likely to cooperate with each other during individual work activities. Program school teachers provide individualized instructions more frequently than teachers of the control schools. The results imply that in program schools, student-teacher relations are more open, student-student relations are more cooperative, and schoolteachers pay more attention to individual differences. The differences are of similar direction in the upper grades, but they are much weaker.

Table 4.13 – Characteristics of individual work in program and control schools

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Students ask for help (%)	66	28	+38**	51	44	+7
Students ask for confirmation (%)	52	38	+14**	47	43	+4
Emergence of peer cooperation (%)	40	21	+19**	32	28	+4
Emergence of individualized instructions (%)	41	19	+22**	30	24	+6

\+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

4.12 Differentiated Education

Increasing differences between students in integrated classes requires strategies that directly take those differences into consideration. Increasing heterogeneity in elementary schools is a phenomenon observed in a number of countries. This evoked the emergence of differentiated education and its international dissemination.[29] The basic idea of differential education is that in the interest of optimal development of children with different capacities, background and attention span, educational techniques (tasks, instructions, motivation etc.) have to take those differences into consideration.

Although many teachers believe in the benefits to differentiated education, there is surprisingly little empirical evidence on this question. Apparently, there is not much relevant, properly comparable data. A main reason may be heterogeneity in the implementation of differential education. In fact, differential education is not an instructional technique but more an approach. At its heart is flexible adaptation to the various situations, requiring a mastery of the profession and innovative thinking on the part of the teacher.[30]

[29] Sizer, T. R.: No two are quite alike. Educational Leadership, 1999/57(1), 6-11.

[30] Tomlinson, C. A.: The Differentiated Classroom: Responding to the Needs of all Learners. Association for Supervision and Curriculum Development, Alexandria, 2000.

The few available empirical studies show a positive impact from differential education, although the same studies draw attention to the difficulties and risks as well.[31] According to some, in order for differentiated education to be successful, not only a few teachers but the entire school should adapt it. The advantage of a school-wide approach is in the professional assistance, mentoring, and consulting potential by which the initial frustration and resistance on the part of (some) teachers can be overcome.[32] This element is important because, in general, the transition to a differentiated education is accompanied with serious difficulties and increased teacher workload.

The conclusions from existing research suggest that in order for differentiated education to be successful, and for teachers to adequately manage heterogeneous and multi-cultural classes, effective and long-term assistance and support is very important. In principle, the OOIH program should provide exactly that kind of support. It seems, however, that differentiated education has not been emphasized in this program. Our classroom observers came across very few instances of differentiated education. Very few were observed in grade 7, somewhat more in grade 3, and overall, program schools are somewhat more likely to differentiate.

In the course of the classroom observations, we paid special attention to possible differentiation with respect to each activity. If some students were instructed in a different way, the general characteristics of those different instructions were recorded separately. In case of more than two parallel activities, the classroom observers examined the numerically more dominant group and the weaker, or the weakest differentiated group. Besides recording the characteristics of the differentiated activities separately, we examined a couple of general issues in relation to the basic principles, the form and the content of differentiation. The results are shown in table 4.14.

[31] Johnsen, S.: Adapting instruction with heterogenous groups. Gifted Child Today, 2003/26(3), 5-6.
[32] McAdamis, S.: Teachers tailor their instruction to meet a variety of student needs. Journal of Staff Development, 2001/22(2), 1–5.

Table 4.14 – Prevalence of differentiated education and its main characteristics in program and control schools

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Substantial differentiation took place during the activity (%)	12	7	+5*	5	4	+1
If there was differentiation, the better students were separated from the dominant group (%)	7	23	-15	0	74	-74**
If there was differentiation, its form was individual work (%)	75	78	-3	71	91	-20

+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

In grade three, 12 per cent of the observed activities in program schools involved explicit differentiation, while the corresponding number is 7 per cent in the control schools. In the upper grades, the corresponding figures are 5 and 4 per cent, respectively. In the program schools, practically all differentiated activities involve separate instructions for a small group of weaker students. Both in the program and the control schools, differentiation emerges in individual work activities. The very few observed instances of differentiation do not provide a possibility for further and more detailed analysis.

4.13 Summary

The results of the classroom observations can be summarized in the following way. We find no significant differences in the general physical conditions and teachers' general competence between program and control schools. However, we do find a shift in the direction of student-centered education in the program schools, in particular in the lower grades. It is manifested in classroom furnishing, the organization of classroom activities, the nature of the activities, and the more frequent personal contact between teachers and students. The shift towards student-centered education is not dramatic, but it affects many dimensions of classroom work. Probably due to the more student-centered education, student autonomy is somewhat higher in program schools, in both the lower and the upper grades.

Group work in the program schools is much more widespread, and it is implemented in accordance with the principles of cooperative learning techniques. As an immediate result, student-student relationships in program schools are more cooperative and collaborative. But cooperative education may also have effects on students' non-cognitive skills and social attitudes, to be analyzed later. Differentiation in the control schools is barely observed, and, although it is more frequent in program schools, it is more of an exception there as well. When observed, differentiation is more likely to involve separate tasks given to a few low-achieving students.

Reading Skills and Admission to Secondary School

The next three chapters examine the effect of the program on the students' skills and attitudes. This chapter focuses on cognitive achievements (reading skills, grades, and directions of education after grade 8). Chapter 6 focuses on positive self-esteem and some related non-cognitive skills, and Chapter 7 looks at ethnic prejudice and related attitudes.

Reading skills, or literacy, are fundamental in modern societies. They are known to have a significant effect on labor market success, welfare, and success in virtually all domains of modern life.[33] Literacy tests, therefore, are probably the most important measures of the cognitive skills learned at elementary schools. The first PISA research focused on literacy as well. One of the two tests in the National Competence Assessment of Hungary, the national standardized test that is carried out on a yearly basis, also focuses on reading skills.[34]

The results of developmental psychology, education research, international experience and the focus group discussions we held with schoolteachers of program schools all presumed that the OOIH program may have a small-scale impact on the students' reading comprehension development at most. Most results in the literature imply that schools, in general, have little influence on students' cognitive processes.[35] To a large extent, children's cognitive skills develop well before the beginning of schooling age, because the development of the brain largely finishes by age six.[36] It seems that when school education has a large impact on the cognitive performance, the impact is generally realized through information processing strategies and through the shaping of some specific skills or abilities.[37]

[33] See, for instance, the OECD PISA-reports (http://www.pisa.oecd.org) and the IALS study (http://www.oecd.org/document/2/0,3343,en_2649_39263294_2670850_1_1_1_1,00.html).

[34] In principle, all students of the fourth and eighth grades examined by us were covered by the National Competence Assessment in mathematics and reading comprehension. Unfortunately, we cannot use the results of those tests in this study. The identity of the students in the National Competence Assessment data are protected to the degree that prohibits linking individual test results to our data. Up to 2008, it was not possible to link individual test results through time, either. Schools can be identified in a restricted use version of the National Competence Assessment data but that would not allow for analyzing the effects on different groups of students (Roma, disadvantaged, etc). Regular, comprehensive competence assessments that allow for linking students' results through time, and, with appropriate provisions, to outside data, would facilitate impact assessment studies similar to ours.

[35] Cole, M.: Cognitive development and formal schooling. In L. C. Moll (ed.): Vygotsky and education: instructional implications and applications of sociohistorical psychology. Cambridge University Press, Cambridge, 1990.

[36] DeBord, K.: Brain development. (Extension Publication) North Carolina Cooperative Extension Service, Raleigh, N. C., 1997.; Shore, R.: Rethinking the brain: New insights into early development. Families and Work Institute, New York, 1997.

[37] Cole, M.: i. m.

The results that emphasize the limits of the school in changing cognitive skills, in particular of the upper grades, complemented with the difficulties faced by disadvantaged individuals in the skills development in their adulthood (see, for instance, the failures of training programs of the long-term unemployed[38]) imply the programs that can bring significant results in the cognitive development of disadvantaged children are the ones that start early in life, well before schooling age.[39]

Of course, the fact that underlying cognitive skills are hard to manipulate in school does not mean that the schools have no role in children's skills development in general, and the development of their reading skills in particular. Children normally learn to read at school. Different school can induce different patterns of development in literacy even if underlying cognitive skills are the same. Nevertheless, there is ample evidence for the inability of school-age education to fully make up for deficiencies in early childhood development. Therefore, we can expect limited success at most from the OOIH program in improving students' reading skills.

We first look at grades. The Hungarian grading system assigns grades 1 through 5, with 1 representing failure and 5 the highest grade. Grades are among the most accessible measures of students' progress. Although they clearly contain subjective elements, they can be a useful piece of information in the effect of program. Table 5.1 shows some of the most important grades.

Table 5.1 – Average grades at the end of the fall semester

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Grade points average	4.0	3.9	+0.1*	3.6	3.6	+0.2**
Mathematics	3.5	3.4	+0.1+	3.0	3.0	0.0
Hungarian literature	3.8	3.6	+0.2**	3.6	3.4	+0.2**
Hungarian grammar	3.6	3.4	+0.2**	3.3	3.2	+0.1
Foreign language	–	–	–	3.6	3.4	+0.2**

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

[38] Péter Hudomiet–Gábor Kézdi: International Experience of the Active Labor Market Programs (Az aktív munkaerő-piaci programok nemzetközi tapasztalatai). 2008 Online: http://www.kormanyzas.hu.

[39] See the James J. Heckman in Skill Formation and the Economics of Investing in Disadvantaged Children. Science, 2006/312. (June 30), 1900-1902. The Hungarian experience of the post-communist transition also supports significance of the pre-school years, see Gábor Kertesi and Gábor Kézdi: Children of the Post-Communist Transition: Age at the Time of the Parents' Job Loss and Dropping Out of Secondary School. The B.E. Journal of Economic Analysis & Policy, 2007/Vol. 7. Iss. 2. (Contributions), Article 8. http://www.bepress.com/bejeap/vol7/iss2/art8.

Grades are somewhat higher in the program schools than in the control schools, both in the lower grades and in the upper grades. (Formal grading in grades 1 through 4 was abolished in Hungary starting with the cohort following the one analyzed here.) Statistically significant differences were found in Hungarian literature, in grammar in the lower grades, and in foreign languages in the upper grades. No difference was found in mathematics grades. Table 5.2 shows that everybody's average is a little better in the program schools. In grade 4, the differences are similar across subsamples, and they are not significant. In grade 8, the program versus control school difference was largest among the Roma students and smallest among the disadvantaged.

Table 5.2 – Grade point average at the end of the fall semester by ethnicity and disadvantage

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Roma students	3.6	3.5	+0.1	3.2	2.9	+0.3**
Non-Roma students	4.2	4.2	+0.0	3.8	3.7	+0.1**
Disadvantaged students	3.6	3.5	+0.1	3.0	2.9	+0.1**
Non-disadvantaged students	4.3	4.3	+0.1	3.8	3.6	+0.2**

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

These results may indicate that students in the program schools develop better, but they may as well reflect different grade-giving habits of program versus control schools.

5.1 Standardized Tests

Cross-school comparison of students' skills is best measured by standardized tests that are independent of the schoolteachers' grading and promotion decisions. As we noted earlier, we could not use the tests used by the National Competence Assessment because of data protection. Instead, we administered tests ourselves within the framework of this research.

When choosing a test, we focused on skills and competences that are valuable outside the world of the school. We used two tests developed by educational scientists in Hungary. In the first wave of data collection, we administered an inductive thinking test (in grade 6) or

a mathematics test (in grade 2), in order to be able to control for "pre-program" cognitive skills (as we noted, the impact assessment research could only start with more than a year's delay due to administrative reasons, therefore the test was administered on cohorts that were already in the program for more than a year). The tests were developed by the research team led by Benő Csapó. Results are in table 5.3.[40]

Table 5.3 – Standardized results of inductive thinking tests (grade 6) and mathematics (grade 2). Cognitive measures taken one and a half years into the program.

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
All students	0.03	-0.03	+0.06	0.01	-0.01	+0.01
Roma students	-0.43	-0.47	+0.04	-0.49	-0.55	+0.06
Non-Roma students	0.22	0.19	+0.03	0.22	0.20	+0.02
Disadvantaged students	-0.43	-0.47	+0.04	-0.55	-0.52	+0.04
Non-disadvantaged students	0.40	0.25	+0.15+	0.18	0.17	+0.01

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

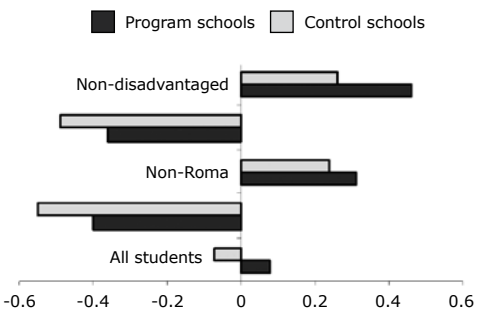
Students in the program schools perform somewhat better than their peers in the control schools, but those differences are not significant at the 5 per cent level in any of the subsamples. The test score gap between Roma and non-Roma and between disadvantaged and non-disadvantaged students are very similar, at two thirds of a standard deviation in grade 2, and three quarters in grade 6. Recall that in Chapter 2, we showed that there is no difference between program and control schools in terms of the family background of the students. The results of table 5.3 show that one and a half years into the program, there are some differences in terms of the cognitive skills, but those differences are negligible. These results are consistent with comparable student background in program and control schools and the limited capacity of schools to affect basic cognitive development of the students.

[40] Benő Csapó (edit.): The school knowledge (Az iskolai tudás). Osiris Publisher, Budapest, 2002, as well as http://www.edu.u-szeged.hu/mtakcs/. We would like to separately express our thanks to Benő Csapó for the assistance provided during the research.

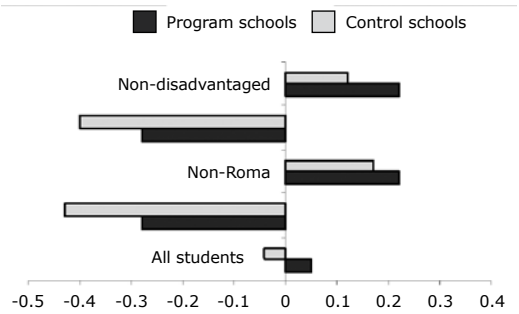
They are also consistent with some positive effects of the program during those one and a half years that these second and sixth-graders spent in the program.

The outcome cognitive measure we analyze is reading skills. The reading tests we used were also developed by the research team led by Benő Csapó. The tests were developed following the principles of the PISA test. The results we standardized within the sample; that is, the average of the sample is set zero, and the standard deviation to one. The results are shown in table 5.4 and the related figure 5.1.

Figure 5.1 – Standardized results of reading comprehension tests



Grade 4



Grade 8

Table 5.4 – Standardized results of reading comprehension tests

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
All students	0.08	-0.07	+0.15**	0.05	-0.04	+0.09+
Roma students	-0.40	-0.55	+0.15+	-0.28	-0.43	+0.16+
Non-Roma students	0.31	0.24	+0.07	0.22	0.17	+0.05
Disadvantaged students	-0.36	-0.49	+0.13+	-0.28	-0.40	+0.12
Non-disadvantaged students	0.46	0.26	+0.20**	0.22	0.12	+0.10+

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

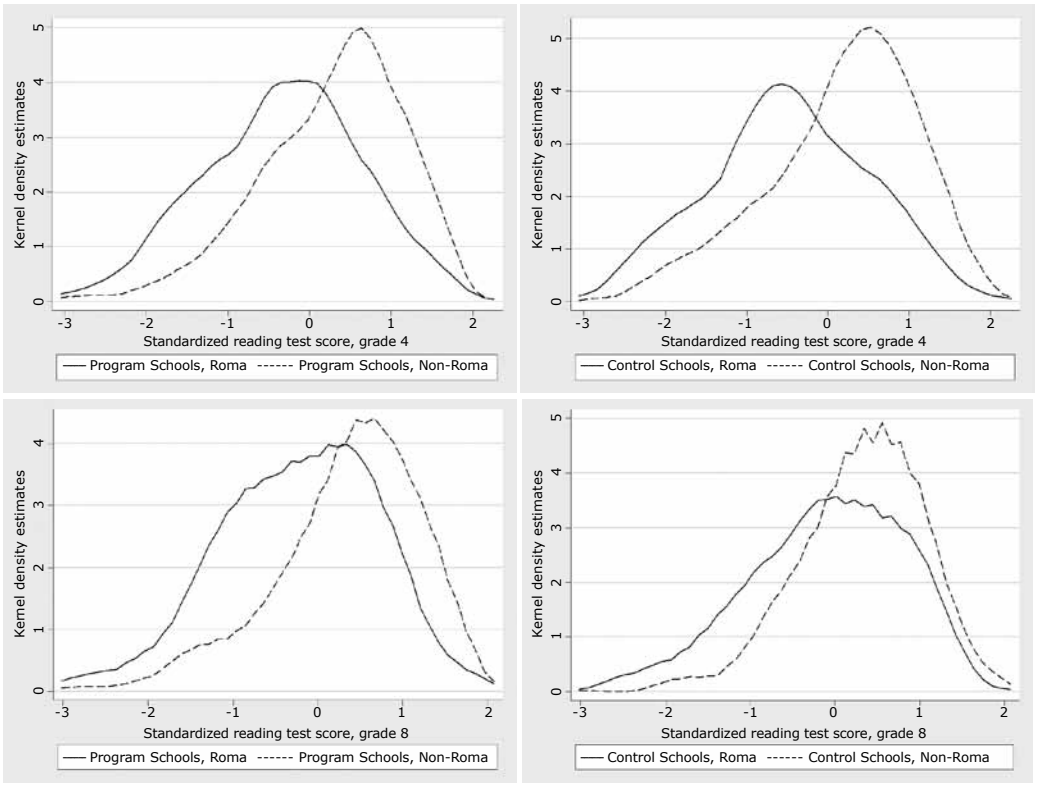
The students of the program schools perform better on the reading tests than the students in the control schools. The average advantage is 0.15 standard deviations in fourth grade (statistically significant at 5 per cent) and 0.09 standard deviations in eighth grade (statistically significant only at 10 per cent). These results are consistent with the notion that the effect of the program should be decreasing with the children's age. Both the effects themselves and their difference between grades are very small, though, which should be kept in mind when interpreting the results.

Similarly to the inductive thinking and mathematics tests measured two years earlier, we see large test score gaps by ethnicity and family background. In control schools, the ethnic test score gap is 80 per cent in grade 4 and 60 per cent in grade 8; the gap between disadvantaged and non-disadvantaged students is similar, at 75 per cent in grade 4 and 52 per dent in grade 8. The test score gap is smaller in grade 8. The ethnic score gap is smaller in program schools and in control schools. Again, we compare relatively small numbers that are estimated imprecisely, which should be kept in mind when interpreting these comparisons.

The heterogeneity of the program versus control difference is broadly consistent across grades. The effect is larger for Roma students than for non-Roma students. They are not necessarily larger for disadvantaged students, though: in grade 4 the difference is actually larger among non-disadvantaged students. The most important result, however, is that the larger gains of the Roma students are experienced not at the expense of the gains of their non-Roma peers. To the contrary: non-Roma students in program school perform better than in control schools, too.

The differences in the averages may hide interesting differences in the distributions. The more detailed examination of the distribution of test score results (figure 5.2) shows non-Roma test scores have a long left tail, while a significantly larger fraction of the Roma test scores are in the (-2,0) range. Both in grade 4 and grade 8 the program school distributions have a significant more mass of the density in the (-1,-2) range than in the (-2,-3) range, and more mass slightly above 0. A causal interpretation of these differences suggests that program schools can raise the reading level of Roma students at the very bottom of the distribution to levels that are still very low, and those schools are also more successful with Roma students that are somewhat above the Roma average but below the non-Roma average.

Figure 5.2 – Distribution of standardized reading test scores in program and control schools



The program school versus control school differences may, in principle, arise from three sources: differences in the student composition, differences between the schools which are independent of the program, and the impact of the program. As we demonstrated in Chapter 2, the student composition is virtually identical in program and control schools regarding all observable characteristics. It is therefore unlikely that compositional differences are the source of the observed differences in reading skills.

Of course, there may be differences in the non-observable characteristics. One way of controlling for such unobservables would be controlling for the cognitive test results made two years before. Assuming that unobserved trends in cognitive development are the same for students with different family background, this difference-in-differences in test scores should be zero in the absence of the second and third mechanisms (program schools add more to the same students either because of or independent of the program). While the earlier test results refer to skills other than reading (inductive thinking and mathematics), controlling for them would make sense under assumption that cognitive skills are either

one-dimensional ("g") or correlated. Note however, that our measure of pre-program skills are taken one and a half years into the program. Controlling for earlier test score results would therefore underestimate the schools' impact.

Deciding whether the schools' impact is due to the program or not is even more difficult. Our take at the problem, as explained in detail in Chapter 2, is repeating the exercise on the narrow sample of program schools that were not integrating before the program (and their control pairs). Recall that, based on interviews by the headmasters, half of the program schools were not integrating before joining the program (where, presumably, "integrating" has a wide meaning, including explicit attention paid to the development of disadvantaged students). If, in the narrowed sample, the program versus control school differences are similar to the overall differences, we can have more confidence in assigning those differences to the impact of the program itself. The results are shown in table 5.5.

Table 5.5 – Average program school versus control school differences in standardized reading tests. Regression results controlled for students' family background as well as test results from two years before. Full sample as well as the narrow sample of program schools that were not integrating before the program (and their control pairs)

	All students	Roma students	Non-roma students	Disadvantaged students	Non-disadvantaged students
Grade 4					
Full sample	+0.10**	+0.12	+0.06	+0.16**	+0.07
Narrow sample	+0.08	+0.21+	+0.03	+0.09	+0.11
Grade 8					
Full sample	+0.09+	+0.13	+0.03	+0.07	+0.04
Narrow sample	+0.13*	+0.09	+0.18*	–0.02	+0.17*

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Controlling for earlier test results induces little change in the program versus control differences. The advantage of program school students dropped from 15 per cent to 10 per cent of a standard deviation in grade 4, while it did not change in the upper grades. The changes are similar in the Roma, non-Roma, disadvantaged and non-disadvantaged groups. By considering the narrow sample of schools that were not integrating before the program, the differences decrease in certain groups (grade 4: overall and disadvantaged students,

grade 8: Roma and disadvantaged students), and their increases in others (grade 4: Roma students, grade 8: overall, non-Roma, non-Disadvantaged students). These comparisons do not show clear tendencies. They are not statistically significant either.

The results are consistent with the notion that, as a result of the program's impact, students in program schools experience a larger increase in their reading skills than their peers in control schools. The positive effects are modest at most, and many times not statistically significant. What is clear, however, is that there was no group whose members suffered disadvantage as a result of the program.

5.2 Admission to Secondary School

After the reading comprehension, we will examine the further education of students. The Hungarian school system is somewhat complicated. Inherited from the communist past is the division of education into an 8-year elementary school and a multi-tier secondary school system of 4-year schools, the upper tiers of which open the way to higher education, while the lower tier schools allow for no such continuation. The dividing line between those who can and those who cannot continue their studies in higher education institutions is the graduating examination taken after grade 12.[41] Some of the most elite secondary schools now recruit students starting at grade 5, but the students in our sample typically stay in their original elementary school until grade 8.

We examine the direction of further studies of the eighth graders in our sample. In these days, virtually all 8th-graders continue their studies somewhere, but a large fraction ends up in lower-tier schools that do not offer (and do not prepare for) a graduating examination. At the same time, the dividing line in terms of labor market success is whether one completes the graduating examination itself or not.[42] Our focus is therefore on the binary event of going to a high-tier secondary school (with maturity exam at the end) or not. Table 5.6 shows the differences between program and control schools in terms of student aspirations.

[41] The graduating examination after grade 12 is called "érettségi vizsga" in Hungarian, which translates as "maturity examination." It is the Hungarian equivalent of the "Matura" or "Baccalaureat" examinations found in many European countries.

[42] See, e.g. Gábor Kertesi and Júlia Varga, "Employment and Educational Attainment in Hungary", and Gábor Kezdi, "Education and Earnings." Both in: K. Fazekas and J. Varga (eds.), The Hungarian Labour Market 2005. In Focus: Education and the Labour Market, Institute of Economics, Budapest, 2005. http://econ.core.hu/doc/mt/2005/en/infocus.pdf.

Table 5.6 – Aspirations for school qualifications in the eighth grade

	Wants to complete secondary school with graduating examination			Wants to get a higher education degree		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
All students	85	80	+5**	47	43	+4
Roma students	73	63	+10*	26	20	+6
Non-Roma students	91	87	+4*	57	53	+4
Disadvantaged students	72	62	+10+	18	21	-3
Non-disadvant aged students	90	88	+2	57	53	+4

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Both the overall results and the results within each group show that the majority of the students want to achieve at least a secondary school degree with a completed graduating examination, and a significant fraction wants to get a college degree as well. There are significant differences across groups: aspirations are lower for Roma students and disadvantaged students. Importantly for our purposes, students in the program schools have higher aspirations than their peers in the control schools. The differences are statistically significant in terms of maturity exam, and they are smaller and insignificant in terms of higher education. The program versus control school difference is largest among the Roma students.

After the intentions, we turn to results. While the intentions were recorded as parts of the non-cognitive questionnaire of the 8th-grade students in May (see in Chapters 6 and 7), the results of the admissions were collected from the school administrations after the end of the school year. Table 5.7 and the related figure 5.3 show actual secondary school admissions. The reported figures are the fraction of eighth graders who were admitted to the upper-tier secondary schools that prepare for a graduating examination.

Table 5.7 – Admission rates in secondary schools that provide a graduating examination, after the eighth grade (%)

	Grade 8		
	Program schools	Control schools	Difference
All students	69	60	+9**
Roma students	49	37	+12*
Non-Roma students	77	70	+7*
Disadvantaged students	44	40	+4
Non-disadvantaged students	77	68	+9**

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

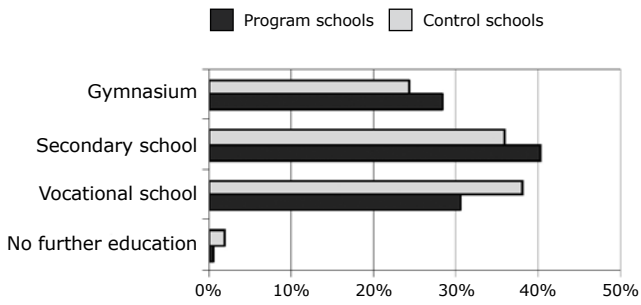
*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Nearly 70 per cent of the students of the program schools are admitted to secondary schools that provide a graduating examination, while the same figure is 60 per cent in the control schools. Admission rates are around 20 per centage points (roughly a third) lower for the Roma and for the disadvantaged students in general. However the Roma – non-Roma gap in admission rates is lower in the program schools. In other words, the program school versus control school difference is largest for the Roma students, both in absolute and in relative terms. Among Roma students, the program-control difference is 12 per centage points (more than a third of the baseline control school rate of 37 per cent), while it is 7 per centage points among the non-Roma students (a tenth of the baseline control school rate of 70 per cent). Similarly to the reading test results, the different is larger among non-disadvantaged students than among disadvantaged ones. Therefore the disadvantaged versus non-disadvantaged gap is larger in program schools. The results by ethnicity and disadvantaged situation suggest that program schools are especially successful in promoting the success of their non-disadvantaged Roma students.

Figure 5.3 demonstrates that the advantage of the program schools originates from higher admission rates both to the most prestigious gymnasium and the other types of upper-tier secondary schools.

Figure 5.3 – Admission into secondary schools after grade 8, per cent.



In table 5.8, we present the program school versus control school differences after controlling for the family background of students as well as their cognitive test scores from two years before. Similarly to table 5.5, we show estimates both for the entire sample and the narrow sample of program schools that were not integrating before the program and their control pairs. These differences after controlling for the cognitive test results are meant to show the differences in the further education opportunities of students of identical pre-program cognitive skills.

Table 5.8 – Program versus control school differences in admission rates to secondary schools that provide a graduating examination, in per centage of 8th grade students. Regression results controlled for students' family background as well as test results from two years before. Full sample as well as the narrow sample of program schools that were not integrating before the program (and their control pairs)

	All students	Roma students	Non-roma students	Disadvantaged students	Non-disadvantaged students
Full sample	+7*	+8	+7*	+4	+7*
Narrow sample	+5	+6	+6+	+7	+5

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Controlling for cognitive skills two years before reduces the program versus control differences a little, but they remain positive if not always statistically significant. The drop in the coefficient is largest among the Roma students. The coefficient remains the largest among the four groups, but it loses its statistical significance. The coefficients on non-

Roma and non-disadvantaged students remain largely intact. The coefficients drop further in the narrow sample. The additional decrease is very small, but it makes all coefficients insignificant at the 5 per cent level (in part because of smaller sample size).

Recall that, as shown above in table 5.3, students in the program schools have somewhat higher results on the inductive thinking test taken in the first wave of data collection, two years before the admission records, in grade 6. In terms of those results, the advantage of program school is larger for Roma and disadvantaged students. Also note that by controlling for test results from two years before, we exclude from the analysis those eight graders students who repeated a grade and were not taking the test two years before because they were not sixth graders then. To the extent that grade repetition is more common in control schools, excluding them shows smaller program-control differences than the real differences. The different results of table 5.7 and 5.8 are due to these two factors.

When interpreting the admission results, one has to keep in mind that had admission depends on the decision of the secondary schools as well. Indeed, one element of the OOIH program is strengthening the relationship between the program schools and secondary schools in order to promote the further education of disadvantaged students. Obviously, the success of such relationships depends in part on the responsiveness of the secondary schools.[43] Therefore, any positive results with respect to admissions are of particular importance.

Summarizing the conclusions of the chapter, we can state that the students of the program schools achieve somewhat higher grades, their reading skills are also somewhat better, and they are more likely to pursue further education in secondary schools that provide a graduating examination than their peers in control schools. The reason is more likely the higher quality of education provided by program schools than differences in student composition. It is also likely that the program schools provide better education in large part because of the program itself is. These latter results are clearer for reading skills and are less robust for successful admission to the better secondary schools.

The most important conclusion of the chapter is that there is no group whose educational, reading comprehension or admission results are worse in the program schools. This suggests that these schools pursue education in an integrated environment in such a way that the skills development of non-Roma and non-disadvantaged students do not suffer any disadvantage.

[43] In a study mentioned earlier entitled "And it is us who give the guiding thread to integration…" (És mi adjuk az integráció vezérfonalát…) Szilvia Németh and Attila Z. Papp highlight the problem with too little focus of the realized OOIH program on the external relationships of the elementary schools.

Chapter 6

Self-Esteem and Other Non-Cognitive Skills

The returns to education on the labor market are significant in all modern economies. In post-communist Hungary, they are especially large.[44] The reasons for the large returns in Hungary are not well understood, but it is more likely due to the extremely low labor market value of people with low educational level than the high quality of higher education in Hungary.[45] It is also difficult to understand in general which skills, embedded in education, are the ones that are valued in the labor market.

As mentioned in Chapter 5, the large differences in cognitive skills are largely set before school age. It is also true that cognitive skills play a relatively small role in the returns to education.[46] In addition to providing useful knowledge, schools are responsible for the development of skills of non-cognitive nature that are valued by the labor market. Some of the most recent results in labor economics show that such elements of personality, often called as "non-cognitive" skills, may be at least as important in labor market success as cognitive skills.[47] There is a lot more disagreement on the relevant dimensions and appropriate measures of those non-cognitive skills than in the case of cognitive skills. In labor economics, the most referred non-cognitive skills include self esteem, the locus of control (the belief in one's ability to manage one's own destiny), and coping (the ability to cope with difficult situations). There is evidence that the importance of some "non-cognitive skills" have significantly increased in the past few decades because of technological change, with the replacement of routine cognitive tasks by the computer.[48] The large returns to education in Hungary suggest that the labor market value of such non-cognitive skills is at least as high in Hungary as elsewhere.

[44] See, e.g. Gábor Kertesi and Júlia Varga, "Employment and Educational Attainment in Hungary", and Gábor Kezdi, "Education and Earnings." Both in: K. Fazekas and J. Varga (eds.), The Hungarian Labour Market 2005. In Focus: Education and the Labour Market, Institute of Economics, Budapest, 2005. http://econ.core.hu/doc/mt/2005/en/infocus.pdf, as well as Flabbi, L.–Paternostro, S.–Tiongson, E. R.: Returns to education in the economic transition: a systematic assessment using comparable data. World Bank Policy Research Working Paper Series, 2007/4225 Online: http://www-wds.worldbank.org/servlet/WDSContentServer/WDSP/IB/2007/05/01/000016406_20070501150829/Rendered/PDF/wps4225.pdf.

[45] Köllő János: Workplace Literacy Requirements and Unskilled Employment in East-Central and Western Europe. Evidence from the International Adult Literacy Survey (IALS). Budapesti Munkagazdaságtani Füzetek, 2006/7. Online: http://www.econ.core.hu/doc/bwp/bwp/bwp0607.pdf

[46] Bowles, S.–Gintis, H.–Osborne, M.: The determinants of earnings: A behavioral approach. Journal of Economic Literature, 2001/39(4), 1137-1176. or Heckman, J. J.–Rubinstein, Y.: The Importance of Noncognitive Skills: Lessons from the GED Testing Program. American Economic Review, 2001/91(2), 145-149.

[47] Heckman, J. J.–Stixrud, J.–Urzua, S. S. (2006): The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. NBER Working Paper, 12006. Online: http://www.nber.org/papers/w12006).

[48] On the increase of the value of social skills see Autor, D. H.–Levy, F.–Murnane, R. J.: The Skill Content of Recent Technological Change: An Empirical Exploration. Quarterly Journal of Economics, 2003/118(4), 1279-1333. Downloadable in the form of a workshop study at: http://www.nber.org/papers/w8337.

The significance of non-cognitive skills, therefore, is comparable to the significance of cognitive skills. Contrary to the latter, non-cognitive skills develop during the school years, especially during adolescence.[49] Student autonomy and cooperative interactions may positively affect the development of self-esteem and other non-cognitive skills. They are also important elements of the OOIH program. As we demonstrated in Chapter 4, the level of autonomy is higher for students in the program schools, and they experience more cooperative interactions than their peers in the control schools. Therefore, the program may have demonstrable impact on the students' non-cognitive skills.

Measuring such skills is difficult. The first problem is the scope of dimensions to be measured; the second one is the availability of tests appropriate for students and can be administered in groups. We also faced an additional difficulty in terms of the Hungarian adaptation and standardization of such tests.

When selecting the non-cognitive characteristics, the expected impact on the later success, primarily that on the labor market, was a primary aspect. The empirical studies cited above demonstrate that positive self-esteem and locus of control (the belief in one's ability to control one's destiny) are significant contributions to success in life. These two dimensions are the primary focus of our non-cognitive test. Positive self-esteem we examine not only in general but also in some specific dimensions such as academic self esteem of physical appearance. Less systematically, we also examine coping, the ability to overcome difficult situations.

6.1 Locus of Control

"Locus of control refers to the extent to which individuals believe that they can control events that affect them. Individuals with a high internal locus of control believe that events result primarily from their own behavior and actions. Those with a high external locus of control believe that powerful others, fate, or chance primarily determine events. Those with a high internal locus of control have better control of their behavior, tend to exhibit more political behaviors, and are more likely to attempt to influence other people than those with a high external locus of control; they are more likely to assume that their efforts will be successful. They are more active in seeking information and knowledge concerning their situation. The propensity to engage in political behavior is stronger for individuals who have a high internal locus of control than for those who have a high external locus of control. One's "locus" (Latin for "place" or "location") can either be internal (meaning the person believes that they control their life) or external (meaning they believe that their environment, some higher power, or other people control their decisions and their life).[50] The concept of control place was developed by Julian B. Rotter.[51]

[49] The successes of the school programs in the area of focussing on the disadvantaged teenagers are summarized in chapter 3.2 of Carneiro, P.–Heckman, J. J. (2003): Human Capital Policy IZA Discussion Paper No. 821, www.iza.org.

[50] Wikipedia article. http://en.wikipedia.org/wiki/Locus_of_control

[51] Rotter, Julian B.: Generalized expectancies of internal versus external control of reinforcements. Psychological Monographs, 1996/80 (609).

Internal control arises as the recognition of the cause-and-effect connection between behavior and its consequences. Internal control is learnt through making free choices and the responsibility taken for them. It is through responsible behavior and its accountability that youngsters learn to seek the reasons for successes and failures in themselves. As a result, an individual who sees him/herself as being capable of performing well in various contexts (generalized self-efficiency) feels that he or she keeps the environment under control. Therefore, there is a strong connection between positive self-esteem and the feeling of internal control.[52] The two, however, can be distinguished from each other both formally and empirically.

Besides introducing the concept, Julian Rotter also developed a test consisting of 13 items. Because of time constraints, we used a shortened four-item version of the test. Locus of control is more of an "adult" concept. As a result, we did not administer the test on fourth graders but focused solely on eighth graders.

We developed the four-item test on a sample of eight-graders (outside the sample of the evaluation study) by maximizing internal consistency reliability. We also standardized the test on a national representative sample, together with the rest of the non-cognitive tests and attitude questions. In what follows, we focus on nationally standardized test results. In other words, the measure means deviation from the national average, and the unit of measurement is the national standard deviation. The differences between the program schools and the control schools are shown in table 6.1 and figure 6.1.

Table 6.1 – Internal locus of control in the 8th grade of the program and the control schools. Nationally standardized measures from a four-item Rotter-scale.

	Program schools	Control schools	Difference
All students	0.23	0.08	+0.15**
Roma students	0.25	-0.02	+0.27**
Non-Roma students	0.22	0.11	+0.12*
Disadvantaged students	0.23	0.10	+0.13+
Non-disadvantaged students	0.23	0.08	+0.15*

Standard errors are robust to heteroskedasticity.

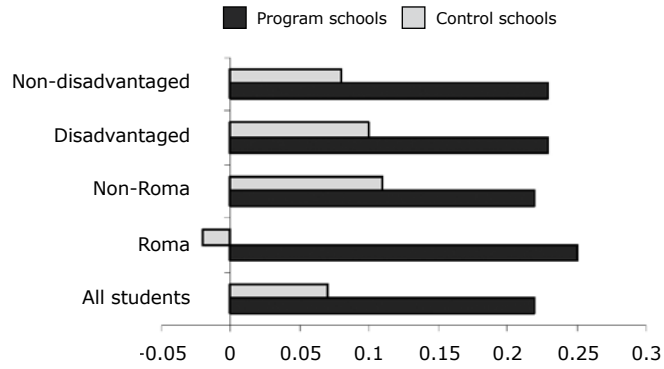
+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

[52] Fitch, G.: Effects of Self-Esteem, Perceived Performance and Choice on Causal Attributions. Journal of Personality and Social Psychology. 1970/44, 419-427.; Klein, J. D.–Keller, J. M.: Influence of Student Ability, Locus of Control, and Type of Instructional Control on Performance and Confidence. Journal of Educational Research, 1990/83(3), 140-46.

Figure 6.1 – Internal locus of control in the eighth grade of the program and the control schools. Nationally standardized measures from a four-item Rotter-scale.



The results show that students of the program have a more internal locus of control. In other words, they are more inclined to think that they themselves control their destiny than their peers in the control schools. The difference is 0.15 standard deviations, statistically significant at the one per cent level. Students in the control schools have results that are close to the national average. In statistical terms, the difference is significant, and is particularly large with respect to Roma students. In the program schools, there are not differences by ethnicity or disadvantaged status in terms of locus of control. The same is true in the control schools with respect to disadvantaged status. However, Roma students in the control schools dispose slightly more external locus of control, i.e. they are less likely to think that they themselves control their own destiny. As a result, the advantage of program schools is largest among the Roma students, at 0.27 standard deviations.

As we discussed in Chapters 2 and 5, simple differences between the students of the program and control schools do not necessarily show the program's impact. In order to get closer to the true impact of the program, we show results from regressions in which we control for students' background and measures of locus of control taken two years in advance (in grade 6). Here, unlike in the case of the cognitive measures analyzed in Chapter, we control for the respondents' social desirability as well, in order to filter out the component of the answers that are there to make a good impression (see more about social desirability in Chapter 2 or later in the next section). Similarly to the analysis in Chapter 5, we repeat the estimates to the narrow sample of program schools that were not integrating before the program (and their control pairs). Table 6.2 shows the estimates of the program versus control differences from those regressions.

Table 6.2 – Program versus control differences in internal locus of control. Regression results, controlled family background, earlier measures of locus of control, and social desirability. Dependent variable: nationally standardized measure from a four-item Rotter-scale.

	All students	Roma students	Non-roma students	Disadvantaged students	Non-disadvantaged students
Full sample	+0.13**	+0.24**	+0.10	+0.15+	+0.11+
Narrow sample	+0.10+	+0.27+	+0.07	0.00	+0.10+

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

The regression estimates in the full sample are virtually identical to the raw differences presented in table 6.1. While it is not surprising that controlling for family background does not make any difference, it is remarkable that neither social desirability nor earlier locus of control test results matter for the program versus control difference. These results suggest that the observed differences in the locus of control are caused by the schools.

Results in the narrow sample of program schools that were not integrating before the program (and their control pairs) are also similar to the full sample estimates. A notable exception is the result on disadvantaged students. The rest of the coefficients are again virtually identical to the earlier ones, although their statistical significance is weaker because of the smaller sample size. Results from the narrow sample are consistent that with the assumption that it is the OOIH program itself that makes program schools increase the locus of control of their students – with the possible exception of disadvantaged students.

6.2 Self-Esteem

In modern psychology, "self-esteem" reflects a person's overall evaluation or appraisal of his or her own worth. Self-esteem encompasses beliefs (for example, "I am competent/incompetent") and emotions (for example, triumph/despair, pride/shame)."[53] Besides emotional processes self-esteem involves cognitive and behavioral aspects.[54] The modern concept of self-esteem originated with the work of Morris Rosenberg.[55]

[53] http://en.wikipedia.org/wiki/Self_esteem

[51] See, e.g. Blascovich, J.–Tomaka, J.: The Self-Esteem Scale. In J. P. Robinson–P. R. Shaver–L. S. Wrightsman (eds.): Measures of personality and social psychological attitudes. Academic Press, New York, 1991, 115-160.

[55] Rosenberg, M.: Society and the adolescent self-image. Princeton University Press, Princeton, NJ, 1965.

While the importance of self-esteem is well established, there is some disagreement about the mechanisms through which self-esteem may develop. Rosenberg used the concept as a general feeling in connection with the value of the self. More recent research, however, looks at self-esteem as a hierarchical system, which is based on specific elements that build on each other.[56] The different approaches lead to different measures. This distinction may be important for our purpose because we look at schoolchildren, who may be in the process of developing some elements of their self-esteem but not yet others.

One of the greatest merits of the Rosenberg study is that it contains a self-esteem scale that became the widest-used test. The Rosenberg Self-Esteem Scale (hereinafter: RSE) was created for a one-dimensional measuring of the comprehensive self-esteem. One of the scale's appeals is its simplicity: it contains ten statements that formulate the opinion related oneself, and the respondent simply has to decide whether he/she agrees with the statements. The score is made up of the total of the positive answers provided to the ten questions. Many psychometric and validity tests have been carried out with the respect to the RSE than any other scale that measures self-esteem. The result of those examinations is largely supportive of the measure, implying that the test deserves to be popular and widely used.[57] On the other hand, the test was developed for adults, and we are not aware of a version for children that works similarly well.

A number of other tests have also been developed for self-esteem, among those several for children as well as for adolescents.[58] We adopted the Harter Self Perception Profile for Children (SPPC) test elaborated by Susan Harter for teenagers. SPPC is widely used for impact assessment studies of programs targeting school-age children.[59] The SPPC is a scale consisting of 36 items. Some of its items look at children's judgments related to their specific skills, while other items measure the extent to which they consider themselves to be valuable in general. The test contains six different sub-scales, five of which measure specific areas: academic competence, social acceptance, competence performed in sports, external appearance, and behavioral discipline. The sixth scale measures general self-esteem. The SPPC can thus provide a more differentiated picture than the measures focusing on a single component.[60]

Here the effect of social desirability may be especially problematic. The problem has already been discussed in the methodology part of the study (Chapter 2). The essence of the problem is social desirability can distort the answers in a positive direction if students want to make a better impression. Recall that the results shown in Chapter 2 suggest that the students of program schools are more disposed to making a good impression. Although, as we indicated,

[56] Blascovich, J.–Tomaka, J.: i.e..
[57] Gray-Little, B.–Williams, V. S. L.–Hancock, T. D.: An item response theory analysis of the Rosenberg Self-Esteem Scale. Personality and Social Psychology Bulletin, 1997/23, 443-451.
[58] Blaskovich, J.–Tomaka, J.: i. m.
[59] Harter, S.: Manual of the Self-Perception Profile for Children. University of Denver: Denver Co., 1985.
[60] Harter, S.: Developmental perspectives on the self-system. In Hetherington E. H. (ed.): Handbook of Child Psychology. Vol. 4. Socialization, personality and social development. John Wiley, New York, 1983, 275-385.

this may be part of the program's impact mechanism and a healthy phenomenon, it may have real distorting impact on the measured characteristics. Therefore, we show all results also with and without controlling for the impact of social desirability.

Recall that we control for social desirability by mixing in questions the purpose of which is to see whether the student wants to make a good impression. But there are other methods for handling the problem. One method is in formulating the questions in such a way that increases the distance from the individual. This method is applied in Harter's SPPC test, and we adopted this format as well. An example is shown in table 6.3. The answers that are positioned from the left indicate the high level of self-esteem, while those from the center to the right indicate a low level of self-esteem. (Of course, the location of positive versus negative self-esteem is reversed for about half of the questions in the complete questionnaire.)

Table 6.3 – Example of the question format in Harter's SPPC

Very true of me	Somewhat true of me			Somewhat true of me	Very true of me
☐	☐	Some children think that they do well in school.	Other children think that they don't do all that well in school.	☐	☐

The 36-item test proved to be too long for our purposes, therefore we had to develop shortened, ten-item version. In addition to the general self-esteem, the short version includes four out of the five items: external, academic, behavioral and social (sport is left out). The short version was reduced from the full test by maximizing internal consistency on a sample of 168 children. The Rosenberg test was also shortened to four items in a similar way. The tests are standardized on our national sample, and here again we use the nationally standardized test results.

Table 6.4 and figure 6.2 show the overall SPPC test scores, the general Rosenberg scores and the five detailed SPPC scores (four dimensions plus general self-esteem).

Table 6.4 – Results of the positive self-esteem test (nationally standardized)

	Grade 4			Grade 8		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
SPPC total	0.09	-0.02	+0.10+	0.25	-0.01	+0.26**
Rosenberg general	0.19	0.18	+0.01	0.03	-0.03	0.06
SPPC general	-0.04	-0.03	-0.01	0.23	0.02	+0.21**
SPPC external	0.01	-0.07	+0.08	0.11	-0.10	+0.21**
SPPC school competence	0.09	-0.02	+0.11*	0.08	-0.08	+0.16**
SPPC good behavior	0.34	0.21	+0.13*	0.21	0.08	+0.13*
SPPC social	0.01	0.01	0.00	0.06	0.11	-0.05

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Students in program schools show more positive self-esteem in virtually all dimensions. The differences are small and not significant statistically in terms of the Rosenberg scores. They are also smaller and often not significant in grade 4. In grade 4, significant differences are measured in terms of school-related items (academic and good behavior).

In grade 8, students in program school have more positive results in all SPPC dimensions except for the social items. In terms of the overall SPPC self esteem and the general and external components, results of the students in program school are 0.20 to 0.26 standard deviations better than their peers in control schools. Differences are larger than 0.1 standard deviations in terms of the school-related dimensions. Given that adolescents of age fourteen are looked at here, external self-esteem is important.

The fact that program versus control differences are larger in grade 8 than in grade 4 is very much consistent with the developmental phases of personality in general, and self esteem in particular. More person-centered educational methods, more cooperative atmosphere and higher student autonomy in program schools are more likely to affect the self-esteem of students in their early teens than below the age of ten. The failure of the Rosenberg scale to show similar differences may be due to the fact that the test was developed for adults. On the other hand, the zero difference between program and control schools may be a bit puzzling. Either this school integration program has no effect on the students' self-esteem, or the measurement is problematic. In the course of this study, we are unable to tell.

Table 6.5 contains the program versus control differences between Roma and non-Roma, as well as disadvantaged and non-disadvantaged eighth grade students in those SPPC dimensions in which statistically significant average differences were found. Figure 6.3. repeats the overall SPPC self-esteem results.

Table 6.5 – Nationally standardized SPPC test results in the eighth grade. Regression results controlled family background, earlier measures of locus of control, and social desirability

	Program schools	Control schools	Difference
SPPC overall			
Roma students	0.18	0.00	+0.18+
Non-Roma students	0.28	-0.02-	+0.30**
Disadvantaged students	0.21	-0.09-	+0.30*
Non-disadvantaged students	0.27	0.04	+0.23**
SPPC general			
Roma students	0.21	0.10	+0.11
Non-Roma students	0.23	0.01	+0.22**
Disadvantaged students	0.26	0.00	+0.26**
Non-disadvantaged students	0.21	0.04	+0.17*
SPPC external			
Roma students	0.09	-0.05-	+0.14
Non-Roma students	0.10	-0.13-	+0.23**
Disadvantaged students	0.16	-0.14-	+0.30*
Non-disadvantaged students	0.09	0.09	+0.18**
SPPC academic			
Roma students	-0.15-	-0.29-	+0.15
Non-Roma students	0.18	-0.02-	+0.20**
Disadvantaged students	-0.32-	-0.30-	-0.02
Non-disadvantaged students	0.21	0.00	+0.21**
SPPC behavioral			
Roma students	0.14	0.09	+0.05
Non-Roma students	0.26	0.10	+0.16**
Disadvantaged students	0.15	0.01	+0.14
Non-disadvantaged students	0.23	0.13	+0.10+

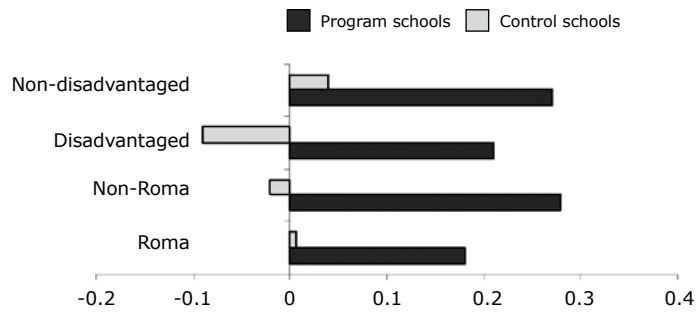
Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Figure 6.3 – Total positive self-esteem in the eighth grade, standardized test results



With few exceptions, the self-esteem of the students in the program schools are always more positive, whether Roma or non-Roma, disadvantaged or non-disadvantaged. The programs versus control differences are larger among non-Roma and non-disadvantaged students in general.

Again, simple program versus control differences do not necessarily show the impact of the program. In principle, the difference may arise from differential student composition, the fact that students of the program schools are more keen on making a good impression. It is also possible that it is the program schools themselves that lead to an increase in their students' self esteem, but this would have happened in the absence of the program as well. Table 6.6 shows the results of regressions that we think can at least partially control for those problems. We show results from regressions that are essentially the same as before with respect to the locus of control. In these regressions we control for students' background and measures of self-esteem taken two years previously, as well as the respondents' social desirability. We repeat the estimates to the narrow sample of program schools that were not integrating before the program (and their control pairs).

Table 6.6 – Program versus control differences in positive self-esteem. Regression results controlled family background, earlier measures of locus of control, and social desirability. Dependent variable: nationally standardized measure of the SPPC total score.

	All students	Roma students	Non-roma students	Disadvantaged students	Non-disadvantaged students
Full sample	+0.24**	+0.28**	+0.24**	+0.38**	+0.20**
Narrow sample	+0.33**	+0.29+	+0.35**	+0.25+	+0.32**

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

The regression estimates are in all cases around the raw differences, and a few times they are larger. In the narrow sample of program schools that were not integrating before the program (and their control pairs), the estimates are even larger. Considering the role of positive self-esteem in the success achievable in the labor market and other areas of life, and that the development of self-esteem can be more expected from an elementary school program than the promotion of cognitive skills, this result is very important.

The results in table 6.6 strongly support the interpretation that the OOIH program has a significant positive effect on students' self-esteem. There is no robust pattern among the different student groups. It seems therefore that with respect to self-esteem, all students benefited from the program in a roughly equal way.

6.3 Coping

The internal control place and the positive self-esteem are the most important non-cognitive skills in this study. In a less systematic way, though, we also looked at the program's impact on the ability to cope with stressful and conflict situations.

The ability which is called simply "coping" in the psychology literature denotes the ability that helps individuals to get over with difficult or conflict situations. Good coping ability enables people to go on without any significant negative consequence even if the root cause of the problem remains.[61] The coping ability may be extremely valuable both at school and workplace, and it can prevent resorting to deviant and self-destructive behavior.

We included four questions connected with the coping ability in the spirit and format of the SPPC test (see above). The questions are formulated in accordance with the coping literature, but this is not an adaptation of an existing tests. (An example: Some children are in distress for a long time if hurt/Other children easily forget insults.) This coping test was standardized together with other tests. The numbers shown below measure the deviation from the national average, and the measurement unit is the national standard deviation.

Table 6.7 and figure 6.4 demonstrate the results of the coping test in the eighth grade on average, as well as in the subsamples by ethnicity and disadvantaged situation. (No statistically significant difference between program and control schools is detectable in grade 4)

[61] Folkman, S.: Personal control and stress and coping processes: A theoretical analysis. Journal of Personal and Social Psychology, 1984/46, 839-852.

Table 6.7 – Nationally standardized results of the coping test in the eighth grade

	Program schools	Control schools	Difference
All students	0.45	0.32	+0.13*
Roma students	0.16	-0.14-	+0.30**
Non-Roma students	0.62	0.51	+0.11
Disadvantaged students	0.19	-0.01-	+0.20
Non-disadvantaged students	0.56	0.44	+0.12

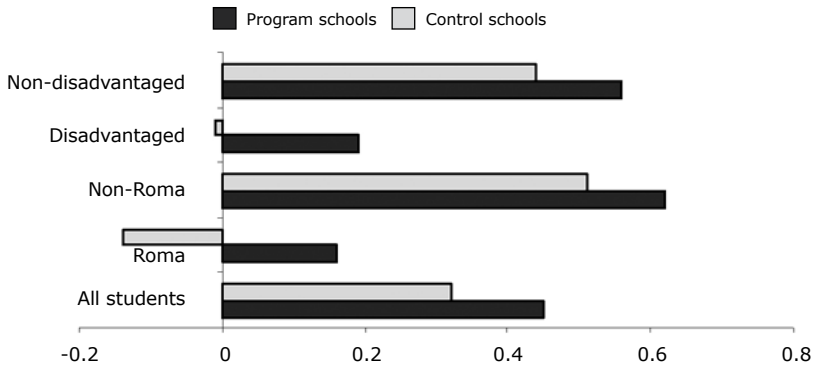
Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Figure 6.4 – Coping in the eighth grade



According to the results, the Roma and the disadvantaged students are significantly worse at coping. Somewhat surprisingly, students in all groups, except for the Roma, both in the control and the program schools, show better coping results than the national average. These differences may have to do with the deviation of the composition of student in our sample from national averages.

More importantly for our purposes, the results show that the students of the program schools have better coping abilities than their peers in the control schools. The difference is particularly significant among the Roma students. According to the results, the Roma students of the program schools are able to cope with stressful situations at a level higher than the national average, while the Roma students of control schools are below the national average with respect to coping abilities.

Table 6.8. shows the usual robustness checks: the regression estimates after controlling for family background, social desirability and the coping test results taken two years earlier, in grade six. Similarly to earlier analyses, we estimated the program versus control differences by regressions that control for family background, social desirability (good impression), and the coping results recorded two years before. We re-estimated the same regressions in the narrow subsample of progam schools that were not integrating before the program and their control pairs. The results are contained in table 6.8.

Table 6.8 – Program versus control differences in coping. Regression results controlled family background, earlier measures of locus of control, and social desirability. Dependent variable: nationally standardized measure from a four-item coping scale.

	ALL STUDENTS	ROMA STUDENTS	NON-ROMA STUDENTS	DISADVANTAGED STUDENTS	NON-DISADVANTAGED STUDENTS
Full sample	+0.08	+0.31*	+0.03	+0.13	+0.17
Narrow sample	+0.16+	+0.41*	+0.08	+0.12	+0.13

Standard errors are robust to heteroskedasticity.

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Controlling for social desirability does not reduce the program versus control differences among the Roma students, but does so among the non-Roma students. This suggests that the program schools help their Roma students better to cope with difficult situations than the control schools. The same may not be true for non-Roma students. Importantly, no student groups sees its coping abilities diminished in program schools relative to control schools. The results in the narrow sample are very similar to the results in the full sample. This suggests that it is the OOIH program itself that makes program schools better prepare their students for coping.

The main results of this chapter can be summarized in the following way. Students of the program schools have better non-cognitive skills in all dimensions analyzed here than students of the control schools. Program school students have more internal locus of control (they are more likely to believe that they themselves are responsible for their successes or failures), they have more positive self-esteem (especially in terms of general, external, and school-related items of self-esteem). Apparently, all these results reflect the causal effect of the OOIH program itself. Some of the effects are larger for Roma, others for non-Roma students, but again, they are positive for all student groups analyzed.

Contrary to the common belief, the characteristics examined here are not secondary to cognitive skills and competences. Their labor-market value is very similar, and, in a number of other areas of life, they are indispensible for success or in order to avoid serious failures. We also know that, as opposed to cognitive skills, they can be significantly developed also at the elementary school age. Therefore, it is an important result that the OOIH program had a positive effect on their development both with respect to the Roma and non-Roma, and disadvantaged and non-disadvantaged students.

Inter-Ethnic Relations

One of the most important goals of integrated education of different ethnic groups is in reducing stereotypes, prejudice and social distance. In a classic book, Allport[62] argued that reduced physical distance, in other words, more "contact" between otherwise hostile groups, can reduce social distance and prejudice under certain conditions. The conditions include equal status of the two groups, inter-group cooperation, common goals, equality of power, and some law or practice that emphasizes equality and supports cooperation. A large literature emerged since with the aim of analyzing the correlation between inter-group contact and prejudices.[63] Despite the nearly half a century of interest among researchers, no agreement has been reached in the subject. Some argue that inter-group contacts result in a significant decrease in prejudice,[64] while others argue for a weak impact at most.[65]

Often-examined forms of distance keeping and conflict are stereotypes, prejudice and discriminative behavior itself. Stereotypes comprise the cognitive component of the attitude towards a group. Prejudice is usually viewed as the affective or emotional aspects of inter-group contact. Measuring behavior is the third direction of the research. Discriminative behavior is usually viewed as the result of stereotypes and prejudice in conflict situations. While the three aspects originate from a common theoretical stem, research fields on the different aspects are often isolated from each other.

In this evaluation study, we analyze all three dimensions. Integrated education increases inter-group contact, and some elements of the OOIH program can be understood as instruments for strengthening the conditions of contact to reduce social distance as conceptualized by Allport. Besides measures on stereotypes and prejudices, we examine social anxiety and social dominance attitudes that are apparently connected with the development of prejudices (see the latter later).

[62] Allport, G. W.: The nature of prejudice. Perseus Books Cambridge, MA, 1954/1979. In Hungarian: Allport, G. W.: Az előítélet. Translated by: György Csepeli. Sociological Library-series, Gondolat Publisher, Budapest, 1977 and Osiris Publisher, Budapest, 1999.

[63] See, for instance, the meta-analysis that summarizes the results of the literature: Pettigrew, T. F.–Tropp, L. R.: A meta-analytic test of inter-group contact theory. Journal of Personality and Social Psychology, 2006/90, No. 5, 751-783.

[64] Jackson, J. W.: Contact theory of inter-group hostility: A review and evaluation of the theoretical and empirical literature. International Journal of Group Tensions, 1993/23, 43-65.; Pettigrew, T. F.: Racially separate or together? McGraw–Hill, New York, 1971.

[65] Amir, Y.: The role of inter-group contact in change of prejudice and ethnic relations. In P. A. Katz (ed.): Toward the elimination of racism. Pergamon, Elmsford, NY, 1976, 245-308.; Ford, W. S.: Favorable inter-group contact may not reduce prejudice: Inconclusive journal evidence, 1960-1984. Sociology and Social Research, 1986/70, 256-258.; Rothbart, M.–John, O. P.: Social categorization and behavioral episodes: A cognitive analysis of the effects of inter-group contact. Journal of Social Issues, 1985/41, 81-104.

Inter-group contact may have a different impact on both the cognitive and emotional components of inter-group attitudes. Pettigrew, in his research on inter-group friendships in Western Europe, found that friendship has a stronger impact on the affective components of inter-group attitudes (for instance, sympathy and admiration for an external group, or the manifested social distance), with smaller effects on the cognitive components (e.g. the acceptance of support programs that target minority groups).[66]

In order to learn about inter-ethnic relations, we need to know the ethnic background of the respondents. As we described above in Section 2, ethnicity was measured separately from the other measures, due to Hungarian data protection regulations. Therefore, all questions related to stereotypes, prejudice and social distance were asked as opinions about both the Roma and non-Roma ethnic group. The respondents' ethnicity was merged to these data afterwards, allowing for identifying opinions about the other ethnic group.[67]

7.1 Stereotypes

A stereotype is usually described as a simplistic, exaggerated and overgeneralized judgment made about the members of a social group. Measuring stereotypes is relatively straightforward: researchers simply ask the respondents about characteristics of a social group in question, or sometimes other people's opinion on those characteristics.

Instead of categorical rejection or agreement, some up-to-date measuring methods use bipolar classification of characteristics which express relative judgments. This more sophisticated method is warranted in part because expressing ethnic stereotypes has become politically less correct. The semantic difference scales that are typically applied present characteristics to the respondents. Respondents then have to make judgments about people belonging to various ethnic groups on a scale where the minimal value means the characteristics displayed on the left side, and the maximum values are on the right side.

Since we had no access to properly validated Hungarian measures of school-age children's ethnic stereotypes, we developed our own measures. Taking tests for adults as a model, our questionnaire included a series of opposite statements. In each case the respondent had to decide which statement of the opposing pair of statements described the members of a given group in the best way. First we held a focus group discussion with 7th and 8th grade students, and then we developed a longer questionnaire, which was filled out by

[66] Pettigrew, T. F.: Generalizing Inter-group Contact Effects on Prejudice. Personality and Social Psychology Bulletin 1997/23:173-85, Pettigrew, T. F.: The Affective Component of Prejudice: Empirical Support for the New View. In: S. Tuch–J. Martin (eds.): Racial Attitudes in the 1990s: Continuity and Change. Westport, CT: Praeger, 1997. 76-90.

[67] Focus groups with students and the trial survey situations showed that the politically correct „non-Roma" label often confused the students, while the exclusionist „Hungarians" label caused neither comprehension nor emotional problems. By adulthood, the situation evidently changes: in the parental statements on ethnic identity (see chapter 2) a possibility was provided to choose a double identity which was chosen by a large number of parents.

a few dozen students. From the long questionnaire we developed a short version with five questions by maximizing internal validity. The series of questions was formulated separately for Roma and non-Roma ("Hungarians", see footnote 65). Table 7.1 shows an example of the posed questions.

Table 7.1 – An example for the semantic difference scale that measures ethnic stereotypes. The question is the extent to which the respondent thinks the statements about a given group groups is true.

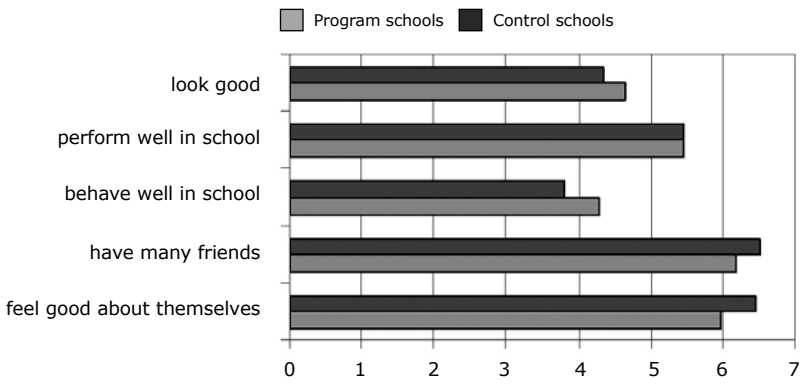
Have all the abilities to perform well at school	1	2	3	4	5	6	7	8	9	10	Do not have all the abilities to perform well at school

The dimensions for the stereotype questionnaire were selected with an eye on students' self-esteem. Self-esteem in our questionnaire was measured using the Harter SPPC four dimensions (plus the general self-esteem); we mapped these with respect to Hungary's various ethnic groups in our stereotype-related questions.

Figure 7.1 shows average stereotype scores, with respect to the other ethnic group, in the five dimensions for program schools and control schools. Table 7.2 shows the overall test results.

Figure 7.1 – Ethnic stereotypes among eight grade students in base and control schools:

Panel A. Roma people (as viewed by non-Roma students)



Panel B. Non-Roma people (as viewed by Roma students)

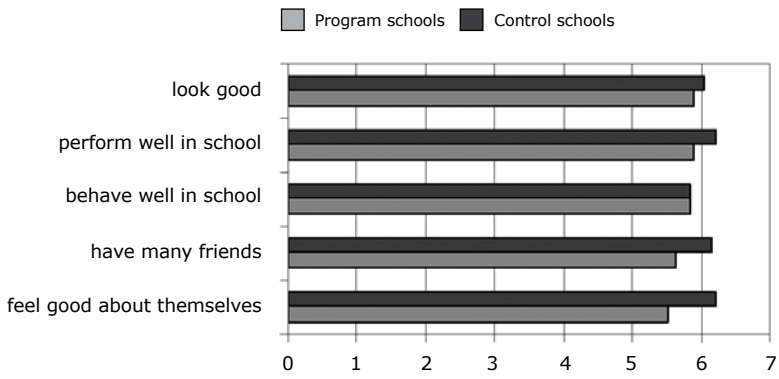


Table 7.2 – Program vs. control differences in stereotypes related to the other ethnic group, standardized differences. Overall score created from the three more objective stereotype measures.

	Stereotypes about non-Roma by Roma students	Stereotypes about Roma by non-Roma students
Entire sample, raw difference	+0.10	-0.14*
Entire sample, controlled for social desirability	+0.11	-0.16*
Narrow sample controlled for social desirability	+0.11	-0.39**

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Striking differences are observed between the two groups of questions. In case of general self-esteem (feeling good about oneself) and sociability (having many friends), the image of each other is very similar. In other dimensions ("they behave well", "they look well", "they perform well at school") Roma think much better of non-Roma than the other way round. These latter dimensions focus on the more objectively defined characteristics (what they are like and not what they think they or their peers are like). The overall measure we constructed using these latter three items.

Stereotypes about non-Roma by Roma students are somewhat stronger in program schools than in control schools. A larger and statistically more significant difference is observed the other way round. Stereotypes about Roma by non-Roma students are weaker in program schools than in control schools.

Controlling for social desirability does not change the program vs. control differences. The measured difference is significantly stronger when the comparison is carried out among program schools that did not integrate before the program (and their control pairs). This finding is consistent with the notion that stereotypes against the Roma were reduced by the program itself.

7.2 Prejudice and Social Distance

Prejudice is defined as the affective aspects of inter-group relations. Measures of ethnic prejudices try to map a large range of emotions related to a given ethnic group. The first prejudice questionnaires in the literature used questions that focused on direct rejection of minorities. In the past decades, measures of prejudiced changed considerably. Questionnaires in use nowadays are less aggressive, and many touch upon fine emotional components as well as or instead of more direct rejection.

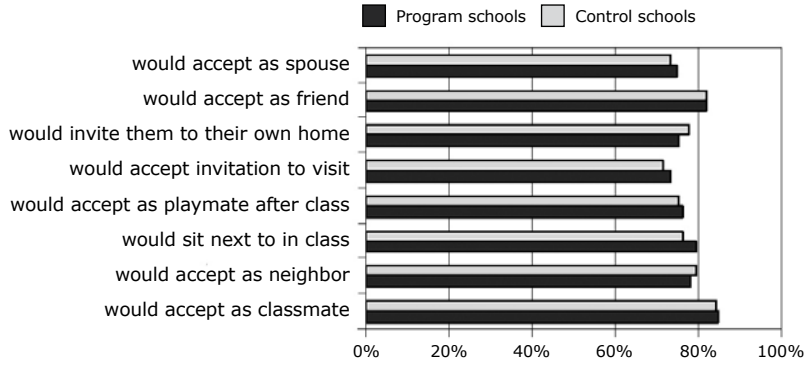
We experimented with several tests. The most conclusive results were received from the Bogardus social distance scale. The scale was developed eighty years ago and has remained a standard tool in the research of inter-ethnic relations ever since.[68] The scale measures the social distance that the individual wishes to keep from the members of a given social group, at various levels of intimacy. Typical questions touch upon marriage, relationships with colleagues and neighbors, or, in case of immigrants, rights to citizenship. In the usual form, answers are binary (yes or no). As a result, a total score of 0 means that the individual does not wish to establish any relationship with the members of the given social group, while higher scores indicate higher level of acceptance and more favorable attitudes towards the group in question. However widespread the use of the scale, there is no standard form of the questionnaire.

In our study we used a scale with eight dimensions. We selected the questions based on the experience of the focus group conversations with eighth grade students. Figure 7.2 shows the proportion of students who provided yes answers to specific questions, that is, those who said yes to a specific contact. The figure shows results separately for Roma and non-Roma students. For each ethnic group, answers referring to the other ethnic group are shown.

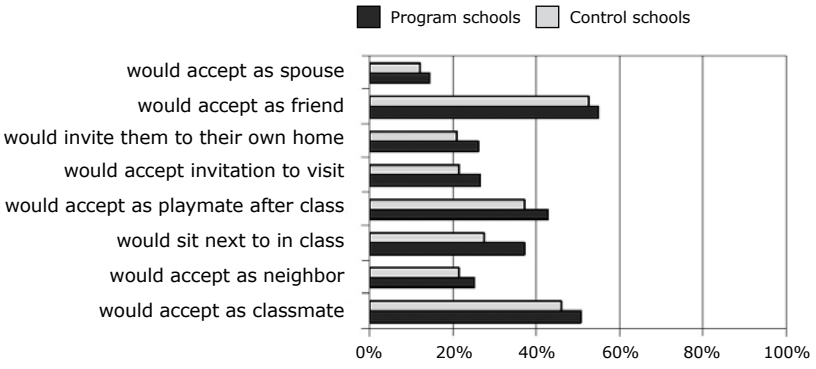
[68] Bogardus, E.: Immigration and Race Attitudes. Heath, Boston, 1928.

Figure 7.2 – Social distance from members of the other ethnic group among eighth grade students in program and control schools. Percentage of students that said yes.

Panel A. How much do the Roma accept the non-Roma?



Panel B. How much do the non-Roma accept the Roma?



The vast majority of Roma students would accept practically any relationship with the non-Roma. There is no significant difference between program schools and control schools. Non-Roma students are significantly less accepting their Roma peers. Half of them are open to friendship and classmate relationship, while less than ten per cent are open to marriage. Although the differences are small, non-Roma students in program schools are more open to relationships to their Roma peers in all dimensions than non-Roma students in control schools.

The program vs. control differences in the cumulative scale are presented in table 7.3, in the usual way (raw differences, controlled for social desirability, and the narrow subsample of

program schools that were not integrating before the program). The unit of measurement used here is not a standardized value. Instead, we present the average value of the 0-1 score across questions.

Table 7.3 – Program vs. control differences in the social proximity against the other ethnic group (differences in average score across questions; result is measured on [0,1] interval)

	Acceptance of non-Roma by Roma students	Acceptance of Roma by non-Roma students
Entire sample, raw difference	+0.02	+0.05**
Entire sample, controlled for social desirability	+0.02	+0.05**
Narrow sample controlled for social desirability	-0.02	+0.12**

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

The first line of table 7.3 repeats the information we saw on Figure 7.2 in a compressed way. Roma students in the program schools and the control schools show very similar levels of acceptance towards their non-Roma peers. Controlling for social desirability or narrowing the sample has no effect on this zero difference. To the contrary, non-Roma students of program schools are more accepting towards their Roma peers than non-Roma students in control schools. The difference is not large (raw difference is 0.05) , but it is not negligible given the low levels of non-Roma answers (overall control school average is 0.30). Moreover, controlling for social desirability does not affect the program-control difference, and narrowing the sample to program schools that were not integrating before the program (and their control pairs) results in larger difference (0.12). Similarly to the results on stereotypes, this last finding is consistent with the notion that prejudice against the Roma is lower in program schools because of the program.

7.3 Social Dominance Orientation

The reasons for stereotypes and prejudice are difficult but not impossible to uncover. Social dominance orientation (SDO) is often thought as an important factor. SDO summarizes is the extent to which individuals accept or believe in social hierarchy and ethnic inequality.[69] A strong SDO means that the individual thinks that some groups are more valuable than others, the hierarchy among social groups is unavoidable and even desirable, and social dominance is a necessary fact of life.

In this sense, therefore, SDO can be conceived as the individual's predisposition for certain ethnic attitudes. Should a suitable opportunity arise, an individual with strong SDO is more likely to develop a negative attitude towards a group of low status. These attitudes are closely related to the belief that the world is based primarily on competition, and that violence is sometimes necessary to make the subordinate groups stay where the individual thinks they should. SDO is also connected to explicit racial stereotypes.

Several versions of measuring the SDO are in use. Since we did not find any measures developed for children, nor any empirical analysis on children, we developed the test from adapted adult tests. The short version of the questionnaire is a six-item scale. The questions included the following: Do you think some people who are more valuable than others, and, Do you think winning is more important than the way we play games? We standardized our SDO test on our national sample as we described in the previous chapters.

The standardized SDO test results are shown in table 7.4.

Table 7.4 – Social dominance orientation, nationally standardized test results

	Program schools	Control schools	Difference
All students	-0.04	0.09	-0.15**
Roma students	-0.00	0.20	-0.20*
Non-Roma students	-0.07	0.02	-0.09+
Disadvantaged students	-0.01	0.14	-0.15+
Non-disadvantaged students	-0.06	0.05	-0.11*

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

[69] Sidanius, J.–F. Pratto: *Social Dominance: An Inter-group Theory of Social Hierarchy and Oppression.* New York, Cambridge University Press, 1999.

The results show that control school students are characterized by higher level of SDO than the national average (likely because the national sample consists of higher status students), while program school students are characterized by lower level. The difference is of the same sign and similar magnitude for Roma and non-Roma students as well as disadvantaged and non-disadvantaged students. It is somewhat larger for Roma and disadvantaged students. Table 7.5 shows the program vs. control differences controlled for social desirability for the entire sample and for the narrowed sample of program schools that were not integrating before the program (and their control pairs).

Table 7.5 – Program vs. control differences in social dominance orientation (nationally standardized results controlled for social desirability)

	All students	Roma	Non-roma	Disadvantaged	Non-disadvantaged
Entire sample	-0.07	-0.16+	-0.03	-0.07	-0.06
Narrow sample	-0.13+	-0.15	-0.15+	+0.01	-0.17*

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

Controlling for social desirability significantly weakens the program versus control differences, while narrowing the sample strengthened those again. The first phenomenon suggests that the students of program schools are more likely to think that social dominance is not a good thing, therefore, they have to suppress such ideas if they want to make a good impression. One interpretation of the finding is that the raw differences presented in table 7.4 may show "real" differences, and social desirability may be part of the mechanism behind the difference. Another interpretation is of course that social desirability leads to a bias, in which case it is table 7.5 that shows the "real" differences. Table 7.5 shows that the programs vs. control differences are more significant in the schools that were not operating in the spirit of integration before the program. Similarly to what we have seen in terms of stereotypes and social distance, the latter is consistent with the causal role of the program in reducing social dominance orientation.

7.4 Social Anxiety

Social anxiety is an important indicator of an individual's behavior in social (interpersonal) relationships, and, as such, it is important in the development of children's later career.[70] Social anxiety may be an important mediation in inter-group conflicts, and it may be interesting in itself as well.

Besides the fear of aliens or outsiders, little research has been done in the subject of children's social anxiety. The important exception is the social anxiety test[71] developed by La Greca and colleagues. We started with our Hungarian adaptation of the original 10-item test, and we created a short, a 5-item version containing the three main dimensions identified by us (fear of a negative judgment, fear of interaction, and retreat). Similarly to earlier personality and attitude tests, social anxiety is better measured for upper-year students. We standardized this one on the nationally representative sample as well.

The program and control school results and their differences are shown in Figure 7.3 and table 7.6.

Table 7.6 – Social anxiety (nationally standardized test results)

	4 grade			8 grade		
	Program	Control	Difference	Program	Control	Difference
All students	0.03	0.11	-0.08*	-0.13	0.07	-0.20**
Roma students	0.03	0.10	-0.07	-0.01	0.11	-0.10
Non-Roma students	0.05	0.13	-0.08	-0.17	0.03	-0.20**
Disadvantaged students	0.11	0.21	-0.10	-0.05	0.21	-0.26*
Non-disadvantaged students	-0.04	0.03	-0.17	-0.18	0.00	-0.18**

+ Statistically significant at the 10% level.

*** Statistically significant at the 5% level.**

**** Statistically significant at the 1% level.**

[70] Leary, M. R.: Social Anxiety, Shyness and Related Constructs. In Measures of Personality and Social Psychological Attitudes. Academic Press, 1991.

[71] La Greca, A. M.–Dandes, S. K.–Wick, P.–Shaw, K.–Stone, W. L.: Development of the Social Anxiety Scale for Children: Reliability and Concurrent Validity, Abstract. Journal of Clinical Child Psychology, 1988/17, No. 1, 84-91.

Anxiety is found to be lower in program schools than in the control schools, in all of the student groups. Quite naturally, the differences are considerably smaller and less significant in fourth grade than in eighth grade. The difference is largest among disadvantaged students in the eighth-grade, followed by non-Roma students. Since social anxiety of disadvantaged students is stronger in general, the results show that not only is social anxiety lower in the program schools, but the difference between groups is also smaller there.

Similarly to our earlier analyzes, we re-estimated the program versus control differences after controlling for social desirability and in the narrow sample of program schools that were not integrating before the program 9and their control pairs). The results are shown in table 7.7.

Table 7.7 – Program vs. control differences in social anxiety (nationally standardized results controlling against making a good impression)

		All students	Roma students	Non-roma students	Disadvantaged students	Non-disadvantaged students
4 grade	Entire sample	-0.09+	-0.13	-0.08	-0.06	-0.14
	Narrowed sample	-0.12+	-0.13	-0.14	-0.18*	-0.06
8 grade	Entire sample	-0.20**	-0.11	-0.18**	-0.15	-0.16**
	Narrowed sample	-0.25**	-0.12	-0.29**	-0.25+	-0.26**

+ **Statistically significant at the 10% level.**

* **Statistically significant at the 5% level.**

** **Statistically significant at the 1% level.**

Social desirability (in other words, the desirability or make a good impression) does not affect the program versus control differences. Similarly to our earlier findings, we see that the difference between program schools that were not integrating before the program and their control pairs is more powerful. This, again, suggests that lower social anxiety of the students of program schools decreased as a result of the program's impact.

Summarizing the results of the chapter, the following conclusions can be drawn. Non-Roma students of program schools see the Roma in a less stereotyped way, they keep a smaller social distance from them, and they think less in terms of social hierarchy. Roma students have a much more positive image of the non-Roma in general, and they are much more accepting towards them than the other way around. In that respect, there are no significant differences between program and control schools. Finally, students of program schools are characterized by lower levels of social dominance orientation and social anxiety.

Chapter 8

What Explains the Impact of the Program?

In Chapter 4 we demonstrated that the classes observed in the program schools are different from those in the control schools in many dimensions. In Chapters 5-7 we showed that the students of the program schools performed better in many dimensions. In this chapter we will examine the relationship between changes in educational methods, classroom management and student behavior on the one hand, and student outcomes on the other hand. Our main question is whether there are some elements in the former that are especially important in the latter. As we shall see, this question is very hard to answer. Before turning to the methodological issues, let's summarize the main findings of the previous four chapters.

One difference we uncovered is the larger prevalence of student-centered education elements in the program schools. A related difference is that a larger fraction of the observed activities are motivating, playful, manual or creative. The most spectacular difference is in terms of work forms. In the program schools, nearly half of the observed activities are group work, whereas the same fraction in control schools is below ten per cent. The proportion of individual work is largely similar, but the dominant work form in the control schools is frontal teaching. We also demonstrated that the teachers of program schools not only apply group work more frequently, but they do that more in accordance with the principles of cooperative education. Differentiated education is rare in the program schools, but still twice as frequent as in the control schools. Besides educational methods, we identified differences in terms of student behavior. The level of cooperation and collaboration among students is higher in program schools, perhaps no surprisingly after having seen such a high prevalence of cooperative group work. Less directly attributable to teaching methods is the fact that medium or high level of student autonomy is more frequent in program schools than in control schools.

We looked at three kinds of student outcomes: cognitive or academic outcomes, "non-cognitive skills", and the social distance kept from members of the other ethnic group. Our results imply that students in the program schools are achieve more positive outcomes in all dimensions. They are better at reading skills, they achieve better grades, and they are more likely to be admitted to better secondary schools. Program school students are also characterized by more positive self-esteem, a better attitude towards their role in their own successes and failure, and a better ability of coping with difficult situations. They are characterized by lower levels of social anxiety and are less likely to think that social dominance and hierarchy are good things. Finally, we find that non-Roma students of the program schools keep smaller social distance from their Roma peers than in the control schools. The differences are sometimes modest (e.g. in cognitive skills), sometimes moderate (e.g. in self-esteem), but they are positive (if statistically significant) in all dimensions and for Roma and non-Roma as well as disadvantaged and non-disadvantaged students.

8.1 Methodological Issues

Impact assessment studies can ask two importat goals: estimating the impact of the program, and if there is one, the mechanisms through which the program exercises that impact. In our impact assessment study, our first question is whether students in the program schools achieve better results, and whether they do so because of the program itself. The second question concerns the elements in the program that are responsible for those better results.

As we discussed in Chapter 2 in detail, we cannot give a completely reassuring answer to the first question because of the non-experimental nature of the program. It is possible that the students of the schools that took part in the program would have achieved better results than the students of the control school without the program. Due to the proper, and fortunate, selection of the control group, the student composition of the control schools is virtually identical to the composition of the program schools in all observable characteristics. We thus have no reason to suppose that the better results of the program school students are due to differences in their family background. The better results therefore reflect value added by the program schools. At the same time, we cannot exclude the possibility that the program schools themselves would have been better without of the program as well. Recall that, in order to gain some additional understanding, we examined the program versus control school differences in the restricted sample of program schools that were not integrating before the program, and their control pairs. In most dimensions the results suggest that the observed program versus control school differences reflect the impact of the program.

If all this is true, it is of particular importance to understand what factors lead to the impacts of the program. It turns out however that answering this question is even more difficult.

One reason is that the program itself is complex, with many elements. In order to uncover the mechanisms, the individual elements need to be separated. But that is practically impossible since approximately 100 classes of the 30 program schools took part in the research, and the program elements have many more possible combinations. Consequently, it is impossible to find a sufficient number of examples for comparisons in which the schools differ from each other only with respect to one program element, but are similar with respect to all others. As a consequence, even if we found that the use of some elements in some program schools is associated with better results, we could not keep other elements fixed. In fact, the use of some elements is apparently associated with the use of other elements. For instance, we will show that the program schools that applied group work added more to the self-esteem of Roma and disadvantaged students than the program schools that did not apply group work. However, if the teachers who apply group work usually have a more student-centered approach, it may easily be the case that what we see is not the effect of group work, but of student-centered pedagogy.

The impacts cannot be separated because of the complexity of the prog-ram and the size of the sample.[72]

Another fundamental measurement problem is due to the fact that schools and schoolteachers are practically free to choose from the elements. This may lead to a reverse causality. The teacher may use some elements due to the students' previous achievements. To the extent that we cannot fully control for those previous achievements but those achievements have a long lasting effect, current achievements may be in part the cause rather than the effect of the observed program element. We tried to circumvent this reverse causality by measuring the program elements, through classroom observations, one year before measuring the students' outcomes. However, since skills and characteristics change slowly (later school achievements are strongly correlated with earlier achievements), the distorting effect may be there anyway even if we control for earlier test results. The problem is all the more severe because we cannot sign the bias: the measurement problems may overstate the estimated effects of some program elements and understate the effects of others.

In spite of the severe problems, we show some estimates that may be informative of the impact mechanisms in this chapter. In particular, we look at whether the program schools that apply certain educational methods achieve better results than the program schools that use the given methods to a lesser extent. The better results are always measured in the students' test results compared to the test results of students of the matched control school pair.[73]

We look at program elements one by one because we cannot analyze them together for the reasons described above. Besides educational methods, we examine the impact of intermediary variables such as student autonomy and the level of interactions. In each case, we estimate the results for all students and also separately for Roma and non-Roma students as well as disadvantaged and non-disadvantaged students.

[72] From the technical point of view, the problem is that we cannot estimate regressions with all measured prog-ram elements on the right side, because there are more explanatory variables than schools (degree of freedom problems). Moreover, inlcuding just two such right-hand side variables would result in very imprecise estimates, because the explanatory variables are strongly correlated (multicollinearity).

[73] In its simplest form, this is a so-called difference-in-differences measurement strategy. We compare the program versus control school difference for pairs where the program school applies the given element to the program versus control difference for pairs where the program schools do not use the element. Since in most cases the use of the program elements is measured in a non-binary scale, the implementation was somewhat more complicated. Formally, in each program-control pair, we standardized the students' test results using the control school average (still measured in the national standard deviation). This way the average results of the program schools are expressed in terms of their advantage to their own control pair. We then looked at the correlation of the standard-ized outcomes and the prevalence of the observed program element, at the level of the classroom observation. Apart from all the students, the comparison was also made separately for Roma and non-Roma students, as well as disadvantaged and non-disadvantaged students. In those cases the stan-dardization was carried out with respect to the comparable students of the control schools.

Students' school achievement was examined in the following dimensions:

- Reading skills (fourth and eighth grades).
- Admission rate to better secondary schools (providing graduating examination called "érettségi" in Hungarian).
- Cumulative SPPC positive self-esteem score (fourth and eighth grades).
- Coping score (eighth grade).
- Social distance from members of the other ethnic group (eighth grade).

Note that social distance will be presented in this chapter in a standardized way (standardized by sample distribution). The higher values of the indicator indicate a larger distance, thus the negative difference shows the better results representing smaller distance.

The following were the criteria based on which the program schools were compared:

- Student-centered pedagogy.
- Work forms.
- Differentiation.
- Student autonomy.
- Cooperation among students.

8.2 Student-centered Education

In this Chapter we measure student-centered elements on a three-grade scale that is derived from five elements. The five elements are the following: there is a chat corner set up in the classroom; the seating order is flexible; there is warm-up activity; there a closing activity; and, the teacher makes personal contact with the students. We set the student-centered variable as weak if zero or one such element was observed, moderate if two or three, and strong if four or five. Of course this is a very crude index. We experimented with several alternatives, and all gave results similar to those presented here.

Table 8.1 presents the distribution of the student-centered index in the lower and the upper grades, separately in program schools and in control schools.

Table 8.1 – Distribution of monitored periods by student-centered elements (%)

	Grade 3			Grade 7		
	Program schools	Control schools	Difference	Program schools	Control schools	Difference
Weakly student-centered	29	55	-26	55	79	-24
Moderately student-centered	36	29	+7	30	14	+16
Strongly student-centered	35	16	+19	15	7	+8
Total	100	100		100	100	

Similarly to what we saw in Chapter 3, the table clearly demonstrates that the education in the program schools is more student-centered, in particular in the lower grades.

Figure 8.1 shows the results of the students of program schools compared to the results of the students of control schools depending on how many student-centered elements were found on the monitored program schools.

Figure 8.1 – Program versus control differences by the level of student-centered education in the program schools

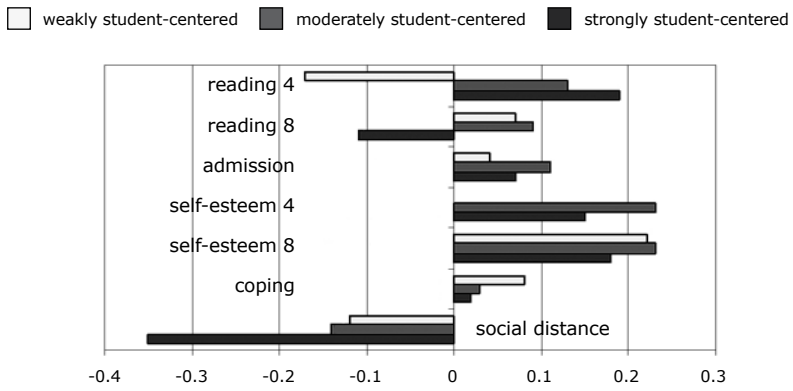


Figure 8.1 shows that the advantage of program schools in terms of reading skills is larger the more student-centered the education in lower grades but not in upper grades. Their

advantage in terms of admission rate does not show any significant pattern. Neither does their advantage in self-esteem in upper grades. The self-esteem results in lower grades provide some support for student-centered education, at least from weak to moderate levels. The only clear case is social distance. There the advantage of program schools is clearly increasing with the level of student-centered education (i.e. the distance is decreasing).

The value of student-centered education in the upper grades was rarely high, even in program schools, that is why the values connected to it should be treated with caution. Still, the results suggest that the impact of student-centered pedagogy was better in the lower grades, at least in the way it was realized in the program schools. However, it is important to note that, with only one exception (reading skills of eighth grade students), the medium and high student-centered education did not worsen the achievements of the students of program schools compared to their control school peers.

Table 8.2 shows the results in the Roma and the non-Roma as well as the disadvantaged and non-disadvantaged subsamples.

The results in the lower grades clearly show the negative role of the weak student-centeredness, but not necessarily to the advantage of high student-centeredness. In the upper grades, the situation with respect to reading comprehension was mixed, and with respect to further education, self-esteem, and coping as far as Roma and disadvantaged students are concerned clearly show a positive correlation. The same was not valid with respect to the non-Roma and non-disadvantaged students for whom it was only the distance-keeping results that clearly improved with the increase of student-centeredness.

Table 8.2 – Program versus control differences by the level of student-centered education in the program schools in ethnic and social subsamples.

Degree of student-centered education in class, by subsample		Reading		Admission	Self-esteem		Coping	Social distance
		4	8	8	4	8	8	8
Roma	Weak	-0.29	+0.26	19%	-0.15	+0.14	+0.38	+0.05
	Moderate	+0.12	+0.07	11%	++0.36	+0.18	+0.38	-0.03
	Strong	+0.27	+0.09	23%	++0.44	+0.35	+0.53	-0.14
Non-Roma	Weak	-0.14	+0.12	5%	++0.07	+0.17	+0.10	-0.08
	Moderate	+0.10	+0.15	17%	++0.22	+0.20	+0.04	-0.04
	Strong	+0.05	-0.14	5%	-0.01	+0.15	+0.01	-0.25
Disadvantaged	Weak	-0.21	+0.07	4%	-0.40	+0.03	+0.19	+0.09
	Moderate	+0.09	+0.09	13%	++0.18	+0.28	+0.18	+0.11
	Strong	+0.27	+0.15	20%	++0.28	+0.37	+0.58	-0.20
Non-disadvantaged	Weak	-0.15	+0.11	5%	++0.25	+0.26	+0.13	-0.14
	Moderate	+0.26	+0.16	14%	++0.29	+0.17	+0.09	-0.16
	Strong	+0.20	-0.05	8%	++0.01	+0.12	-0.04-	-0.39

Note: The figures are constructed the following way. In each program school – control school pair, we standardized the individual outcomes using the control school average (for the test scores, the national standard deviation was kept as a unit). These standardized values show the advantage relative to the average control school student. Thkis table shows the outcome advantage measure, averaged within classroom observations, by the index of student-centered education. For the subsamples, the program school advantages are measured in comparison with the respective group in the control schools.

The patterns within each subsample are very similar to the overall patterns. The relation of student-centered education and student outcomes is positive in lower grades, and mixed or negative in upper grades, with the important exception of social distance. It is remarkable that the social distance between ethnic groups is significantly smaller in program school classes that experience strongly student-centered education than in program school classes that are characterized by weakly or moderately student-centered education (all measured as the advantage relative to the matched program school). If we were to give a causal interpretation, we could conclude that student-centered education

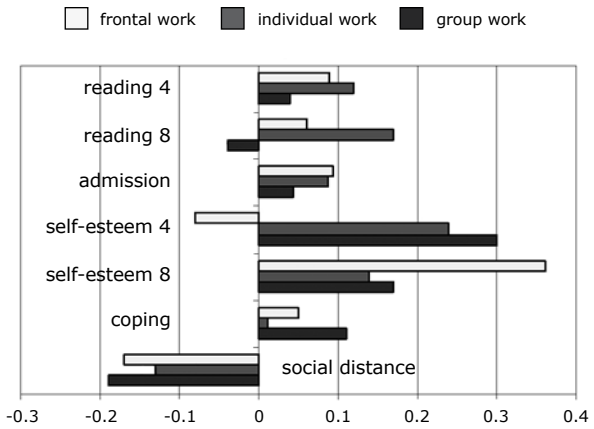
is an important element in the success of the program in lower grades and with respect to social distance in upper grades as well. With respect to other upper-grade outcomes, student-centered education does not seem benefitial.

8.3 Work Forms

There are significant differences in terms of the work form chosen for the observed classes between program schools and control schools (table 4.10 and figure 4.2). Group work is a lot more emphasized in program schools, while frontal work is a lot less frequent. The prevalence of group work is weaker in lower grades and stronger in upper grades.

Figure 8.2 shows the advantage of the students of program schools (relative to the average student of the matched control school) by the observed work form.

Figure 8.2 – Program versus control differences on the basis of work forms that were frequently used in the program schools



It seems that the work form chosen by the program school teachers is not strongly related to the advantage of the students of the program schools. If there is any pattern, it shows that students whose observed activities were more likely to be group work perform worse on cognitive tests and have lower self-esteem in upper grades. Since endogeneity in the choice of work form by schoolteachers is likely to be strong, these results may be more informative of what schoolteachers try to do with classes of different average levels of skills than the effects of those work forms on the skills themselves (the reverse causality problem mentioned above). Nevertheless, it is quite surprising that even social distance is not significantly lower in program school classes characterized by group work than those characterized by frontal work.

Table 8.3 shows the same results in the subsamples of Roma students, non-Roma students, disadvantaged students and non-disadvantaged students.

Table 8.3 – Program versus control differences by work form in the program schools, in ethnic and social subsamples

Work form of activity, by sub-sample		Reading		Admission	Self-esteem		Coping	Social distance
		4	8	8	4	8	8	8
Roma	Frontal	+0.01	+0.15	+15%	-0.04	+0.16	+0.37	-0.07
	Individual	+0.24	+0.36	+12%	+0.35	+0.18	+0.19	-0.02
	Group	+0.06	+0.13	+21%	+0.52	+0.20	+0.51	-0.03
Non-Roma	Frontal	+0.08	+0.19	+15%	-0.11	+0.33	+0.10	-0.10
	Individual	+0.01	+0.19	+14%	+0.18	+0.14	+0.08	-0.04
	Group	+0.00	+0.06	+0%	+0.29	+0.11	+0.06	-0.14
Disadvantaged	Frontal	+0.15	+0.02	+2%	-0.15	+0.28	+0.07	+0.08
	Individual	+0.07	+0.16	+19%	+0.21	+0.11	+0.09	-0.02
	Group	+0.00	+0.09	+10%	+0.28	+0.11	+0.42	+0.05
Non-disadvantaged	Frontal	+0.12	+0.15	+10%	+0.01	+0.37	+0.08	-0.13
	Individual	+0.11	+0.24	+13%	+0.21	+0.10	+0.10	-0.16
	Group	+0.15	-0.02	+ 6%	+0.37	+0.20	+0.11	-0.24

Note: The figures are constructed the following way. In each program school – control school pair, we standardized the individual outcomes using the control school average (for the test scores, the national standard deviation was kept as a unit). These standardized values show the advantage relative to the average control school student. This table shows the outcome advantage measure, averaged within classroom activity observations, by the observed work form of the activity. For the subsamples, the program school advantages are measured in comparison with the respective group in the control schools.

The results show that there are no general patterns in the different subsamples either. In some dimensions, Roma students whose observed activities are more likely to be group work perform somewhat better (admission, self-esteem, coping), but the reading test results

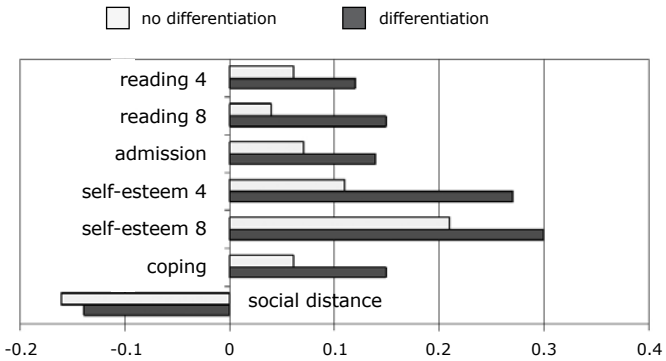
are best for the individual work form. For the non-Roma students, the results for group work are rarely better than the other work forms, and even the social distance kept from the Roma is only marginally lower. The patterns by social disadvantage are even less clear. While some results are better for the group work among Roma students, the figures by subsample do not provide support for group work to significantly enhance skill formation or decrease social distance. As we mentioned above, reverse causality is likely to prevent any strong causal interpretation. In other words, the figures showed in this subsection do not support the positive effect of group work, but they do not contradict that either.

8.4 Differentiation

In Chapter 4, we documented that in lower grades, differentiation was found in 12 percent of the observed activities in program schools, compared to 7 percent of the activities in the control schools. The corresponding numbers in upper grades were 5 and 4 percent, respectively. Differentiation is very rare in both program and control schools, but somewhat more frequent in program schools.

Figure 8.3 shows the advantage of the students of program schools (relative the average student of the matched control school) by whether the observed classroom activity involved differentiated education.

Figure 8.3 – Program versus control differences by whether differentiation is experienced in the class activities of program school students



The students of program schools who take part in differentiated education have better outcomes than the students of program schools who do not take part in differentiated education (where, as usual, all outcomes are compared to the average outcomes in the matched control schools). The differences are modest, but they are uniformly present in cognitive skills, non-cognitive skills and admission rates, both in upper grades and, when measured, in lower grades. If we allow for a causal interpretation, it seems that differentiated

education brings better cognitive and non-cognitive outcomes on average. Perhaps not very surprisingly, differentiation makes no difference in terms of social distance.

As we have shown in Chapter 4 (table 4.14) differentiation is rarely practiced, and when it is practiced, it constitutes primarily of giving individual tasks for slower students. It is therefore especially interesting to see the potential differences behind the average positive effects of differentiation shown above. Table 8.4 shows the results in the Roma versus non-Roma and disadvantaged versus non-disadvantaged subsamples.

Table 8.4 – Program versus control differences by whether differentiation is experienced in the class activities of program school students, in ethnic and social subsamples.

Whether differentiation is observed, by subsample		Reading		Admission	Self-esteem		Coping	Social distance
		4	8	8	4	8	8	8
Roma	No differentiation	+0.08	+0.19	+18%	+0.25	+0.19	+0.40	+0.00
	Differentiation	+0.03	+0.00	-32%	+0.22	+0.18	+0.82	-0.12
Non-Roma	No differentiation	+0.02	+0.09	+8%	+0.08	+0.17	+0.07	-0.09
	Differentiation	+0.07	+0.01	+34%	+0.24	+0.25	+0.03	-0.18
Disadvantaged	No differentiation	+0.09	+0.09	+11%	+0.06	+0.16	+0.26	+0.04
	Differentiation	+0.04	-0.11	-40%	+0.08	+0.05	+0.62	-0.29
Non-disadvantaged	No differentiation	+0.10	+0.11	+8%	+0.15	+0.26	+0.09	-0.18
	Differentiation	+0.19	+0.04	+24%	+0.39	+0.29	+0.11	-0.17

Note: The figures are constructed the following way. In each program school – control school pair, we standardized the individual outcomes using the control school average (for the test scores, the national standard deviation was kept as a unit). These standardized values show the advantage relative to the average control school student. This table shows the outcome advantage measure, averaged within classroom activity observations, by whether the observed activity involved differentiation or not. For the subsamples, the program school advantages are measured in comparison with the respective group in the control schools.

The results of the table are striking. They show that in terms of reading skills, admission rates, and self-esteem it is the outcomes of the non-Roma and the non-disadvantaged students to which differentiation makes a positive difference. The reading and admission results

of Roma and disadvantaged students are worse in classes that experience differentiated education, while they experience no robust difference in terms of self-esteem. Coping skills, on the other hand, show the opposite results: there Roma and disadvantaged students show significantly better coping skills in classes with differentiated activities. Again not surprisingly, the social distance results show no significant differences.

One interpretation to the findings of table 8.4 is that differentiation, in the form it is practiced in program schools, helps the more able students and hurts the less able students. Recall that differentiation in this sample constitutes mostly of giving individual tasks for slower students, among whom the Roma and disadvantaged are over-represented. According to this interpretation, they perform worse in these classes because they are given tasks that may not help their development. At the same time, the rest of the students may achieve better outcomes because their development is not held back. The overall result is positive (see figure 8.3), but it comes at the expense of less able students.

Another, equally plausible interpretation is that classes with more unequal skills (between Roma and non-Roma or disadvantaged and non-disadvantaged students) make teachers more likely to use differen-tiated education. This is the problem of reverse causality, described above in detail. It is possible that the few observed program school activities with differentiated education happen in those program school classes where skill inequality is the largest. According to this interpretation, in such classes schoolteachers choose to give individual tasks to the slowest students because they are less able than the slowest students in program school classes where differentiated education is not observed. Again, it is the correlation of ability and ethnicity as well as social disadvantage that transform the mechanism into lower outcomes for Roma and disadvantaged students in classes characterized by differentiated activities.

It is impossible to tell which of the two interpretations is correct (or stronger). In any case, the evidence does not support that the kind of differentiation practiced in the few program schools would bring strong benefits to the Roma and disadvantaged students. This is not a surprising result given even the few observed differentiated activities are far from the ideal of differentiated education.

8.5 Student Autonomy

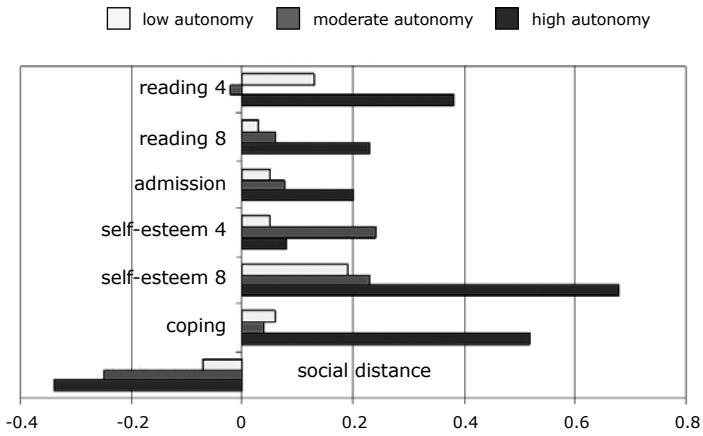
Student autonomy was judged medium or high level in more than 50 per cent of observed program school classes, compared to less than 40 per cent of control school classes. The difference is modest but not negligible (see table 4.5). It is possible that such modest differences are at least in part responsible for the differences in student outcomes because those differences are not very large either.

We consider student autonomy to be such an intermediary variable. On the one hand, it may be a mechanism towards better (or worse) student outcomes. On the other hand, it is an

outcome in itself that can be affected by educational methods and classroom management in general, and by changes induced in them by the program in particular. It is not easy to manipulate but a long-term result of sustained practice of certain kinds of education.

Figure 8.4 shows the results of the students of program schools compared to those of control schools depending on the level of student autonomy observed in the program schools.

Figure 8.4 – Program versus control differences based on the level of autonomy of students of program schools



The robust result shown by figure 8.4 is that student outcomes are better (their advantage over the matched control school is larger) in program school classes that are characterized by high levels of student autonomy. In addition, the relationship between outcomes and student autonomy is monotonic in many dimensions: outcomes are better at moderate levels than at low levels, and even better at high levels. Even more encouraging to see is that the differences are large and monotonic in upper grades where we think our measures work better. A causal interpretation of these results is that student autonomy is an important mechanism towards better outcomes of students, be it cognitive skills, admission to better secondary schools, self-esteem, coping ability, or the distance kept from the other ethnic group.

Table 8.5 presents the same results in the Roma versus non-Roma and the disadvantaged versus non-disadvantaged subsamples.

Table 8.5 – Program versus control differences by the observed level of student autonomy in program school classes, in ethnic and social subsamples

Level of student autonomy by subsample		Reading		Admission	Self-esteem		Coping	Social distance
		4	8	8	4	8	8	8
Roma	Low	+0.12	+0.09	+7%	+0.15	+0.21	+0.30	+0.00
	Moderate	+0.08	+0.25	+22%	+0.30	+0.13	+0.14	+0.01
	High	+0.85	+0.44	+48%	+0.24	+0.72	1.08	-0.13
Non-Roma	Low	+0.07	+0.10	+11%	+0.01	+0.17	+0.11	+0.02
	Moderate	-0.03	+0.09	+7%	+0.24	+0.17	+0.00	-0.22
	High	+0.09	+0.25	+4%	+0.11	+0.80	+0.33	-0.47
Disadvantaged	Low	+0.13	-0.07	-10%	+0.01	+0.16	+0.24	+0.01
	Moderate	-0.02	+0.22	+20%	+0.09	+0.06	+0.18	+0.06
	High	+0.56	+0.31	+41%	+0.26	1.22	1.75	+0.06
Non-disadvantaged	Low	+0.15	+0.13	+10%	+0.06	+0.14	+0.10	-0.12
	Moderate	+0.05	+0.07	+7%	+0.36	+0.30	+0.08	-0.24
	High	+0.43	+0.35	+14%	+0.07	+0.66	+0.09	-0.50

Note: The figures are constructed the following way. In each program school – control school pair, we standardized the individual outcomes using the control school average (for the test scores, the national standard deviation was kept as a unit). These standardized values show the advantage relative to the average control school student. This table shows the outcome advantage measure, averaged within classroom observations, by whether the level of student autonomy observed in the classroom. For the subsamples, the program school advantages are measured in comparison with the respective group in the control schools.

The results shown in Table 8.5 provide further support for the causal role of student autonomy in better school achievements. The outcomes of Roma and disadvantaged students are systematically better in the program schools (relative to their matched control school) when student autonomy is higher. The relationship is somewhat less clear for the non-Roma and the non-disadvantaged students (except for the sharp pattern in social distance). Even in their case, though, medium or high levels of student autonomy correspond to better outcomes in general but not always.

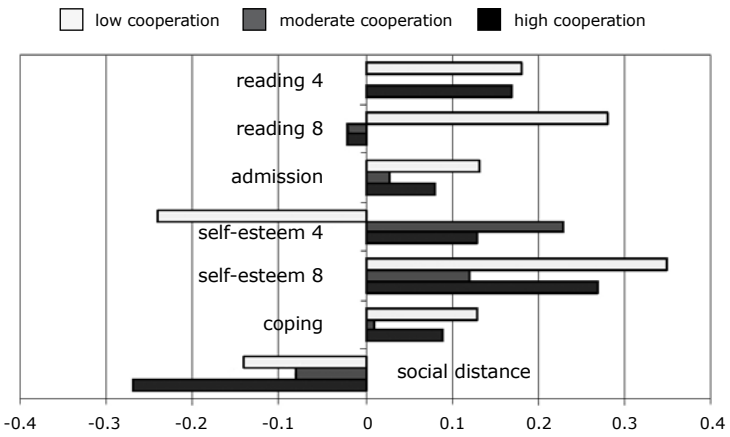
8.6 Cooperation among Students

Observed cooperation and collaboration among students may be an intermediary variable, similarly to student autonomy. On the other hand, they are more likely to be direct consequences of a teaching method: cooperative group work. It is therefore likely that the patterns of program school advantage with respect to observed student cooperation are similar to the patterns with respect to the prevalence of group work. While cooperative group work automatically involves cooperation, it can be evoked by other means as well. Therefore it receives separate examination.

As we showed in Chapter 4, student cooperation was observed to be of medium or high level in close to 90 per cent of program school classes, compared to 40 per cent in control school classes (table 4.5). On the one hand, it is plausible that such large differences may be responsible for the differences in student outcomes. On the other hand, the level of student cooperation observed in the classroom cannot be a very strong mechanism towards student outcomes: otherwise large differences in cooperation should translate to more than the observed differences in student outcomes.

Figure 8.5 shows the results of the students of program schools compared to those of control schools by the level of student cooperation observed in the program schools.

Figure 8.5 – Program versus control differences based on the level of cooperation of program school students



According to the figure, higher levels of student cooperation in program schools are not strongly related to better student outcomes, with the exceptions of self-esteem in grade 4 and social distance. In general, grade 8 results are worse in classes with higher observed cooperation. In general, the patterns are similar to the patterns observed in the case of group work, in accordance to our expectations.

Similarly to the estimated effects of group work, or any other mechanism in this Chapter, reverse causality may drive many of the results. It is quite possible that the teachers of program school classes with lower skills on average are more likely to choose the cooperative learning technique than the program school teachers with better skilled students. And similarly to the other estimates in this chapter, we are not able to identify the direction of the causality.

Table 8.6 presents the same results in the Roma versus non-Roma and disadvantaged versus non-disadvantaged subsamples.

Table 8.6 – Program versus control differences by the level of observed student cooperation in program school classes, in ethnic and social subsamples

Level of student cooperation by subsample		Reading		Admission	Self-esteem		Coping	Social distance
		4	8	8	4	8	8	8
Roma	Low	-0.05	+0.44	+9%	-0.60	+0.19	+0.48	-0.15
	Moderate	+0.05	+0.10	+18%	+0.40	+0.15	+0.37	+0.15
	High	+0.18	+0.22	+25%	+0.51	+0.25	+0.42	-0.19
Non-Roma	Low	+0.17	+0.40	+22%	-0.11	+0.40	+0.22	-0.13
	Moderate	-0.04	+0.00	+1%	+0.15	+0.05	+0.04	-0.03
	High	+0.12	+0.00	+6%	+0.14	+0.19	+0.15	-0.13
Disadvantaged	Low	-0.04	+0.12	-6%	-0.55	+0.26	+0.35	-0.23
	Moderate	+0.06	-0.06	+6%	+0.12	-0.01	+0.23	+0.29
	High	+0.27	+0.26	+26%	+0.14	+0.38	+0.29	-0.20
Non-disadvantaged	Low	+0.24	+0.30	+19%	-0.02	+0.38	+0.19	-0.06
	Moderate	+0.04	+0.10	+5%	+0.23	+0.12	+0.02	-0.20
	High	+0.34	-0.05	+7%	+0.25	+0.27	+0.14	-0.19

Note: The figures are constructed the following way. In each program school – control school pair, we standardized the individual outcomes using the control school average (for the test scores, the national standard deviation was kept as a unit). These standardized values show the advantage relative to the average control school student. This table shows the outcome advantage measure, averaged within classroom observations, by whether the level of student cooperation observed in the classroom. For the subsamples, the program school advantages are measured in comparison with the respective group in the control schools.

The results are similar to those with group work. Among the Roma and the disadvantaged students, higher levels of student cooperation are associated with better outcomes. The relationship is more of the opposite for the non-Roma and the non-disadvantaged subsamples.

Reverse causality is of course very likely to hold. Therefore, one should be very cautious in making any causal interpretation. If we were to do so anyway, the results would imply that higher levels of student cooperation are substantially mo-re likely to help the Roma and the disadvantaged students. In any case, non-Roma and non-disadvantaged student keep the positive program versus control difference even at high levels of cooperation. In other words, while cooperation may help the Roma and the more disadvantaged students, it does not seem to hurt the non-Roma and the non-disadvantaged.

The main conclusions of the chapter can be summarized as follows. Student-centered education is accompanied with positive results in the lower grades, and it seems to strengthen disadvantaged (not necessarily Roma) students in the upper grades as well. Group work and cooperation among students are accompanied with more positive impact for Roma and disadvantaged students, they do not seem to improve the results of the others, but they do not worsen them either. Differentiated education, rare in the sample, is accompanied with improvements in all areas, but not for all students. The way differentiation is practiced in the observed classes seems to help the non-Roma and the non-disadvantaged, and it may actually hurt the Roma and the disadvantaged students. The level of student autonomy is significantly positively associated with all outcomes in all examined groups, in the upper grades in particular.

The results of this chapter should be taken with a large grain of salt. Reverse causality may be driving many of our results (teaching methods are chosen partly because of the skill level of students), which is virtually impossible to control for. With all those caveats in mind, our most important result is that student autonomy, perhaps helped by student-centered education, seems to contribute to the better outcomes of program school students, for Roma and non-Roma, disadvantaged and non-disadvantaged students alike.

Chapter 9

Class Composition and Student Outcomes

The goal of the OOIH program is promoting quality education in integrated classes. In the previous chapter, we focused on the quality of education. In this section, we look at integration at its narrowest sense, the composition of students in the classroom. In particular, we examine the correlation of student composition and student outcomes, i.e. whether classes that differ in terms of their composition also differ in terms of the average outcome of the students in the class. We are especially interested in whether the correlation is different in the program schools versus the control schools. Besides providing an additional aspect to the effect of the program, these comparisons can help us understand the effect of student composition on outcomes in an environment focusing on integrated quality education, in comparison to a traditional environment. We focus on two aspects of student composition: the fraction of Roma students and the fraction of disadvantaged students (see Chapter 2 for the definitions). Similarly to the previous chapters, we look at the outcomes of Roma and non-Roma students as well as disadvantaged and non-disadvantaged students separately.

Student achievements in most dimensions are likely to be worse in classes with a higher proportion of minority and disadvantaged students, for several reasons. One reason is the impact of the peer group: children's motivation may be significantly affected by the motivation of others, which is often lower among disadvantaged students. This can have a direct negative effect on the outcomes of the entire class, and the effect may be stronger with a higher number of disadvantaged students.

The second factor is the likely increase in the educational problems with the increase of the proportion of disadvantaged students. The incentives within the school system in Hungary do not compensate for the extra effort and frustration that may arise in such classes. As a result, classes characterized by a larger number of disadvantaged students may experience lower quality of education. When disadvantaged family background coincides with minority status, these effects may be strengthened further.

Finally, the correlation between student composition and student outcomes can be significantly reinforced by selective admissions into schools or selective assignment to classes within schools. If everybody believes that a higher proportion of Roma and disadvantaged students causes worse outcomes, students that are considered more talented will avoid such classes. This in itself leads to a situation in which the observed outcomes of all students in classes with a higher proportion of Roma and disadvantaged students may be worse, whether the previous two mechanisms work or not.[74] The three mechanisms imply the

[74] Such mechanisms are called self-confirming stereotypes. See, for instance, Glenn Loury: „The Anatomy of Racial Inequalities", Harvard University Press, 2002; chapter 2.

same correlation: student outcomes are lower in classes with a larger fraction of Roma and disadvantaged students.[75]

Note the important exception of ethnic prejudices: if Allport's conditions are satisfied (see Chapter 7), the level of tolerance is likely to be higher (the level of prejudice lower) in classes with a larger fraction of Roma students because of higher level of inter-group contact.

For all those reasons, an important question of our impact assessment study is the extent to which the program can break the correlation when negative and reinforce it when positive. The OOIH program has a direct effect on student composition, and it is likely to have an effect on all three mechanisms described above. There are no classes with a very large fraction of Roma students in the program schools, which is not true for some control schools (see Chapter 3). The program supports the educational methods that are more effective in integrated classes, therefore making schoolteachers better prepared to deal with educational problems (Chapter 4). In addition, the larger emphasis on student-centered education and student cooperation can decrease the negative peer group effects.

The results described in Chapters 5 to 7 suggest that, on average, the effect of the program is positive in all dimensions for all students, albeit not always large. The analysis in this chapter can reveal whether the effect of the program depends on the fraction of Roma and disadvantaged students in the class.

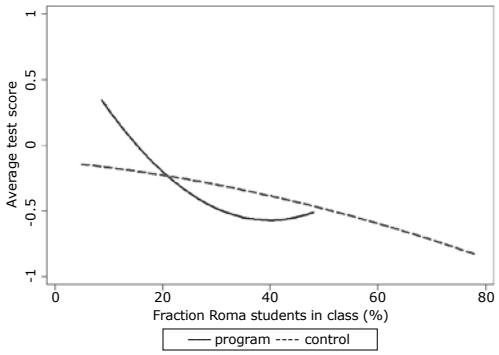
The examined outcomes are reading comprehension (fourth and eighth grades), admission to secondary schools that provide a graduating examination ("érettségi" in Hungarian; eight grade), the overall SPPC score of positive self-esteem (fourth and eighth grades), coping abilities (eighth grade), and the score of social distance (eighth grade). With respect to each outcome variable, we show graphs with the composi-tion of the class on the horizontal axis and the average outcome on the vertical axis. Each graph contains two lines, one for program schools, and one for control schools. The outcomes we show separately for Roma and non-Roma students (in the first set) or disadvantaged and non-disadvantaged students (the second set). Note that the fraction of Roma students varies less among classes in the programs schools (Figure 3.1). As a result, the lines for program schools span a shorter interval on the horizontal axis. We present the graphs with respect to the fraction of Roma students first and disadvantaged students second.

[75] These arguments are elaborated, in Hungarian, in Gábor Kertesi–Gábor Kézdi: "Elementary School Segregation (part I-II)" (Általános iskolai szegregáció (I–II. rész)). Közgazdasági Szemle, 2005/4, 317-355. és 2005/5, 462-479.

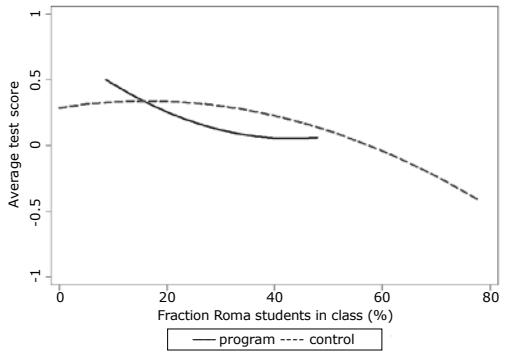
9.1 The Fraction of Roma Students

Based on table 5.4 we know that the reading comprehension results of both Roma and non-Roma students are better in program schools. Figure 9.1 shows the distribution of the achievement surplus by class composition.

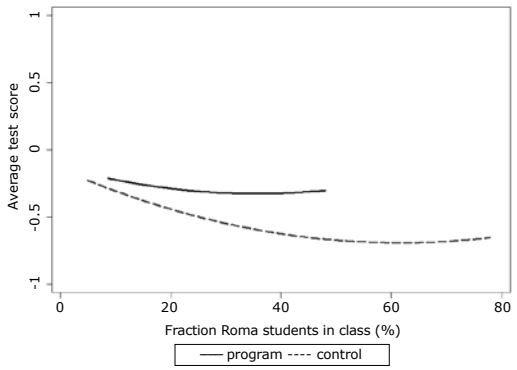
Figure 9.1 – Fraction of Roma students in the class and reading scores of Roma and non-Roma students (program schools and control schools separately)



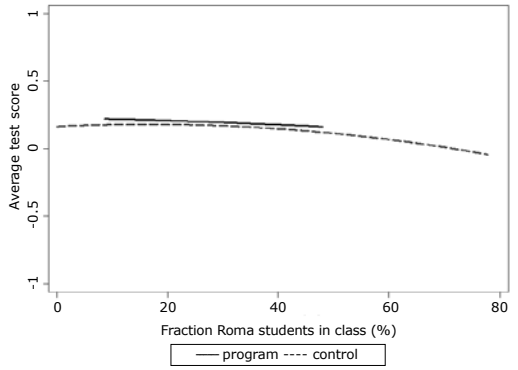
Reading score of Roma students, grade 4



Reading score of non-Roma students, grade 4



Reading score of Roma students, grade 8



Reading score of non-Roma students, grade 8

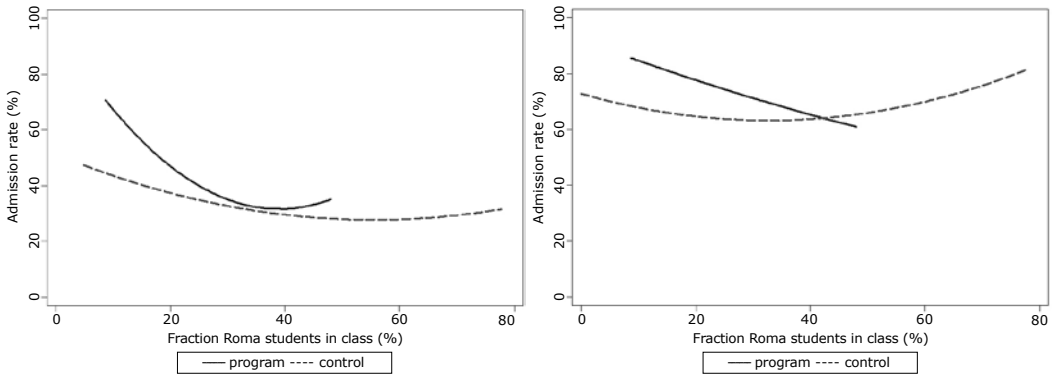
Control school lines are marked with a grey line. As expected, they show a negative correlation, which is particularly powerful in grade 4. This suggests that outcomes are worse in classes with a higher fraction of Roma students. Somewhat surprisingly, in grade four, this negative correlation continues to be present in program schools. If anything, it is even stronger there. One way of interpreting these results is that in lower grades, the program schools can achieve significant success in classes where the fraction of Roma students remains below 20 per cent.

In the upper grades, the correlation of the fraction of Roma students and reading outcomes is weaker to begin with. Contrary to what we see in lower grades, here that correlation is weakened further in the program schools. As a result, reading outcomes and the ethnic composition are independent in program schools. These outcomes are also uniformly above the outcomes in the control schools, with a difference insignificant for non-Roma students but larger for Roma students (see also table 5.4). Recall that in table 5.5, we demonstrated that better reading scores for Roma students do not necessarily reflect the impact of the program itself. In any case, whether due to the program, or for other reasons, program schools seem to have overcome the negative effect of class composition on reading outcomes in the upper grades.

The admission results are similar to the lower-grade reading results in terms of the effect of classroom composition. On the other hand, the results are virtually uniformly better for program schools, similarly to the upper-grade reading test results. Here the vertical axis shows the percentage of students who were admitted to a secondary school that prepares for the maturity examination (figure 9.2). The conclusion we can draw is that the program schools can achieve significant success in classes where the fraction of Roma students remains below 20 per cent.

One should keep in mind that the admissions results are in part shaped by the decisions of the secondary schools, which are not directly targeted by the program. This can be an important explanation for the larger success of classes (and schools) with fewer Roma students. Many of the classes with higher fraction of Roma students are located in regions and neighborhoods with a higher proportion of Roma population. Those regions and neighborhoods tend to have fewer secondary schools in general, and upper tier secondary schools in particular. In light of such difficulties, it is remarkable that the program schools manage to place a larger fraction of their Roma and non-Roma students in those secondary schools even in the poorer regions and neighborhoods.

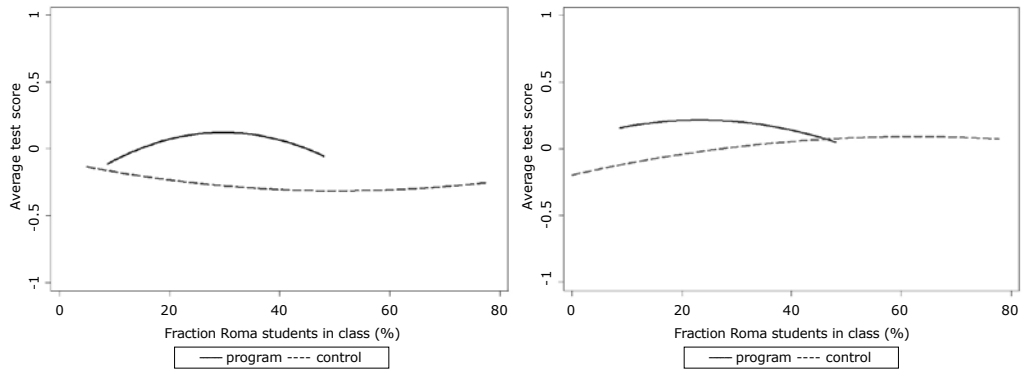
Figure 9.2 – Fraction of Roma students in the class and admission rates to upper-tier secondary schools (%), of Roma and non-Roma students (program schools and control schools separately)



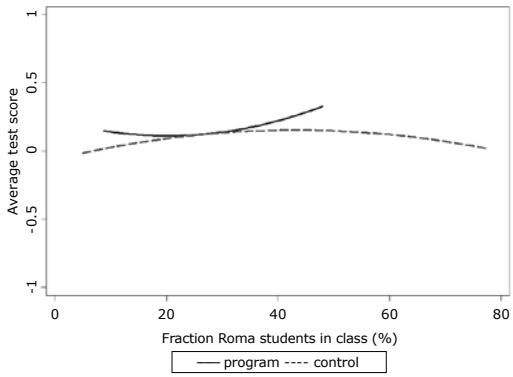
Reading score of Roma students, grade 8 Reading score of non-Roma students, grade 8

Figure 9.3 shows the results for positive self-esteem, by looking at the nationally standardized test scores aggregated from all items in the SPPC test.

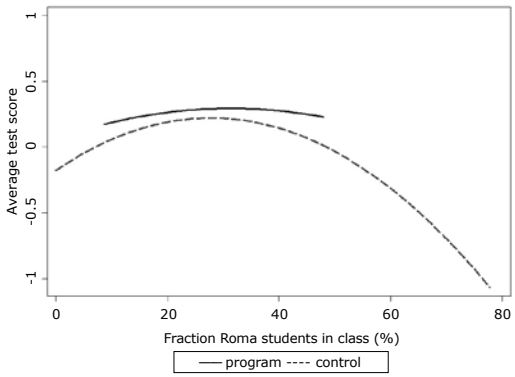
Figure 9.3 – Fraction of Roma students in the class and self-esteem scores of Roma and non-Roma students (overall SPPC scores; program schools and control schools separately)



Self-esteem score of Roma students, grade 4 Self-esteem score of non-Roma students, grade 4



Self-esteem score of Roma students, grade 8



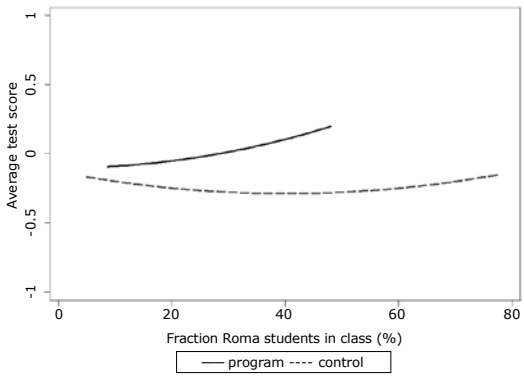
Self-esteem score of non-Roma students, grade 8

The grey lines indicate that there is no negative correlation between self-esteem and the fraction of Roma students, except for non-Roma scores in grade 8 in majority-Roma classes. The shape of the lines is not very different in program schools. Self-esteem results are uniformly better in program schools. The improvement is sometimes larger in classes with few Roma students (non-Roma results in grade 4), sometimes in classes with more Roma students (Roma results in grade 4), and often there is no such a monotonic relationship.

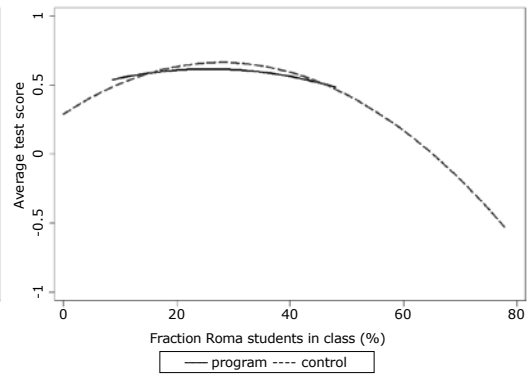
An important result is that the negative correlation between the fraction of Roma students and non-Roma self-esteem in majority-Roma classes disappears in the program schools. The trivial reason is that there are virtually no majority-Roma classes in program schools. Moreover, it seems that even with classes above 40 per cent Roma, non-Roma self-esteem is no worse in program schools than in classes with few Roma students.

Figure 9.4 shows the relationship of classroom composition and coping abilities.

Figure 9.4 – Fraction of Roma students in the class and coping scores of Roma and non-Roma students (program schools and control schools separately)



Coping score of Roma students, grade 8

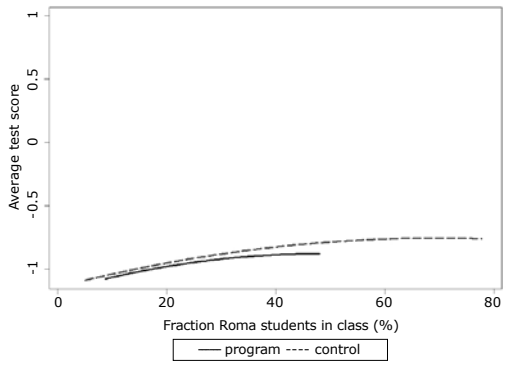


Coping score of non-Roma students, grade 8

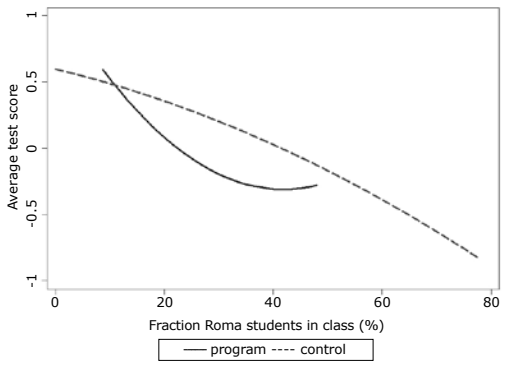
In light of the results of table 6.7, it is not surprising that results in program schools are significantly better among the Roma students but not the non-Roma students. Otherwise, the picture for non-Roma students is very similar to the picture of self-esteem in grade 8. For Roma students, a remarkable novelty here is the advantage of program schools seems to be increasing with the fraction Roma students. As we noted above, small sample size should make one cautious about drawing strong conclusions from the shape of the figures. If one is willing to do so anyway, the results suggest that the program strengthened the coping abilities of Roma students mostly in classes with many Roma students.

The last outcome we look at is ethnic prejudice. We show the nationally standardized values of the social distance scale. One should keep in mind that here negative values indicated smaller distance and thus a higher level of tolerance. Figure 9.5 shows the distance keeping of Roma students from non-Roma, and distance keeping of non-Roma students from Roma.

Figure 9.5 – Fraction of Roma students in the class and the social distance kept by Roma students from their non-Roma peers, as well as the social distance kept by non-Roma students from their Roma peers (program schools and control schools separately)



Social distance kept by Roma students from their non-Roma peers, grade 8



Social distance kept by non-Roma students from their Roma peers, grade 8

With respect to levels, the results of figure 7.2 are present here as well. Roma students keep little social distance from their non-Roma peers both in program schools and in control schools. The non-Roma students keep a substantial distance, but the distance is smaller in program schools. There is no significant relationship between classroom composition and social distance kept by the Roma students.

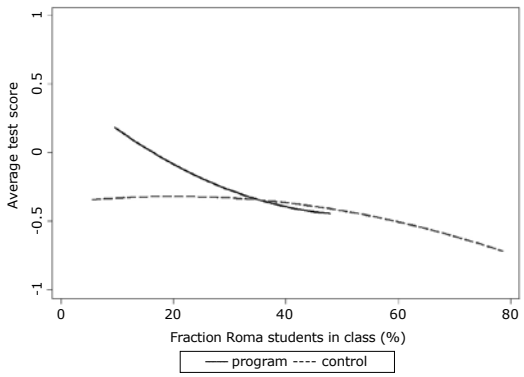
The most important result we see on figure 9.5 is the strong negative effect of the fraction of Roma students on the social distance kept from them by non-Roma students. The relationship strong and negative in control schools. It is even stronger in program schools. These results provide a strong support for the Allport contact hypothesis that was reviewed in Chapter 7. Recall that the essence of the hypothesis is that more "contact" among groups that are otherwise hostile toward each other can reduce the prejudices towards each other. The right panel of figure 9.5 shows exactly such a relationship in control schools. In program schools, the reduction of social distance is even steeper in the lower range of the fraction of Roma students. In the middle range, no further reductions are observed, but the level of the distance is kept below the control school levels. The results show that the conditions necessary to make the Allport contact hypothesis are satisfied both in the program and the control schools, and they are somewhat stronger in the program schools. As a reminder, four such conditions are emphasized: equal status of the two groups when entering into a relationship; intergroup cooperation; common goals; and the supporting role of some authority, law or habit. We can only guess the relative importance of those factors in the program versus control differences, but higher levels of student-student cooperation in program schools may be an important factor.

9.2 The Fraction of Disadvantaged Students

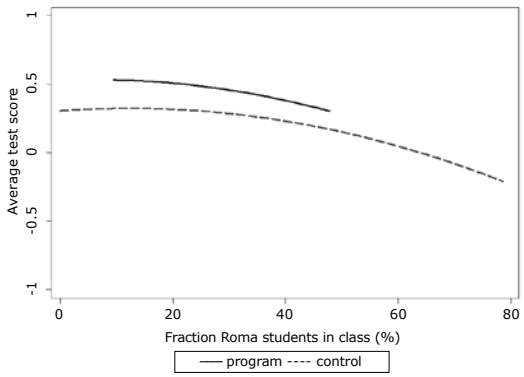
Except for the effects on ethnic prejudice, the effect of classroom composition on student outcomes is likely to be very similar if, instead of the fraction of Roma students, we look at the fraction of disadvantaged students. In this section we do just that. This exercise can provide a robustness check for the results shown in section 9.1. If the results are similar here, our interpretation with respect to the underlying mechanisms receives further support. If the results are very different here, those interpretations are weakened. In our sample, two thirds of Roma students are disadvantaged, and similarly the other way round, two thirds of disadvantaged students are Roma. The correlation of the fraction of Roma students and the fraction of disadvantaged students in the class is 0.75. This is another reason why we expect similar pictures here than in section 9.1.

We reproduce the same figures as in figures 9.1 through 9.4, but we replace the horizontal axis to represent the fraction of disadvantaged students, and we show results separately for disadvantaged students and non-disadvantaged students. We start with the reading test scores shown in Figure 6.

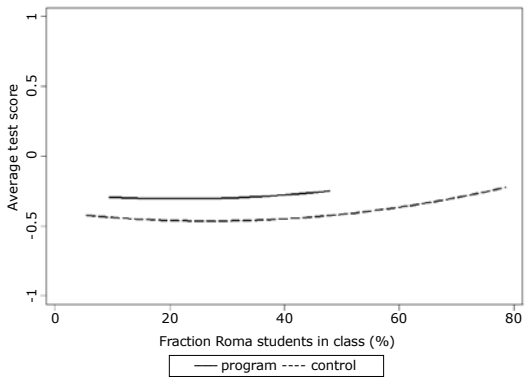
Figure 9.6 – Fraction of disadvantaged students in the class and reading scores of disadvantaged and non-disadvantaged students (program schools and control schools separately)



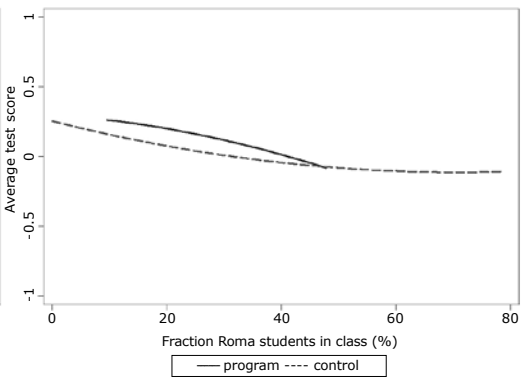
Reading score of disadvantaged students grade 4



Reading score of non-disadvantaged students grade 4



Reading score of disadvantaged students, grade 8

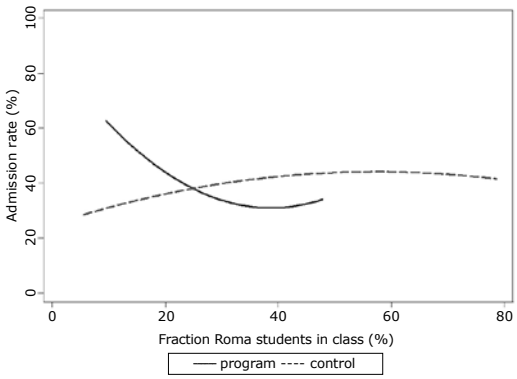


Reading score of non-disadvantaged students, grade 8

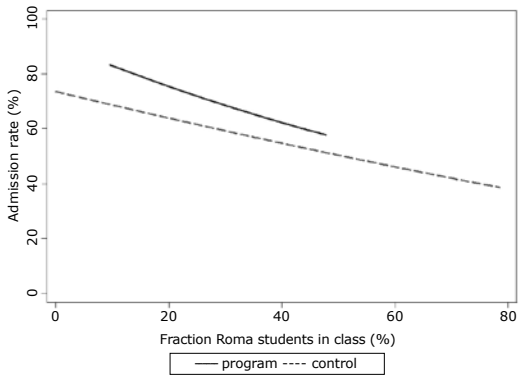
The grade 4 pictures are very similar to those in figure 9.1. The negative correlation between the fraction of disadvantaged students and the reading score is not weaker in program schools. If anything, it is stronger. In grade 8, there is no strong negative relationship either in program schools or control schools.

As Figure 9.7 shows, the admission results are again very similar to Figure 9.2, the analogous relationships with respect to the fraction of Roma students.

Figure 9.7 – Fraction of disadvantaged students in the class and admission rates to upper-tier secondary schools (%), of disadvantaged and non-disadvantaged students (program schools and control schools separately)



Reading score of disadvantaged students grade 8

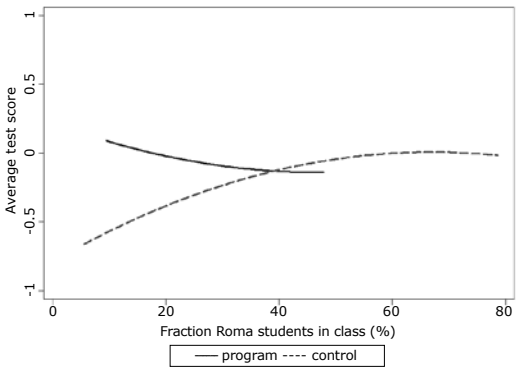


Reading score of non-disadvantaged students grade 8

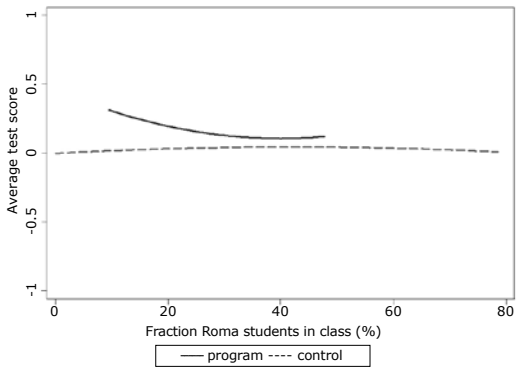
Taken together, while positive altogether, the reading and admission results are negative in terms of the ability of the program in breaking the negative relationship between student composition and student outcomes. It seems that the program was most successful in classes with less than 30 per cent of Roma students.

Figure 9.8 shows the results for self-esteem, measured by the total SPPC score standardized at the national level.

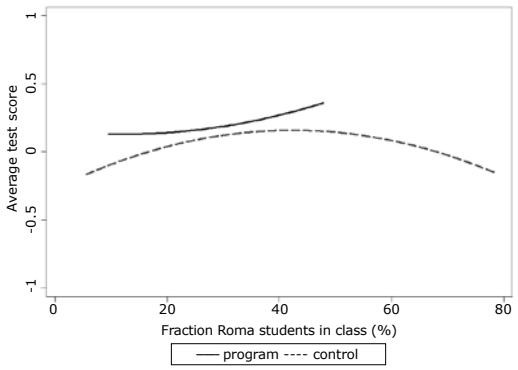
Figure 9.8 – Fraction of disadvantaged students in the class and self-esteem scores of disadvantaged and non-disadvantaged students (overall SPPC scores; program schools and control schools separately)



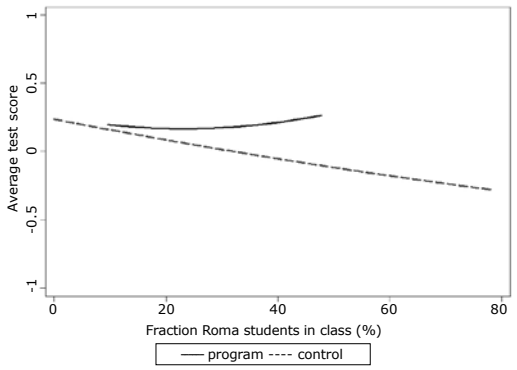
Self-esteem score of disadvantaged students, grade 4



Self-esteem score of non-disadvantaged students, grade 4



Self-esteem score of disadvantaged students, grade 8

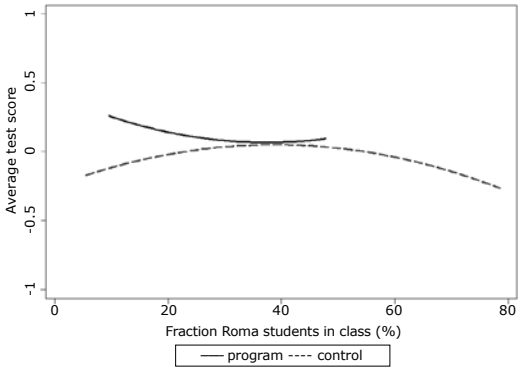


Self-esteem score of non-disadvantaged students, grade 8

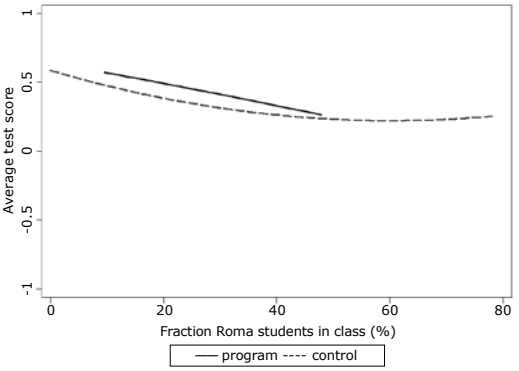
The self-esteem figures are little different here than with respect to the ethic composition (figure 9.3). The correlation of self-esteem and the fraction of disadvantaged students is never strongly negative. In lower grades, program schools show very weak negative relationship, while in upper grades control schools do, especially with respect to the results of non-disadvantaged students. Most importantly, the figures here support our previous conclusion about the prog-ram breaking the somewhat negative correlation there.

Lastly we turn to the results on coping. Figure 9.4 shows the relationship of classroom composition and coping abilities of disadvantaged and non-disadvantaged students. The picture is similar with respect to the results of the non-disadvantaged students, but it is less encouraging with respect to the results of the disadvantaged students.

Figure 9.9 – Fraction of disadvantaged students in the class and coping scores of disadvantaged and non-disadvantaged students (program schools and control schools separately)



Coping score of disadvantaged students, grade 8



Coping score of non-disadvantaged students, grade 8

Taken all results together, the effect of the program on the correlation of classroom composition and student outcomes is mixed. In terms of reading skills and admission to better secondary schools, program schools seem to reproduce the negative correlations. In other words, the advantage of program schools is largest in classes with less than 20 per cent of Roma or disadvantaged students.

On the other hand, the results with respect to non-cognitive skills show just the opposite. In terms of self-esteem and coping, program schools seem to bring the largest benefits in classes with larger fraction of Roma and disadvantaged students. Perhaps most importantly, ethnic prejudice against the Roma shows a steady decrease as the fraction of Roma students increases in the class, and the relationship is steeper in program schools than in control schools.

What does all this mean with respect to our original question? On the one hand, it seems that eliminating classes with Roma and disadvantaged majority is an important objective in itself. In many important dimensions, the benefits of the program decreased with larger fraction of Roma and disadvantaged students. On the other hand, it seems that the OOIH program in itself was not able to eliminate all the mechanisms through which the correlation between student composition and effectiveness comes about.

Concluding Remarks

The goal of the research summary was to provide a comprehensive assessment of the impact the OOIH program. Our impact assessment study examined the effects of that program on classroom work (teaching methods, classroom management, interactions, student autonomy etc.), the students' ethnic attitudes, and their skills. We interpret skills in a broad sense, including the cognitive dimension (reading comprehension in particular) and non-cognitive dimensions as well. Recent research results demonstrate that, besides cognitive skills and competences, positive self-esteem, the sense of controlling one's own destiny and the ability to cope with difficult situations are also important determinants of labor market success. Those dimensions were therefore also looked at in this study. We examined the effects on the development of Roma and non-Roma as well as disadvantaged and non-disadvantaged students.

The analysis was based on a matched sample of 30 participating schools and 30 control schools that are very similar to the program schools. The student composition of the program schools and the control schools are virtually identical. The study estimated the effect of the program in various ways: by comparing outcomes in the program and the control sample; controlling for student background and stu-dents' results recorded in the beginning of the program; and restricting the sample to the schools that were not integrating before the program (and their control pairs). The different estimation methods yield very similar results, and this suggests that the measured differences are indeed caused by the program.

In summary, the study shows that the program increases the level of integration within schools, and it leads to a shift in the direction of student-centered education, higher levels of student autonomy and a wide-spread use of cooperative group work. The study finds that the students of the program schools achieve somewhat higher grades, their reading skills are also somewhat better, and they are more likely to pursue further education in secondary schools that provide a graduating examination ("érettségi" in Hungarian) than their peers in control schools. The effects on cognitive and academic development are largest for the Roma students, it is positive, if often modest, for all student groups.

The effect on non-cognitive skills is also positive and larger than on cognitive skills and achievements. Program school students have more internal locus of control (they are more likely to believe that they themselves are responsible for their successes or failures), they have more positive self-esteem (especially in terms of general, external, and school-related items of self-esteem. Some of the effects are larger for Roma, others for non-Roma students, but again, they are positive for all student group analyzed. Ethnic prejudice against the Roma is also positively affected by the program. Non-Roma students of program schools see the Roma in a less stereotyped way, they keep a smaller social distance from them, and they

think less in terms of social hierarchy. Finally, students of program schools are characterized by lower levels of social dominance orientation and social anxiety.

In many respects the program was quite imperfect in the way it affected the schools' operation and the work in the classrooms. The positive effects of the program across the analyzed dimensions are all the more remarkable. The mechanisms in the background of the success are hard to identify. On the one hand, integration seems to be a necessary requirement. On the other hand, the main driver of the positive effects seems to be in the use of modern, student-centered educational methods that reinforce student autonomy and cooperation.

Since participation in the program was not a result of a controlled random experiment, one cannot be one hundred per cent sure that the observed positive phenomena are all caused by the program itself. There are several arguments for the impact of the program – these were documented in detail in the report. However, it is important to see that even if not always due to the program itself, the program schools, with integrated education, achieve better results than the control schools that have students of identical family background.

The results of this study demonstrate that it is possible to achieve integrated education in which both Roma and non-Roma students achieve better skills and attitudes. Naturally, it is not an easy task for the schools. It may necessitate significant changes in the organization of education, it may require new teaching methods and a new approach from the schoolteachers, and it surely involves extra work by all teachers and managers. Those teachers as well as managers need appropriate training, support, and incentives. The OOIH program seems to have put those elements in place in a small scale. All those elements should be adopted if we want to enjoy similar effects on a national scale.

Németh Szilvia – Szira Judit

ANNEX
on the Hungarian Educational Integration/Desegregation Programme

Legal and administrative framework 2002-2005

Annex on the Hungarian Educational Integration/Desegregation Programme[1]

The Hungarian education system is a decentralised public education system, in which the role of the central government is very limited. Independent local self-governments have significant autonomy in decision making; they are the maintainers of kindergartens and local primary schools. Elected county governments are responsible for the secondary education system and for those institutions following a special curriculum, dormitories and state care. Although the scope of obligations is clear, all school maintainers have the right to open any kind of institutions.[2] Education at a glance data[3] shows also the high degree of school autonomy. This autonomy allows schools to adjust their local curriculum and services to the local needs and expectations.

These arrangements mean that it is difficult for central policy makers to implement their goals. Since 1994 governments paid increasing attention to promoting equal chances in the education system, especially the mandatory integration of Roma and children with special needs. But research has shown that in spite of these actions the gap between Roma and non Roma community has been continuously growing. Not just state funds but European sources were used to try to improve the social, labour and educational situation of the Roma community (health and housing were less prior). The largest programme in terms of size, aiming at the social integration of the most disadvantaged youth and Roma in particular, was a Phare programme (HU9904-01), which started in 1999. The total budget of the programme was 12.52 million Euro, cofinanced by the Ministry of Education and the Phare funds of the European Union. The programme concentrated on three areas: (1) the reduction of the dropout rate of Romani children in education, (2) the development of training programme and student support systems on secondary level, and (3) the development of talent promotion programme in order to enhance the opportunities of disadvantaged children to continue their studies at the tertiary level. Although the total budget available for the programme has been spent, according to the follow-up monitoring report, it was not an effective or successful programme. The report suggested that the beneficiary schools continued using the same inefficient and in several cases segregated educational practices. Out of the nineteen examined schools there are only two schools without segregated Romani classes; only one school managed to assure opportunity for the

[1] Authors of the paper: Nemeth Szilvia researcher, TARKI-TUDOK, Center for Knowledge Management and Educational Research and Szira Judit senior adviser, Roma Education Fund.

[2] In the 2007/08 schoolyear, out of 489 vocational schools 220 were maintained by the local government, 115 by county government (it is their responsibility), 19 by the central budget, 22 by church, 86 by foundations or individuals, and 27 by other.

[3] Decisions made on primary education issues: 4% central government, 29% local government, 68% school.

children to continue on the secondary level on equal footing; four schools had explicitly positive attitudes towards Roma; and five schools indicated that there are serious ethnic conflicts in the schools, while according to the opinion of the teachers such problems existed in seven more schools. According to the report the main lesson of the survey is that Phare support was the catalyst of reforms only in very rare cases and that schools applied for these funds were not really interested in the ultimate intentions and goals of the programme, but instead were rather keen to gain extra funding.[4]

During the period 1998-2002 the type and amount of scholarships for Roma pupils and students increased and various new forms of normative per capita financial support for schools and for their maintainers were introduced.

Administrative Steps Aiming Social and Educational Integration of Roma[5]

The government has appointed in 2002 a minister responsible for equal opportunities, who[6] developed the governmental program promoting the social integration of Roma. This program mentions the necessity of decreasing segregation in education and increasing the quality of education.[7] The Minister has introduced the Act for "Promoting equal treatment and equal opportunities."[8] The Act makes up for important deficiencies related to exploration and legal redress of discrimination cases. These are among others the definition of direct and indirect negative discrimination, the specification of protected groups, and that during the legal redress process it puts tools – not previously present in the Hungarian legislation – in the hands of the aggrieved and the legal protection organisations that can make the law enforcement against discrimination and legal redress efficient at last.

A political state secretary responsible for Roma matters has been appointed in the Prime Minister's Office and an Office for Roma Matters has been created. In the meantime – based on the directives of the European Council[9] – the Authority for Equal Treatment came into existence.

First in the government[10] the Minister of Education has nominated a ministry commissioner[11] responsible for the integration of socially disadvantaged and Roma children. The office of

[4] This paragraph was written based on a manuscript that had been prepared by Orsolya Szendrey.
[5] Summary based on: Discrimination in Education, 2006, UNESCO-report, Hungary. (ed. Szilvia Nemeth), National Institue for Public Education, Budapest, and Equal Access to Quality Education for Roma, Hungary, 2007. Open Society Institute, EU Monitoring and Advocacy Program, (reporter Lilla Farkas, ed. Specialist Szilvia Nemeth).
[6] Lévay Katalin, Member of the European Parliament from 2003.
[7] Government Decree1021/2004. (III.18.) on „The Government Programme Facilitating the Social Integration of Roma People and Related Provisions", Annex No. 1. Chapter IV.
[8] Act CXXV of 2003 on Promotion of Equal Treatment and Equal Opportunities.
[9] Directive no. 2000/43/EC.
[10] In other ministries, except for NKÖM, Roma rapporteurs are operating.
[11] Mohacsi Viktoria, 2002-05, member of the European Parliament from 2005, Daroczi Gabor 2005-06.

the ministry manages a separate budget and it is entitled to submit independent proposals. The fight against segregation started with legislative modifications, professional programmes and network building, and then continued with follow-up monitoring of the modifications, programmes, developments and with execution of the necessary modifications. All steps were taken with active Roma participation, the majority of the office employees were Roma,[12] there were a high number of Roma among the external experts working for the office, and the more substantial documents, developments and programmes were prepared through consultation with different Roma organisations.

The mid-term strategy of the Ministry of Education[13] construed the question of educational effectiveness and the development of basic competences as the highest priorities for intervention. It also analyses the topic of inequalities and the following sub-goals have been defined towards the solution:

» Extension of kindergarten attendance among the socially disadvantaged groups of children, integration of Roma and children with special needs.
» Integration of Roma children.
» Inclusion of children with special educational needs.
» Modernisation of the network of vocational schools.
» Elimination of negative discrimination.
» Integration of children with specific nurturing needs.

In line with the mid-term strategy a National Development Plan has been created for using the resources available from the structural funds of the European Union. The Operative Programme for Human Resources (Humán Erőforrás Operatív Program – HEFOP) has secured significant financial resources for the educational sector. The integration and support for educational success of Roma children appears as a horizontal goal and as a separate programme.

The Ministry of Education has elaborated a separate strategy for reducing educational inequalities.[14] The case of Roma integration has appeared in other professional materials and strategies of the ministry, e.g. in the strategy of Life Long Learning[15] and in the strategy for the mobility of the Hungarian educational and training branch.

[12] At the creation of the office the ratio of Roma and non-Roma employees was 50-50%, then it improved significantly in favour of Roma, 80% of the employees being Roma.
[13] Mid-term strategy of the Ministry of Education, March 2003.
[14] Centre for Educational Policy Analysis, NIPE, 2004.
[15] The EU has obliged the countries to elaborate a Life Long Learning strategy.

Main Actions of the Ministry of Education 2002-2006

Date of introduction	Action
2003.	**Prohibition of negative discrimination** (Based on this all such actions can be pronounced null that negatively discriminate a group of children, pupils or even one single child or pupil based on their gender, age, origin, familial situation or any other reason.)
	Free course books for disadvantaged children – books are provided free in public education for students from socially and economically disadvantaged families. (From 2003 in grades 1.-4. of primary schools, from 2004 also in grades 5.-8. of primary schools. From 2005 in grades 9. and 10. and in vocational grades.)
	Meal costs of schoolchildren are reduced for those from socially and economically disadvantaged families
	Mandatory admission to day-nursery, boarding-school (According to the modified public education act, this becomes mandatory if during admission the endangered status of the child, pupil or the guardianship court arrangements made for the child surface as a problem. [Act on Public Education 65.§])
	The school of the 21st century credit program and the NDP ROP have issued tenders for creating **new kindergarten places** in the socially disadvantaged regions.
	All children with special needs may participate in a **capability development training** backed by a normative subsidy of 17 thousand, later 20 thousand Ft.
	Integration normative can be utilised in grades 1, 5 and 9 of primary schools in an ascending system. This is three times the capability development normative, 60 thousand Ft.
	Creation of a national integration network supporting the integration of Roma and disadvantaged children.
	Roma family coordinators in primary schools, their employment and training: in the framework of this programme unemployed Roma people work in primary schools as youth helpers, assistant teachers. Besides working they also participate in adult education and vocational training.

2003.	**"Tanoda" programme (Act on Public Education 95.§)** The definition of the "tanoda" method has appeared in the Act of the Public Education: it is a method facilitating the school success of children with special needs through activities outside the school. Good results can be achieved in the afternoon activities outside the school. Since 2003 about 57 afternoon schools have been established.
	Definition of the project method 45. § Instead of the usual 45 minutes routine subject processing of the schools this is a learning and upbringing method building on the children's previous experiences, based on experiences gained during the common work with the teacher and on the cooperation of children.
	The "Leave the Last Bench" programme has been launched, because the Roma pupils are unjustifiably misdiagnosed and over-represented in the schools for the slightly mentally disabled.
	By modifying the decree number 32/1997 of MKM there is a possibility for effective appearance of the **romani and beash** languages in the school education.
	"Cohesive society programme" – campaign programme (The government has launched a communication campaign that tries to change prejudices of the wider society against Roma)
2004.	**Review of the National Core Curriculum** The National Core Curriculum includes a requirement that all children participating in public education have to learn about the culture of the Roma and our common history.
	Arany János Boarding school programme: 5 boarding schools, 5 grammar schools and 124 children participate in the programme.
	Vocational school development programme: In vocational schools the drop-out rate has been 30% for years now. The goal of the programme is to halve this value. The majority of the students in the vocational schools are socially disadvantaged.
2005.	**Positive discrimination:** The applicants who reach the score of those admitted to paid education may be admitted to the first university training.
	"Útravaló"-programme: all applicant children from poor families receive scholarship and special support from mentor teachers. The tender was nurtured by the Tempus Public Foundation.

1. Per Capita Normative Support for integrated Education[16]

In 2002 the Office of the Ministerial Commissioner for Integration of Roma and Disadvantaged Children started its operation by transforming the ministerial decree (11/1994.MKM) and by creating the decree [OM 57/2002 (XI.18)] of the Ministry of Education in order to create the legal and financial background (39/E: per capita normative support for integrated education) that can pave the way for many disadvantaged children to reach the level of education desired in the Hungarian society. This regulation introduces the concept of preparatory training for the realization of potential and integration. New forms of assistance are aimed at making it possible for children with different social and cultural backgrounds to be taught together and receive the same level of education. The decree does not order schools to implement integration, but it does provide guidance to schools that choose to integrate. The new integration programme was introduced in the 2003/2004 school year. Institutions undertaking to implement new approaches launched integration programmes in a concerted manner for children in their first, fifth, and ninth year of education.

According to this decree and the integration programme, those disadvantaged students who have special educational needs and are therefore currently participating in a preparatory training programme (usually in separate "catch-up" classes) are eligible to participate in an integration programme (which is supported with supplementary per capita allowance). Through the integration programme, these students study in the same class or, when a class is split, in the same group with students not participating in the training programme.
Pupils who can benefit from this programme are those students:

- Whose parents attended only elementary school and find it difficult to understand the modern requirements of mainstream education.
- Whose family is eligible for supplementary family allowance, i.e., they come from an economically disadvantaged[17] environment.
- Who have special social background according to the head of the school.

Since it is the maintainer (in most cases this means the local government) who decides how many classes the school can start and how many children they can admit, it was an especially important goal to make the local governments understand: Roma children need to be educated in the schools as other children, allocated in equal numbers in an integrated way. The school director decides which child gets into the school and the composition

[16] Summary based on Country Assesment series of the Roma Education Fund; Advancing Education of Roma in Hungary; Authors: Nemeth Szilvia and Szira Judit; www.romaeducationfund.org.

[17] Act No. 79 of 1993 on Public Education (PEA) defines socially disadvantaged children as follows: Children who are taken into protection by the notary pursuant to their family conditions or social status and/or children whom the notary declares eligible for regular child protection benefits. Multiple disadvantage results from parents' education level not exceeding eight grades – including unsuccessful further education – and also from placement in long-term State care. (Ministry of Education and Culture uses this definition becuse official data collection on pupils' etbnic identiy is forbidden in Hungary.)

of classes is his responsibility as well. Therefore there was a need to plan actions against segregation not only between schools but also within schools.

The integration normative provides support for the schools since the sum of the normative support for primary school children was in September 2003 187 thousand Ft for classes 1-4, and 194 thousand Ft for classes 5-8, starting from September 2004 it was 193 and 202 thousand Ft.

According to the legal regulations, children with special needs "receive capability development training from the school, which means a personal, individual development program for each child, and the responsible teachers and the parents make an assessment of the progress together with the students every three months."[18]

The next point of the legislation that aimed the creation of integrated schools and classes was even more important for the office. The teachers educating children with special needs have to develop their capabilities as described above; at the same time it is not possible to separate poor and rich children from each other.

2. Desegregation

If in a school there are parallel classes in a grade than the difference between the ratios of poor children attending different classes can not be more than 25 percent. This means that if there are two classes with headcount 24 and in class A there are 12 poor children the number of poor children in class B can range from 6 to 18. This principle is also valid in the schools: in a settlement disadvantaged children should be present proportionally in each local school.

After two years the commissioner's office has initiated modifications based on follow-ups and inspections. The main directive of this was that the segregation should apply for all children living in the given locality and the common learning and upbringing of poor and non-poor children should affect all of them not only public, but private schools, run by churches or foundations,

3. The National Development Plan[19] (NDP)

In the framework of measure 2.1 of the Operative Programme for Human Resources Development – "Securing equal opportunities for pupils with special needs in the educational system" – in the time period 2004-2006, nearly 7.7 billion HUF were allocated for projects supporting equal opportunities for pupils with special needs. The programme consists of central pedagogical developments and applications.

[18] 11-1994. (VI.8 MKM decree 39/D and 39/E.).
[19] Reference on the report.

The developments of the core programme 2.1 started off in the framework of the following five projects:

» Support of the introduction and application of complex pedagogic development programmes at universities.
» Programmes concerning the development and implementation of further teacher training and training for specialists.
» Development and implementation of programmes that increase the support regarding integrated education, and increase the society's acceptance and understanding of these programmes. These programmes target the local decision-makers, the local self-governments, the financers, and non-teacher specialists.
» Development of know-how on integrated education, implementation of a methodological database and of service provider programme packages.
» Development of research projects, prevention methods and models, preventing early drop-out and early recognition of risks for drop-out.

The direct goal of the planned developments is that more than 10,000 teachers – and people supporting pedagogical work – should be able to sustain the effective integrated education of pupils having different backgrounds and having special needs. The goal is that this should become widespread in public education,[20] and that between 2004-2006 the winner of the application should support the professional aid on pedagogical and institutional development in at least 270 public institutions.

The components of the application regarded the following programmes:

2.1.2 Integral education of pupils with special needs
2.1.3 Support for the preparation of pupils with special needs
2.1.4 Support for extracurricular activities of type "tanoda" (informal place of learning),
2.1.5 Integral education of cumulatively disadvantaged pupils
2.1.7 Decrease in school segregation
2.1.8 Development of schools in localities where only one school operates and where cumulatively disadvantaged pupils learn

4. National Educational Integration Network (NEIN)

The core tasks of the network (NEIN):

» Promotion of an inclusive, integral pedagogical culture.
» Help the abolishment of detected segregation forms in the Hungarian educational institutions.
» assist institutions that receive educational normatives for Roma pupils

[20] The 2.1 programme does not deal only with disadvantaged and Roma pupils, but also with pupils with special educational needs.

On the basis of applications, 45 elementary schools in four regions (South Transdanubia, North Hungary, North Great Plain, Central Hungary) were selected in order to become models for other schools. Specialised regional and local coordinators were working with these institutions. The mandate of these specialists was to develop professionally the selected model-institutions (development of integrated education) and to provide know-how to other institutions (promotion of an integrated education attitude) in order to shape the social environment.

The network organises integral pedagogical trainings, professional forums and suitable workshops. In Hungary, there is no notable expertise in how teaching staff learns from each other.[21] The complete teaching staff should understand integrated education; however, this is not sufficient either. The neighbourhood of the school and the nearby schools would have to become familiar with integral education. The National Institute for Public Education (Országos Közoktatási Intézet), on behalf of the Ministry of Education and Culture was monitoring the institutional effectiveness of the NEIN. They investigated various factors: the use of pedagogical-methodological tools of the model-institutions, their everyday pedagogical practice, forms of cooperation between schools and parents, and the way these institutions were prepared to transfer the accumulated knowledge to other non-model schools.[22]

The Ministry of Education and Culture also commissioned research on the usage of the per capita normative support for integration at the school level.[23] The outcome showed that institutions located in cities and in small villages, respectively from the South-Western region requested the normative for integrated education at rates higher than the national average. At the same time, schools located in the North-West and Central regions (regions where the percentage of the Roma is lower) the requests were under the national average. Consequently, the integration normative reached the targeted groups. The researchers surveyed 568 schools, from which 554 school-directors expressed their opinions about the introduction of the integration normative. According to the school-directors, in 27 per cent of the schools everybody agrees with the necessity of introduction of integration normative; in 29 per cent of the schools there is dissatisfaction; and in 24 per cent of the schools there was rejection from non-Roma parents. The researchers observed that the directors of those schools show dissatisfaction, where the number of Roma pupils is high, and because of the ethnic composition of the pupils they have no possibility to introduce integrated education. According to the results of the research, the integration normative did decrease segregation, but could not stop the process of segregation between schools.

[21] In the Self-Generative School programme of the Soros Foundation, there was a successful attempt – based on a collaboration between schools and financers – where they succeeded to learn from each other.

[22] Research results see in: Integration in Practice (ed. Szilvia Nemeth), Országos Közoktatási Intézet, Budapest, 2006, www.ofi.hu/tudastar.

[23] Research results see in: Final research report, Segregation in Education of Roma Students in Primary Schools, (Ilona Liskó-Gábor Havas) Felsőoktatási Kutatóintézet, Budapest, 2004.

5. Multicultural Elements in Education

The Act on Public Education prescribes the teaching of the minorities' culture and history for those belonging to the majority; however, effective steps were not taken. Hungarian society has extremely limited knowledge on the Roma. The most effective means to decrease prejudice – as the results of a "deliberative poll" demonstrate – is to increase the knowledge of both the majority and minority pupils. An indispensable condition to reach this is integrated education, in order to achieve that everyone should know about the Roma and about the shared history.

During the review of the National Core Curriculum an amendment was introduced: "At all levels of education, each pupil has to acquire familiarity on the culture and shared history of minorities that compose the nation."

6. Educational Materials and Curriculum Policy[24]

Textbooks are free to eligible children – such as those who receive the permanent child protection allowance. It is a State obligation to provide teachers and textbooks in minority languages.[25] The Minister has the power to fix the price of textbooks printed in fewer than 1,000 copies. Notably, however, the latest ministerial information sheet on the maximum price of textbooks in minority languages does not contain a single textbook written in Romanes.[26]

Ministry of Education officials argue that certain national minorities (e.g. Germans and Jews) are more frequently referred to in the many textbooks that are available on the market, which is completely liberalised. Since no authoritiative studies or analysis has been done of the large existing body of schoolbooks in Hungary, the extent of references to Roma, in history or literature, cannot be tracked.

The National Core Curriculum (NCC) contains the framework curriculum on the basis of which individual schools have to adopt their pedagogical programmes (the actual curriculum taught in the school). The NCC prescribes that "every child at every level of education shall be acquainted with the culture and common history of the minorities constituting the nation." In the chapter "Man and Society" the NCC stipulates that in grades 9-12 pupils

[24] Shortened extract from monitoring report: Equal Access to Quality Education for Roma, Hungary, 2007. p. 237.

[25] Pursuant to Art. 44 of Act No. 77 of 1993 on the Rights of National and Ethnic Minorities, the additional costs of minority mother-tongue education shall be borne by the State. Pursuant to Art. 46 (2), teacher training and retraining with a view to ensuring minority language education is a State duty.

[26] Ministry of Education Information Sheet, listing the highest price of school books on ethnic and national minority language, literature and history printed in less than 1,000 copies in 2006-2007, 28 February 2006, available at http://www.om.hu/main.php?folderID=723&articleID=6969&ctag=articlelist&iid=1 (accessed on 24 February 2007).

shall be provided with information on the history of national and ethnic minorities living in Hungary, including that of the Roma. Some textbooks on "Man and Society" contain such information. However, some reflect deeply rooted anti-Roma stereotypes and bias.[27]

According to Government Decree 243/2003 (XII.17) on the Publication, Introduction and Application of the National Core Curriculum,[28] minority education aims at preserving and strengthening minority self-identity. Its objectives therefore relate to the use of minority languages, the cultivation of minority culture, the knowledge of the history of the so-called "Mother Country" (which is not supposed to be applicable to Roma, regardless of their much earlier Indian provenance), tolerance and the social advancement of Roma. Undoubtedly, as the mandate of the Ministerial Commissioner in charge of the Integration of Roma and Socially Disadvantaged Children (Integration Commissioner) demonstrates, the objectives relating to Roma have somewhat changed in recent years. In his latest activity report the Integration Commissioner noted that contrary to an express obligation in Article 48 (1)b PEA, the history and culture of Roma has not been taught to majority children.

7. Minority Education Policy[29]

Ministerial Decree No. 32/1997 (XI. 5) MKM on the Guidelines of National and Ethnic Minority Education in Pre-Schools and Primary Schools governs education in the minority language. The Act on the Rights of National and Ethnic Minorities No. 77 of 1993 recognises both Romanes and Beash as ethnic minority languages spoken by Roma.[30] Procedural laws allow for the use of one's mother tongue, regardless of citizenship, but official forms are missing in both languages spoken by Roma. No information as to how many children using Romanes or Beash as their mother tongue and who are also proficient in the majority language is available.

[27] Interview with Gábor Daróczi, March 2006.

[28] Ministry of Education website Information sheet 3 December 2000 on the role of the National Core Curriculum] http://www.om.hu/main.php?folderID=391&articleID=2337&ctag=articlelist&iid=1 (accessed on 24 February 2007).

[29] Shortened extract from monitoring report: Equal Access to Quality education for Roma, Hungary, 2007. p. 229.

[30] Art. 42 MA. This provision is located under the heading of the cultural and educational self-governance of minorities.

8. Establishing Institutions Monitoring Discrimination[31]

Administrative Procedures

The Equal Treatment Authority has authorization to act against any discriminatory act irrespective of the ground of discrimination (including race and social status) or the field concerned (including education).

Pursuant to Article 80 PEA, notaries at the local or county level are in charge of controlling the legality of operations in private schools. Parallel to suspending the transfer of normative budgetary support, pursuant to complaints by parents or NGOs, notaries can challenge an illegal action, decision or omission in court. If a private school continues to fail to comply with the law, notaries have the power to revoke the permission and strike the school out of the registry.

In case teaching violates public order, public health, public morals, is directed against the fundamental rights of others or if the school lacks the conditions necessary for its operation, the competent notary may order it to terminate such a situation. Should the school fail to comply, the notary has the power to suspend its activities and challenge its action in court. The procedure following such a suspension is as described above.

Articles 83 and 84 of the Public Education Act provide for an administrative complaint mechanism against unlawful decisions of a school (to the maintainer) or the maintainer (to the notary or in case of local government run schools to the Office of Public Administration). Decisions that discriminate are null and void. Judicial review is available against such decisions, but in the latter case can only be brought by the Office. It shall be noted that sanctions available under Article 80 of the Public Education Act against unlawful acts of private schools seem far more effective than those available against public schools. Ultimately, public schools cannot be closed down, nor can State funding be withheld from them.

The National Public Education Evaluation and Examination Centre (OKÉV) can also investigate discrimination in schools and impose fines for petty offence or as a result of its administrative review pursuant to Article 95/A(5) PEA.

Conciliation Procedures

Mediation by the Equal Treatment Authority

The ETA does not explicitly authorize the Authority to mediate between parties, but under Article 64 of the GPSA, the Authority, as a public administrative organ, is authorized

[31] Shortened extracts from monitoring report: Equal Access to Quality education for Roma, Hungary, 2007. and Discrimination in Education – UNESCO report, 2006.

to try to resolve the conflict through forging an agreement between the parties, if the circumstances of the case seem to allow it. Pursuant to Article 75 of the Act No. 140 of 2004 on the general procedure and services of state authorities (GPSA), if the parties reach an agreement in the course of the complaints procedure, the Authority includes the agreement in a formal decision.

If the attempt to have the parties reach an agreement is not successful, the Authority continues its proceeding, and – depending on the result of the investigation – imposes a sanction or rejects the complaint.

Education Commissioner

Decree 40/1999 of the Minister of Education established the Commissioner for Educational Rights. Under Article 1 of the Decree, the Office of the Commissioner for Educational Rights is an independent, internal organizational unit of the Ministry of Education that promotes citizens' rights concerning education. The Decree establishes a special conciliation procedure.

Parents, students, teacher etc. have the right to complain, provided that all available administrative remedies are exhausted and less than a year has elapsed since the measures complained of (Article 5). Complaints relating to Articles 70/F and 70/G of the Constitution, public education, higher education and vocational education and training can be brought to the Commissioner (Article 3). The explicit inclusion of Article 70/A of the Constitution (on non-discrimination) in the scope would be highly advisable. In 2001 the Commissioner examined discrimination only in relation to disabled students.

Complaints not dismissed by the Commissioner undergo the conciliation procedure. The Commissioner sends the petition to the institution complained of for a declaration and initiates that consensus be reached with the petitioner. In case of an agreement the Commissioner prepares a report and sends it to the parties concerned. If no consensus is reached, the Commissioner prepares a report on the results of the conciliation and calls on the institution to terminate the infringement. In case of non-compliance the Commissioner sends a recommendation to both the institution and its supervisory organ. The latter have the duty to respond within 30 days. The Commissioner reports to the Minister of Education (Article 7). In 2001 the Office issued initiatives and recommendations on 51 occasions. Following changes in the Ministry's leadership, the Commissioner's office has been restructured. Information as to his present practices is not available.

The Ministry has established another conciliation procedure: the *Education Mediation Services*, which has a handful of staff.[32]

[32] Its mandate is based on Act No. 55 of 2002 on mediation in civil law disputes. It is not expressly mentioned in the PEA.

The Parliamentary Commissioners

Under Article 32/B of the Constitution, the Parliamentary Commissioners investigate violations of constitutional rights and initiate general or individual measures to remedy such violations. There are currently four such commissioners in Hungary: the Parliamentary Commissioner for Civil Rights (General Commissioner), the Deputy Commissioner for Civil Rights, the Parliamentary Commissioner for the Rights of National and Ethnic Minorities (Minorities Commissioner) and the Commissioner for Data Protection.

Under Act No. 59 of 1993 parliamentary commissioners are appointed by a two-thirds parliamentary majority vote. Financial independence and immunity are provided for. Any victim of acts or omissions of public authorities or public service providers can complain to their office, provided that all administrative remedies are exhausted or none exist. Commissioners can proceed ex officio and the Minorities Commissioner has done so on many occasions in relation to the segregation of Roma children in education.

They can investigate into any authority, including local governments. They may request information, a hearing, written explanation, declaration or opinion from the competent official or demand that an inquiry be conducted by a superior. When finding a violation, they issue recommendations, to which perpetrators must respond within 30 days. Further, they may:

» Petition the Constitutional Court.
» Initiate that the prosecutor issues a protest.
» Propose that a legal provision be amended, repealed or issued.

Commissioners may initiate disciplinary or criminal proceedings. Their main publicity weapon is their annual report submitted to Parliament. Further, they can request parliamentary investigations and debates.

Accessibility, though a concern, is greatly promoted by the good level of cooperation among these bodies.

Inspections

The form and content of measures that inspections can impose are left up to school maintainers, and may include disciplinary sanctions against teachers and school directors. Maintainers are also free to report crimes – such as the endangering of a child – to the police. Although OKÉV has the right to review school activities, its sanctioning powers are extremely limited. It can either fine – up to the equivalent of € 400 – teachers and other

officials for intentionally discriminating against children[33] in the course of public education or the Ministry can deny the registration of schools that are in serious breach of the law.

There are no county inspectorates, the system is central (OKÉV) and school maintainer based (local or county government appointed educational experts). No public data is available either on Roma public education experts, or on OKÉV equal treatment inspectors. NIN employs several Roma monitors and colleagues who undertake inspection like visits to notoriously segregating schools and feed their findings back into the system. Their reports seem to show a significantly different attitude to segregation than the reports by official inspectors.

[33] Pursuant to Article 142(5) of Government Decree No. 218/1999. (XII.28.) it is a petty offence to discriminate against a child by intentionally breaching legal rules on public education. No imposition of such fines has so far been reported.

The second volume in the Roma Education Fund Working Paper Series, addresses a key topic: what is the impact of desegregation on children's educational outcomes? This is obviously an important issue, but it is a key topic for REF since we have often supported improvements in Roma education through desegregation programs. But the pursuit of desegregation should not be based just on belief that it is the right thing to do; our actions must be grounded in evidence. And the central piece of evidence is whether desegregation contributes to more success in school for Roma children.

Moreover, in pursuing desegregation, as a tool for improving the educational outcomes for Roma, we must also consider the impact on other children. Serious policy (and political) questions would be raised if desegregation only helped Roma children.

This study, whose quasi-experimental design ensures robust conclusions, demonstrates that this is emphatically not the case: both Roma and non-Roma children studying in schools participating in the Hungarian Government's desegregation program did at least as well and usually better than children in comparable schools across a range of different outcomes. Moreover, the study begins to identify why the program is successful.

This study is important reading for all those responsible for policies and programs that promote social inclusion for a better society and more productive economy.

Toby Linden
Director
Roma Education Fund

The title of Working paper №1 that was published in 2006:
"Expected Long-Term Budgetary Benefits to Roma Education in Hungary". If you wish to download this publication, please visit the website of the Roma Education Fund at:
http://www.romaeducationfund.hu/documents/Kertesi-Kezdi-BudgetaryBenefits.pdf

ROMA EDUCATION FUND
Mark Center ■ Teréz körút 46 ■ H-1066 Budapest, Hungary
Tel. +36-1-235-8030 ■ Fax. +36-1-235-8031
info@romaeducationfund.org ■ www.romaeducationfund.org