Immigrant Children's Cognitive Outcomes and the Effect of Ethnic Concentration in Danish Schools*

Peter Jensen† Astrid Würtz‡
CIM and Department of Economics,
Aarhus School of Business, University of Aarhus

April, 2007

Preliminary, please do not quote without permission

Abstract

This study uses a unique PISA-dataset from Denmark to investigate whether immigrant children and children in schools with high ethnic concentration have lower cognitive outcomes than native children in schools with low concentration of immigrants. We find that immigrant children from non-Western countries tend to have lower reading test scores than ethnic Danish children, and that children in schools with a high ethnic concentration score significantly lower in a reading test than children in schools with low ethnic concentraion. These results are robust across estimation methods. Immigrant children's lower outcome is related to the ethnic concentration in the schools they attend and their relatively low socio economic status. Instrumenting for ethnic concentration reveals that even after taking into consideration that individuals may sort across neighbourhoods, ethnic concentration in the school and the child's own ethnicity are still important factors in determining child outcome. School fixed effects estimations show that also relative age and gender are important factors in determining child outcomes. Finally, there is a strong positive effect on children's cognitive outcome of speaking Danish at home.

JEL Classification: .
Keywords:

*We would like to thank Nina Smith and Anna Piil Damm for helpful suggestions and Pia Elisabeth Kristensen and Ulrik Brandt Jørgensen for competent research assistance. Finally, we would like to thank the Rockwool Foundation for providing the data.

†E-mail: pje@asb.dk
‡E-mail: awu@asb.dk

1 Introduction

Denmark has one of the most expensive educational systems in the world. It is therefore important to target policy efforts and resources where they have an effect, especially with respect to optimizing inputs into schooling. There is a lack of studies on causal effects between children's ethnic background and their educational outcome and we will therefore focus on determining these effects. We will answer whether immigrant children in Denmark are worse off than native Danish children with respect to educational outcomes and we will suggest possible explanations for the relationship. The focus will be on the effect of ethnic concentration in the schools and children's own background. Finally, based on the estimations we will give some policy suggestions.

One can think of several explanations for why immigrants in Denmark and native Danes do not have the same educational outcomes. First, children's cognitive outcome may be affected by e.g. role models where they live. This is referred to as neighborhood effects. One can easily imagine that if a child grows up in a neighborhood dominated by individuals with a low socio economic status then the effect on the child may be negative because the child cannot mirror itself in positive role models. As seen in Figure 1, the employment rate is lower among immigrants from non-Western countries compared to native Danes. Especially the share of non-Western immigrants out of the labor force is much higher than for native Danes[1]. Therefore, if neighborhood effects are important for determining child outcomes then the share of immigrants and the composition of ethnic Danes in the neighborhood is potentially very important.

Peer effects, i.e. the mix of students in the child's class, may create or shape incentives or disincentives for studying. Family background, values, and tradition for prioritizing school work will probably also affect children. Resources may also vary between schools and have an impact on children's outcome. There can for example be variation in teacher quality, teacher/student ratio, class size, and quality of books between and within schools. All these different school related effects are referred to as school effects in short.

[1] Immigrants from Western countries tend to have a socio economic status much closer to that of ethnic Danes. The difference between 2nd generation immigrants from Western countries and ethnic Danes is minor whereas the difference between 1st generation immigrants from Western countries and ethnic Danes is somewhat bigger but not as big as for immigrants from non-Western countries. These numbers are not shown here but are available from the authors upon request.

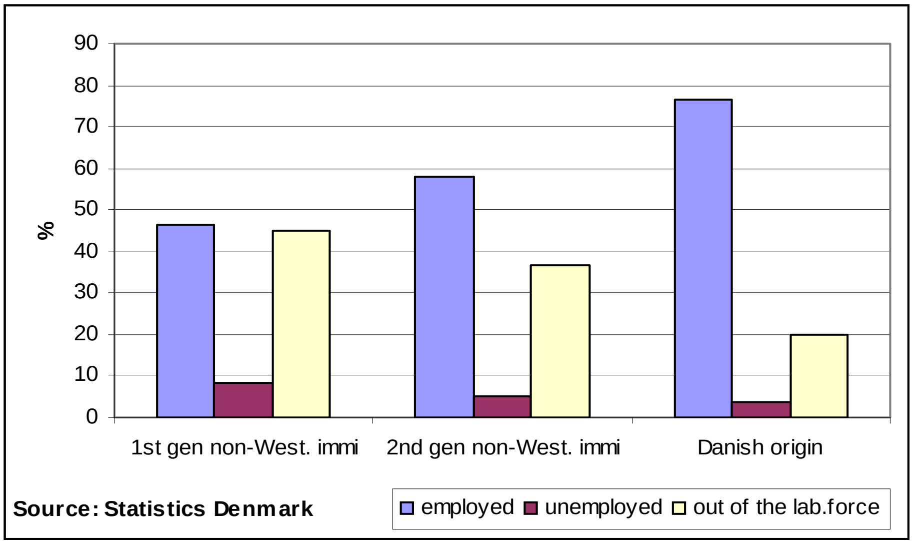

Figure 1: Percentage of individuals aged 16-64 who are employed, unemployed, or out of the labor force within 1st generation immigrants from non-Western countries, 2nd generation immigrants from non-Western countries, and individuals with Danish origin, Denmark 2005.

We show in the empirical section that immigrant children from non-Western countries tend to have a lower cognitive outcome than ethnic Danish children. These results are robust across estimation methods. The lower outcome of immigrant children is related to the ethnic concentration in the schools the children attend. It is clear from the estimation results that the low socio economic status of the immigrants is also important. Instrumenting for ethnic concentration reveals that even after taking into consideration that individuals sort across neighborhoods, ethnic concentration in the school and the child's own ethnicity are still important factors in determining child outcome. Most of the results from the simple OLS estimations generally seem to be robust even though the coefficients in most cases decreases slightly in the school fixed effects estimation. The significant negative effect of having non-Western immigrant background and being 16 years old, and the positive effect of speaking Danish at home seem to be stronger after taking school fixed effects into account, though.

In section 2 we take a look at a selection of the literature relevant for this study. Section 3 introduces the three different estimation methods and section 4 describes the data used in the estimations. Estimation results are provided in section 5 and, finally, section 6 concludes.

2 Literature

The literature on immigrants and immigrant outcomes suggests that there are several possible endogeneity problems to be aware of. First, there is a potential problem with determining neighborhood effects. Ethnic concentration is probably endogenous because individuals sort into neighborhoods because they want (or don't want) to live in a ghetto area. Since we do not have random assignment to housing in general in Denmark we have to consider this potential endogeneity problem in the empirical analysis[2].

Further, when investigating children's cognitive outcomes one has to recognize that peer effects may also be endogenous. The peer group can be the result of individual choices, e.g. unobserved school quality. Several recent studies are addressing the issue of peer effects, see e.g. Hoxby (2000), Hanushek et al. (2003), Entorf and Lauk (2006), and Ammermueller and Pischke (2006), and it is argued that peer achievement has a positive effect on achievement growth. In addition, different school types are also likely to determine the size of peer effects[3]. In many cases it is difficult to separate neighborhood effects and peer effects since most children attend schools in the proximity of their residence. Parents' housing decision may also simultaneously depend on choices about preferred neighborhood and school.

Lastly, as Hoxby (2000) emphasizes, school resources are probably endogenous since they are not randomly allocated to municipalities, schools or students. Therefore, parents may choose housing based on schools in the area. Furthermore, school administrators, teachers and politicians affect school resources by allocating more resources to schools with many "weak" students, or create smaller classes for weak students.

Omitted background variables related to outcomes from schooling might be correlated with school resources and cause bias (up or down) in single equation estimates. Hanushek et al. (2003) argue that issues of omitted and mismeasured variables are likely to be more important than for example whether peer interactions are simultaneously determined. Omitted variables bias can be addressed using instruments of class size, teacher/student ratio

[2] For a small group of individuals, refugees, we have seen random assignment to housing as a result of the governments's refugee placement policy in the period 1986 to 1998, see e.g. Damm and Tranæs (2006).

[3] In general, it can be very difficult to identify peer effects because of the socalled reflection problem, see Manski (1993).

etc., see e.g. Browning and Heinesen (2003). Another way to deal with omitted variables bias is by including a large set of controls, and this is the approach we will take. Finally, reverse causality is also a potential source of mismeasurement.

Todd and Wolpin (2003) emphasize that for an empirical investigation of children's educational outcomes one optimally needs information on both family and school inputs, current and past. This ensures a correct specification of the child's cognitive achievement production function when considering child development a cumulative process depending on the history of family and school inputs as well as on inherited endowments.

The early childhood development (ECD) model and the education production function (EPF) model are production function models often used for the study of child cognitive achievement. In many studies ECD or EPF models are estimated without having a dataset with rich enough information. In ECD studies the datasets lack information on institution and school inputs, whereas in EPF studies there is often little information on family inputs available in the datasets. Furthermore, there seems to be a lack of consensus over which variables to include in the different specifications. This implies that researchers working on the same data source can find completely opposite effects of e.g. maternal employment on children's cognitive achievement, see also Todd and Wolpin (2003).

The data sources that will be used in this study do not suffer from the usual problems with missing information on either family or school inputs. The register information covers families back to the year of child birth and the surveys contain extra background information not included in registers. Furthermore, some school information is also included in the registers and the surveys. Using a dataset with rich information will improve identification and reduce omitted variables bias and allow for several different estimation methods. We take a closer look at these methods in the next section.

3 Empirical Model

To address the issue of whether immigrant children are worse off than native Danish children with respect to cognitive outcomes and to investigate whether ethnic concentration in schools affect children's outcomes several different estimimation methods can be used. For the purpose of identifying causal effects on child outcome one can either use exogenous variation in school input variables, "natural experiments", longitudinal data, e.g. take school fixed effects into account, or use instrumental variables. The three latter approaches are used in this study.

We base estimations on the EPF model as defined by Todd and Wolpin (2003). The general (or cumulative) expression for the empirical model is given as

$$T_{ija} = T_a \left[F_{ij}\left(a\right), S_{ij}\left(a\right), \mu_{ij0}, \varepsilon_{ija} \right], \tag{1}$$

where T_{ija} is a measure of achievement for child i from household j at age a, F is a vector containing the history of parent-supplied inputs up to the given age a, and S is a vector containing the history of school-supplied inputs. μ_{ij0} is the child's endowed ability, and ε_{ija} is a measurement error in test scores. The impact of inputs and of the genetic endowment is allowed to depend on the age of the child. The biggest challenge of using this model is that μ_{ij0} is unobservable, and information on both contemporaneous and historical family and school inputs is needed for consistent estimation[4].

Another estimation strategy is to specify traditional production functions for immigrants but it seems more reasonable to use education production functions and then further control for immigrant background as done by Rangvid (2005) in a PISA-Copenhagen study with focus on immigrants.

Three main estimation methods are employed in this study. We first use simple OLS estimation even though we know that including endogenous variables in an OLS estimation might imply biased estimates. This is to get a reference point for the other estimation methods. Second, instrumental variables (IV) estimation is used where instruments for ethnic concentration in the child's neighborhood are included. Also, an extensive set of family, school, and background variables from both the survey and the registers is included. Third, we use school fixed effect estimation (FE) to remove observed and unobserved fixed school effects that may affect children in the same school in the same way[5].

In all estimations it is assumed that T_{ija} can be expressed as a linear function of the explanatory variables. We simplify the notation and in the

[4]Specifications as e.g. the value-added specification or the contemporaneous specification have not been chosen here because they both require very strong assumptions for identification. In the contemporaneous specification, for example, only contemporaneous measures on school and family inputs are used and this of course demands strong assumptions for identification - e.g. assuming that only contemporaneous inputs matter to the production of current achievement. Our data sources provide information not only on contemporaneous variables and therefore it is logical to use a less stringent estimation method.

[5]We also would like to use FE on individuals to remove individuals' unobserved heterogeneity but since we only have one PISA test score per child this method is not feasible with our data. Further, we do not have school class indicators so we cannot within schools do FE estimations.

OLS regression we therefore estimate

$$T_i = \beta_0 + \beta_1 S_i + \beta_2 F_i^1 + \beta_3 F_i^2 + \varepsilon_i, \tag{2}$$

where family inputs are split into ethnic concentration in the neighborhood, F_i^2, and all other family and background variables, F_i^1. Some of the variables included are only available for the year the child is tested whereas e.g. information on parental employment and work experience is know back in time. The error term, ε_i, is assumed not to be correlated with the explanatory variables to ensure consistent OLS estimates.

In the IV estimation[6] we recognize that ethnic concentration in the neighborhood, F_i^2, might be endogenous, i.e. that the OLS estimator might be biased because F_i^2 and ε_i are correlated. We therefore introduce an instrument for ethnic concentration, Z_i. As instrument we use the share of children living in non-profit rental housing among 9th graders in the school. Many immigrants live in this type of housing and it therefore appears to be a good instrument for ethnic concentration. Second, following Dustmann and Preston (2001) we also use ethnic concentration in a larger geographical area as instrument for local ethnic concentration. Arguing that mobility is geographically limited by employment and family history this seems to be a reasonable instrument for local ethnic concentration. We test the strength of these instruments and argue for their validity in section 5.2. It is important to notice that we may encounter problems with the estimator if the instruments are weak, see e.g. Wooldridge (2002), Bound et al. (1995), and Staiger and Stock (1997), since standard errors using IV methods tend to be large, especially for weak instruments. This consequently makes it harder to find significant effects. In finite samples the IV estimator is biased towards the OLS estimator but the bias decreases the stronger the instrument is and the more observations we have.

Formally, when using IV techniques the endogenous regressor's value is predicted and the predicted value is used as a regressor in the original model, i.e. first the endogenous variable is regressed on the instrument and all exogenous variables in the model using OLS,

$$F_i^2 = \delta_0 + \delta_1 S_i + \delta_2 F_i^1 + \delta_3 Z_i + r_i \tag{3}$$

and thereafter we insert Equation (3) in Equation (2) and estimate

$$T_i = \alpha_0 + \alpha_1 S_i + \alpha_2 F_i^1 + \lambda_1 Z_i + \nu_i,$$

[6] Actually, we employ 2SLS estimation because we have more than one instrument. In the following we will refer to the 2SLS estimation as IV estimation.

where $\nu_i = \varepsilon_i + \beta_3 r_i$, $\alpha_j = \beta_j + \beta_3 \delta_j$, and $\lambda_1 = \beta_3 \delta_3$. If the instrument is truly exogenous we get consistent OLS estimates in the second stage, otherwise we risk getting large inconsistencies in the IV estimates even if the correlation between the instrument and child outcome is weak, see Bound et al. (1995).

Finally, when taking school fixed effects into account we remove all observable and unobservable factors that are fixed and common to children in the same school. This is done without distinguishing between observed and unobserved school constant factors. It enables us to remove effects caused by resources per child, teacher quality, and the effect the principal has on the school. It also removes e.g. effects of ethnic concentration in the school since this factor is fixed within schools. We estimate the following equation

$$T_{is} = \beta_0 + \beta_1 S_{is} + \beta_2 F^1_{is} + \beta_3 F^2_{is} + \gamma_s + \varepsilon_{is},$$

where s denotes the school and γ_s is the fixed effect in school s.

4 Data

The data used for studying children's cognitive achievement is a combination of a unique survey data and Danish administrative register data.

The survey data consists of the Danish subsample of the OECD Programme for International Student Assessment (PISA) study from the year 2000 which is combined with a special PISA study from 2005 with an oversampling of immigrants in Denmark (PISA-I). The children in PISA-I are given the exact same cognitive test as the random sample of Danish children participating in the PISA-2000 study. For information on the PISA-2000 study, see OECD (2002).

Children participating in the PISA studies are equipped with identifiers to combine PISA information with register data from Statistics Denmark. This gives us information on the child and its parents from child birth to the year 2005 (2000 for the children participating in PISA-2000). Furthermore, PISA-children have answered questionnaires about their family background, and school principals have answered a questionnaire about school resources etc. This information is also linked to the child. We include a wide variety background variables from questionnaires and registers in our estimations to decrease possible omitted variables bias.

The dependent variable in this study is the child's PISA reading score[7]. The PISA-2000 study focuses on children's reading abilities but also tests

[7] We use the WLE score since we are not comparing between countries. The WLE score is calculated by the ACER institute in Australia which is responsible for all PISA analyses

some of the children in mathematics and science. We focus on the reading score since all children are tested in reading. As mentioned, the cognitive test for PISA-I children is the same as for PISA-2000 children and we therefore compare reading scores from the two studies.

As part of the sampling criteria all children participating in PISA-2000 are 15 years old. The PISA-I sample, on the other hand, is collected from 9th grades and the children are not sampled by age. Therefore, we have children in the PISA-I sample who are older than 15. 80% of the sample is 15 years old whereas 95% is 15 or 16 years old. We keep children in the ages 15 and 16 from the PISA-I sample but note that this might bias test scores slightly[8].

Different sampling in 2005 compared to 2000 implies that we might have to deal with sample selection issues. First, the schools chosen for participation in PISA-I are chosen from the top based on the number of immigrants in schools in 2003. This in itself poses sample selection issues. Further, as just mentioned, the children are not all the same age as in the PISA-2000 study but are instead sampled on the basis of the grade they attend.

On the other hand, ignoring potential sample selection issues, we have a unique sample of children where a huge share of the children have non-native Danish background. We therefore have a sample that is big enough to actually provide useful information about immigrants in Danish schools. Comparing with children from PISA-2000 who are in "Danish" schools we obtain relevant and unique information which is not available in most other countries.

based on OECD's concept. WLE scores are not simple observed test statistics but instead based on estimated models for all countries participating in the analysis under study. Questionnaires from PISA-2000 are used along with estimated models from PISA-2000 to calculate PISA-I test scores. Standard deviations are calculated as simple standard deviations which do not take into account that each WLE score has been calculated, i.e. is not directly observed. In regressions and for statistical tests we will use the PISA test scores *as if* they were observed test scores. This is the only practical solution for performing statistical analyses using this type of data and it is the same method used in other Danish and international studies. For a critical discussion of the PISA measures see Allerup (2005).

[8] In 2000 a separate survey was also completed among a large group of 16 year olds in Denmark. They were given the exact same PISA-test as the 15 year olds and it turned out that the results were a little better but still very similar to the results for 15 year old Danes, see Andersen et al. (2001). In our sample with many immigrant children we might instead find that keeping 16 year-olds in the sample will bias test scores downwards. These older children may namely be a selected group in the sense that they have delayed schooling because of language difficulties, for example. Descriptive evidence suggests that 16 year-olds in the PISA-I sample to a larger degreee are 1st generation immigrants than 2nd generation.

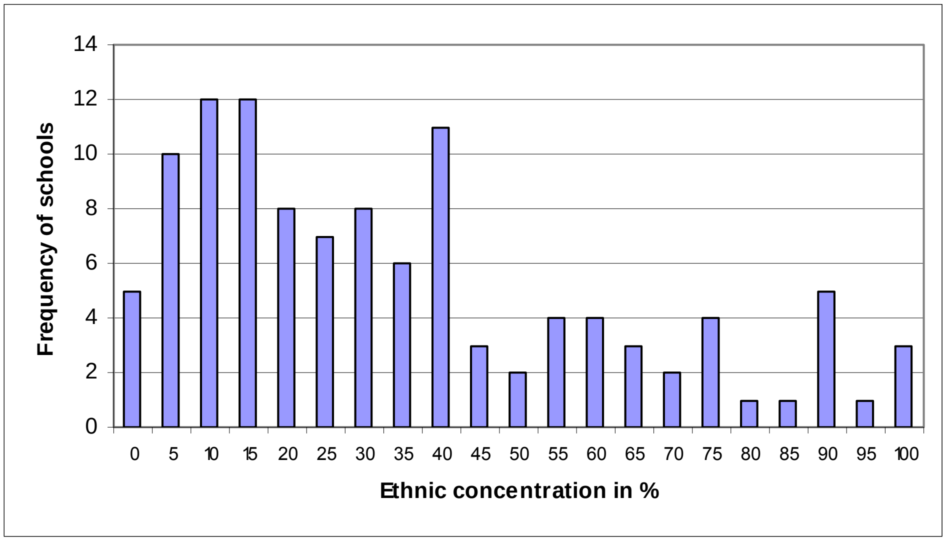

Figure 2: Distribution of ethnic concentration in schools in the PISA-I sample (in groups of 5%).

The estimation dataset is a combination of parts of the PISA-I sample and a part of the PISA-2000 sample. We keep immigrants from non-Western countries from PISA-I along with their ethnic Danish school mates. Immigrants from Western countries are dropped since they only consist of 1.5% of the sample and they have very different characteristics than the non-Western immigrants. Further, we keep ethnic Danes from the PISA-2000 dataset but only those without immigrant school mates[9]. This group of ethnic Danish children is the comparison group along with the small group of individuals from the PISA-I dataset where the share of non-Western immigrants in the 9th grade is below 5%[10]. This amounts to about 4% of the PISA-I sample.

Figure 2 shows the distribution of ethnic concentration in schools in the PISA-I sample. The degree of ethnic concentration in the schools varies a lot. Most of the schools have less than 40% non-Western immigrants but we also see that some schools have more than 90% non-Western immigrants.

[9] We define a 9th grade to be "ethnic Danish" if less than 5% of the children are non-Western immigrants. 47.25% of children in the PISA-2000 sample does not have any non-Western immigrant school mates, 67.39% has 5% or less.

[10] In the original PISA-2000 dataset only 1.2% of the sample is Western immigrants. These individuals are dropped in our estimations.

Table 1 shows means and standard errors for all variables included in the estimations. For comparison, means for individuals in the PISA-2000 and PISA-I samples are shown separately along with means for the main estimation sample, the pooled PISA-2000 and PISA-I sample. The datasets are restricted as described above.

It is clear that the PISA test score in reading is much lower for individuals in the PISA-I sample than for the Danish children in the reduced PISA-2000 sample. The ethnic concentration is of course much higher in PISA-I children's schools and so is the ratio of school mates living in a problem housing estates. This latter finding is not surprising since immigrants tend to live in non-profit rental housing and only this type of housing can be categorized as a problem housing estate[11]. We further notice that ethnic concentration in the county is higher for PISA-I children which indicates that they dominantly live around the big cities.

16% of children in the PISA-I sample are 16 years old[12] whereas all children in PISA-2000 are 15 years old because of the sampling. We control for age in the estimations in order to take this into account. PISA-I children's parents are worse off than PISA-2000 parents with respect to length of education, occupation, labor market attachment, family gross income[13], and work experience. This might be one explanation for PISA-I children's lower test scores so we look more into this in the estimations below.

If we look at the study environment for PISA-2000 and PISA-I children it seems to be quite similar. The amount of books in the households is almost the same and the availability of a quiet place to study is almost the same as well. The biggest differences between households seem to be in the number of siblings and in the language spoken. Fewer PISA-I children speak Danish with their parents which is not surprising.

Finally, there seems to be some systematic differences between PISA-2000 children's schools and PISA-I children's schools[14]. PISA-I children attend bigger schools and the schools are more often public schools than for the PISA-2000 children. The number of students per class is quite similar, though. The amount of time the children spend on doing homework is similar across the two samples but PISA-2000 children tend to be more absent from school.

[11] The concept of problem housing estates is explaned in section 5.2.

[12] 1st generation immigrants in 9th grades tend to be older than 2nd generation immigrants.

[13] Gross income is deflated to 2005 dkk.

[14] Given the systematic differences between schools the FE estimation might provide some useful information.

Table 1. Means for all variables included in estimations.

	Pooled sample		PISA-2000		PISA-I	
	mean	std.err.	mean	std.err.	mean	std.err.
WLE-test score, reading	483.32	105.32	503.91	93.53	468.92	110.60
ethnic concentration	18.02	22.46	0.77	1.33	30.09	22.40
ethnic conc. 0-5 (0/1)	0.44	0.50	1.00	0.00	0.04	0.21
ethnic conc. 5-25 (0/1)	0.29	0.45	0.00	0.00	0.49	0.50
ethnic conc. 25-50 (0/1)	0.18	0.38	0.00	0.00	0.30	0.46
ethnic conc. 50-100 (0/1)	0.10	0.30	0.00	0.00	0.17	0.38
ratio sch.mates in probl. hous. estate	12.62	21.41	0.90	2.83	20.92	24.72
ethnic conc. in the county	7.96	2.78	6.75	2.17	8.82	2.85
non-West. immigrant (0/1)	0.17	0.38	0.00	0.00	0.29	0.46
girl (0/1)	0.51	0.50	0.50	0.50	0.51	0.50
student 16 y.o. (0/1)	0.09	0.29	0.00	0.00	0.16	0.37
father's length of edu. (years)	11.79	3.27	12.06	2.77	11.60	3.58
mother's length of edu. (years)	11.53	3.43	11.88	2.77	11.28	3.81
family gross income (/100,000 dkk)	4.96	4.14	5.51	5.09	4.61	3.36
father's occ., self-empl. (0/1)	0.10	0.31	0.15	0.35	0.08	0.26
father's occ., high level (0/1)	0.19	0.39	0.21	0.41	0.17	0.38
father's occ., medium/low level	0.55	0.50	0.57	0.49	0.53	0.50
father's occ., out of lab.force (0/1)	0.16	0.37	0.07	0.26	0.22	0.42
mother's occ., self-empl. (0/1)	0.04	0.19	0.05	0.22	0.03	0.17
mother's occ., high level (0/1)	0.12	0.33	0.14	0.35	0.11	0.31
mother's occ., medium/low level	0.62	0.49	0.68	0.47	0.58	0.49
mother's occ., out of lab.force (0/1)	0.22	0.41	0.12	0.33	0.29	0.45
father's work exp. (years)	17.56	9.10	19.61	7.42	16.10	9.87
mother's work exp. (years)	13.77	8.38	15.57	6.87	12.50	9.08
speak Danish at home (0/1)	0.87	0.33	0.98	0.14	0.79	0.41
number of siblings	2.02	1.36	1.91	1.27	2.09	1.42
index: cultural possessions	-0.17	0.99	-0.09	0.97	-0.22	1.00
more than 500 books (0/1)	0.12	0.33	0.13	0.34	0.11	0.32
quiet place to study (0/1)	0.84	0.37	0.88	0.33	0.82	0.39
living in problem housing estate (0/1)	0.12	0.33	0.01	0.09	0.20	0.40
private school	0.16	0.36	0.29	0.45	0.04	0.20
number of students	479.31	206.46	384.76	201.50	564.57	170.67
number of students per class	20.04	2.20	19.80	2.10	20.19	2.25
index: absence from school (0-4)	1.48	0.56	1.58	0.63	1.41	0.49
index: time spent on homework (0-4)	2.58	0.61	2.58	0.60	2.58	0.61
obs	6192		2549		3643	

5 Estimation Results

In the following sections estimation results from OLS, IV, and FE estimations are provided. If one of the explanatory variables is missing for an individual we employ the commonly used practise to add a dummy variable with the value 1 and change the value of the original variable to 0 in order to not lose the observation. The dummy variables are not included in any of the tables below.

5.1 OLS Estimation

We start out with OLS estimations to have a benchmark. We know that OLS estimates are likely biased when including potentially endogenous variables and we will look more into this issue in the next section. All estimations in Models 1-5 in Table 2 are based on the pooled sample of Danish children from PISA-2000 with immigrant children from PISA-I and their Danish school mates.

In the estimations in Table 2 more explanatory variables are included each time we change model. Models 1-3 include variables which seem to be exogenous to the dependent variable, child test score, except for ethnic concentration in the school. Models 4 and 5 include more variables that potentially are endogenous to child outcome, namely whether the child attends a private school, the number of students per school and per class, and in Model 5 also indexes for child absence from school and time spent on homework. These indexes are very likely endogenous because the amount of time spent on homework depends on the child's ability and interest for school work. Absence from school is probably related to the child's interest in the school as well. In the following we will mostly focus on Models 1-3.

If we first look at ethnic concentration in the school and ethnic background we see that a higher ethnic concentration is significantly related to lower child outcomes in the PISA reading test. This result is robust across specifications. Parental and family background information seems to explain some of the negative effect of ethnic concentration in the school since the negative effect of ethnic concentration on the reading score decreases from Model 1 to 2. If the child has a non-Western background the test score is significantly lower than for children in ethnic Danish schools. The reduction is 15 points according to Model 3 compared to a decrease of 74 points in Model 1.

The negative effect of ethnic concentration is mainly driven by a negative effect in schools with a very high share of children with a non-Western background. This is seen in appendix Table A1 where dummies for ethnic concentration between 5 and 25%, 25 and 50%, and above 50% are included in the estimation. The reference category is schools with less than 5% ethnic concentration. Including dummies for ethnic concentration instead of the actual level of ethnic concentration does not change coefficients for the other explanatory variables much.

A very robust result is that girls score about 27 points more in the reading test than boys. We also see that if the child is 16 years old he/she scores about 30 points lower than 15 year olds. This is probably related to the fact that 16 year olds in the sample tend to be 1st generation immigrants, i.e. children who in many cases have not been in the Danish school system throughout their entire "school-life". They therefore might have to struggle more with language issues etc. In relation to that, we see that it has a significantly positive effect if the child speaks Danish with the parents at home, and Table 1 showed that this is more often the case for children in the PISA-2000 sample, i.e. for ethnic Danish children.

The more education and the higher income the parents have, i.e. the higher the socio economic status, the better the child outcome. We saw in Table 1 that immigrant children's parents tend to have less education and lower income than parents of children with an ethnic Danish background. The parents' occupational level also seems to be important for child outcome, especially if the mother is in high level occupation. This increases child reading score with about 20 points. On the other hand, if the mother is out of the labour force, which dominantly is the case for immigrant children, then there is a significantly negative effect (on a 10% level) on child outcome. After controlling for parental education and occupation, parental work experience does not explain child outcome. PISA-I children tend to have more siblings which might be related to the higher fraction of mothers out of the labor force. The effect on the reading score of more siblings is slightly negative.

Finally, we see that the better the study environment, i.e. the more books at home, more cultural possessions, and a quiet place to study, the better the child's reading test score. Each of these factors seem to increase the reading score with about 10 points but at the same time reduce the coefficients of parental education, occupation, and gross income compared to Model 2. As shown in Table 1, the amount of books at home and the availability of a quiet place to study is almost the same for PISA-2000 and PISA-I children.

Table 2. OLS estimation results using WLE reading score as dependent variable. Pooled sample of PISA-2000 and PISA-I.

Explanatory variables	Model 1	Model 2	Model 3	Model 4	Model 5
constant	493.86**	370.25**	383.05**	353.12**	393.92**
ethnic concentration	-0.41**	-0.26**	-0.22**	-0.21**	-0.26**
non-West. immigrant (0/1)	-74.07**	-14.81**	-15.02**	-17.56**	-22.50**
girl (0/1)	25.75**	27.63**	26.68**	25.63**	25.08**
student 16 y.o. (0/1)	-35.76**	-31.78**	-29.33**	-30.00**	-29.01**
father's length of edu. (years)		3.38**	2.94**	2.81**	2.65**
mother's length of edu. (years)		3.08**	2.49**	2.34**	2.31**
family gross income (/100,000 dkk)		1.51**	1.20**	1.02**	0.80*
father's occ., self-empl. (0/1)		-1.34	-4.23	-3.58	-3.47
father's occ., high level (0/1)		18.83**	13.03**	12.22**	12.36**
father's occ., out of lab.force (0/1)		1.39	0.67	1.31	2.35
mother's occ., self-empl. (0/1)		3.05	-0.52	1.17	1.69
mother's occ., high level (0/1)		26.57**	19.99**	19.49**	19.60**
mother's occ., out of lab.force (0/1)		-5.47	-6.95*	-6.71*	-7.04*
father's work exp. (years)		0.27	0.27	0.21	0.17
mother's work exp. (years)		0.23	0.10	0.03	-0.06
speak Danish at home (0/1)		24.72**	21.56**	21.71**	19.34**
number of siblings		-2.65**	-2.58**	-2.17**	-1.86*
index: cultural possessions			10.68**	10.06**	10.35**
more than 500 books (0/1)			9.83**	11.02**	10.58**
quiet place to study (0/1)			12.92**	13.32**	11.91**
living in probl. housing estate (0/1)				-6.33	-5.13
private school				-0.01	1.96
number of students				0.02**	0.02**
number of students per class				1.53**	1.42**
index: absence from school (0-4)					-24.24**
index: time spent on homework (0-4)					3.14
R-squared	0.15	0.23	0.25	0.26	0.28
obs	6192	6192	6192	6192	6192

**: significant on a 5% level, *: significant on a 10% level.

Including school information in Model 4 does not have a huge impact on most of the coefficients from Model 3. The most interesting results is the counter intuitive result that the number of students per class seems to have a small positive effect on children's reading scores. One explanation might be that children who have problems with school work, e.g. with reading, are moved to smaller classes where they can get more help and attention from the teacher. That is, the positive effect of having more students per class might be a reverse causality result.

In Model 5 two extra indexes are also included and they show that if the child has a high level of absence from school then the reading score is significantly lower. As mentioned earlier, these to indexes are highly endogenous so we can only somewhat trust these coefficients.

In Appendix A we also include OLS estimates of reduced samples for comparison. Model 3a in Table A2 includes only the PISA-I sample, Model 3b includes 1st generation immigrants (from PISA-I), and finally Model 3c includes 2nd generation immigrants (from PISA-I). These estimates from the reduced samples are less significant than estimates in Table 2 because of much fewer observation in Models 3a-c.

5.2 IV Estimation

To address the issue of potential endogenous neighborhood effects caused by individuals' sorting across neighborhoods, instrumental variables can be used. A good instrument is a variable that is correlated with ethnic concentration but uncorrelated with children's educational outcome.

The first instrument we suggest to use is the share of non-profit rental housing in the area since immigrants often live in this kind of housing. Intuitively, this instrument seems valid since the share of non-profit rental housing in the area does not directly affect the child's educational outcome. Since most children attend schools in their neighborhood the share of school mates who live in non-profit rental housing is a good proxy for the share of non-profit rental housing in the area. We do not have direct information about all individuals' type of housing but we have an indicator for whether an individual lives in a so-called problem housing estate. The definition of a problem housing estate is a housing estate which has received money from the "Government's City Committee" (Regeringens Byudvalg) in 1994 for social activities or for hiring resident consultants[15]. There is a high concentration of ethnic minorities and individuals with social problems in these estates. They are aggregated into 135 housing estates with an average of 1.500 residents (about 200.000 individuals in total). Housing estates are then further aggregated into 5 groups based on the severity of social problems[16]. We use the ratio of school mates living in problem housing estates

[15] Most of the housing estates receiving money are non-profit rental housing estates.

[16] The final aggregation is based on a "problem-indicator" which depends on the share

as instrument for ethnic concentration in the neighborhood.

Table 3 shows the distribution of children from PISA-I and PISA-2000 in problem housing estates. About 20% of the full PISA-I sample live in problem housing estates. Of the problem housing estates, 1 represents the 20% most problematic housing estates, 5 the 20% least problematic housing estates, etc. It is clear from Table 3 that there is an overrepresentation of PISA-I children in the most problematic housing estates and an underrepresentation in the least problematic housing estates. In comparison, only 4% of children in the full PISA-2000 sample live in problem housing estates and the distribution within problem housing estates is less skewed towards the most problematic housing estates than for the PISA-I sample. When focusing only on ethnic Danes in the PISA-2000 sample we have less than 1% living in problem housing estates.

Table 3. Distribution of children in problem housing estates (phe), only for those living in problem housing estates.

Problem housing estates	1	2	3	4	5	% in phe
% of PISA-I children	30.3	24.7	14.3	20.0	11.1	20.8
% of PISA-2000 children	22.1	21.4	25.3	14.9	16.2	3.9

A second instrument for ethnic concentration in the neighborhood is the ethnic concentration in a larger geographical area. This instrument is used in a similar way in Dustmann and Preston (2001). The idea is that individuals' housing decisions are endogenous locally but individuals are bound to certain larger areas because of job, family etc., so ethnic concentration in the larger geographical area is exogenous to child educational outcome. The instrument is valid if ethnic composition of larger areas is highly correlated with ethnic concentration in smaller areas but is beyond the control of individuals. For ethnic concentration in a larger geographical area we use the county. Table 4 shows the ethnic concentration in the 14 Danish counties in 2005 and there is a relatively big variation between counties. Bornholm has the lowest ethnic concentration with only about 3.5% individuals with a non-Danish origin whereas Copenhagen has the largest with 12.5%.

of residents from ethnic minorities, the unemployment rate, the share of inhabitants aged 18 to 66 who receive cash assistance from the municipality but without being listed as unemployed, the share of inhabitants aged 18 to 66 who are early retired, the share of households with single parents, the average disposable income per person in the housing estate compared to the average disposable income per person in the municipality (individuals aged 18 to 66), and the size of the housing estate. For more on the definition and grouping of problem housing estates see Appendix 1 in Hummelgaard et al. (1997).

Table 4. Ethnic concentration in the Danish counties in 2005.

County	Ethnic concentration
Copenhagen	0.125
Frederiksborg	0.082
Roskilde	0.071
Western Zealand	0.059
Storstroem	0.050
Bornholm	0.036
Funen	0.070
Southern Jutland	0.072
Ribe	0.058
Vejle	0.064
Ringkjoebing	0.052
Aarhus	0.078
Viborg	0.039
Northern Jutland	0.048

In Table 5 we have included both of the above mentioned instrumental variables to instrument for ethnic concentration in the school's 9th grades. Including both instruments at the same time increases the strength of the instruments since an optimal linear combination of the two instrumental variables can be found.

Compared to Table 2, Model 3 with OLS estimates we see that the magnitude of the coefficients is very stable. The coefficient of the instrumented variable, ethnic concentration, decreases somewhat from the OLS estimates to the IV estimates. According to the IV estimates a 10% increase in ethnic concentration significantly reduces child reading scores with 2 points. This does not sound of much but moving a child from a school with few non-Western immigrants to a school where the ethnic concentration is almost 100% reduces the child's test score with about 20 points.

The stability of the coefficients between estimation methods point in the direction that we might be better off using OLS estimates since they by definition have smaller standard errors. Instrumenting for ethnic concentration does not change results much.

Table 5. IV estimation results using both instrumental variables to instrument for ethnic concentration. WLE reading score is the dependent variable, pooled sample of PISA-2000 and PISA-I.

Explanatory variables	coefficients	t-statistics
constant	382.45**	32.38
ethnic concentration	-0.19**	-2.11
non-West. immigrant (0/1)	-15.91**	-2.26
girl (0/1)	26.66**	11.46
student 16 y.o. (0/1)	-29.51**	-6.56
father's length of edu. (years)	2.95**	6.18
mother's length of edu. (years)	2.50**	5.96
family gross income (/100,000 dkk)	1.21**	2.38
father's occ., self-empl. (0/1)	-4.11	-0.85
father's occ., high level (0/1)	13.05**	3.50
father's occ., out of lab.force (0/1)	0.65	0.16
mother's occ., self-empl. (0/1)	-0.50	-0.08
mother's occ., high level (0/1)	20.03**	4.95
mother's occ., out of lab.force (0/1)	-7.01*	-1.87
father's work exp. (years)	0.27	1.33
mother's work exp. (years)	0.10	0.46
speak Danish at home (0/1)	21.65**	3.31
number of siblings	-2.59**	-2.69
index: cultural possessions	10.69**	7.85
more than 500 books (0/1)	9.81**	2.43
quiet place to study (0/1)	12.99**	3.86
First stage		
ratio school mates in phe	0.67**	79.29
ethnic conc. in the county	1.11**	19.04
obs	6192	

**: significant on a 5% level, *: significant on a 10% level.

5.2.1 Strength of Instruments

An instrument is strong if the instrumental variable is highly correlated with the endogenous variable, ethnic concentration. From the raw correlations in Table 6 the first instrument seems to be quite strong. There is a clear relationship between living in a problem housing estate and having immigrant background. As we would expect, the correlation between ethnic concentration in the county and ethnic concentration in the neighborhood is weaker but there is still a fairly strong positive correlation. The second instrument therefore seems to be weaker than the first instrument based on the correlations in Table 6.

Table 6. Raw correlation between instruments and ethnic concentration in the schools.

	Pooled sample
ratio of school mates in phe	0.81
ethnic concentration in the county	0.41

The first stage t-statistic is reported in Table 5 for the case where both candidates for instruments are included in the estimation. Both candidates are positively related to ethnic concentration and they are both highly statistically significant. When the first stage t-statistics is high the instrument is strong. These t-statistics are therefore convincing us about the strength of the instruments, see also Staiger and Stock (1997). Further, a F-test for joint significance of the instruments in the first stage regression reveals that they are jointly different from zero. The key identifying assumption for the instruments is therefore fulfilled.

We still need to ensure that the instruments are not only strong but also valid. Otherwise, the instruments do not make sense and the IV estimates are unreliable.

5.2.2 Validity of Instruments

It is well known that the exclusion restrictions cannot be directly verified, that is, we cannot test whether both instrumental variables are uncorrelated with the error term in Equation (**??**), see e.g. Angrist et al. (1996). We will therefore argue for the validity of the instruments in words, more specifically as in Dustmann and Preston (2001) when using ethnic concentration in a larger geographical area as instrument for local ethnic concentration.

To take the latter instrument first it seems reasonable that individuals have no control of the ethnic concentration in a larger geographical area, e.g. a county, and that choice of housing is not based on the general level of ethnic concentration in the county. It is also possible that individuals are geographically limited in their choice of housing by employment and family history which restrict them to live in a certain geographical area. Further, ethnic concentration in a larger geographical area is very unlikely to affect child outcomes. On the other hand, it is not very likely that individuals do not choose housing *locally* based on unobserved attributes also affecting child outcomes, e.g. the ethnic concentration in the neighborhood, which means that local ethnic concentration is endogenous to child outcome.

Ethnic concentration in larger areas turn out to be fairly highly correlated with ethnic concentration in smaller areas as shown in Table 6 so ethnic concentration in the county seems to be a good instrument for local ethnic concentration. Sorting within the county does not alter the overall ethnic composition of the county and therefore does not alter the validity of the instrument. The main problem with this instrument is that individuals might not be restricted to living in certain geographical areas. If individuals' choice of geographical area is driven by unobserved components the instrument is not valid. Then we should be concerned that unobserved individual attributes across larger geographical areas might be systematically different.

Many immigrants live in non-profit rental housing and the share of school mates in this type of housing therefore appears to be a good instrument for ethnic concentration. As a proxy for whether school mates live in non-profit rental housing the ratio of school mates in problem housing estates is used. We see from Table 6 that there is a strong positive correlation between the ratio of school mates in problem housing estates and the ethnic concentration in the school's 9th grade. Further, it also seems reasonable to argue that the type of housing does not directly have an effect on child outcomes. This instrument therefore also seems to be valid.

5.3 School Fixed Effects Estimation

We now take school fixed effects into account in the estimations. This allows us to remove any observable as well as unobservable effects that are fixed within schools. The descriptive statistics in Table 1 suggsted possible systematic differences between PISA-I and PISA-2000 children's schools and these (fixed) systematic differences are taken into account int the school FE estimation. The effect of ethnic concentration is e.g. taken into account since ethnic concentration is constant across schools. One can also imagine that a particular school principal may affect child outcomes through his goals for the school and the teachers. This potential effect will also be removed in the school fixed effects estimation.

Table 7. Estimation results from school fixed effects estimation using WLE reading score as dependent variable. Pooled sample of PISA-2000 and PISA-I.

Explanatory variables	coefficient	t-statistic
constant	390,00**	33,69
non-West. immigrant (0/1)	-20,27**	-2,95
girl (0/1)	25,38**	10,95
student 16 y.o. (0/1)	-31,14**	-6,85
father's length of edu. (years)	2,56**	5,39
mother's length of edu. (years)	2,28**	5,48
family gross income (/100.000 dkk)	0,89*	1,85
father's occ., self-empl. (0/1)	-3,62	-0,75
father's occ., high level (0/1)	10,05**	2,66
father's occ., out of lab.force (0/1)	0,97	0,23
mother's occ., self-empl. (0/1)	0,42	0,07
mother's occ., high level (0/1)	19,24**	4,80
mother's occ., out of lab.force (0/1)	-5,59	-1,51
father's work exp. (years)	0,24	1,17
mother's work exp. (years)	0,00	-0,02
speak Danish at home (0/1)	24,23**	3,74
number of siblings	-2,25**	-2,33
index: cultural possessions	9,24**	6,79
more than 500 books (0/1)	8,66**	2,16
quiet place to study (0/1)	11,45**	3,40
R-squared within	0,18	
R-squared between	0,40	
R-squared overall	0,25	
obs	6152	

**: significant on a 5% level, *: significant on a 10% level.

Table 7 shows the estimation results from the FE estimation. Compared to OLS estimates in Table 2, Model 3 most of the FE estimates are slightly smaller in absolute value. The exceptions are the indicators for having non-Western immigrant background, speaking Danish at home, and for being older than 16 years. The indicator for being of non-Western immigrant background is now bigger. Test scores are reduced with 20 points for these individuals. Further, if children speak Danish with their parents at home it has a significantly positive effect on their reading score, somewhat bigger than what was suggested in Table 2. Finally, when school fixed effects are taken into account the effect of being 16 years old compared to being 15 years old is even more negative than what was suggested earlier. Older children tend to get more than 30 points less than other children in the PISA reading test.

6 Conclusion

This paper uses PISA-data from Denmark to investigate whether immigrant children and children in schools with high ethnic concentration have lower cognitive outcomes than native children in schools with low concentration of immigrants. We find that immigrant children from non-Western countries tend to have lower reading test scores than ethnic Danish children, and that children in schools with high ethnic concentration score significantly lower in a reading test than children in schools with low ethnic concentration. Schools with ethnic concentration higher than 50% seem to drive this result. The negative effects of having non-Western immigrant background and from being in a school with high ethnic concentration are robust across estimation methods.

Immigrant children's lower outcome is related to the ethnic concentration in the schools they attend and their relatively low socio economic status. Instrumenting for ethnic concentration reveals that even after taking into consideration that individuals may sort across neighborhoods, ethnic concentration in the school and the child's own ethnicity are still important factors in determining child outcome. School fixed effects estimations confirm this and reveal that also age and gender are important factors in determining child outcomes. Finally, there is a strong positive effect on children's cognitive outcome of speaking Danish.

[Policy suggestions]

[Further research]

References

Allerup, P. (2005). PISA præstationer - målinger med skæve målestokke? *Dansk Pædagogisk Tidsskrift 53*(1), 68–81.

Ammermueller, A. and J.-S. Pischke (2006). Peer Effects in European Primary Schools: Evidence from PIRLS. *IZA Discussion Paper No. 2077*.

Andersen, A. M., N. Egelund, T. P. Jensen, M. Krone, L. Lindenskov, and J. Mejding (2001). *Forventninger og færdigheder - danske unge i en international sammenligning*. Copenhagen, Denmark: AKF, DPU and SFI-survey.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of Causal Effects Using Instrumental Variables. *Journal of the American Statistical Association 91*(434), 444–455.

Bound, J., D. A. Jaeger, and R. M. Baker (1995). Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak. *Journal of the American Statistical Association 90*(430), 443–450.

Browning, M. and E. Heinesen (2003). Class Size, Teacher Hours and Educational Attainment. *CAM working paper No. 2003-15*.

Damm, A. P. (2007). Determinants of Recent Immigrants' Location Choices: Quasi-Experimental Evidence. *Forthcoming in Journal of Population Economics*.

Damm, A. P. and T. Tranæs (2006). Påvirker det flygtninges løn og beskæftigelse at bo i en kommune med mange flygtninge? In *En befolkning deler sig op?*, Chapter 6, pp. 198–215. Denmark: The Rockwool Foundation Research Unit.

Dustmann, C. and I. Preston (2001). Attitudes to Ethnic Minorities, Ethnic Context and Location Decisions. *Economic Journal 111*(470), 353–373.

Edin, P.-A., P. Frederiksson, and O. Åslund (2003). Ethnic Enclaves and the Economic Success of Immigrants - Evidence from a Natural Experiment. *Quarterly Journal of Economics 118*(1), 329–357.

Entorf, H. and M. Lauk (2006). Peer Effects, Social Multipliers and Migrants at School: An International Comparison. *IZA Discussion Paper No. 2182*.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003). Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics 18*, 527–544.

Hoxby, C. (2000). Peer Effects in the Classroom: Learning from Gender and Race Variation. *NBER Working Paper Series No. 7867*.

Hummelgaard, H., B. K. Graversen, D. Lemmich, and J. B. Nielsen (1997). *Udsatte boligområder i Danmark*. Copenhagen, Denmark: AKF Forlaget.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies 60*(3), 531–542.

OECD (2002). PISA 2000. Technical report, OECD, Paris.

Rangvid, B. S. (2005). Sources of Immigrant's Underachievement: Results from PISA-Copenhagen. *akf working papers, August, 9*.

Staiger, D. and J. H. Stock (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica 65*(3), 557–586.

Todd, P. E. and K. I. Wolpin (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *Economic Journal 113*, F3–F33.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts, London, England: The MIT Press.

A Appendix

OLS estimations using dummies for levels of ethnic concentration.

Table A1. OLS estimation results using WLE reading score as dependent variable. Pooled sample of PISA-2000 and PISA-I.

Explanatory variables	Model 1	Model 2	Model 3	Model 4	Model 5
constant	491.87**	369.88**	382.17**	347.03**	388.92**
ethnic conc. 5-25 (0/1)	-0.76	-5.86**	-2.84	-11.72**	-15.54**
ethnic conc. 25-50 (0/1)	-10.70**	-8.08**	-4.19	-9.86**	-13.81**
ethnic conc. 50-100 (0/1)	-30.48**	-20.72**	-17.23**	-14.87**	-19.49**
non-West. immigrant (0/1)	-74.73**	-15.02**	-15.69**	-17.88**	-22.66**
girl (0/1)	25.66**	27.65**	26.66**	25.72**	25.24**
student 16 y.o. (0/1)	-36.79**	-31.36**	-29.51**	-28.66**	-27.15**
father's length of edu. (years)		3.39**	2.95**	2.84**	2.68**
mother's length of edu. (years)		3.12**	2.50**	2.42**	2.41**
family gross income (/100.000 dkk)		1.52**	1.21**	1.02**	0.80*
father's occ., self-empl. (0/1)		-1.52	-4.09	-3.98	-3.99
father's occ., high level (0/1)		18.78**	13.06**	12.16**	12.27**
father's occ., out of lab.force (0/1)		1.57	0.80	1.57	2.70
mother's occ., self-empl. (0/1)		2.98	-0.47	1.01	1.44
mother's occ., high level (0/1)		26.57**	20.10**	19.49**	19.54**
mother's occ., out of lab.force (0/1)		-5.25	-6.80*	-6.26*	-6.44*
father's work exp. (years)		0.27	0.27	0.21	0.17
mother's work exp. (years)		0.24	0.10	0.05	-0.03
speak Danish at home (0/1)		24.77**	21.71**	21.60**	19.15**
number of siblings		-2.70**	-2.59**	-2.24**	-1.93**
index: cultural possessions			10.60**	10.15**	10.50**
more than 500 books (0/1)			9.82**	10.87**	10.39**
quiet place to study (0/1)			13.03**	12.85**	11.28**
living in probl. housing estate (0/1)				-7.59	-6.55
private school				-0.83	0.89
number of students				0.03**	0.02**
number of students per class				1.77**	1.71**
index: absence from school (0-4)					-25.10**
index: time spent on homework (0-4)					2.68
R-squared	0.15	0.23	0.25	0.27	0.28
obs	6192	6192	6192	6192	6192

**: significant on a 5% level, *: significant on a 10% level.

Estimation results for individuals from the PISA-I sample and separately for 1st and 2nd (and higher) generation immigrants.

Table A2. OLS estimation results with the same variables included as in Table A1, Model 3 for the sample of PISA-I children and separately for 1st and 2nd generation immigrants from PISA-I. WLE reading score is the dependent variable.

Explanatory variables	Model 3a	Model 3b	Model 3c
	PISA-I	PISA-I, 1.gen	PISA-I, 2. gen
constant	410,64**	429,13**	495,41**
ethnic conc. 5-25 (0/1)	-8,90	-95,50**	-60,04**
ethnic conc. 25-50 (0/1)	-11,34	-92,97**	-64,65**
ethnic conc. 50-100 (0/1)	-24,47**	-90,71**	-81,80**
non-West. immigrant (0/1)	-20,15**	-	-
girl (0/1)	24,08**	28,43**	8,95
student 16 y.o. (0/1)	-30,08**	-43,29**	-47,15**
father's length of edu. (years)	2,77**	4,52**	1,13
mother's length of edu. (years)	1,51**	0,01	0,10
family gross income (/100.000 dkk)	1,62**	7,16	1,63
father's occ., self-empl. (0/1)	-16,53**	-21,44	-29,53*
father's occ., high level (0/1)	8,19	47,49*	68,76**
father's occ., out of lab.force (0/1)	-0,46	6,77	5,45
mother's occ., self-empl. (0/1)	-0,99	12,39	10,66
mother's occ., high level (0/1)	20,79**	139,57**	39,28
mother's occ., out of lab.force (0/1)	-8,97*	-12,77	-20,03*
father's work exp. (years)	0,49*	2,14	0,17
mother's work exp. (years)	-0,36	3,33	-0,43
speak Danish at home (0/1)	18,21**	18,16	8,49
number of siblings	-3,63**	-3,46	-4,80
index: cultural possessions	11,27**	2,96	5,42
more than 500 books (0/1)	16,44**	20,06	36,88
quiet place to study (0/1)	13,83**	17,72	6,18
R-squared	0,27	0,23	0,15
obs	3643	463	606

**: significant on a 5% level, *: significant on a 10% level.

One could use the sample from Model 3b to test whether the governments' refugee placement policy could work as a strong instrument. The problem with this exogenous source of variation is that it is only valid for refugees, i.e. 1st generation immigrant from certain countries and in a certain time period. Our sample of 1st generation immigrant children is quite small and reducing it to only include refugee children limits the sample further. Limiting it even more to children who immigrated in the relevant time period, 1986 to 1998, leaves us with a sample too small to use for IV estimations. The method of using this "natural experiment" to find exogenous variation in residence for refugees has been used by e.g. Edin et al. (2003) for Sweden and Damm (2007) for Denmark.