Achievement, Schooling and Family background: Evidence for Sweden

Mikael Lindahl

SOFI, Stockholm University

This version: June, 2007

ABSTRACT

This study investigates whether schooling compensates for disadvantageous family background among Swedish children. Using a sample of sixth grade pupils, we associate pupil and family background characteristics with math test score changes during summer, when schools are out of session, and during the school years, when schools are in session. Results are that when schools are in session, the test score gap between Swedish and immigrant children decreases, whereas this is not the case during the summer vacation. The results hold whether or not we condition on previous test scores. Hence, immigrant children, who on average scores low on tests, gain relatively more from schooling. We also found that learning is unrelated to parents' socioeconomic level during both the summer and the school year. Hence, Swedish education does a good job in mitigating the existing math skill gap between Swedish and immigrant children, but not between children from different socioeconomic backgrounds. These results differ from previous research conducted on US data.

*This is a much revised version of a longer paper previously circulated under the title: "Summer Learning and the Effect of Schooling: Evidence from Sweden." I thank Anders Björklund and Alan Krueger for valuable comments on this earlier version. I also thank Peter Björklund and Ossian Wennström for excellent help with the data collection. I am most grateful to the pupils and teachers who participated in this study. I also thank HSFR and Handelsbanken for financial support.
**Address: Swedish Institute for Social Research (SOFI), Stockholm University, 106 91 Stockholm. E-mail: mikael.lindahl@sofi.su.se

I. Introduction

How can a society equalize opportunities for its citizens? This question is of central importance for policy makers in nearly all societies and a classic issue of study for social science scholars. The purpose of this paper is to learn whether there is a role for schooling in decreasing inequality of opportunities in Sweden. More specifically, we ask whether schooling compensates for "disadvantaged" family background by particularly increasing those children's scholastic achievement, and thereby diminish the test score gap that exists between children with different backgrounds.

To analyze this issue, we follow a path from the educational sociology literature and associate demographic and family background characteristics with test score changes during the school year, when schools are open, and with test score changes during the summer, when schools are closed. More specifically, we look at how socioeconomic and immigrant status of the parents and the gender of the child correlate with seasonal learning.

Previous studies using this methodology have all used US data. Heyns (1978) found that for pupils in Atlanta schools, learning during summer was positively related to parental income.[1] Entwisle, Alexander and Olson (1997) followed 790 children from 20 public Baltimore schools from first to sixth grade. They found that school-year learning was unrelated to socioeconomic level but that for math and reading, summer learning was positively associated with a pupil's socioeconomic level. Cooper et al. (1996) performed a meta-analysis of 39 summer-learning studies, a few of which analyzed summer learning by background characteristics. They concluded that low-income children always lost in reading skills during summers,

---

[1] Another early reference is Murnane (1975), who found that inner-city children in US gained reading skills during the school-year, whereas the development of reading skills was non-positive during the summer. O'Brien (1998) found that pupils who were non-white and from low-income families, lost math and reading skills during the summer, but that they gained during the school year.

whereas middle-class children either increased their skills or lost less. Evidence from such studies led Alan Krueger, in a New York Times column in August 2000, to state that such pattern “suggests public schools are doing more to help poor children overcome the obstacles they face in their homes and neighborhoods than is commonly appreciated.” In a recent study, Entwisle et al. (2007) followed the same children in Atlanta further. They conclude that differences in achievement in ninth grade across SES-groups mainly are due to differential summer learning in elementary school years. They also argue that these differentials account for a large part of SES-differences in high-school and college achievement.

An interpretation of these results is that in the US, schools prevent children from disadvantaged backgrounds to lose more compared to peers with more advantaged backgrounds. This is an important result, because this means that schools make a good job in preventing the achievement background gap to increase even further, something that, one might interpret, would have happened had children spent even more time outside time of schooling. It is time outside schools, either before they start school or during summer vacations, that creates the observed test score background gap. The policy challenge is then to prevent the gap from increasing during these times.

However, not all studies agree with the findings for U.S. outlined above. An important exception is the study by Fryer and Levitt (2004), who analyzed a nationally representative sample of about 5,200 children starting kindergarten in 1998. Their data includes scores from math and reading tests performed at the start and end of kindergarten and first grade. They found that the black-white test score gap in math was unchanged over the summer but that it increased somewhat during the school years. For reading, there is a small decrease in the gap over summer and a large increase in the gap the first school year. Hence, there is no evidence that black children fare worse than white children during the summer and strong evidence (especially for reading) that black children lose ground during the first school year. Hence, these results points toward schools increasing black-white differences, instead of decreasing them, which one would expect based on the evidence for income discussed above.

In this study we contrast the results from existing US studies with results from a new study we perform on a sample of Swedish pupils. We believe this be of great interest for at least two reasons: First, Swedish school policy has deliberately used schools as an instrument to equalize opportunities of individuals (Erikson and Jonsson, 1996), probably to a larger extent than in the US. For instance, in Sweden educational resources are still to date redistributed towards schools with a high fraction of pupils from disadvantaged backgrounds (see National Agency for Education, 1999). Second, the income inequality in Sweden is (and has been for long) much smaller than in the US (Björklund, 1998), and hence childrens' social background are more homogenous in Sweden. Due to these different features of societies, conclusions from the US studies might not be transferable to Sweden.

This paper uses data from a sample of pupils from schools in Stockholm, Sweden's capital. The sample contains information on scores from math tests for the same children in the spring of the fifth grade and in the fall and spring of the sixth grade, making it possible to calculate individual test-score changes when school is in and out of session. The sample also contains socioeconomic measures from register data and data on childrens' demographics from a questionnaire answered by teachers.

Section II presents and discusses the data. Section III compares test score levels and changes during the school year and during the summer for pupils with different background characteristics. Section IV conducts sensitivity analysis and Section V concludes.

II. Data

The Swedish National Agency for Education organizes a national math test to be distributed to all schools in Stockholm for pupils to take during the spring semester of the fifth grade. I contacted schools at the start of the fall 1998 semester.[2] I selected four parts of this test, which I then distributed to the pupils at the start and end of the sixth grade. The spring of the fifth grade test was given during the period February-June in 1998, with the four test parts conducted at separate occasions. The national math test in the spring of 1998 was administered and conducted without my participation, and these tests were collected from the schools in fall 1998 and corrected by me. Note that the pupils and teachers were then not aware that the test would be distributed again in the sixth grade. The fall of the sixth grade test were given from the last week in September to the first week in November in 1998, and the spring of the sixth grade test were given during the last four weeks of the term (in May-June) in 1999. Hence, the pupils were tested at three occasions, spring 1998 (5^{th} grade), fall 1998 (6^{th} grade) and spring 1999 (6^{th} grade). The summer vacation in Sweden lasts 10 weeks, from early June to late August. We, not the teachers, graded the tests on all three occasions.

In total, 556 pupils did the test on all three occasions, and took the test under similar conditions regarding time allowed and teacher help.[3] These 556 pupils were from 33 classes from 16 schools. The same test was used on all three occasions. The

[2] For a more thorough description of the data collection, see Lindahl (2002).
[3] According to the sixth grade class lists, 701 pupils were available for testing in the fall semester in sixth grade, in those classes that participated in this study. In the sensitivity section 4, we investigate whether our results are sensitive to this sample selection issue.

four test parts included in the test at each occasion all tested math skills, but were of different types, with questions ranging from simple counting exercises too more advanced problems. In total, each pupil should have conducted 12 test parts. However, not all pupils took all four parts of the test at each occasion. Of the 12 test parts, the average pupil misses scores on 1.4 test parts (11.9 percent). Missing scores are imputed, using the non-missing scores on each test parts, separately at each test occasion (see Appendix 1 for a discussion of this). The sum of the scores on each test part, for the three test occasions (spring of the fifth grade, fall of the sixth grade and spring of the sixth grade), are presented in Table 1. The minimum possible score is 0, and the maximum possible score is 72. Since no pupil has the minimum or maximum score, the test seems to allow for improvements and reductions of the scores from one test occasion to another.

The test scores increased with test occasions. The summer gain is 1.80 raw points (0.14 standard deviations), and the school year gain is 6.62 raw points (0.49 standard deviations). For an illustration of this, see figure 1, which uses actual data. Adjusting for the time difference between the tests does not change the results very much: the weekly gain in test scores is still a bit over one-third higher during the school year. Thus, clearly less improvement is seen between the dates including the summer vacation. It also looks as though pupils actually gain over the summer, but this might also be because the tests were taken several weeks before and after summer break or because there are re-test effects from taking the same test twice. In figure 1, we also show a hypothetical situation, showing why non-ideal test dates (in the presence of linear learning) might overstate the average summer gain. Correlations between test scores at the different test occasions are between 0.73-0.78 (Table 2). To facilitate interpretation in later estimations, we standardize the test scores at each occasion to have mean zero and variance one (also shown in Table 1).

Data on school, class, and teacher characteristics were gathered with a questionnaire distributed to the teachers at the time of the fall sixth grade test. Teachers were asked to answer questions about themselves (their teaching experience in total and with their present class and their education) and their students (pupils' genders and nationality of pupils' parents) and to provide information about their class sizes.[4] To get information on pupils’ social background, the addresses of the pupils (from the class lists) were matched with block data on education and family income. These data are taken from Statistics Sweden databases and were partly calculated by them for the purpose of this project.

The demographic and family background variables are GIRL, IMMIG, parents’ years of education, and the logarithm of family income. GIRL equals one if the pupil is a girl and zero if the pupil is a boy. Parents’ years of education and family income are the average years schooling and the average family income (before tax) for the adults with children in the relevant age range in the block where the pupil lives. IMMIG is an indicator variable that equals one if the pupil’s parents do not have Swedish as their native language, and zero otherwise. Hence, IMMIG shows whether or not the pupil lacks Swedish parents in the household, and is thereby a proxy for whether the child is a first- or second-generation immigrant to Sweden. A socioeconomic index, SES, is created as the average of the standardized parents’ years of education and the standardized logarithm of family income. Hence, parents’ years of education and the logarithm of family income are weighted equally.

[4] This information is used in Lindahl (2005) for estimating the effect of class size on pupil achievement.

Table 3 shows correlations among the background variables. We see that IMMIG and SES are negatively correlated. Among reasons to use SES in the analysis, instead of parents' years of education and the logarithm of family income, are that the last two are highly correlated. This makes any separate effect, while controlling for the other, hard to disentangle. We are also not primarily interested in separate effects of these variables on pupils' learning. If any of these are separately included, the results stay very similar.

III. Results

We start by regressing test score levels (spring of fifth grade, fall and spring of sixth grade) and changes (during summer and school year periods) on IMMIG, GIRL and SES, separately, without controlling for the other variables. We then regress test score levels and changes on these three variables included at the same time. Since parents' income and education were collected at the block aggregation level, standard errors are corrected for blocks-clustering when SES is included as a regressor. Note that we do not include any school variables in the regressions, since we expect these characteristics to be affected by the background characteristics.[5]

Table 4 show the results, where the dependent variables are test score levels at spring fifth grade and fall and spring sixth grade (columns 1,3 and 5) and the change in test scores between first and second and second and third test occasion (columns 2 and 4). Results are that immigrant children score on average 0.7 standard deviations less then Swedish pupils at the end of fifth grade. One year later, the difference has

[5] For instance, school resources in Stockholm are re-distributed towards areas with a high fraction of foreign pupils. If school characteristics would be included, this would hide part of the total effect of being a foreign pupil on learning.

decreased to 0.6 standard deviations.[6] This difference is entirely due to the effect during the school period. When GIRL and SES are included as controls, the estimate of the association between IMMIG and test score levels decreases quite a lot, something which is entirely due to controlling for SES. However, the association between IMMIG and summer and school-year learning are unaffected. Hence, immigrant children gain during the school year and do not gain (make equivalent gains or loose) during summer. An interpretation of these results is that the school helps immigrant children overcome lower skills in math at and earlier point in time.

The effect of parents' SES on test score levels is also strong. A one standard deviation higher SES is estimated to increase test scores by 0.25-0.35 standard deviations. Interestingly, SES does not affect test score changes during the summer and the school-year, whether or not IMMIG and GIRL are included as controls. Hence, low-SES pupils make about equivalent gains as high-SES pupils during both summer and school year periods. Therefore, the school does not appear to help low-SES children overcome their lower math skills. This is in contrast with the results from the US literature.

The results by gender show that test score levels are insignificantly different between girls and boys. The estimate for GIRL on learning through the school year is positive, but not significantly different from zero. Hence, there is no statistical evidence that boys gain less math skill during the school year.

The seasonal estimations underlying columns 2 and 4 of Table 4 assumes that achievement starting levels are unrelated to achievement changes. This assumption might be invalid due to regression-to-the-mean or because of improper design of the

[6] In order to interpret a standard deviation change in the math test score, one can note that in Lindahl (2002), about 1 standard deviation higher score on math test in sixth grade is associated with about 10 percent higher earnings (for an earlier cohort of sixth grade pupils) controlling for background characteristics and (in some specifications) IQ scores.

test.[7] We therefore re-estimated the seasonal specifications from Table 4, by adding test scores at the end of fifth grade (to the specification used for the estimates in column 2) and test scores at the beginning of sixth grade (to column 4), to allow pupils with different starting achievement in the beginning of the periods to learn at different speeds. When this is done the results do appear to change quite a bit, as shown in columns 1 and 3 of Table 5.[8]

However, including lagged test scores makes the estimations more complicated since then even classical measurement errors in test scores bias all estimates, whereas without lagged test scores included (as in columns 2 and 4 of Table 4), no adjustment for classical measurement error in the test scores is necessary.[9] If we believe that test scores imperfectly measure real achievements we need to deal with this issue. This means we need to know the reliability of the test used in this study. We therefore estimate the so-called Cronbach's alpha reliability, by utilizing that we have scores on four separate parts of the test at each test occasion.[10] Doing this give results that the reliability ratio is around 0.8, which we then use to

---

[7] Regression to the mean could exist because of ceiling and floor effects in the tests or because it is easier in general for weak pupils to improve on tests, due to their low starting knowledge. By improper design of the test we mean that we cannot know a priori that the test design is such that an absolute improvement in scores is translated into a comparable absolute improvement in mathematical knowledge, which we want to capture, in all parts of the test-score distribution.

[8] In columns 1 and 3, the estimates for the effect on previous test scores on test score changes are much less than one, about –0.30 and –0.25, respectively, indicating strong regression-to-the-mean.

[9] Classical measurement error in an explanatory variable will always bias the effect of this variable towards zero, even in the presence of additional (measurement error-free) controls. Note that when the change in a variable is regressed on the initial value of the variable this no longer holds, since the variable measured with error also is part of the outcome variable. Instead the effect of lagged test scores on the change in test scores will be biased towards showing a negative value, i.e towards showing regression-to-the-mean. The direction of the bias in a control variable depends on the direction of association between lagged test score level and this control variable. If this association is positive, the effect of the control variable on the change in test scores is too positive, if no measurement error adjustment is made.

[10] Assuming classical measurement error, we can consistently estimate the reliability of the average test score on each test occasion by estimating the alpha reliability (see Cronbach [1951]), using the formula $\alpha = Nr/[1+(N-1)r]$, where N is the number of test parts and r is the average of all two test parts-correlation. Using the test score parts from spring of the fifth grade it is estimated to be 0.79 (N=4 and r=.49). Note that in calculating alpha we do not use the imputed test part scores. The alpha estimate give a lower bound estimate of the true reliability ratio if the test parts are non-parallel, i.e. they capture

correct the estimates for measurement error. When this is done (in Columns 2 and 4 of Table 5), the estimates fairly closely resemble the earlier ones in Table 4. Another way to correct for bias due to measurement error in test scores is to use earlier test scores as an instrument. Here, the spring fifth-grade scores as an instrument for the fall sixth-grade scores, in the school-year learning regressions. If there is no serial correlation in test scores, this will also correct for this type of bias. Estimates are shown in column 5 of Table 5. Results are very similar as in column 4.[11]

The tests should for every pupil, ideally, have been taken at the exact end and start of the semesters. This was not the case here and is hardly possible to administer in practice.[12] We first attempt to control for the fact that there exists a difference among school classes in the number of weeks between performed tests during both seasons. We control for this by including variables capturing the number of weeks between spring fifth and fall sixth grade tests, in the summer regressions, and between fall of sixth and spring of sixth grade, in the school-year regressions. The results are basically unchanged compared to the ones reported in columns 2 and 4 of Table 4. Second, we attempts to control for this issue, as well as for the fact that the tests were not taken at exact dates in beginning and end of school years (and hence the time between tests during the summer period involves some time of the school-year). If learning is linear during the school year, we can simply predict scores at the end and

different types of math skills. We estimate alpha to be 0.79 in spring of fifth grade, and 0.82 and 0.83 at fall and spring of sixth grade.

[11] Using these ways to adjust the estimates for measurement error in test scores greatly diminishes the regression-to-the-mean found earlier. As pointed out above, measurement error in test scores will bias the lagged test score estimate toward showing regression-to-the-mean.

[12] There were, on average (st.dev) , 26.9 (2.9) weeks between the spring of the fifth-grade and the fall of the sixth-grade test occasions, and an average (st.dev) 31.1 (2.1) weeks between the fall and spring of the sixth grade test occasions. The spring fifth-grade test was given in the February-June period The fall sixth-grade test was given from the last week of September to the first week of November. And the spring sixth-grade test was given within the last three weeks of the school year. On average, the spring fifth-grade test was taken 9.7 weeks before the summer break and the fall sixth-grade test 7.2 weeks after the summer break. Entwisle, Alexander and Olson (1997) state that the tests were administered in May and October. So between 17-26 weeks passed between their test occasions.

start of the sixth grade.[13] Further, if we assume that the individual learning rate is the same in the fifth and the sixth grade, conditional on observed school and class variables, we can also predict scores at the end of the fifth grade. How this is done is shown in Appendix 2. Results are shown in Table 6, where we see that some of earlier results are reinforced. IMMIG now has a stronger positive effect on learning during the school year. Regarding the effects of SES on summer and school-year learning, these are basically unaffected by this adjustment.

We therefore conclude that, relative to Swedish pupils, immigrant pupils gain more math skill during the school year and make equivalent or lose math skill during the summer vacation. There is no evidence that learning is associated with SES, either during the summer or school year.[14]

IV. Sensitivity analysis

This section investigates the robustness of the results in the previous section to relaxation of several assumptions. We investigate the issue of how taking the same test several times might impact our estimates. We also investigate sample selection and attrition.[15]

First, we have ignored the issue of whether the estimated background-effects are affected by so-called re-test bias, which might occur here since the same test were distributed to the pupils at spring fifth grade and fall and spring sixth grade. Re-test bias means that observed test score changes partly reflect pupils learning the questions

[13] We tested the assumption of linear learning during the school year, by regressing weekly learning rate on the number of weeks between tests. If number of weeks between tests lack influence on the learning rate, it means that learning is linear, since the weekly learning rate is the same over time. The weeks between the fall and spring of the sixth grade test occasions vary between 24 and 31.5 weeks, hence the assumption that pupils learn linearly during the school year is tested by using that there is some variation in the time at which the fall sixth-grade tests were taken.

[14] Given how small the estimates are when we associate SES and seasonal test score changes, these results are unlikely to be explained by measurement error in SES.

and how to solve them, instead of reflecting true achievement changes. We do however argue that it is unlikely that the results are driven by the same test being used. First, re-test effects are just additional measurement errors in test score levels, here at fall and spring of sixth grade. Hence, if re-testing affects pupils in an heterogenous way, and not just affect all pupils in the same way so that just the mean test score gains are estimated incorrectly, we would expect it to lead to decreased reliability ratios across test occasions. This do not appears to really be the case here. The reliability ratios are 0.79, 0.82 and 0.83 at spring of fifth grade, fall of sixth grade and spring of sixth grade. Hence we only observe a slight increase, which also can be explained by higher correlated measurement errors on the test parts during the last two test occasions, since the four parts of the test then were taken at only two occasions. Second, the estimated background effects in Table 4 will be consistently estimated if re-test errors in test scores are uncorrelated with these background variables. This can be investigated since the four test parts included in each test are different and not everybody took all test parts at the spring of the fifth grade occasion. We could therefore compare test score changes over the summer on test parts were pupils did take the parts at both occasions (this estimate is partly due to re-test bias), with score changes over the summer on test parts were the pupils did not take the part before summer but took the test part for the first time after the summer (this estimate is not due to re-test bias). We found no evidence that re-test bias varied with background characteristics.[16]

[15] Although the authors of some of the earlier studies of seasonal learning are aware that these issues might be present, very few explicitly tests for whether they bias their estimates.

[16] This last comparison can only be done if we somehow impute scores on these missing test parts. Since scores on the test parts pupils actually took during spring fifth were taken for the first time and therefore unaffected by re-test effects, we impute (using regression) the missing test score parts, by using all existing scores on the test parts they did take before summer. We regress test score part changes during summer (where missing values at spring fifth grade were imputed), on a dummy for whether or not test part in spring of fifth grade was done. We then interacts this dummy with the

Second, there are pupils who should have been available for testing at all occasions, but for different reasons were not. A total of 701 pupils should have been available for testing in the sampled schools, but only 556 pupils took all tests. A comparison between the pupils in the sample with the 145 pupils not in the sample, shows that sample attrition is more common among pupils with parents who are non-Swedish and who have lower-than-average education and income. Since sample selection can bias the estimates and since we have information of background variables for all pupils, we need to address this issue. One way is to use that some of these pupils took the test on two (but not all three) test occasions. Hence we can check the sensitivity of attrition on the results by adding these pupils in the estimations. Adding the 35 pupils in the summer regressions and the 60 pupils in the school year regressions, for which test scores are available, give very similar results as before.

Third, all pupils did not take all parts of the tests at all occasions. Whether this influences the results can be tested by adding variables capturing the number of test parts taken to each summer and school period regression in Table 4. Adding a variable of the sum of the number of test parts taken at the first and second test occasion to the summer regression leaves the results basically unchanged. The same is true when we add a variable capturing the sum of the number of test parts taken at the second and third test occasion.[17]

background variables to see whether re-test effects differ by background characteristics. We also controlled for the other background variables.

[17] The average (standard deviation) of the sum of the number of test parts taken during the 1st and 2nd test occasions is 7.38 (0.97), and the average (standard deviation) of the sum of the number of test parts taken during the 2nd and 3rd test occasions is 6.97 (1.22). The maximum sum of the number of test parts that could be taken at two occasions was 8.

V. Conclusion

We have found evidence that in Sweden, schooling help children with non-Swedish parents to overcome an (for whatever reason) existing math skill deficit relative to Swedish children. This skill deficit is estimated to be about two-thirds of a standard deviation at the time when pupils finish fifth grade. During the summer this relative test score gap is basically unchanged, whereas during the sixth grade school-year, about one-fifth of this gap is filled. It therefore seems that the school does a good job in narrowing this gap in test scores. This can be interpreted as a beneficial effect of the distribution of school resources that exists in Stockholm, where resources are strongly allocated towards neighborhoods with a high fraction of immigrant pupils.

We also estimated one standard deviation lower SES to be associated with about one-third lower score on math test at the end of fifth grade. We did however not find schooling too diminish this strong role of SES as a predictor of math test score. This is contrary to results for the US. These results can perhaps be interpreted in light of the differences between the Swedish and the US societies, whose full effect is at work during the summer holiday. In Sweden, the social environment is equal enough to prevent pupils with disadvantaged social backgrounds to significantly lose skills when schools are closed. Instead, the higher inequality in the US could manifest itself in terms of significantly different summer learning experiences for children from different social backgrounds.

Appendix 1: Imputation of missing test-part scores (not intended for publication)

Among the 556 pupils, not all pupils had done all parts of the test on all test occasions. Even though all pupils took at least one test part at spring fifth grade, fall sixth grade and spring sixth grade, not all pupils took all four parts of the test at each occasion. Of the 4x3=12 test parts conducted 222 pupils did all 12 parts, 225 pupils did 10-11 parts, 97 pupils did 8-9 parts and 12 pupils did 5-7 parts. The average number of test parts done was 10.57. As long as the pupil's knowledge of the specific math skill tested in a particular test part is unrelated to whether or not this pupil took this test part, this is not a problem. This was definitely the case for the fall and spring of the sixth grade test occasions, since whether or not pupils did not took some test parts then was not based on any pupil- or teacher selection. In some instances it was the case that some classes would not have enough time to do all test parts, which then was known to me before and the selection of what test parts to use was within my control. Whether this is also the case at the first test occasion (in fifth grade) is more uncertain.

An potential measure to use would have been to standardize the existing scores on each separate test parts, and then take the average of these standardized scores as the measure of each pupil's achievement each test occasion. Hence, we would then have just ignored the missing test part scores. If missing values are randomly determined, this should give very similar results as the imputation process used above. However, as can be seen in the table below, where we regress number of non-missing test parts on background characteristics at spring fifth grade and fall and spring of sixth grade, the number of missing test part scores are related to some background characteristics.

Dependent variable:	Number of non-missing test parts at spring of fifth grade	Number of non-missing test parts at fall of sixth grade	Number of non-missing test parts at spring of sixth grade
	(1)	(2)	(3)
IMMIG	-.20	.07	-.49
	(.12)	(.07)	(.21)
SES	.02	.00	-.17
	(.05)	(.03)	(.06)
GIRL	.08	.01	-.08
	(.06)	(.06)	(.08)
R^2	.021	.002	.029

Notes: Number of observations is 556. Estimation method is OLS. An intercept is always included. Standard errors (in parentheses) are adjusted for blocks-clustering on SES. The dependent variable in column 1 has a mean (std) of 3.60 (0.72). The dependent variable in column 2 has a mean (std) of 3.78 (0.63). The dependent variable in column 3 has a mean (std) of 3.19 (1.05).

Based on this, we believe that it is important not to ignore the missing test part scores, but at the same time limit the potentially adverse effects of sample attrition and to be able to use as a large sample as possible. We therefore instead deal with this by imputing scores for the missing test parts on each occasion. In other words, I regress the score on one test part on the scores on the other test parts on one occasion and predict scores for the test parts with missing values.

It should be remembered however that quite few missing test parts are imputed. In total each pupil should have conducted 12 test parts. However, not all pupils took all four parts of the test at each occasion. Of the 12 test parts, the average pupil miss scores on 1.4 test parts (11.9 percent). So only about 12 percent of the missing test part scores are imputed.

The imputation is done in the following quite standard way, exemplifying with spring of fifth grade: First, we regress score on test part 1 on scores on test parts 2, 3 and 4, for all pupils with observations on all these test parts (407 pupils): $y_{1i} = c_1 + c_2 y_{2i} + c_3 y_{3i} + c_4 y_{4i}$. We then predict score for those 13 pupils who miss score on test part 1 and have test part scores on 2,3 and 4. Second, we regress score on test part 1 on scores on test parts 3 and 4, for all pupils with observations on all these test parts (412 pupils): $y_{1i} = b_1 + b_3 y_{3i} + b_4 y_{4i}$. We then predict score for those 9 pupils who miss score on test part 1 and have test part scores on 3 and 4.

And so forth, until the 29 pupils with missing score on test part b is imputed. Hence test part score for those 29 pupils is a weighted average of the score on the other 3 test parts (13 pupils) two of the other test parts (11 pupils) and one of the other test parts (5 pupils).

These steps are then repeated for test part 2, 3 and 4 at spring of fifth grade, as well as for test part 1, 2, 3 and 4 at both fall of sixth grade and spring of sixth grade. We then take the sum of the scores on all four test parts at each test occasion (spring fifth grade and fall and spring of sixth grade), and standardize the scores on each test occasion.

Appendix 2: Prediction of test scores at start and end of school years (not intended for publication)

To predict scores at the start and end of the sixth grade, we first express the weekly test-score change for pupil *i* in sixth grade as:

$$(1) \qquad w_{6,i} = \frac{T_{6_2,i} - T_{6_1,i}}{t^a_{6_2} - t^a_{6_1}},$$

where $T_{6_2,i}$ and $T_{6_1,i}$ are standardized test scores in spring and fall of sixth grade, for pupil *i*; $t^a_{6_2}$ and $t^a_{6_1}$ are number of weeks into the sixth grade that the tests were taken, in spring and fall of the sixth grade, for pupil *i*.

Assuming that weekly learning rate is constant during the school year, the predicted test scores at the first and last week of sixth grade, can be calculated as:

$$(2) \qquad T^p_{6_2,i} = T_{6_2,i} + (w_{6,i} \times (t_6 - t^a_{6_2,i}))$$

$$(3) \qquad T^p_{6_1,i} = T_{6_1,i} - (w_{6,i} \times t^a_{6_1,i}),$$

where t_6 is the length of the school year, in weeks, which is set to 38.

Because we do not have two test occasions in fifth grade, we cannot use the same method to calculate predicted test scores at the end of fifth grade. One alternative is to assume that the rate of weekly learning is the same in fifth grade, as in sixth grade. But even though many school characteristics are constant between these grades, some are not. So I try to account for this by correcting the fifth-grade learning rate for the changes in class size between the grades. I first estimate the following equation:

$$(4) \qquad w_{6,i} = c_6 + b_6 CS_{6,i} + \varepsilon_{6,i},$$

where $CS_{6,i}$ is class size in sixth grade for pupil *i* and c_6 and b_6 are parameters to be estimated. Estimation of (4) gives estimates $\hat{c}_6$, $\hat{b}_6$ and $\hat{\varepsilon}_{6,i}$, where $\hat{\varepsilon}_{6,i}$ is the estimated residual. I then calculate the weekly learning rate in fifth grade as:

$$(5) \qquad \hat{w}_{5,i} = \hat{c}_6 + \hat{b}_6\, CS_{5,i} + \hat{\varepsilon}_{6,i},$$

where $CS_{5,i}$ is the class size in fifth grade, for pupil *i*. Then predicted test scores the last week in fifth grade is calculated as:

$$(6) \qquad T^p_{5_2,i} = T_{5_2,i} + (\hat{w}_{5,i} \times (t_6 - t^a_{5_2,i})),$$

where $T_{5_2,i}$ is the standardized test scores in spring fifth grade for pupil *i*.

References

Björklund, Anders. 1998. "Income Distribution in Sweden: What is the Achievement of the Welfare State." *Swedish Economic Policy Review* 5(1): 39-80.

Cooper, Harris, Barbara Nye, Kelly Charlton, James Lindsey, and Scott Greathouse. 1996. "The Effect of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review." *Review of Educational Research* 66(3): 227-268.

Cronbach, Lee J. 1951. "Coefficient alpha and the internal structure of tests," *Psychometrika* 16: 297-334.

Entwisle, Doris R., Karl L. Alexander and Linda Steffel Olson. 1997. *Children, Schools and Inequality*. Boulder, CO: Westview Press.

Entwisle, Doris R., Karl L. Alexander and Linda Steffel Olson. 2007. "Lasting Consequences of the Summer Learning Gap," *American Sociological Review*, vol 72 (April: 167-180).

Erikson, Robert, and Jan O. Jonsson. Eds. 1996. *Can Education Be Equalized?*. Westview Press.

Fryer Jr., Roland G. and Steven D. Levitt. 2004. "Understanding the Black-White Test Score Gap in the First Two Years of School," *The Review of Economics and Statistics*. 86(2): 447-464.

Heyns, Barbara. 1978. *Summer Learning and the Effects of Schooling*. New York: Academic Press.

Krueger, Alan B. 1998. "Reassesing the View That American Schools Are Broken." *Federal Reserve Bank of New York Economic Policy Review* 4(1): 29-43.

Lindahl, Mikael. 2002. "Summer Learning and the Effect of Schooling: Evidence for Sweden," IZA WP 262.

Lindahl, Mikael. 2005. "Home vs. School Learning: A New Approach to Estimating the Effect of Class Size on Achievement," *Scandinavian Journal of Economics*, 107(2), p. 375-394.

Murnane, Richard J. 1975. *The Impact of School Resources on the Learning of Inner-City Children*. Cambridge, MA: Ballinger.

O'Brien, Daniel M. 1998. "Does Minority and Disadvantaged Children Suffer Summer Fallback in Cognitive Skills?" Paper prepared for presentation at the Meetings of the Southern Economic Association in November 1998.

Table 1: Descriptive statistics

	Mean	St. Dev	Min	Max
Test scores (raw scores)				
Fifth-grade, spring	42.32	12.93	8.63	68.87
Sixth-grade, fall	44.12	13.43	5.00	70.50
Sixth-grade, spring	50.74	12.28	12.00	71.00
Test scores (in standard deviation units)				
Fifth-grade, spring	0.00	1	-2.61	2.05
Sixth-grade, fall	0.00	1	-2.91	1.96
Sixth-grade, spring	0.00	1	-3.15	1.65
Demographic and family background variables				
GIRL	0.50	0.50	0	1
IMMIG (non-Swedish parents)	0.23	0.42	0	1
Parents' years of education	12.36	1.96	7.53	19.67
Log (family income) [Family income in SEK]	12.60 [345,300]	0.54 [240,434]	11.19 [72,499]	14.75 [2,557,953]
SES (Socioeconomic index in standard deviation units)	0.00	0.93	-2.51	2.52

Notes: Number of observations is 556. Test scores pertain to the national mathematics tests. Missing values are imputed. SES is calculated as the average of the standardized values of parents' years of education and log of family income. In 1998, 10 SEK was worth about 1 US $.

Table 2: Correlation matrix of the test scores

	Fifth-grade, spring	Sixth-grade, Fall	Sixth-grade, spring
Fifth-grade, spring	1.00		
Sixth-grade, fall	0.73	1.00	
Sixth-grade, spring	0.73	0.78	1.00

Table 3: Correlation matrix for demographic and family background variables

	GIRL	IMMIG	Parents' education	Log (family income)	SES
GIRL	1.00				
IMMIG	-0.03	1.00			
Parents' education	0.03	-0.50	1.00		
Log(family income)	0.01	-0.55	0.72	1.00	
SES	0.02	-0.57	0.93	0.93	1.00

Table 4: Associations of test score levels and changes with background variables

Test score levels and changes:	Spring 5th grade	Summer change	Fall 5th grade	School period change	Spring 6th grade
	(1)	(2)	(3)	(4)	(5)
IMMIG	-.68 (.10)**	-.02 (.07)	-.70 (.10)**	.12 (.07)+	-.58 (.10)**
IMMIG, controlling for other background variables	-.36 (.11)**	-.03 (.10)	-.39 (.14)**	.15 (.08)*	-.24 (.13)*
SES	.35 (.05)**	-.00 (.04)	.35 (.05)**	-.01 (.03)	.33 (.05)**
SES, controlling for other background variables	.26 (.05)**	-.01 (.05)	.25 (.06)**	.03 (.04)	.27 (.05)**
GIRL	-.01 (.08)	-.01 (.06)	-.02 (.08)	.07 (.06)	.04 (.08)
GIRL, controlling for other background variables	-.03 (.08)	-.01 (.07)	-.04 (.07)	.07 (.06)	.03 (.08)

Notes: Number of observations is 556. Standard errors (in parentheses) are adjusted for blocks-clustering when SES is included in the estimation. The dependent variable in column 2 (Summer change) is the difference between standardized test scores in fall of sixth grade and spring of fifth grade. This variable has mean (st.dev.) 0.00 (0.74). The dependent variable in column 4 (School period change) is the difference between standardized test scores in spring of sixth grade and fall of the sixth grade. This variable has mean (st.dev.) 0.00 (0.67). The dependent variables in columns 1, 2 and 3 (standardized test scores at spring of fifth grade, fall of sixth grade and spring of sixth grade), all have mean (st.dev) 0.00 (1.00). + significant at 10%; * significant at 5%; ** significant at 1%.

Table 5: Associations of test score changes with background variables, controlling for earlier test score levels: Without and with adjusting for measurement error in test scores

Test score changes:	Summer change		School period change		
	Unadjusted	ME adjusted	Unadjusted	ME-adjusted	ME-adjusted
		R=0.8		R=0.8	IV
	(1)	(2)	(3)	(4)	(5)
IMMIG	-.22 (.07)**	-.09 (.06)	-.04 (.07)	.11 (.05)*	.13 (.07)+
IMMIG, controlling for other background variables	-.14 (.10)	-.07 (.07)	.06 (.08)	.14 (.06)*	.16 (.09)+
SES	.10 (.04)*	.03 (.03)	.07 (.03)*	-.00 (.02)	-.02 (.04)
SES, controlling for other background variables	.07 (.05)	.02 (.03)	.08 (.04)*	.03 (.03)	.02 (.04)
GIRL	-.02 (.06)	-.02 (.05)	.06 (.05)	.07 (.04)+	.07 (.06)
GIRL, controlling for other background variables	-.02 (.06)	-.02 (.05)	.06 (.05)	.07 (.04)+	.07 (.06)

Notes: Number of observations is 556. Standard errors (in parentheses) are adjusted for blocks-clustering in columns 1,3 and 5 when SES is included in the estimation. The dependent variable in column 2 (Summer change) is the difference between standardized test scores in fall of sixth grade and spring of fifth grade. This variable has mean (st.dev.) 0.00 (0.74). The dependent variable in column 4 (School period change) is the difference between standardized test scores in spring of sixth grade and fall of the sixth grade. This variable has mean (st.dev.) 0.00 (0.67). The estimations in column 5 are from regressions of standardized test scores in spring of sixth grade on background variables and first-stage predicted standardized test scores in fall of sixth grade, where the variables included in the first stage regression are background variables and standardized test scores at spring of fifth grade. The effects of standardized test scores at spring of fifth grade in first stages are always highly significant (p-values always below 0.01).
+ significant at 10%; * significant at 5%; ** significant at 1%.

Table 6: Associations of predicted test score changes with background variables:

Test score changes:	Summer change	School period change
IMMIG	-.09 (.10)	.16 (.09)+
IMMIG, controlling for other background variables	-.12 (.13)	.20 (.11)+
SES	.01 (.05)	-.02 (.04)
SES, controlling for other background variables	-.02 (.06)	.03 (.05)
GIRL	-.06 (.09)	.10 (.08)
GIRL, controlling for other background variables	-.06 (.09)	.11 (.08)

Notes: Number of observations is 556. Standard errors (in parentheses) are adjusted for blocks-clustering when SES is included in the estimation. The dependent variable in column 2 (Summer change) is the difference between predicted standardized test scores in fall of sixth grade and spring of fifth grade. This variable has mean (st.dev.) 0.01 (1.00). The dependent variable in column 4 (School period change) is the difference between predicted standardized test scores in spring of sixth grade and fall of the sixth grade. This variable has mean (st.dev.) 0.00 (0.94). + significant at 10%; * significant at 5%; ** significant at 1%.

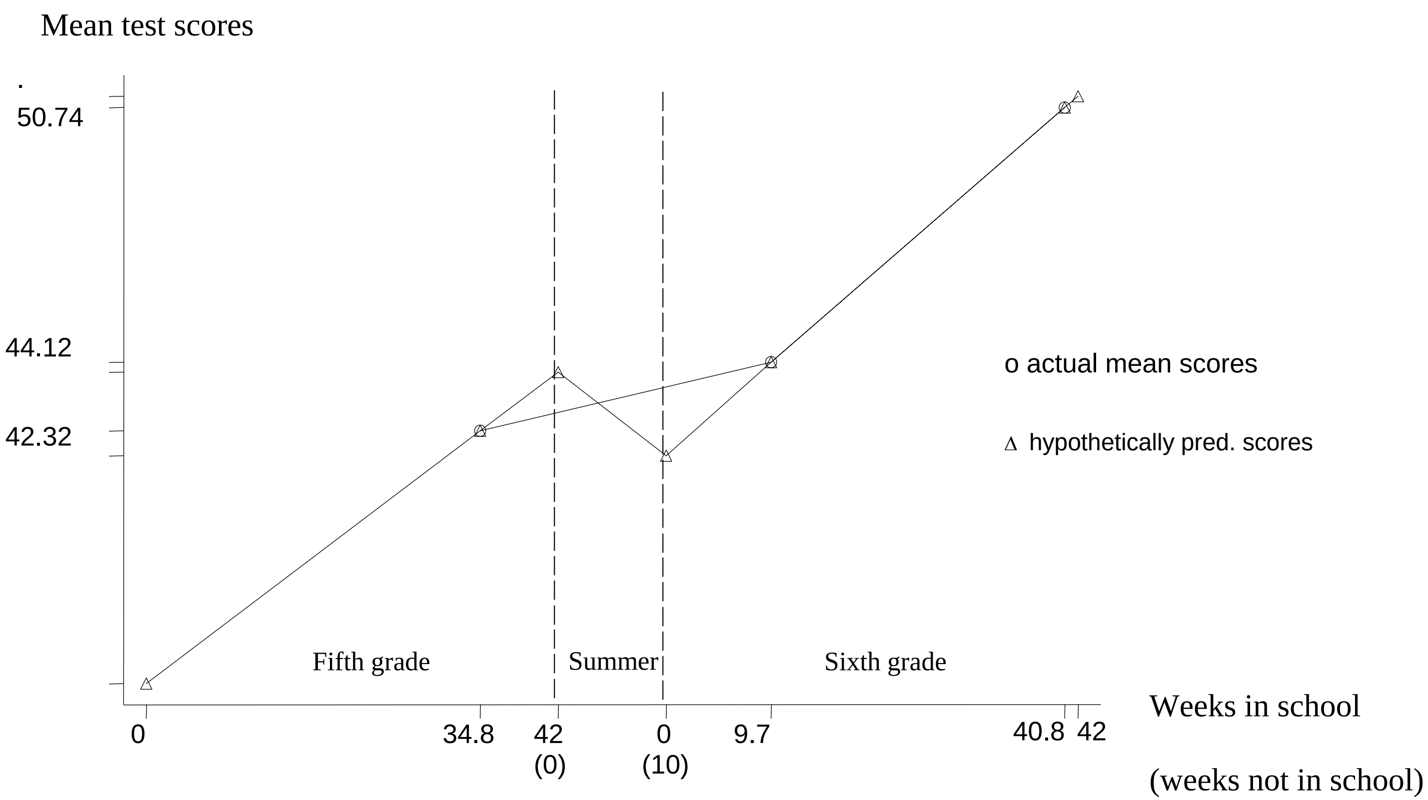

Figure 1: Learning profiles for the mean pupil