WORKING PAPER

Parental education and schooling outcomes: Evidence from panel data on overall development and within year patterns in primary school

Stefa Hirsch

Paul Jungbluth

ABSTRACT

There is no doubt that our family shapes us. But at the latest when children start school they are exposed to a formal structure that allows for comparison and enables competition. Using a series of mathematics tests from Dutch primary education we investigate how parental education relates to student achievement over time and within school years. We find an increasing correlation of parental education level and test scores. Our results suggest that in the first four years the main driver is more precise measurement. While the influence of parental education is first constant or even decreases, it steeply increases towards the transition to secondary school. Furthermore within year patterns suggests an equalizing influence of school during the first three years.

Stefa Hirsch
School of Business and Economics, Economics (AE2)
Maastricht University
PO Box 616, 6200 MD Maastricht
The Netherlands
s.hirsch@maastrichtuniversity.nl

Paul Jungbluth
School of Business and Economics, Economics (AE2)
Maastricht University
PO Box 616, 6200 MD Maastricht
The Netherlands
p.jungbluth@maastrichtuniversity.nl

June 2015

I. Introduction

School is not the first time that children learn. But for many it is the first formal learning environment with the opportunity of comparison. Each child starts school with a unique endowment, shaped by their innate abilities and personality as well as their experiences in the family. School serves all of them and puts them in a comparable setting and provides them with the same information. Still the parental background most likely continues to affect the child's skill development and interacts with the effect school has. Parental pressure and teacher's unconscious expectations based on the parental background are only two of those channels.

In this article we focus on parental education and analyze its relationship with children's schooling outcomes. Our empirical context are twelve mathematics test scores for five cohorts of primary school students in the Dutch province of Limburg. We start our analysis with the observation that the correlation between our measure of parental education and the students' test scores increases throughout the six years that we observe. The same holds for the correlation between scores in consecutive tests. Predictions of a theoretical model show that this pattern could be either driven by increasing importance of parental education for predicting test scores or by decreasing variance in the error term of the test results. A latent variable model, estimating the influence of mathematic ability as well as parental education on the observed test scores, allows us to distinguish these two factors.

During the first three years of school a within-year pattern, suggesting an equalizing effect of school, is dominating the relationship between parental education and test scores. From year four onwards and specifically the last two years leading up to the final assessment test the influence of parental education increases strongly from test to test. The results in this final assessment test itself, which is the basis for secondary school track placement, are less dependent on the parental background than those of the last regular test. This latter result is consistent across all analyses; simple correlations as well as regressions and different model specifications.

A decomposition of the total variance in a mathematic ability, a parental education and an error component allows to compare the different factors: All model specifications show a clear decrease of the errors' standard deviation until grade four, some even until grade five. In the first year it accounts for multiple times as much of the total variance than the component referring to parental education. Even with the decrease the error term still explains a sizable part of the variance, exceeding that explained by the parental education level.

The basis for the analysis are panel data on test scores and family background from more than 6.000 students of five cohorts from the Dutch province Limburg. The data consist of administrative records as well as questionnaire-based data, collected within the ongoing research project , within an ongoing regional education monitor conducted as a cooperative project between schools and Maastricht University since 2009. An advantage of this data set for this particular analysis is the large number of comparable tests over the course of six years. All but the last test are validated to be comparable over different school years and to measure mathematic ability consistently. The last test differs from the others because it is constructed as an assessment test at the end of primary education. It is still validated to be comparable across cohorts. The test result is a main determinant of the teacher's binding advice for the secondary schooling track placement.

This article is related to the literature dealing with the reproduction or even amplification of inequality through intergenerational dependency of educational attainment – in short: the rich get richer, the poor get poorer. The educational or sociological field calls this the Matthew Effect, referring to a verse in the biblical Gospel of Matthew (Luyten, Cremers-Van Wees, & Bosker, 2003; Sammons, 1995). In this context increasing correlations of measures for the socio-economic background with schooling outcomes over time within one cohort are considered an observation supporting an increasing influence of socio-economic factors. The economic field takes a different approach. They leave these terms aside and mostly examine test score gaps and how they develop over time (Fryer Jr & Levitt, 2004; Polidano, Hanel, & Buddelmeyer, 2013). This method has been criticized for treating test scores as being measured on an interval scale. Theoretical considerations have shown the low robustness of common methods in this field to transformations of test scores (Bond & Lang, 2013; Lang, 2010).

In addition to the methodological problems pointed out by Bond and Lang another difficulty lies in the fact that each test score is also influenced by an individual error term. In case the size of the error term systematically evolves over time this might influences the observed correlations as well as the comparison of test score gaps. We address this concern by estimating the predictive power of mathematic ability and parental education as well as the error term for each test simultaneously. The unique data set of Dutch primary student outcomes allows us to analyze the relationship of parental education, measured as the highest parental education level, with twelve standardized tests over the course of six years of primary school.

The remainder of this article is structured as follows: In section 2 we introduce the data set. As a starting point for the analysis in section 3 we take a naïve glance at the data and provide the correlation of our parental education measure with the series of test scores as well as the correlation of consecutive test scores. Based on a simple model, in which the test result depends on mathematic ability as well as parental education level, in section 4 we distinguish two hypothesis of how an increase of the correlation could be explained. In section 5 the results of some regression approaches as well as our latent variable model are presented, including some robustness checks and a variance decomposition. Section 6 includes a discussion of the results and its implications and section 7 concludes.

II. Data and empirical context

a. Data

The basis for the analysis is a panel of test score data for more than 6.000 primary school students of five cohorts. The data consist of administrative records as well as questionnaire-based data, collected within the ongoing regional education monitor (Onderwijsmonitor Limburg, short OML), conducted as a cooperative project between schools and Maastricht University since 2009. All regular primary schools in the southern part of the Dutch province Limburg were asked to participate in the project. [1]

An advantage of this data set for this particular analysis is the large number of comparable tests over the course of the six years of Dutch primary education.[2] We use test scores for two mathematics tests each year, one taken in the middle and one taken at the end of the year. Only for the final year in place of an end-year test we use the score in the mathematic section of the nation-wide skill assessment test at the end of primary school. Usually this test is taken shortly after the mid-year test. Officially this test is called “Cito Eindtoets”, in the following we will refer to it simply as final assessment test.[3] If nothing else is stated explicitly the term “final assessment test” always refers to only the mathematics section.

All tests used are constructed by Cito, “Centraal Instituut voor Toetsontwikkeling”, meaning central institute for test development. The institute was founded by the government

[1] Limburg is one of twelve provinces of the Netherlands covering the southeastern part of the country. It shares a long boarder with Germany in the east and Belgium in the west and south. The project focusses on all schools south of the city of Sittard.

[2] In the Netherlands education starts at the age of four and primary school runs for eight years (“groep 1” to “groeop 8”). The first two years consist mostly of preparatory activities and the types of tests taken during those years are not comparable to the tests considered here. In accordance with the international classification “groep 3” is hereafter referred to as grade one and primary school is considered to run from “groep 3” till “groep 8” (Central Bureau voor de Statistiek, 2011).

[3] The schools are required to use some test at the end of primary school. In 2013 about 86 per cent of the schools used the “CITO Eindtoets” for this purpose (Inspectie van het Onderwijs, 2014). From 2014 onwards schools are obliged to use a standardized test at the end of primary school. There are also other providers, but Cito is most likely to remain the dominant one.

in 1968 to develop and maintain standardized tests. Since 1999 Cito is a fully privatized testing and assessment company, active also internationally.[4] Through the mother foundation it still indirectly receives public subsidies for the development of new tests or for pilot studies.[5] Schools have to buy all tests they want to use. The more students they are testing, the more they have to pay. In addition Cito also offers a software for analyzing student's test scores and their development.

In this study we use all mathematics tests available for grade one thru six. The regular mid- and end-year tests are validated to be comparable over the years and to measure mathematic ability consistently. They are administered and graded by the schools themselves, using a standardized coding scheme. While those tests can be considered low stakes tests, the final assessment test at the end of grade six, administered by the schools, but graded externally, is of high importance. The result is a main determinant of the teacher's binding advice for the secondary schooling track.

For the regular mid- and end-year tests raw scores as well as skill scores are provided. The former correspond to the number of correct items in the specific test and are therefore a relative performance measure for everyone participating in that test. The latter are a measure for the mathematics skills, allowing also comparisons across different grades. Most of the analyses provided are based on percentiles of raw scores. Also because schools can choose between two levels of detail for reporting the skill scores. Nonetheless robustness checks using the skill scores support the claim that the results are independent from this choice.

Our measure for socio-economic background are self-reported parental educational level. It is assessed when the children are in grade six. The answers are coded according to the five levels of the International Standard Classification of Education. In the Dutch education system this corresponds to the following categories: ISCED 1 for at most primary education,

[4] Cito B.V. (2014).

[5] Buitelaar, Ros, Vink, and van der Kroft (2013).

ISCED 2 for lower secondary education (vmbo), ISCED 3 for upper secondary or lower vocational education (havo/vwo), ISCED 4 for upper vocational education and ISCED 5 for higher education at the university level.

b. Sample

The aim of the regional education monitor OML is to collect panel data for all education levels (pre-school, primary school and secondary school) in Limburg. The data used here refer to the sub-project focussed on children in the last year of primary school. In all years more than 90 per cent of the regular primary schools in South Limburg participated. The participation rate even increased until 2013.

The cooperating schools provided access to administrative data, including students' test scores for the final assessment test as well as for earlier regular tests if used by the respective school. Since test results are stored digitally we also have the complete records for students of the early cohorts in the sample as long as they did not change school.

Test score data are complemented by questionnaire data assessed when the children were in their last year of primary school. In April, in the last third of the school year, the schools administered a survey taken by the children themselves and sent out questionnaires to the parents. In case the parents objected to the data of their child being used this child was removed from the data set completely. Each year between 1 and 2 per cent of all parents made use of this opportunity. Even though this is most likely a specific group of parents, the number of cases concerned is small enough to be disregarded. The overall parental response rate was between 66 and 71 per cent. Concerns regarding potential selection bias will be addressed as far as possible by providing descriptive statistics for the children based on the availability of parental information and test scores.

Table 1 Descriptive statistics according to availability of test scores conditional on parental information

	(1) Balanced sample			(2) Unbalanced sample (excl. balanced sample)			(3) At most final assessment test & parental information			(4) Unknown parental education level		
Variable	N	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd
Age in years in grade six	1,359	11.81	0.43	4,989	11.86	0.49	10,543	11.81	0.47	15,680	11.89	0.50
Gender (% males)	1,356	0.48	0.50	4,990	0.49	0.50	10,962	0.48	0.50	16,306	0.51	0.50
Weight / extra funding (in %)	1,359	0.10	0.30	4,991	0.09	0.29	10,776	0.10	0.30	16,117	0.17	0.37
Migration background (in %)	1,355	0.14	0.35	4,979	0.14	0.35	11,070	0.14	0.35	2,224	0.07	0.26
Repeated a school year (in %)	1,224	0.09	0.28	4,443	0.11	0.32	9,983	0.09	0.29	230	0.17	0.38
Highest parental education (1-5)	1,359	3.23	1.09	4,991	3.29	1.09	11,089	3.39	1.05			
Total score on IQ test (0-43)	1,332	32.52	4.80	4,830	32.61	4.66	10,671	32.47	4.61	5,420	31.60	5.11
Final assessment test score (500-550)	1,359	537.60	8.24	4,991	537.10	8.67	10,544	537.30	8.49	15,709	535.70	9.36
Final assessment math section (% correct)	1,359	75.64	15.87	4,991	73.45	17.30	10,544	73.53	17.09	15,709	71.86	17.83

Note: The table reports the number of observations, the mean values and the standard deviations for the listed variables by group. The groups are mutually exclusive. The values reported for the unbalanced sample therefore refer to the unbalanced sample excluding students from the balanced sample.
The age variable is based on the students' month-exact age at the beginning of grade six, their final year in primary education. The variable "weight" refers to a dummy variable based on whether the school receives extra funding for that student based on an expected disadvantage. This is the case either based on a low socio-economic background of the parents or in the past also the migration background of the child. A child is considered to have a migration background if at least one parent was not born in the Netherlands. The dummy variable on repeating a school year is coded one in case the child repeated one of the years in primary school as defined here. Spending an additional year in pre-school does not count as a repeated year. The highest parental education level is measured in terms of the five ISCED levels as throughout the analysis. The score on the IQ test refers to the number of correct items in a non-verbal intelligence test conducted in the final year of primary education around the same time the final assessment test was taken by the children. The final assessment test score reflects the overall score including sections on language, mathematics and study skills. The percentage of correct answers in the mathematics section reported last in the table is the measure for the last test considered in the analysis.

c. Descriptive statistics

In order for a student to be in the sample that is used for the analysis the following criteria have to be fulfilled: (1) the school has to use the tests provided by Cito and had to agree to provide us with the test scores, (2) the parents did not object the data use and (3) at least one parent had to participate in the survey and provide information on his or her highest attained education level. For the analysis we use two different samples. The unbalanced sample includes all 6350 students for which the parental education level, the result of the final assessment test and at least one additional test result are available. The balanced sample in contrast is restricted to the 1359 students for which all twelve test scores are available. Table 1 provides some descriptive statistics for the two samples used (column 1 and 2) and compares them to students with known parental education only taking the final assessment test or no test (column 3) as well as to students with unknown parental education level (column 4).

In many respects the balanced (column 1) and the unbalanced sample (column 1 and 2 together) that are the basis for our analysis are similar. The most notable difference is that the fraction of children who repeated a school year is two percentage points lower in the balanced sample than in the unbalanced sample. This also leads to a lower average age in grade six and might explain the slightly lower test scores in that group. Nonetheless the unbalanced sample offers the advantage of a larger number of observations. So that it is a good source of information for the overall developments in primary school. But due to selective testing[6] within year patterns will not be interpreted for this sample.

Comparing the mean values and the standard deviations of the balanced sample with those of the third group shows almost no differences. Only the average parental education level is a little higher for the latter group. With about twice as many observations as the unbalanced sample this group accounts for the majority of students for whom parental

[6] In appendix A we provide information on the respective number of observations and the composition of participating students for each of the tests in the unbalanced sample.

information on education are available. Since the balanced sample barely differs from this group with respect to the observed outcome variables we however conclude that the selection with respect to being tested or not should not have much influence on our findings.

Selection with respect to the availability of parental information in contrast seems to be more important: The fourth group for which Table 1 reports descriptive statistics differs strongly from the others. In order to look at this difference in more detail Table 2 reports the same descriptive statistics as presented in Table 1, but based on another group distinction. Conditional on the availability of the final assessment test result and at least one more test result the students are distinguished according to whether their highest parental education level has been reported or not. As a result Table 2 compares all children that are either in the balanced or the unbalanced sample described above to a sub-group of the fourth group in Table 1, those for whom no parental information but some test score data are available.

Table 2 Descriptive statistics according to parental information, conditional on test score data availability

	(1) Unbalanced sample (incl. balanced sample)			(2) No parental education level		
Variable	N	mean	sd	N	mean	sd
Age in years (Jan. last school year)	6,348	11.85	0.48	6,338	11.90	0.50
Gender (% males)	6,346	0.49	0.50	6,336	0.50	0.50
Weight / extra funding (in %)	6,350	0.10	0.29	6,353	0.17	0.37
Migration background (in %)	6,334	0.14	0.35	721	0.08	0.28
Repeated a school year (in %)	5,667	0.11	0.31	87	0.23	0.42
Highest parental education (1-5)	6,350	3.27	1.09			
Total score on IQ test (0-43)	6,162	32.59	4.69	1,863	31.65	5.29
Final assessment test score (500-550)	6,350	537.20	8.58	6,353	535.90	9.19
Final assessment math section (% correct)	6,350	73.92	17.03	6,353	72.41	17.41

Note: The table reports the number of observations, mean values and standard deviations for variables according to groups with different availability of parental information, conditional on the availability of test score data. For a detailed description of the variables see the note of Table 1.

Table 2 shows that children whose parents did not participate in the survey or did not report their education level are a lot more often assigned a weight for extra funding. The means for repeating classes and the migration background also differ widely. However based on the large number of missing observations for the second group this last comparison is not informative. As the fraction of children with extra funding already suggests, the achievement in the IQ test as well as in the final assessment test and the mathematic sub-section are on average higher for children for whom the parental education level is known. At the same time the variance of the skill measures is larger than in the other group. This implies that there are a lot more children with low socio-economic background, but also some more children with high socio-economic background in this group compared to our samples. Unfortunately we are unable to further investigate this selection based on the available data.

III. A naïve glance at the correlation

In the debate about intergenerational mobility and inequality, in the media as well as in the scientific literature (Torche, 2015), authors commonly refer to observed correlations between schooling outcomes and parental characteristics to support their claims or theories. And with only one outcome measure at hand this is often the best that can be done. Based on a series of available tests scores we are able to go beyond that. We take this widely used measure as a starting point, before introducing our model. This allows for easy comparisons to other studies and contexts. In that sense this section can be seen as an extension of the descriptive statistics. In the results section we will compare the results based on the latent variable model with the analysis of correlations and discuss the added value of the former.

a. Magnitude of correlations between mathematics skills and parental education

First, to see whether the observed correlations are of comparable magnitude to other data, we make a comparison with test scores from PIAAC, the OECD's Programme for the International Assessment of Adult Competencies. In general the correlation of parental education with the mathematics score in the final assessment test are very similar to the comparable correlations calculated based on the corresponding sub-scores that Dutch adults attain in PIAAC (see Table 3). Even though one has to be careful in drawing conclusions from this comparison across age groups and generations this similarity shows that the magnitude of the correlation observed in the data is not out of the ordinary.

Table 3 Correlations of test results with highest parental education level, comparing mathematics sections of final assessment test and PIAAC

	Language / literacy	Mathematics	Problem solving / study skills
Final assessment test, OLM	0.3392	0.2734	0.3269
PIAAC test results for Dutch adults	0.3157	0.2651	0.3287

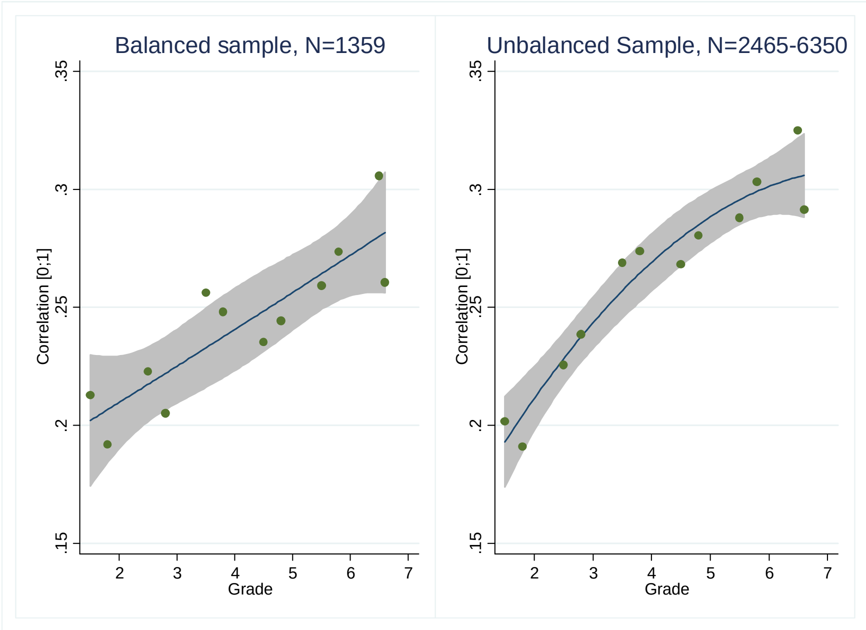
Note: In order to compare the correlations the coding of parental education in the OLM data was first adjusted from five to three categories, as it is the case in the PIAAC data.

b. Correlations with parental education over the course of primary school

As a starting point for our analysis we take a first naïve glance at how the correlations between the mathematics test scores and the highest parental education level develop over time. Figure 1 shows this correlation separately for each test from grade one thru grade six for the balanced sample as well as for the larger unbalanced sample. The graphs further include quadratic prediction plots and confidence areas. In both cases the correlation clearly increases over time, but there are some notable differences.

Both prediction plots start with or slightly below a correlation of 0.2 in grade one. But the correlations in the unbalanced sample rise steeper, therefore leading to a higher end point. Converting it into percent of explained variance of the linear relationship between test scores and parental education level this development corresponds to an increase of about 4.3 or respectively 5.5 percentage points.[7]

Figure 1 Correlation of mathematics test results with highest parental education level



Note: The unbalanced sample includes groups (1) and (2) from Table 1. Since not every student took every test in this sample, the number of observation varies. The respective number of observations for each of the tests can be found in Appendix A. The balanced sample only includes group (1) from Table 1.

The variable "grade" captures the grade as well as the time during the school year, when the respective test is usually taken. For the result of the mid-year test in grade one it takes on the value 1.5 and for the end-year test in grade one it is 1.8. The same scheme applies for grades two thru five. In the final year the grade for the mid-year test is 6.5 as usual and since the final assessment test is only taken shortly after the mid-year test the grade assigned is 6.6.

Furthermore the within year patterns are quite different between the samples. The correlations in the balanced sample show a decreasing within year pattern for the first years and a monotonously increasing pattern across years from grade four onwards. The only within year decrease of the correlation in the unbalanced sample is observed in the first year, which is also the year when the number of observations for the mid- and end-year test is most comparable in magnitude (see Appendix A). Apart from those differences there are also

[7] This difference is not significant. Still this points to potential differences between schools that test their students consequently and non-selectively and schools that test infrequently or selectively. Those differences could refer to many factors, for example on student composition or teaching quality, just to name two.

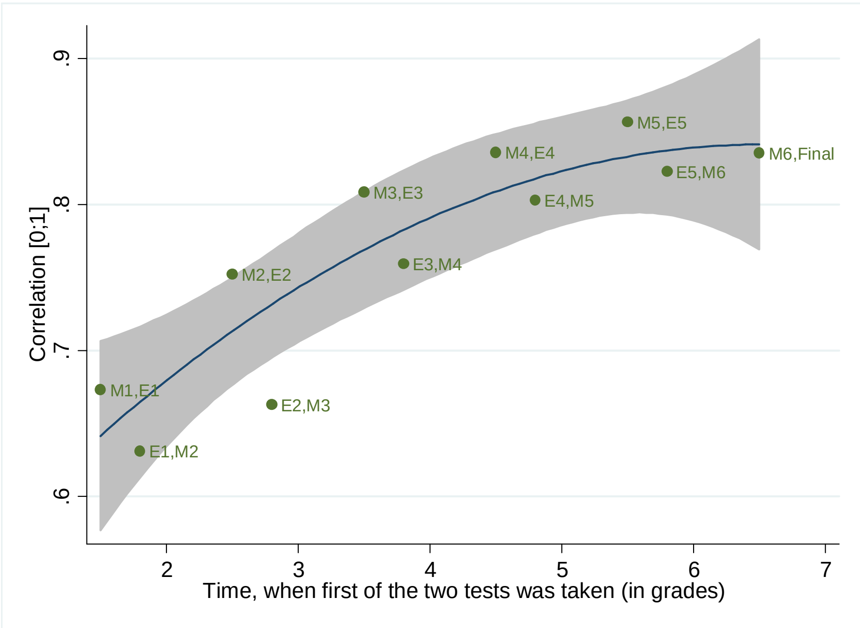
similarities: Firstly, in both cases the correlation peaks at the mid-year test in grade six. Secondly, the correlation of the final assessment test with parental education level is clearly lower and closer to the correlations observed in grade five. These observations will be discussed in more detail and related to the main analysis in the results section. To conclude, while observed within year patterns are less reliable in the unbalanced sample the overall development is relatively comparable in both samples.

c. Current test scores as predictors of future test scores

Another way of looking at how the test scores develop over time is to analyze how well one test score predicts a later one. In Figure 2 we therefore plot the correlations of consecutive test scores (test one and two, test two and three, and so on) as well as a quadratic prediction plot of how they develop. Overall the correlation between tests following each other increases by more than a third. Naturally, due to teacher effects or similarities in content, test results within one year are higher correlated than across years. This holds for all grades except for the first year of schooling.

The figure shows that as the children grow older and reach higher grades their current test results become increasingly better predictors of future test results, reaching more than 65 per cent of explained variance in grade five. One potential explanation for this observation could be an increase in the precision of the tests, which would correspond to a decrease in measurement error. Based on a simple model that we present in the next section we will discuss this and examine whether there could be alternative explanations.

Figure 2 Correlations of consecutive tests with quadratic prediction plot and confidence interval



Note: The labels show which test scores have been correlated for that data point. M (for middle) and E (for end) denote the time of the school year when the test was taken and the number refers to the grade. "Final" stands for the mathematics section of the final assessment test. The correlations displayed are calculated based on the balanced sample. Using the unbalanced sample leads to a very similar graph.

IV. Theoretical considerations

a. Model

We propose a simple model in which the schooling outcome in each time period is determined by the individual's mathematic ability as well as highest parental education (PE). Equation 1 shows this for a single period. Y_i denotes the individual i's test score, $MATH_i$ the individual mathematic ability and PE_i the highest parental education level. The individual error term is denoted ε_i.

$$Y_i = c + \beta_1 MATH_i + \beta_2 PE_i + \varepsilon_i \qquad (1)$$

Considering several periods changes the equation in the following way: The test score varies with time, but MATH and PE are assumed to be constant. For PE this is straight forward, since it is unlikely that the parental education level changes during the child's time in primary school and even if it did, we only observe it at one point in time. That the underlying general mathematics ability does not change is a stronger assumption. But the results of the literature studying children's skill development (Cunha & Heckman, 2008; Heckman, 2006) supports the claim that, at least after the age of eight, not much change in cognitive abilities occurs. Later in a robustness check we will relax this assumption.

Even though the ability and the parental education level do not change over time, still their influence on the test score might differ from year to year. Practical reasons for this could be different types of tasks in the tests or changes in how the teacher reacts to the child's parental background. Therefore we allow the parameters to vary across periods, in our case across tests taken it different points in time. The resulting model is presented in Equation 2. The parameters to be estimated are the coefficients of MATH and PE for each time period as well as the error term, which may also vary over time.

$$Y_{it} = c_t + \beta_{1t} MATH_i + \beta_{2t} PE_i + \varepsilon_{it} \qquad (2)$$

Here MATH is a latent variable, based on what all twelve test scores have in common. It can therefore be interpreted as an intra-individual constant, determining success in mathematics tests. In that sense it includes not only mathematic ability, based on innate ability and on out-of-school training, but also test taking ability, influenced by non-cognitive skills and motivation as well as additional factors that are constant over time within the individual. This being said, for reasons of simplicity we will refer to it as MATH or mathematic ability. Estimates of β_{1t} are measures of how important this composite is for success in that specific test and β_{2t} captures the explanatory power that parental education has for the mathematics test score Y in test t across individuals.

b. Explaining observed correlations

In section III we have shown that in the data set of Dutch primary school students the correlation of PE_i and Y_i as well as the correlation between Y_{it} and Y_{it+1} increase over time.

For a correlation to increase the model outlined above provides two alternative reasons: (1) Either a common determinant of both measures or the coefficient linking the measures increases, or (2) a determinant they do not share decreases in magnitude. For both pairs (Y_{it}, PE_i) and (Y_{it}, Y_{it+1}) a common or connecting factor is β_2 and a distinguishing factor is ε_i. Therefore this framework provides two possible explanations for the observation of an increase in the correlation of PE with the test scores and an increase in the correlation between consecutive test scores: Either the parental education level becomes increasingly important for educational success over time or the measurement error decreases.

We are aware that a changing measurement error could be driven by different factors, e.g. systematic changes with age or increasing precision of the tests. We will mainly concentrate on the decomposition of the total variance and the general development of the different factors. In the final section potential underlying mechanisms are discussed.

V. Results

a. Influence of parental education on mathematic test scores over the course of primary school

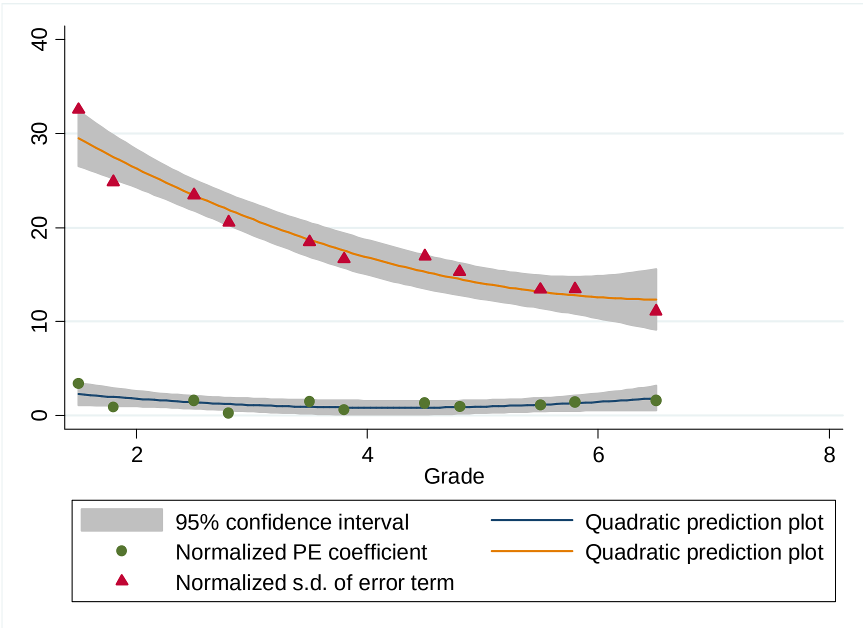
According to the outlined model the increased correlation is either driven by an increase in the PE coefficient or by a decreasing measurement error. The empirical problem in estimating the model is that the students' mathematic ability is not directly observed. In the following we discuss different ways to handle this.

In a simple approach the final assessment test can be used as a measure for student ability. Run twelve regressions, one for each test, provides then a series of estimates for the coefficients. The development of the resulting PE coefficients as well as the standard deviation of the error term, both normalized by the coefficient for MATH, are presented in Figure 3.[8] Striking is that, first, the standard deviation of the error term strongly decreases.

[8] We report the normalized coefficients since the coefficients for MATH might also differ across tests.

And second, it is still always more than twice as large as the respective PE coefficient. Compared to the magnitude of the changes in the standard deviation of the error term the estimated influence of parental education stays almost constant, showing only a slight u-shape.

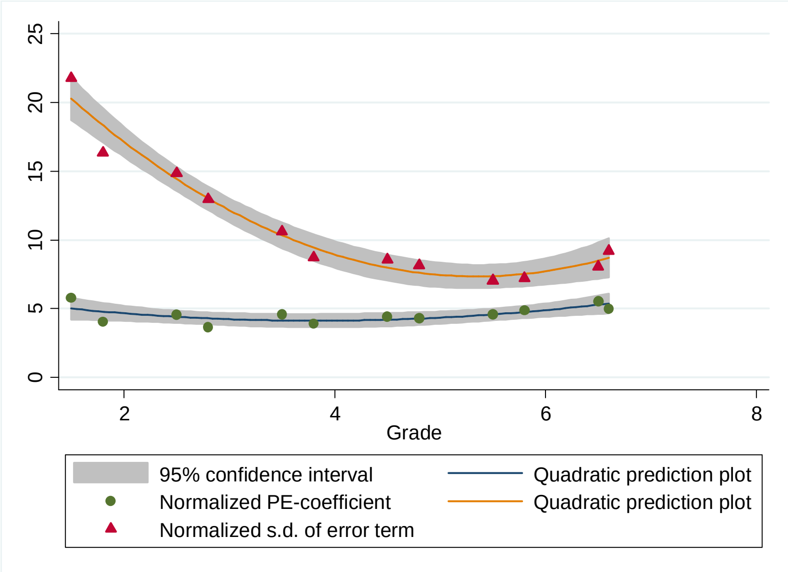
Figure 3 Mathematic ability defined as percentile rank the final assessment test, normalized coefficients of PE and standard deviation of error for separate regressions per test



Note: Final assessment test left out, since it was used as a measure for mathematic ability.

This approach provides us with a first idea of the magnitude and development of the PE coefficient and the error term. But using the result in the final assessment test as a measure of ability is somewhat critical. Maybe content-wise the final assessment test comes closest to measure overall mathematic ability, but time-wise it is like putting the cart before the horse. The larger measurement error in the beginning might only reflect the longer time difference between the points of testing. Moreover this measure of mathematic ability makes no use of the fact that the data set offers twelve independent measures of mathematic ability.

Figure 4 Mathematic ability defined as latent variable estimated based on a structural equation model, normalized coefficients of PE and standard deviation of error for separate regressions per test



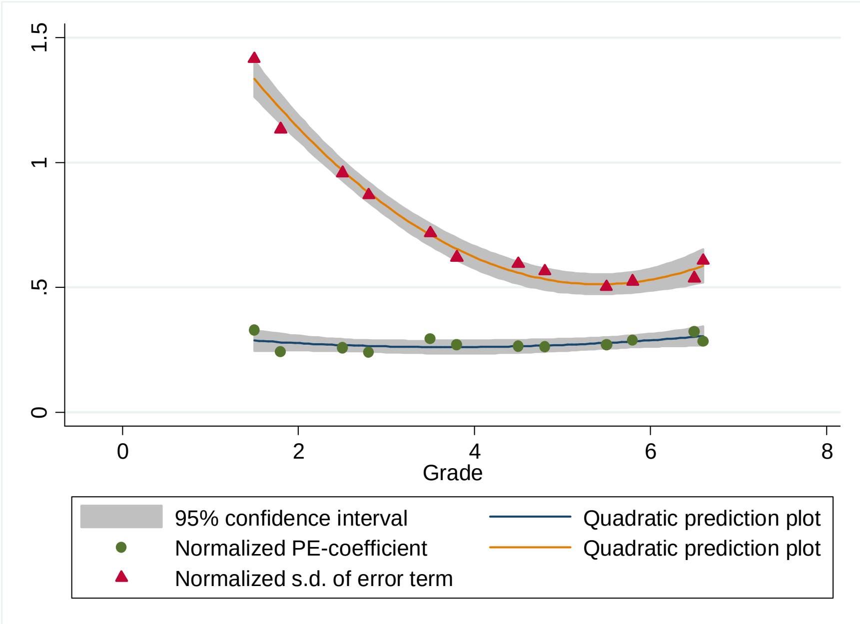
Alternatively a structural equation model can be used to estimate a hypothetical mathematic ability for each child out of what all the tests have in common, using also the covariance between tests.[9] In Appendix B we explain for a simple case of three tests how the estimation procedure of the structural equation model works by only using the variance-covariance matrix. Figure 4 shows the results in a graph comparable to Figure 3. In this case the normalized standard deviation of the error term is generally lower and the normalized PE coefficient consistently higher. But the pattern is the same. In addition the R^2 is higher for each of the underlying regressions. All in all this measure of mathematic ability seems to fit the model better than the results on the final assessment test.

Figure 4 is still based on twelve separate calculations. The information from the covariance between tests is used to predict values for the latent variable MATH, but not for fitting the model itself. The coefficients can also be directly based on the structural equation model introduced above. So while the regression approach estimated twelve times Equation 1,

[9] As already stated above we do not claim that this is a measure of innate mathematic ability. The measured ability is most likely a product of innate ability as well as experiences or stimulation during early childhood, non-cognitive skills and motivation.

for each test separately, the structural equation model estimates the twelve equations simultaneously, which is equivalent to Equation 2. Figure 5 presents the normalized PE coefficients as well as the standard deviation of the error term of the different tests for our preferred structural equation model (sem). The results shown are based on the balanced sample. The comparable results for the unbalanced sample (not shown) are almost indistinguishable.

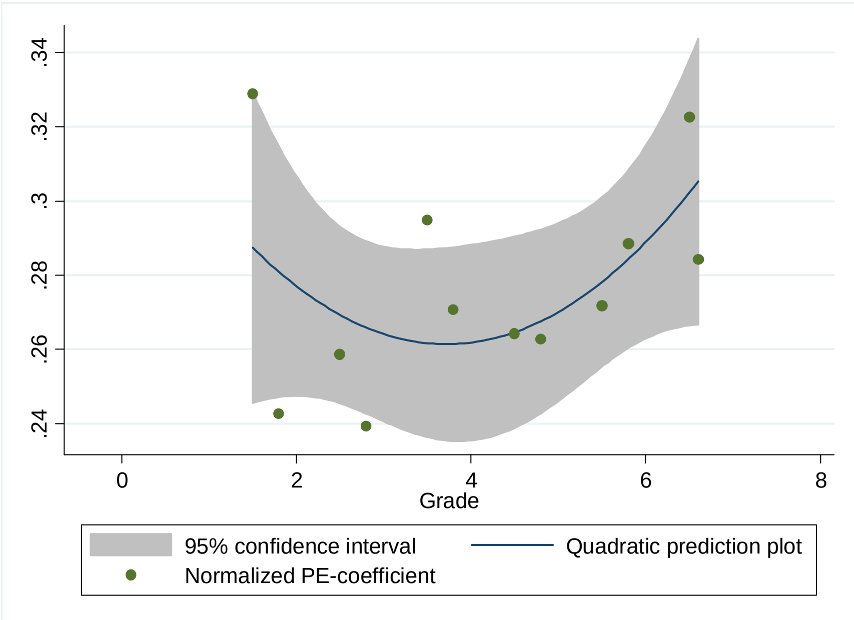
Figure 5 Mathematic ability defined as latent variable, normalized PE coefficients and the errors' standard deviations based on structural equation model, balanced sample



Note: In order to account for teacher specific components of the error term, the model allows the error terms of tests to correlate within each school-year. Furthermore the mid-year test scores of consecutive years are may also correlate. Only the error term of the final assessment test is assumed to be uncorrelated with all other error terms since the test differs in many respects from the others and it is graded externally. Standard errors are clustered at the school and year-combination level.

While the pattern of correlations are very different for the balanced and the unbalanced sample the coefficients calculated based on the sem are very similar. With regard to the overall development over the course of primary education a clear decrease of the errors' standard deviation of test scores until the mid-year test in grade five becomes apparent. The development of the normalized PE coefficient from Figure 5 is shown in more detail in Figure 6. In the first year a sharp decrease, almost from the highest to the lowest of all of the values, stands out. In year two and three the within year pattern is similar. The coefficients in the third year stand out in the level, the coefficients for both tests are higher than those in the second and fourth year.[10] A monotone increase can only be observed for the last four regular tests, starting with the end-year test in grade four and continuing until the mid-year test in the final grade, when it reaches a level similar to the starting point. But in general the magnitude of the error's standard deviation exceeds that of the PE coefficient by far.

Figure 6 Development of normalized PE coefficient (detail from Figure 5)



b. Robustness checks

The results shown for the development of the PE coefficient and the standard deviation of the error term in Figure 5 are based on percentiles of raw scores and on a specific structural equation model. The assumptions of the sem presented above are that mathematic ability and

[10] Grade three stands out in most of the figures, but it is not surprising. There are three ways in which the tests of that year differ from the other ones: Firstly, the biggest change in the primary school curriculum happens at that time. New tasks, such as multiplication and division as well as calculating with units, are introduced. Secondly, according to experts it is the only year with major discrepancies in the commonly used textbooks. Therefore, if some children are confronted with tasks they did not learn to solve at school, their prior knowledge and thereby their home environment is likely to play a larger role in determining success. Thirdly, it is likely that the children read the questions themselves for the first time. Until the mid-year test in grade two all tasks are read out by the teacher. The end-year test in grade two already includes texts for the children, but it is likely that the teacher still supports the children during this first time.

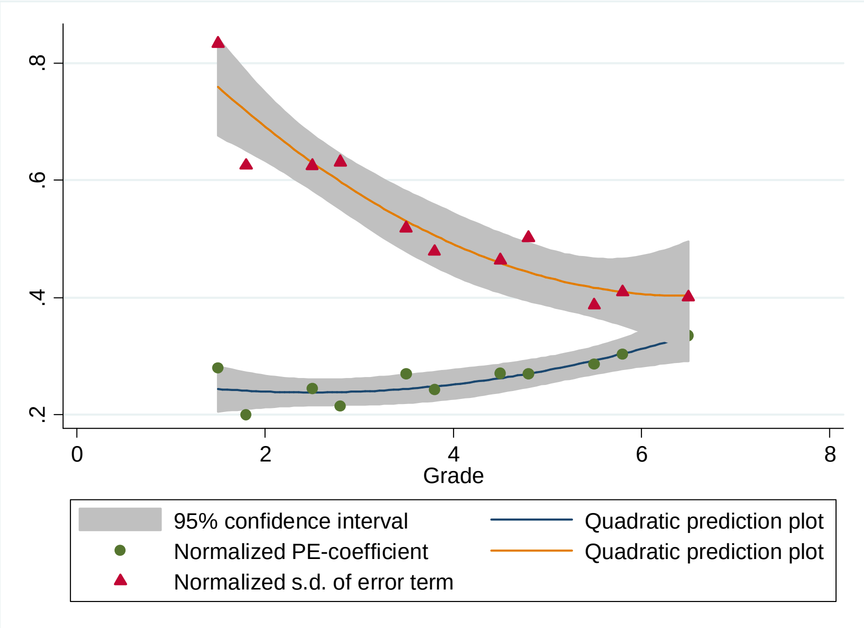
the highest parental education level are uncorrelated, that the error term of tests correlate within one year and the error terms of the mid-year tests additionally correlate across consecutive years. Furthermore clustered standard errors are used. Holding the definition of the test scores constant the results are very robust against different specifications of the sem. Allowing for correlation between the two independent variables as well as not allowing any correlations, neither between the independent variables nor between error terms, both yield almost indistinguishable results to the ones shown in section V.a.

An additional robustness check focusses on the underlying test scores. Based on the preferred model it does not matter whether raw scores or skill scores are used to build percentiles. And using directly the raw scores or skill scores instead of the percentiles only changes the scale of the y-axis but not the relative development of the coefficients and standard deviations.

Another concern might be that the results are influenced by the composition of the sample. Apart from selective testing, which is already addressed by comparing the balanced and unbalanced sample, students might be tested at different points in time than it is meant to be. If students lack behind or if they are already ahead teachers sometimes use tests that are supposed to be taken in a different grade. In our balanced sample this concerns 70 of the 1359 students, so a little above 5 per cent. Excluding those students only changes the results minimally. In the graph comparable to Figure 5 this is indistinguishable.

As the robustness checks show, the structural equation model produces consistent results across samples, different definitions of the test scores and across a range of model specifications. But all approaches discussed so far are based on the crucial assumption that the mathematic ability is constant over time. In order to keep the advantages of the structural equation model that estimates the latent variable based on several tests, but allowing for changes in mathematic ability we use a “rolling” structural equation model. This means, that we use one test before and one test after the test of interest to estimate the model. For the second test for example we use test one thru three, for the third test we use test two thru four, and so on. In consequence based on this method no estimates can be generated for the first and last test in the series. Since the development in the beginning is of special interest we use tests on basic calculation and sorting skills to calculate a prior measure.[11] Unfortunately there is no similar opportunity for the last data point based on the final assessment test.

Figure 7 PE coefficient and standard deviation of error, normalized by MATH coefficient, based on rolling structural equation model



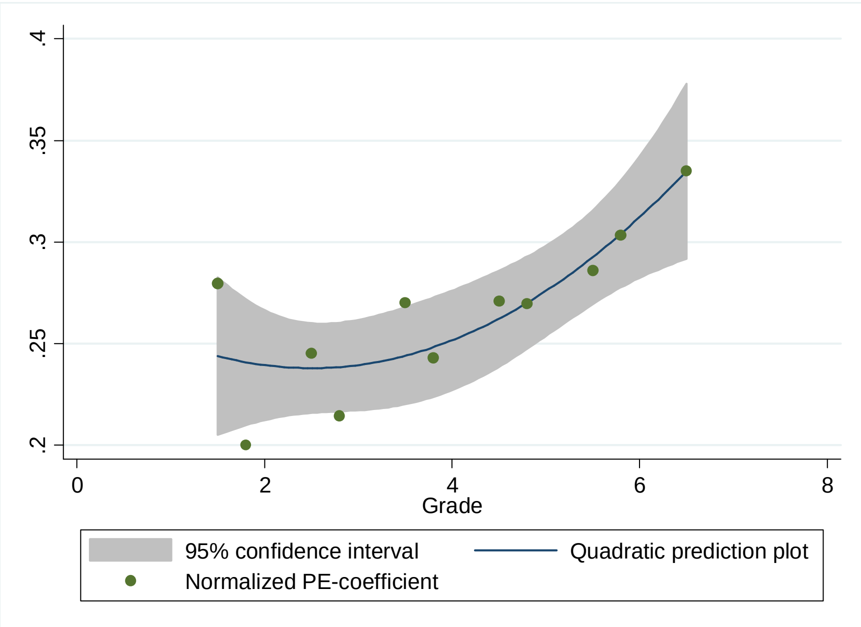
Note: Based on sub-sample to balanced sample, including results from last year of pre-school N=823.

The results of the rolling sem are shown in Figure 7. The increase in the magnitude of the normalized PE coefficient is more pronounced here. At least partly this is caused by the missing last point in the series of test scores. As the other graphs before show the PE coefficient is relatively lower for the last than for the second last test. So leaving out this test changes the picture. Further the u-shape of the coefficient on parental education is less pronounced (see Figure 8). The magnitude of estimated influence of parental education in the

[11] Those test results are available only for a subset of our balanced sample. In order to keep the number of observations as large as possible we combine the two available tests. We use the percentile rank of a weighted sum of both test scores. If only one was available, that score was used instead of the weighted sum.

first test is now on the same level as in year four not as before in the last year of primary school. What stays are the qualitative implications that in the first three years within year changes in the influence of parental education are dominating the development and only during the two years leading up the last regular test the coefficient strictly increases. In grade six the normalized PE coefficient is of similar magnitude as the standard deviation of the error, but the coefficient's increase over the complete six years is still only half of the decrease of the error's standard deviation.

Figure 8 Development of normalized PE coefficient based on rolling sem from Figure 7 in detail

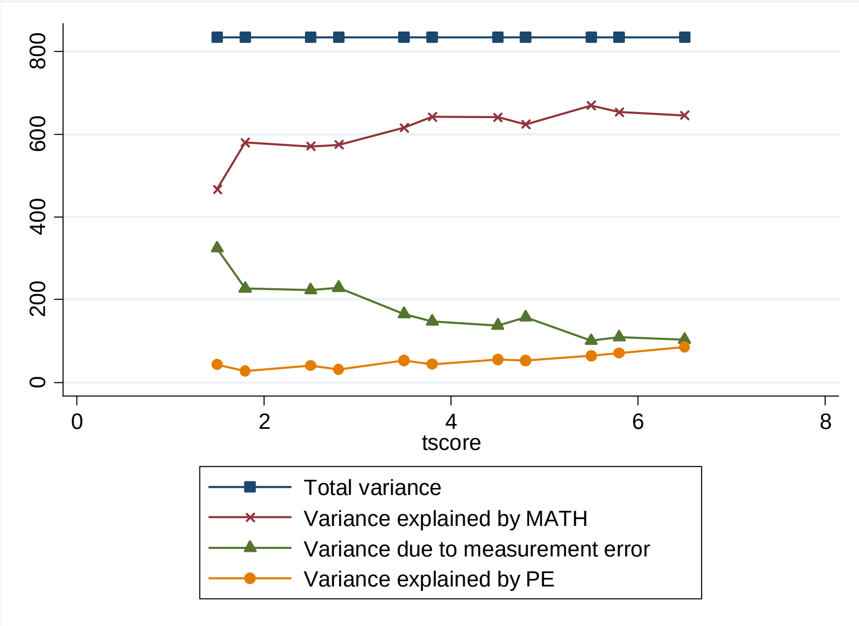


c. Variance decomposition

The approach of the rolling structural equation model further allows us based on the calculations shown in Appendix B to make the results more tangible by providing a decomposition of the overall variance in test scores. Figure 9 shows the contributions to the variance of the three components: mathematic ability, parental education background and the error term of this model. Figure 10 depicts the variance explained by parental education as well as the unexplained variance relative to what MATH can explain.

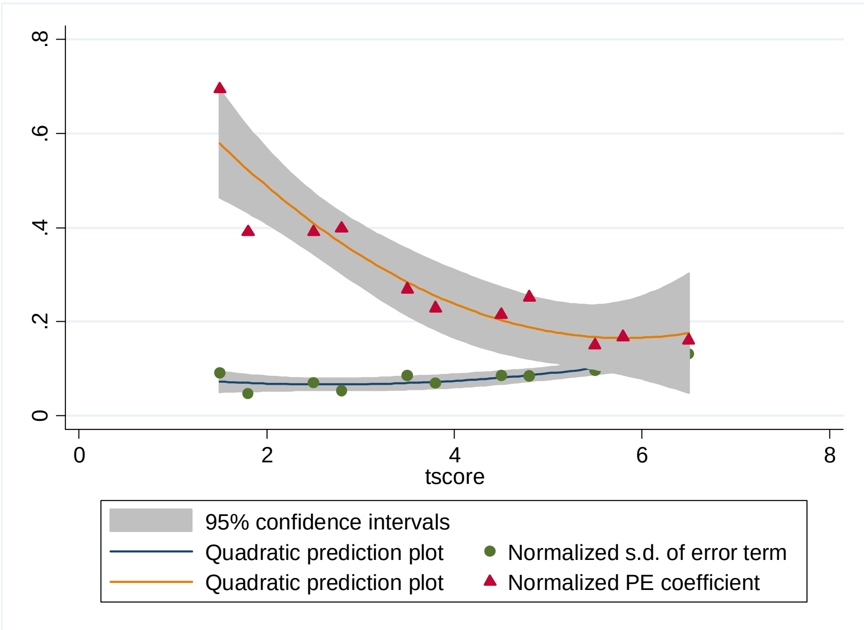
The overall variance is constant due to using percentile ranks. Naturally the variance explained by both explanatory variables increases, but distinctively. In the first half of primary school the variance explained by MATH increases relatively more,[12] in the second half the relative fraction explained by parental education background increases more strongly.

Figure 9 Decomposition of test score variance



[12] This development also coincides well with the findings from Cunha and Heckman (2008) as well as Heckman (2006) of widely constant cognitive ability after the age of eight.

Figure 10 Variance explained by PE component and unexplained variance relative to MATH component



d. About within year patterns and the special role of the final assessment test

Over the course of the first three years the coefficient of PE is always larger for the test score of the mid-year test than for that of the end-year test. This observation is the same for all model specifications. In grade four the coefficient is almost the same for both tests and from year five onwards the coefficient increases consistently from test to test across all model specifications again. This pattern is very similar to the development of the correlation based on the balanced sample in Figure 1.

The pattern in grade one to three is consistent with empirical evidence comparing the development of children from different backgrounds during school breaks and during school times. Downey, Von Hippel, and Broh (2004) find support for school having an equalizing effect. In our case this relation turns around grade four and five, which could be related to specific preparation or training of children with higher educated parents with regard to the final assessment test. The incentives to score high on this test are strong, since it is a main determinant of the teacher's binding recommendation for secondary schooling track. The different tracks in turn lead to specific qualifications for the labor market. Even if the teacher deviates from what the final assessment test result suggests, some secondary schools only accept students with a score above a certain threshold. During the first three years after the transition a quarter of children changes the track at least once. And among them only 40 per cent change to a higher schooling level (Inspectie van het Onderwijs, 2014). In this respect the first track placement is very important.

While the tests leading up to the final assessment test show an increasing trend in the relationship between the children's parental educational background it peaks in the mid-year test in grade six and is clearly lower for the final assessment test. This makes sense when considering that the final assessment test is not only based on the subject taught most recently. It focusses on a longer time span and is intended to measure the overall mathematic skills acquired during primary school. Furthermore it is the only test of the ones used that is graded externally.

VI. Discussion

One of the major questions based on the analysis of test scores is how meritocratic an education system is, meaning how much success is determined by ability compared to social class. Empirical studies (e.g. OECD, 2013) show big differences between countries with respect to the variance in performance that can be explained by differences in parental socio-economic status (PE). So in aiming for meritocratic fairness it is crucial for policy makers to understand the underlying mechanisms leading to the observed social dependency of outcome measures. But as we show the development of the correlation alone can be misleading.

Using a series of comparable mathematics test scores over the course of six years of Dutch primary education we study how the influence of PE, measured by the highest parental education level, develops and how it relates to the measurement error. Our three main results can be summarized as follows: (1) The relative relevance of PE in explaining the variance in test scores is small compared to the variance explained by mathematic ability as well as compared to the variance not explained by the model. Especially in the first three years the measurement error accounts for multiple times as much variance as PE. In the models assuming mathematic ability to be constant the normalized standard deviation of the error term is twice as large as the normalized coefficient for PE. But even though the share of variance explained by PE is small it slightly increases in relevance compared to mathematic ability, also after accounting for the measurement error. (2) For the first test in the middle of grade one we find one of the highest, in some models the highest, coefficient capturing the relationship between parental education background and test scores. From the second test onwards it starts on a very low level and increases on average every year until the last regular test in grade six. Assuming constant mathematic ability the influence of PE never exceeds the starting level in grade one. Under less restrictive assumptions it is only exceeded in the last two regular tests.[13] (3) Finally throughout all graphs we consistently observe a specific within year pattern and a change in that pattern midway through primary education. During the first three years the influence of PE always decreases from the mid-year test to the end-year test. In year four it is stable and from then on it increases within and also across the school years. But this pattern only holds for the regular tests. The result of the final assessment test is less dependent on PE then the regular test taken only shortly before. In the following we will discuss potential mechanisms and implications of all three results.

Already the correlation presented in the beginning shows that PE accounts for at most between five and ten percent of the variance in test scores. This is small compared to the error term and even smaller compared to the variance due to mathematic ability[14], as captured by what all twelve tests have in common. Allowing this measure of mathematic ability to be correlated with PE did not change our results. A restriction is that with our data set all the

[13] Though due to some selection with respect to parental non-response in our sample we might underestimate the influence of PE.

[14] The concept as we use it in this analysis can be seen as mathematic ability plus test taking ability, where the latter of the two is likely to be influenced also by non-cognitive skills (Borghans & Schils, 2012).

influence that PE has on mathematic ability before the children start school at the age of five cannot be measured. The ability that they bring to school might well be partly dependent on parental education. This is in line with the literature on the development of cognitive skills (Cunha & Heckman, 2008; Heckman, 2006). Our analysis of the development over time suggests that on average school seems to reduce social-dependency of test outcomes during the first few years. Later, in the last one and a half to two years before the final assessment test, it increases again.

Using a series of tests we also focus on the development of the error term. The huge observed changes in the error component of the variance can have different reasons. It might be due to an increasingly stable performance of children or only an increasing experience in test taking. Furthermore the construction of the tests is likely to play a role. As one can see from the reported maximum raw score in the table in Appendix A the number of items increases almost from test to test. In order to make the tests as comparable as possible and to reduce distortions the variation in the number of items should be kept to a minimum. Of course it is not reasonable to start with very long tests in the first grades, but we doubt that it provides a benefit to add four or five questions every test during the last three years.

Additional insights come from a closer look at the within year pattern. For the first three years we find support for the claim of an equalizing influence of school with respect to PE during the school year. The change in this pattern around the fourth year could be explained by distinct reactions to approaching the high stakes test at the end of primary school along the lines of parental educational background. It is consistent with additional support or pressure from parents with higher education levels or alternatively with increased effort by the children themselves. Further the regular tests allow for teaching to the test. They did not change for several years and are therefore known by many teachers. Differential effects of this teaching to the test, based on parental education, could therefore also explain the observations. Alternatively the changes could also be related to the increasing test length. As Borghans and

Schils (2012) show, the longer a test is, the more important are non-cognitive skills, which might also be correlated with parental educational background. These are of course only three possible explanations. More research regarding the underlying mechanisms is needed in order to support or reject these.

The advantage of children with a higher parental educational background does not seem to transfer completely to the final assessment test. Compared to the other tests the mathematic component of the final assessment test seems to be less dependent on parental education. This raises concerns about the recent decision to change the timing and take the final assessment test later. Since this year the recommendations for secondary schooling tracks have to be based more on the earlier regular tests. Our results suggest that this change might increase the dependency of the teacher's track recommendation on social background characteristic.

VII. Conclusion

In this article we applied a simple structural equation model to a series of comparable mathematics test scores over the course of Dutch primary education, distinguishing the influence of mathematic ability and parental education level. Our results suggests that the steady increase in the correlation of test scores with parental education can be explained by an interplay of a clear decrease in the errors' standard deviation in the first three to four years and a slight increase of the coefficient on parental education in later years. Our results are consistent across regression approaches as well as different specifications of the structural equation model and robust against varying definitions of test scores.

Specific within year patterns suggest that depending on their education level parents and / or their children themselves react differently to similar results in mid-year tests and in preparation towards the final assessment test in grade six. During the first few years the results are consistent with an equalizing effect of school, while towards the final assessment test parental education becomes increasingly important in determining success in mathematics tests. This final test itself, which determines the binding recommendation for the secondary schooling track, is less dependent on the children's parental educational background than the final regular test.

Overall parental educational background plays a relatively small role in determining the test score outcomes during the age range of six to twelve years compared to mathematic ability and the error term. The variance explained by parental education never exceeds the fraction of unexplained variance in our model. Though to interpret our results in terms of intergenerational mobility we have to be careful since additional measures, such as income, might play a crucial role.

What we can say is that it is worth to look behind pure correlations between parental background and schooling outcomes. Especially measurement error of tests might distort the picture provided by the development of those correlations over time. Measures to make tests more comparable over years, like keeping the number of questions similar are also helpful for comparisons across years. Finally, externally graded final assessment tests seem to be less vulnerable to training related to parental background.

References

Bond, T. N., & Lang, K. (2013). The evolution of the Black-White test score gap in grades K–3: The fragility of results. *Review of Economics and Statistics, 95*(5), 1468-1479.

Borghans, L., & Schils, T. (2012). *The leaning tower of Pisa: decomposing achievement test scores into cognitive and noncognitive components.* Paper presented at the 2012 NBER Summer Institute in Education Economics, Cambridge.

Buitelaar, R., Ros, B., Vink, A., & van der Kroft, L. (2013). Eindtoets PO: Het verdriet van Cito. Retrieved 11.12.2014, from http://www.didactiefonline.nl/actueel-dossier-cito/47-uncategorised/11688-eindtoets-po-het-verdriet-van-cito

Central Bureau voor de Statistiek. (2011). Inpassing van het Nederlandse onderwijs in ISCED 2011. *Internatnional Standard Classification of Education.* Retrieved 17.7.2014, 2014, from http://www.cbs.nl/nl-NL/menu/methoden/classificaties/overzicht/isced/default.htm

Cito B.V. (2014). De geschiedenis van Cito. Retrieved 17.7.2014, from http://www.cito.com/about_cito

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources, 43*(4), 738-782.

Downey, D. B., Von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review, 69*(5), 613-635.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics, 86*(2), 447-464.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *science, 312*(5782), 1900-1902.

Inspectie van het Onderwijs, Nederland. (2014). De staat van het onderwijs, onderwijsverslag 2012/2013. Utrecht, Netherlands.

Lang, K. (2010). Measurement matters: Perspectives on education policy from an economist and school board member. *The Journal of Economic Perspectives, 24*(3), 167-181.

Luyten, H., Cremers-Van Wees, L., & Bosker, R. (2003). The Matthew effect in Dutch primary education: Differences between schools, cohorts and pupils. *Research Papers in Education, 18*(2), 167-195.

Polidano, C., Hanel, B., & Buddelmeyer, H. (2013). Explaining the socio-economic status school completion gap. *Education Economics, 21*(3), 230-247.

Sammons, P. (1995). Gender, ethnic and socio-economic differences in attainment and progress: a longitudinal analysis of student achievement over 9 years. *British Educational Research Journal*, *21*(4), 465-485.

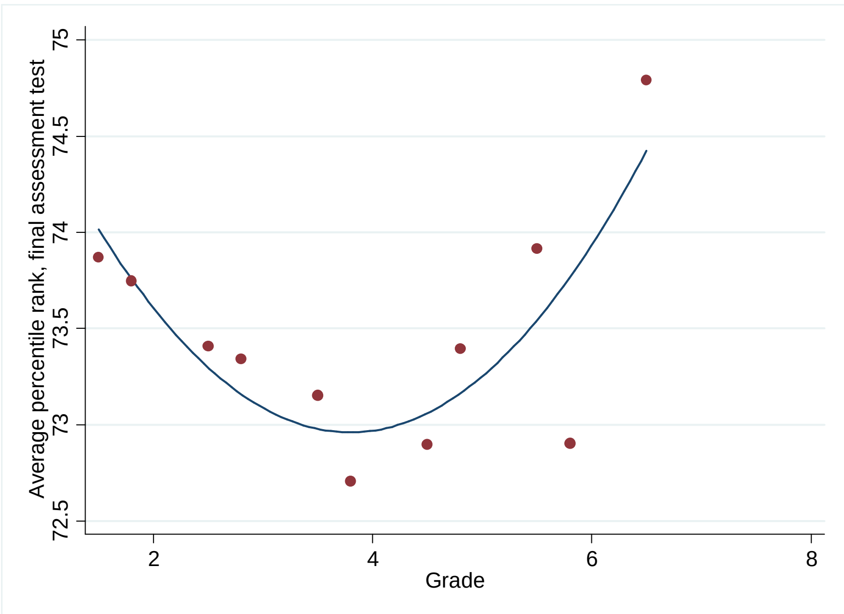
Torche, F. (2015). Analyses of Intergenerational Mobility An Interdisciplinary Review. *The ANNALS of the American Academy of Political and Social Science*, *657*(1), 37-62.

Appendix A

Table 4 Number of observations and descriptive statistics for tests in unbalanced sample

Test	N	mean	sd	min	max
Grade 1, mid-year	5,216	36.79	4.868	10	42
Grade 1, end-year	5,214	45.06	6.309	11	53
Grade 2, mid-year	4,292	47.18	7.842	0	58
Grade 2, end-year	4,164	46.55	8.322	0	60
Grade 3, mid-year	4,259	53.00	9.606	6	79
Grade 3, end-year	3,899	58.39	11.21	11	78
Grade 4, mid-year	3,868	74.43	14.36	14	100
Grade 4, end-year	3,614	77.37	14.94	18	104
Grade 5, mid-year	3,570	81.51	15.78	12	109
Grade 5, end-year	2,367	76.85	15.29	11	105
Grade 6, mid-year	3,126	79.85	17.07	0	107
Grade 6, final math assessment test	6,350	73.92	17.03	15	100

Figure 11 Changes of student composition in unbalanced sample (balanced sample excluded), according to mathematic performance in final assessment test



Appendix B

The structural equation model (SEM) introduced to estimate the coefficients for all twelve equations simultaneously uses in addition to the within test score variance also the covariances between the different dependent variables. In our case this allows the estimation of the latent variable MATH based on all available tests, and improves the estimates of the magnitude of the error term. For the simple case of three tests with some assumptions the coefficients can be easily calculated by hand, only using the observed values from the variance-covariance matrix.

For convenience we simplify Equation 2 as follows: The subscript i for the individual is left out and the variable names were simplified to m for mathematic ability and s for highest parental education level.

$$Y_t = \alpha_t + \beta_t m + \gamma_t s + \varepsilon_t \qquad (3)$$

In addition we assume mathematic ability and parental education background not to be independent and we also do not allow the error terms of different tests to be correlated. Irrespective of whether these assumptions are realistic or not, here they serve the purpose of having a simple model for demonstrating the underlying calculations of the structural equation model. In the model used in the analysis we relaxed those assumptions.

Table 5 shows how the variance-covariance matrix of the test scores and the variable for parental education level relate to the coefficients, error terms and variances of m and s under these assumptions.

Table 5 Variance-covariance matrix of test scores and s

	Y1	Y2	Y3	S

Y1	$\beta_1^2 var(m) + \gamma_1^2 var(s) + var(\varepsilon_1)$			
Y2	$\beta_1\beta_2 var(m) + \gamma_1\gamma_2 var(s)$	$\beta_2^2 var(m) + \gamma_2^2 var(s) + var(\varepsilon_2)$		
Y3	$\beta_1\beta_3 var(m) + \gamma_1\gamma_3 var(s)$	$\beta_2\beta_3 var(m) + \gamma_2\gamma_3 var(s)$	$\beta_3^2 var(m) + \gamma_3^2 var(s) + var(\varepsilon_3)$	
S	$\gamma_1 var(s)$	$\gamma_2 var(s)$	$\gamma_3 var(s)$	$var(s)$

Setting the variance of m to unity and observing the variance of s directly nine variables remain to be calculated based on nine equations.

$$var(m) = 1 \tag{4}$$

$$var(s) = var(s) \tag{5}$$

$$\gamma_1 = \frac{cov(Y_1, s)}{var(s)}, \gamma_1 = \frac{cov(Y_1, s)}{var(s)}, \gamma_1 = \frac{cov(Y_1, s)}{var(s)} \quad (6.1 - 6.3)$$

$$\beta_1\beta_2 = cov(Y_1 Y_2) - \frac{cov(Y_1, s)cov(Y_2, s)}{var(s)} \tag{7}$$

Under the assumption of $var(m) = 1$ Equation 7 denotes the part of $cov(Y_1, Y_2)$ that is based on the variance in mathematic ability. This holds for $\beta_1\beta_2$ and $\beta_1\beta_2$ respectively. Those three measures can then be used to calculate the mathematic ability component of the test variance $var(Y_1)$ (see Equation 8). The square root of this measure provides is the coefficient for mathematic ability.

$$\beta_1^2 = \frac{\beta_1\beta_2 * \beta_1\beta_3}{\beta_2\beta_3} \tag{8}$$

The parental education component of the test variance can be calculated directly from the covariance of the test with s and the variance of s, which both can be found in the variance covariance matrix. The remaining error component is what is left of the overall test variance after accounting for the components going back to mathematic ability and parental education level.

$$\gamma_1^2 var(s) = \frac{cov(Y_1, s)^2}{var(s)} \tag{9}$$

$$var(\varepsilon_1) = var(Y_1) - \gamma_1^2 var(s) - \beta_1^2 \tag{10}$$

Applying those formulas to the data yields exactly the same results as running a structural equation model under the assumptions made.