



ELSEVIER

Contents lists available at ScienceDirect

Economics of Education Review

journal homepage: www.elsevier.com/locate/econedurev



Expanding schooling opportunities for 4-year-olds

Edwin Leuven^{a,*}, Mikael Lindahl^b, Hessel Oosterbeek^c, Dinand Webbink^d

^a ENSAE-CREST, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France

^b Uppsala University, Sweden

^c University of Amsterdam & Tinbergen Institute, Amsterdam, The Netherlands

^d CPB Netherlands Bureau for Economic Policy Analysis, The Netherlands

ARTICLE INFO

Article history:

Received 12 February 2008

Accepted 18 June 2009

JEL classification:

I21

I28

J24

Keywords:

Early childhood education

Achievement

Evaluation

ABSTRACT

We use a novel quasi-experimental strategy to estimate the effect of expanding early schooling enrollment possibilities on early achievement. It exploits two features of the school system in The Netherlands. The first is rolling admissions; children are allowed to start school immediately after their 4th birthday instead of at the beginning of the school year. The second is that children having their birthday before, during and after the summer holiday are placed in the same class. These features generate sufficient exogenous variation in children's enrollment opportunities to identify its effects on test scores. Making available one additional month of time in school increases language scores of disadvantaged pupils by 6 percent of a standard deviation and their math scores by 5 percent of a standard deviation. For non-disadvantaged pupils we find no effect.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

Impact estimates of early childhood interventions are of particular interest as it has been argued that early interventions have dynamic complementarities in human capital production; early learning makes subsequent learning easier (Heckman, 1999). Most evidence on the effectiveness of early interventions comes from targeted early childhood programs in the US.¹ This paper investigates whether these results carry over to the more general setting of increasing the amount of regular education for young children.

Many developed countries provide free high-quality education for 4-year-olds.² This is also the case for The Netherlands where children can start school when they are 4 years old, and where compulsory schooling begins when children turn five. The program consists of 24 h per week during 41 weeks per year. The staff in front of the classroom are certified primary school teachers and are paid according to standard primary school wage scales. The curriculum consists of structured learning activities, and typically children will have started to read and write by the age of six. Over the period 1995–2001 the annual expenditure per pupil was about 3500 Euro. This is in the same order of magnitude as state spending on public prekindergarten programs in the United States (Barnett, Hustedt, Robin, & Schulman, 2005).

We estimate the impact of expanding enrollment opportunities (i.e. the intention to treat) around age 4 on subsequent achievement at age 6. An important contribu-

* Corresponding author.

E-mail address: edwin.leuven@ensae.fr (E. Leuven).

¹ Currie (2001) summarizes the beneficial short-term and long-term effects of small-scale intensive interventions such as the Perry Preschool project, the Chicago Child-Parents Centers and the Carolina Abecedarian Project. Large-scale targeted programs with lower per pupil expenditures also appear to have positive short-term and long-term effects. Garces, Thomas, and Currie (2002) report beneficial effects on various later outcomes from participation in Head Start.

² Some countries, like France for example, provide state funded preschooling for 3-year-olds.

tion of our study is that we are able to separate this effect from (linear and quadratic) age effects. For identification we exploit institutional features of the Dutch early schooling system that generate arguably exogenous variation in the time young children can spend in school. The first feature is that children are allowed to start school immediately after their fourth birthday as opposed to the beginning of the school year as is the case in most other countries. The second feature is that children having their birthday before, during and after the summer holiday are placed in the same class. These features generate a difference of up to 11 weeks in the time children can spend in school which is not collinear with variation in age. Eleven weeks constitutes about 15 percent of the amount of schooling they may have had at the moment of the tests. The key identifying assumption is that birth patterns are independent of other factors that affect test scores. We present evidence that birth patterns in our sample (and subsamples) are unrelated to observed background characteristics, suggesting that the identifying assumption is satisfied.

Our main finding is that making available one additional month of time in school increases language scores of disadvantaged pupils by 6 percent of a standard deviation and their math scores by 5 percent of a standard deviation. For non-disadvantaged pupils we find no effect. For comparison: the gap between the average scores of disadvantaged and non-disadvantaged pupils amounts to 0.7 standard deviation for language and 0.6 standard deviation for arithmetic.

In the next section we briefly review the related literature and discuss how these analyses relate to our analysis. Section 3 documents the details of the Dutch regulations regarding school enrollment age and the scheduling of holiday periods and describes how we use this in our estimation framework. Section 4 describes the data and presents descriptive statistics. Section 5 presents the estimated effects of increasing early school availability on achievement. Section 6 concludes.

2. Related literature

Our analysis is most closely related to studies that examine the impact of age at school entry, age and schooling on (early) achievement. To put our contribution into perspective it is useful to assume that a person's amount of human capital (Y) is an additive function of two components. The first component is solely determined by age. The second component relates to human capital accumulation while in school. A general specification capturing this is the following:

$$Y(\text{age}, s, \text{age}_e) = Y^0(\text{age}) + Y^1(s, \text{age}_e)$$

where Y^0 captures how human capital depends on age (age) when a person is not in school. For young people it seems reasonable to assume this function to be increasing and possibly concave. The amount of human capital that has been accumulated in school, Y^1 , obviously depends on the amount of time spent in school, s . In addition it may depend on the age at which pupils enroll in school, measured by their age at school entry, age_e . Various papers argue that

the effectiveness of time in school depends on age at school entry, for instance because a pupil's span of attention varies with age.

In addition to the effect of biological age on human capital, a pupil's age relative to that of its classmates may also play a role. Relative age can only matter when a child is in school and thus necessarily enters Y^1 . In school systems with one annual entry date, the relative age effect operates through age_e , while in case of rolling admissions (such as in The Netherlands), this effect causes heterogeneity in the impact of s .

When estimating the impact of the various components on outcomes, one important complication arises from the fact that age, age at entry and length of schooling are linearly dependent³:

$$\text{age} = \text{age}_e + s$$

The second complication is that age at entry and length of schooling are choice variables, and this endogeneity needs to be accounted for in the analysis. The key challenge is thus to identify sources of exogenous variation that affect one of the determinants of Y without affecting at the same time one of the other determinants. How do previous papers deal with this?

2.1. Age

Bedard and Dhuey (2006), Strøm (2004), Puhani and Weber (2007) have attempted to estimate the effect of age on test scores. These papers use data from school systems characterized by one annual school entry date.⁴ Moreover data are from pupils placed in the same grade level and tested at the same date. By implication, pupils will all have identical amounts of time in school (denoted by s^*). Consequently, these studies estimate functions of the following form:

$$Y(\text{age}, \text{age}_e | s = s^*) = Y^0(\text{age}) + Y^1(\text{age}_e | s = s^*)$$

and find that older children perform better. The problem faced by these studies is that with the data at hand, age and age_e are perfectly correlated. Hence, it is impossible to disentangle the effects from these two variables. Strøm interprets his results in terms of differences in school entry age, while Bedard and Dhuey frame their interpretation in terms of relative age.⁵ Note that testing pupils with different birth dates at different dates in order to keep age at the test date constant, does not help identification because

³ For another discussion of the consequences of this identity for identification of schooling effects, see Hansen, Heckman, and Mullen (2004).

⁴ For a survey of older papers, see Stipek (2002).

⁵ Bedard and Dhuey (2006) never make explicit what they mean by relative age effects and how these can be distinguished from biological age effects. Cascio and Schanzenbach (2007) separate biological from relative age effects using data from the random assignment to classes in experiment STAR. Their findings point to heterogeneous impacts across social groups. Disadvantaged children have worse outcomes when they are the youngest in class, while such a relative age effect is not present for children from more advantaged families. In contrast, disadvantaged children experience a negative impact of their biological age on outcomes, while the opposite holds for non-disadvantaged children.

this results in a perfect (negative) correlation between age at school entry (age_e) and amount of time in school (s).⁶

The estimates in [Bedard and Dhuey \(2006\)](#) are for 4th and 8th graders in 19 OECD countries. The mean estimate across countries indicates about 0.2 unit of a standard deviation (S.D.) higher test score for the oldest in a class. These estimates capture the combined effects of maturity, relative age and starting school at an older age. [Mayer and Knutson \(1999\)](#), [Datar \(2006\)](#) and [Elder and Lubotsky \(2009\)](#) use data from the US, which also has one school entry date per year, and thus face the same problem as the previous papers. However, since school cutoffs differ across U.S. states, the variation in age across pupils is due to difference in birth dates and school-cohort cutoffs. Hence, they break the perfect correlation between entry age and age at test date and can therefore include controls for age.

Mayer and Knutson include a linear age term and quarter of birth dummies. The quarter of birth dummies are interpreted as the effect of age at school entry. Starting school a year younger (but having the same amount of schooling and age at the test date) results in an average test score increase of 0.2–0.4 S.D. Datar uses exogenous variation in birth dates and kindergarten entrance age policies to construct instrumental variables for kindergarten entrance age. Using panel data she is able to disentangle the pure age effect from the entrance age effect, and finds that children that enter kindergarten at an older age have higher test scores at kindergarten entry (a pure age effect) and also have steeper test score gradients during the first 2 years in school.⁷ Elder and Lubotsky instrument for actual kindergarten entrance age with a constructed measure based on (interactions between) birth date and state cutoff laws. They estimate entrance-age effects for children affected by the state entry cutoffs, also controlling for quarter-of-birth dummies. They find that children who enter kindergarten at a later age have higher test scores at entry, but these effects decline sharply during the first years of enrollment.

A similar strategy is used by [Bedard and Dhuey \(2006\)](#) when they perform pooled cross-country estimations utilizing that school-entry age regulations varies across countries. Controlling for month-of-birth, they get estimates of similar sizes as for within countries (about 0.2 S.D.). A potential complication with this approach is that the quarter of birth dummies may pick up non-linear age effects.

2.2. Time in school

Another strand of papers has attempted to estimate the effect of time in school on test scores.⁸ An early example of

such a study is [Cahan and Cohen \(1989\)](#). These authors collected test score data for over 12,000 pupils in grades 4–6 in Israel, where children born in the same calendar year also start school at the same day. Since they have data from adjacent grades and pupils are tested at the same time using the same test, they can estimate effects of both age and time in school. Pupils placed in the same grade level have the same amount of schooling but differ in age, whereas pupils born in adjacent months but placed in different grade levels have (almost) the same age but differ in their amounts of schooling (and in their school starting age). The findings indicate that the effect of an additional year of schooling on test scores is about twice the effect of being 1 year older (about 0.30 S.D. versus 0.15 S.D. on average over the 12 tests). Note that in this design, the effect of age captures both the effect of chronological age and the effect of age at school entry. Cahan and Cohen discuss their age effects in terms of the effects of chronological age, not in terms of effects of age at school entry. Since the effect of schooling captures both effects from time in school, school entry age, as well as relative age effects, the 0.3 S.D.-estimate is likely an overestimate of spending an additional year in school as a consequence of starting school 1 year earlier.

Recently [Gormley and Gayer \(2005\)](#), and [Gormley, Gayer, Phillips, and Dawson \(2005\)](#) used the same design as Cahan and Cohen and estimate the impact of Oklahoma's pre-K program for 4-year-olds in Tulsa on cognitive/knowledge test scores, motor skills and language scores. Attendance increases test scores by approximately 0.4 S.D. Because there is no full compliance (as in Cahan and Cohen's analysis for Israel) the analysis recovers impact estimates for the treated.

[Berlinski, Galiani, and Gertler \(2009\)](#) estimate the impact of pre-primary education on 3rd grade test scores in Argentina. They exploit variation in treatment intensity across regions and cohorts stemming from a large-scale program aimed at increasing pre-school attendance for children, and find that 1 year of pre-primary school increases test scores by 0.23 S.D.⁹

[Hansen et al. \(2004\)](#) and [Cascio and Lewis \(2006\)](#) estimate the impact of time in school on scores at military qualification tests (AFQT) in the US. The test is administered at the same date for one birth cohort, even though students attend different grades. Cascio and Lewis extend the approach in Cahan and Cohen by also using variation in the states' school entry cutoffs to estimate the effect of time in school. Hence, they estimate the schooling effect that is due to variation in school entry age across states. This makes it possible to allow for independent effects of age by including quarter-of-birth dummies. They find that an additional year of formal education raises scores for minority groups with about 0.3 S.D.

⁶ [Angrist and Krueger \(1992\)](#), [Dobkin and Ferreira \(2006\)](#), [Fredriksson and Öckert \(2006\)](#) look at long-term effects of school-start age. If cohort-effects in completed education and earnings are negligible and relative age effects are small, these estimates can more likely be interpreted as true school-age entry effects on long-term outcomes.

⁷ These estimates are conditional on schooling. Datar assumes that age effects are linear and that there are no interaction effects between age and schooling.

⁸ For surveys of the development psychology literature on the estimation of age and schooling effects, see [Ceci \(1991\)](#) and [Stipek \(2002\)](#).

⁹ [Berlinski, Galiani, and Manacorda \(2008\)](#) finds long-term effects of pre-school attendance (retrospectively collected) for Uruguay, exploiting variation in pre-school attendance between siblings. [Dhuey \(2007\)](#) uses U.S. state-variation in adoption of a policy to publicly subsidizing kindergarten to estimate the effects of enrollment in kindergarten on later outcomes. She finds positive effects on academic and labor market outcomes, especially for black individuals and those from low SES backgrounds.

Taken together, the results are remarkably similar across samples in different countries. An additional year of education raises test scores with about 0.2–0.4 S.D.

The next section explains how the timing of school holidays in Dutch primary schools causes variation in the amount of time in school independent of variation in age and allows us to estimate the effect of making (pre)school available at age 4.

3. Background and identification strategy

Dutch primary schools consist of 8 grades covering the age groups of 4–12-year-old children. While in most countries children typically enter primary school at the same date, in the Netherlands the rule is that children are *allowed* to enroll in primary school the first school day after their 4th birthday, while enrollment is *compulsory* from the first school day of the month after the child reached the age of 5 onwards. About 98 percent of the children start school before their 5th birthday. When exactly between their 4th and 5th birthday a child actually enrolls is up to the parents. The total number of schooldays a child has attended at a given date is therefore to some degree a choice variable. The rule that enrollment is permitted at age 4 and compulsory at age 5 determines the maximum and minimum amounts of time a child can spend in primary school.

The second important feature where our identification builds on, is that a school year cohort in the Netherlands consists of everyone born between October 1 of a given year and September 30 of the next year. At the same time, a school year runs from summer holiday to summer holiday. The formal rule is that a child who enrolls in school on the first school day after its 4th birthday spends the period until October 1 in grade 1. Then it spends the period from October 1 until the (next) summer holiday again in grade 1. After the summer holiday the child continues in grade 2.¹⁰ Together these features produce variation in the amount of time a child can spend (maximum length of schooling) in school which is not collinear with age (or its square).

Fig. 1 shows the relationship between a child's birthday and its maximum length of enrollment in school. The flat segments correspond to holiday periods and the downward sloping segments represent school periods.¹¹ For children having their fourth birthday on the same downward sloping segment, maximum length of schooling in school varies one-to-one with age; being one day older adds one day to the maximum length of schooling. Differences in test scores between two otherwise identical pupils

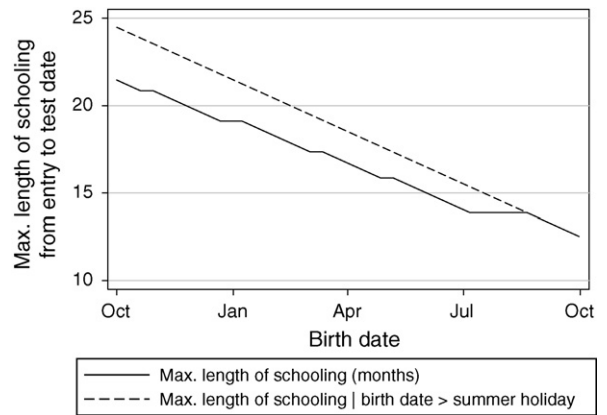


Fig. 1. Relation between birth date and max. length of schooling for a given cohort.

within a segment are attributable to their difference in age as well as to their difference in maximum length of schooling. Maximum length of schooling does not vary across children having their fourth birthday on the same flat segment. Consequently, differences in test scores between two otherwise identical pupils from these segments are solely attributable to differences in their age.

Children in the after-summer-group can have, correcting for their age, at most 11 weeks more time in school than children born in other periods. This is most easily seen when we compare the extrapolation of the line segment for the after-summer-group (the dashed line in Figure 1 which ignores holidays) with the solid line segments for the before-summer-group. The vertical distance between these two lines is the difference in maximum length of schooling corrected for differences in age.

Maximum length of schooling at the test date is by definition equal to the difference between age at the test date, age at school entry and the length of holiday between entry and test: $s^{max} = age - 4 - h$. Variation in the amount of holidays between age 4 and the moment of the test (h), breaks the perfect correlation between maximum length of schooling and chronological age.

Since the identification in our analysis comes from a non-linearity there are limits to the extent to which we can control for age. Given that we only consider children within the same cohort this might not be an important problem. We show below that moving from a linear to a quadratic age specification does not affect our results, higher order terms however capture the non-linearity we are exploiting and our identification breaks down.

We will estimate the following model

$$Y = Y(s^{max}, age)$$

From the figure it is clear that we use cross-sectional variation within the entire school year cohort. Since the variation we exploit is conditional on age, we need to control sufficiently flexible for the effect of differences in age on test scores. Once we do this any remaining differences in test scores between children is attributable to differences in maximum length of schooling. In the analysis we will

¹⁰ Relative age is a fuzzy concept in this setting when measured halfway grade 2. During their time in first grade, pupils started as the youngest and belonged to the oldest towards the end. Hence, Dutch pupils experience a changing composition of peers by age in their first period in primary school. This is different from school systems with only one annual intake where the oldest (youngest) pupil is always the oldest (youngest). This implies that our estimates of the effect of age are probably best understood as the effect of biological age.

¹¹ In practice the exact timing of the summer holidays varies somewhat from year to year and between three different regions (North, Middle and South). In all cases, however, the summer holiday ends well before October 1, and hence there are always children at a grade level who have their 4th birthday between the end of the summer holiday and October 1.

estimate the following specification

$$y_i = \alpha + \beta_{RF} \cdot s_i^{max} + \delta \cdot age_i + \lambda \cdot age_i^2 + x_i' \gamma + \varepsilon_i \quad (1)$$

where y_i is the 2nd grade test score, s_i^{max} is the maximum length of schooling for pupil i , x_i are year and region indicators and their interactions, and individual characteristics.¹² The identifying assumption is

$$E[s_i^{max} \cdot \varepsilon_i | age_i, age_i^2, x_i] = 0$$

For the interpretation of our estimates it is important to understand what the learning environment would have been in the absence of enrollment in school. The exogenous variation in maximum length of schooling is caused by the incidence of up to 11 weeks of school holidays. Six of these 11 weeks are in the summer, the other 5 are during the rest of the year. Most Dutch families spend 3–4 weeks of the summer holiday away from their home (often abroad). The other holiday weeks are typically spend at home, with one of the parents (most often the mother) looking after the child(ren). This is possible due to the relatively low labor force participation of women in the Netherlands. This is especially true for women with low levels of education and with a minority background.¹³ In fewer cases, grandparents, other family or professional child care will play a role. This implies that around a third of the holiday weeks is spent away from home and the remainder in many cases at home with one of the parents or another family-member. This will often also be the situation for children whose school attendance is postponed.

Some have argued that the timing of births during the year may depend on unobserved characteristics of the parents which have an effect on children's achievement. This point was raised by Bound, Jaeger, and Baker (1995) in their comment on the use of quarter of birth as an instrumental variable for years of schooling by Angrist and Krueger (1991). For the countries in their data set Bedard and Dhuey (2006) find no evidence that more educated mothers target birth so that their children are the oldest in the class. In Section 5 we will present similar evidence for our data set which indicates that there is no systematic relation between parents' levels of education and children's maximum length of schooling.

One might be interested in the results from a 2SLS procedure where actual length of schooling is instrumented with s_i^{max} . Unfortunately our data set contains only limited and unreliable information on actual enrollment. For this reason we will report estimates of the reduced form model,

¹² Substituting $s_i^{max} = age - 4 - h$ into Eq. (1) gives

$$y_i = (\alpha - 4\beta_{RF}) - \beta_{RF} \cdot h_i + (\delta + \beta_{RF}) \cdot age_i + \lambda \cdot age_i^2 + x_i' \gamma + \varepsilon_i.$$

This expression makes clear that the identification of the impact of s_i^{max} comes from variation in holidays (only the sign of the coefficient is reversed). Note that in this expression the coefficient of age captures both the pure linear age effect and the effect of potential time in school.

¹³ Among women with (without) a partner and their youngest child between 0 and 5 years old, labor force participation equalled 57 (37) percent in 2001, with those participating working on average 20.9 (25.3) h per week. Participation rates and working hours increase with level of education and are lower for women with a minority background than for Dutch women (Merens and Hermans, 2009).

$\hat{\beta}_{RF}$, which as we argued are interesting from a policy point of view in their own right.

We can, however, infer something about β_{IV} even without information on actual enrollment since the IV estimate is given by the following expression

$$\hat{\beta}_{IV} = \frac{\hat{\beta}_{RF}}{\hat{\pi}_{FS}} \quad (2)$$

where $\hat{\pi}_{FS}$ is the (unknown) first-stage regression coefficient of the instrument on actual length of schooling. The question therefore is whether it is possible to bound $\hat{\pi}_{FS}$? Suppose we increase enrollment opportunities by 1 month. Those who enroll immediately when they turn 4 before the expansion takes place – the constrained group c – can increase their enrollment by at most 1 month. For this group we can thus infer that $0 \leq \hat{\pi}_{FS,c} \leq 1$. Those who delay enrollment before the expansion – the unconstrained group u – can increase their enrollment by 1 month plus the delay. While possible in principle, this seems an unlikely response because it requires that people respond to the loosening of a constraint that was not binding. A more likely scenario is that those who enter at, say, age 4 years and 3 months when they are allowed to start at age 4 years and 0 months, will also enter at age 4 years and 3 months when they are allowed to start at the age of 3 years and 11 months. For this group we then would have $\hat{\pi}_{FS,u} = 0$.

Under these behavioral assumptions the overall estimate of $\hat{\pi}_{FS}$ is thus a weighted average of 0 and a value between 0 and 1, so that we have that $0 \leq \hat{\pi}_{FS} \leq 1$. It then follows from (4) that

$$\hat{\beta}_{IV} \geq \hat{\beta}_{RF}$$

and our reduced form estimate is a lower bound on the IV effect of actual schooling.

4. Data and descriptive statistics

4.1. Data

This paper uses data from five waves of the so-called PRIMA survey.¹⁴ This bi-annual survey contains information on Dutch pupils who were enrolled in grades 2, 4, 6 and 8 in the school years 1994/1995, 1996/1997, 1998/1999, 2000/2001 and 2002/2003. Several survey instruments have been used for collection of the data: administrative sources, tests, and questionnaires for teachers, parents and school headmasters. Each wave contains information of about 600 primary schools and around 55,000 pupils which is approximately 10 percent of the relevant age population. The survey design is such that it samples pupils from grades and not from cohorts. This is unfortunate because grade repeating is a fairly common phenomenon in Dutch primary schools and thus introduces substantial selection issues. Only for advancement from grade 1 to grade 2, grade repeating is not an issue.¹⁵ We therefore restrict the analy-

¹⁴ PRIMA is an acronym of "Primair Onderwijs" which is Dutch for primary education. The Dutch word "prima" also translates as excellent.

¹⁵ Less than 3 percent of the second graders in the survey are older than second graders should be, indicating that they repeated a grade. The

Table 1
 Descriptive statistics.

	Non-disadv. (1)	Disadvantaged		
		Dutch (2)	Minority (3)	All (4)
Age (months)	70.33 (3.40)	70.55 (3.40)	70.63 (3.35)	70.59 (3.38)
Maximum length of schooling (months)	16.67 (2.58)	16.82 (2.59)	16.89 (2.56)	16.86 (2.57)
Education mother				
Missing	0.09	0.06	0.09	0.07
Primary	0.01	0.12	0.53	0.34
Lower secondary	0.16	0.73	0.23	0.47
Upper secondary	0.50	0.08	0.12	0.10
Higher	0.23	0.01	0.03	0.02
Education father				
Missing	0.11	0.15	0.18	0.16
Primary	0.01	0.10	0.40	0.26
Lower secondary	0.19	0.70	0.27	0.47
Upper secondary	0.41	0.04	0.11	0.07
Higher	0.28	<.01	0.05	0.03
Not disadvantaged	1	0	0	0
Disadv. Dutch	0	1	0	0.47
Disadv. minority	0	0	1	0.53
Girl	0.49	0.50	0.50	0.50
Boy	0.51	0.50	0.50	0.50
Language	0.32 (0.95)	-0.04 (0.89)	-0.69 (0.83)	-0.39 (0.92)
Arithmetic	0.28 (0.98)	-0.14 (0.92)	-0.52 (0.86)	-0.34 (0.91)
Number of obs	28,942	11,149	12,744	23,893

Note: Mean values with standard deviations in parentheses.

sis to 2nd graders who are for the first time in grade 2 and, as a consequence, estimate the short term effect.

The outcome measures we use in the analysis are scores on two cognitive tests; one related to counting, classification and order and one related to the understanding of words, sentences and placement of words in sentences.¹⁶ These tests were developed for the Dutch government by testing agency “Centraal Instituut voor Testontwikkeling” (Central Institute for Test Development) in order to measure pupils’ readiness for arithmetic and reading, respectively. We will therefore refer to these tests as “arithmetic” and “language”. The tests are administered around February, which is more or less halfway the school year. The raw scores on these measures are based on tests which are especially designed for this data collection. From year to year the tests for the same grade levels are identical. The purpose of this is to compare achievement levels over time. Because the scales of the raw scores have no clear meaning, we transformed these scores for each test into wave specific standardized scores, having mean zero and standard deviation one.

In the estimations we control for gender, education levels of father and mother, year and region. We report results separately for disadvantaged pupils and for non-

disadvantaged pupils. The reason for this is that the counterfactual of not going to school is likely to be quite different across these groups because the situation at home is very different. Children in non-disadvantaged families may face a home environment providing them with ample opportunities to develop their cognitive and language skills whereas children in disadvantaged families are more likely to face a home environment providing fewer learning opportunities. The Dutch funding scheme for primary schools distinguishes two main groups of disadvantaged pupils: native Dutch pupils with both parents having at most a degree from low-level vocational school and pupils with an ethnic minority background with both parents having at most a degree from low-level vocational school.¹⁷ Schools receive extra funding for pupils from these groups.

As mentioned above, the PRIMA survey contains only limited and unreliable information on actual enrollment.¹⁸ No questions with regard to actual enrollment have been asked in the 1994, 2000 and 2002 waves. In 1996 and 1998 parents were asked how old their child was when it entered school. Here parents are supposed to report age in years (4, 5 or 6) and months (0–11). Only 40.1 percent of the obser-

survey contains no information, however, that identifies whether they repeated first or second grade. Three percent is therefore an upper bound on grade repetition from grade 1 to grade 2.

¹⁶ Documentation containing the test items can be found here: <http://www.dans.knaw.nl/en/>.

¹⁷ Until 1995 native Dutch children were classified as disadvantaged if only one of the parents had at most a degree from low-level vocational school, provided that the breadwinner worked as a manual laborer.

¹⁸ Unfortunately there are also no other sources with information on precise age at which young children start school. For funding purposes schools only have to administer how many pupils attend at October 1st of each year.

Table 2
Maximum length of schooling and background characteristics.

	Non-disadv. (1)	Disadv			Pooled (5)
		Dutch (2)	Minority (3)	All (4)	
Education mother (reference category = Missing)					
Primary	0.005 (0.014)	−0.002 (0.011)	−0.005 (0.007)	−0.005 (0.006)	−0.002 (0.005)
Lower secondary	0.006 (0.007)	−0.003 (0.009)	0.007 (0.008)	0.002 (0.006)	0.004 (0.004)
Upper secondary	0.006 (0.007)	0.000 (0.011)	0.002 (0.009)	0.002 (0.007)	0.004 (0.004)
Higher	0.011 (0.007)	0.028 (0.022)	0.005 (0.014)	0.010 (0.011)	0.009* (0.005)
Education father (reference category = Missing)					
Primary	−0.010 (0.013)	−0.004 (0.009)	0.006 (0.006)	0.003 (0.005)	0.002 (0.004)
Lower secondary	−0.006 (0.006)	−0.004 (0.006)	−0.003 (0.006)	−0.004 (0.004)	−0.004 (0.004)
Upper secondary	−0.007 (0.006)	−0.007 (0.011)	−0.005 (0.008)	−0.006 (0.006)	−0.005 (0.004)
Higher	−0.007 (0.006)	−0.033 (0.030)	−0.007 (0.011)	−0.011 (0.009)	−0.006 (0.004)
Disadv. Dutch				0.0002 (0.003)	−0.003 (0.003)
Disadv. minority					−0.003 (0.003)
Girl	0.003 (0.002)	−0.006 (0.004)	0.004 (0.003)	0.0002 (0.003)	0.002 (0.002)
N	28,942	11,149	12,744	23,893	52,835
F-test joint sign.	0.834	0.819	0.490	0.660	0.369

Note: All regressions include 4 year dummies, 2 region dummies and their interactions, age and age squared.

vations in 1996 and 1998 have non-missing values on both year and month. For the disadvantaged groups, this figure is even worse; 31.2 percent for disadvantaged native pupils and 19.3 percent for disadvantaged minority pupils. Schools have also been asked to report for each pupil the year and the month in which they started to attend school. We thus have two measures of the same variable. Regressing the parents measure on the school measure and vice versa gives the reliability ratios of both measures. The reliability ratio of the parents measure equals 0.62 for all groups together, but reduces to 0.26 for minority pupils. The reliability ratio of the school measure equals 0.23 for all groups together (and 0.27 for minority pupils). The low response rates together with the low reliability ratio's among those who responded make the information on actual enrollment useless for further analysis.

4.2. Descriptive statistics

Table 1 shows descriptive statistics for the three groups of non-disadvantaged pupils, Dutch disadvantaged pupils and minority disadvantaged pupils. In column (4) we pool the two disadvantaged groups. The average age of the children in second grade at the moment of the test is almost 6 years. On average they could have spent almost 17 months in school at the day of the test. Note that there are no differences between disadvantaged and non-disadvantaged children in this respect, as expected if birth patterns are comparable.

Not surprisingly, the parents of disadvantaged pupils have lower levels of education than the non-disadvantaged.

The missing values for mother's education is much lower than for fathers, which could be because mothers are perhaps more likely to have filled in the parent questionnaire. Moreover, mother's education is missing as often for disadvantaged as for non-disadvantaged children.

The bottom rows of Table 1 presents average test scores for the various groups. In column (1) we see that as early as in second grade, non-disadvantaged children score about 1/3 of a standard deviation above average. The difference between non-disadvantaged and disadvantaged minority children is 1 standard deviation on the language test, and a bit less (0.80) on the arithmetic test. Comparing non-disadvantaged to disadvantaged Dutch pupils, we observe a difference on both tests of around 0.40.

5. Results

5.1. Exogeneity of maximum length of schooling

Before discussing the estimates of the intention to treat effects that are the focus of this paper, we first examine the exogeneity of maximum length of schooling by regressing this on four dummies for mothers' education, four dummies for fathers' education, a gender dummy, as well as on age, age squared and dummies for years, regions and year-region interactions. We do this for each group separately. Table 2 reports results from regressions of maximum length of schooling on background characteristics controlling for other covariates. These results show that there is no systematic relation between maximum length of schooling and observed background characteristics. This

Table 3
Language.

	Non-disadv.		Disadvantaged					
	(1)	(2)	Dutch		Minority		All	
			(3)	(4)	(5)	(6)	(7)	(8)
Max. length of schooling	-0.037* (0.021)	-0.021 (0.027)	0.047 (0.032)	0.053 (0.039)	0.074*** (0.029)	0.066* (0.034)	0.061*** (0.021)	0.060** (0.026)
Age	0.086*** (0.016)	0.168** (0.078)	0.024 (0.024)	0.057 (0.116)	-0.003 (0.022)	-0.048 (0.102)	0.010 (0.016)	0.008 (0.075)
Age-squared/100		-0.067 (0.063)		-0.027 (0.092)		0.036 (0.080)		0.002 (0.060)
Marginal effect age		0.074*** (0.020)		0.019 (0.030)		0.003 (0.026)		0.010 (0.020)
R-squared	0.082	0.082	0.084	0.084	0.090	0.090	0.195	0.195

Note: All regressions include 4-year dummies, 2 region dummies and their interactions. The background variables are: 4 dummies for mother's education, 4 dummies for father's education, and 1 gender dummy variable. The final two columns also include a dummy for disadvantaged Dutch. The standard errors are corrected for clustering at the school level and are heteroscedasticity robust. The "Marginal effect age" is the derivative with respect to age of the estimated model, evaluated at the mean age. Standard errors of the marginal effect of age are calculated using the delta-method.

Table 4
Arithmetics.

	Non-disadv.		Disadvantaged					
	(1)	(2)	Dutch		Minority		All	
			(3)	(4)	(5)	(6)	(7)	(8)
Max. length of schooling	-0.027 (0.020)	-0.007 (0.025)	0.069** (0.035)	0.065 (0.042)	0.052* (0.029)	0.031 (0.037)	0.060*** (0.022)	0.047* (0.027)
Age	0.089*** (0.015)	0.192** (0.084)	0.014 (0.026)	-0.009 (0.122)	0.019 (0.022)	-0.095 (0.105)	0.017 (0.017)	-0.049 (0.078)
Age-squared/100		-0.083 (0.067)		0.018 (0.097)		0.092 (0.086)		0.054 (0.063)
Marginal effect age		0.075*** (0.019)		0.017 (0.032)		0.035 (0.029)		0.026 (0.021)
R-squared	0.093	0.093	0.087	0.087	0.086	0.086	0.123	0.123

Note: See Table 3.

holds for all groups. The *F*-test for the joint significance of the background variables have *p*-values ranging from 0.490 to 0.834. Although our exclusion restriction only needs to hold conditionally on socio-economic status, we also checked whether our instrument is independent of observed characteristics in the full sample. These results are reported in column (5) and the *p*-value on the *F*-test equals 0.369. This evidence therefore supports the identifying assumption that maximum length of schooling is orthogonal to unobserved characteristics related to achievement. We can therefore reject that maximum length of schooling varies systematically with these background variables which affect test scores.

5.2. Effects of expanding schooling opportunities

Table 3 presents estimates of the effect of maximum length of schooling on pupils' language scores. Results are presented separately for the various groups of pupils and for two different specifications. The first specification includes a linear age term, the second specification also includes age squared as control variable. All specifications control for the full set of background characteristics.¹⁹

Notice first that the differences between the results from the specifications without and with the squared age term are small. We can never reject the hypothesis that the effect of this squared age term is equal to zero. This indicates that a linear specification of the effect of age on language scores is accurate for the given population. For the non-disadvantaged pupils the maximum length of schooling has no impact on language scores. The point estimates are even negative and significantly so in the first specification. This suggests that an expansion of schooling opportunities will not be beneficial for this group of pupils. The results are very different for pupils from the disadvantaged groups. For disadvantaged Dutch pupils the impact of an extra month of schooling opportunities equals 5 percent of a standard deviation, but this estimate lacks precision. For disadvantaged minority pupils an extra month of schooling opportunities raises their language scores by 7 percent of a standard deviation. These estimates are significantly different from zero. Since the effects for the two disadvantaged groups are of fairly similar magnitudes (equality cannot be rejected), we pool the two groups in order to gain precision. This gives us the results in final two columns. The estimate equals 0.06 and is significantly different from zero.

Table 4 reports comparable results from regressions in which the arithmetics score is the outcome variable. Again

¹⁹ Excluding the background characteristics gives very similar results.

the results from the specifications with linear and quadratic age terms are very similar, and again we find negative point estimates for non-disadvantaged pupils and positive effects for the two groups of disadvantaged pupils. For the two disadvantaged groups, only the impact estimates from the specifications with the linear age term are significantly different from zero. The estimates from the specifications with the squared age term lack precision. When the two groups of disadvantaged pupils are pooled, the impact estimates from both specifications are significantly different from zero. Disadvantaged pupils gain around 5 percent of a standard deviation on their arithmetics score when schooling opportunities are expanded by one month. The similarity of the results for two independent outcome measures makes it less likely that the results can be attributed to randomness.

For both outcome measures, the effect of age is significantly positive for non-disadvantaged pupils, while it is small and not significantly different from zero for pupils from the two disadvantaged groups. Interestingly, the sum of the (linear) age effect and the effect of the maximum length of schooling is of the same order of magnitude for each of the groups: 0.49 for non-disadvantaged pupils and 0.71 both for disadvantaged Dutch and disadvantaged minority pupils. Within each group, older pupils perform better than younger pupils; for non-disadvantaged pupils this is independent of potential school exposure, while for disadvantaged pupils this is the case because older pupils have higher values for maximum time in school.

The different findings for non-disadvantaged and disadvantaged children can probably be attributed to the differences in counterfactuals between these groups. As we mentioned in Section 3, identification comes from different exposure to holidays. The environment to which non-disadvantaged children are exposed during holidays is in terms of cognitive development apparently a close substitute to the school environment. For disadvantaged children this is not the case: the environment to which they are exposed during holidays is detrimental to their cognitive development relative to spending time in school. Cascio and Schanzenbach (2007) propose the same explanation for their finding that “disadvantaged school entrants who are biologically older are less likely to take the ACT or SAT than their biologically younger counterparts”. Unfortunately, we have no information on activities during the school holidays which would allow us to create a clearer picture of the counterfactual.

We find a significant effect for minority pupils on language and not on arithmetic. This can be explained by the fact that the two largest minority groups in the Netherlands are from Turkish and Moroccan origin. Many parents of these pupils still use Turkish or Moroccan as language at home.

As mentioned above, our identification strategy assumes that a quadratic specification of the age-achievement profiles is sufficient. While this is restrictive it is important to note that the related studies discussed in Section 2, typically include only a linear age term. To further probe the issue of the age controls, we regressed outcome measures on (1) year dummies and other controls but excluding region dummies and the region-year

interactions, and on (2) region dummies and other controls but excluding year dummies and region-year interactions. The first specification also uses variation in the timing of holidays between regions to identify the effect of maximum length of schooling (but does not control for region effects), the second specification also exploits variation in the timing of holidays between years to estimate the effect of interest (but does not control for year effects). For all groups the estimates from these alternative specifications are very similar to the estimates reported in Tables 3 and 4, and in no single case could we reject equality of the effects from the alternative specification and from the reported specification. This is reassuring and adds to the credibility of the findings.

6. Conclusions

This study introduced a novel way to estimate the effect of expanding enrollment opportunities on test scores and identify this separately from the age effect. This was possible due to the specific feature of the Dutch schooling system that allows children to start school when they turn 4. Together with the incidence of school holidays and the fact that a school year cohort consists of children born between October 1 and September 30 of the next year, this generates exogenous variation in enrollment opportunities which is not collinear with age (and age squared).

For disadvantaged pupils we find that increasing enrollment opportunities by one month increases language scores on average by 0.06 standard deviation and arithmetic scores on average by 0.05 standard deviation. Non-disadvantaged pupils do not benefit in test scores from expanded enrollment opportunities. Given that the achievement gap between disadvantaged pupils and non-disadvantaged pupils amounts to 0.6 to 0.7 standard deviation units, one additional month of enrollment opportunities, closes the gap by almost 10 percent. This suggests that at age 4 school and home environment are close substitutes in the production of achievement for non-disadvantaged children, whereas for disadvantaged children school provides better learning opportunities than the home environment. Although these effects are reduced form effects and as such do not estimate the causal effect of enrollment, we argue that they are lower bounds on the effects of actual schooling in a 2SLS analysis.

The test scores are measured around 2 years later and the effects we measure are therefore relatively short-term effects. Yet, as the results of Garces et al. (2002) show, even if intervention effects on test scores fade out over time there may be long-term effects on other outcome variables.

The 0.05–0.06 standard deviation increase in test scores reported here come at a cost of (depending on the type of disadvantaged pupil) 354 to 541 euro per pupil (Statistics Netherlands). Our effect estimate is at the high end compared to the studies reviewed in Section 2, and also compares favorably to the costs and effects of Head Start. Currie and Thomas (1995) report an effect of Head Start participation on early test scores of 0.203 of a standard deviation for disadvantaged white children. For Afro-American children they find no significant effects. Participation in Head Start costs approximately \$3500

per child per year. Increasing opportunities to enroll into primary school at younger ages are therefore an interesting policy alternative to targeted programs such as Head Start.

Acknowledgments

An earlier version of this paper circulated as Leuven, Lindahl, Oosterbeek, and Webbink (2004). We gratefully acknowledge valuable comments from seminar participants and from two anonymous referees.

References

- Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics*, 106(4), 979–1014.
- Angrist, J. D., & Krueger, A. B. (1992). The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples. *Journal of the American Statistical Association*, 87(418), 328–336.
- Barnett, W. S., Hustedt, J. T., Robin, K. B., & Schulman, K. L. (2005). *The state of preschool: 2005 state preschool yearbook*. New Brunswick, NJ: The National Institute for Early Education Research.
- Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *Quarterly Journal of Economics*, 121(4), 1437–1472.
- Berlinski, S., Galiani, S., & Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93, 219–234.
- Berlinski, S., Galiani, S., & Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, 92, 1416–1440.
- Bound, J., Jaeger, D. A., & Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430), 443–450.
- Cahan, S., & Cohen, N. (1989). Age versus schooling effects on intelligence development. *Child Development*, 60, 1239–1249.
- Cascio, E. U., & Lewis, E. G. (2006). Schooling and the armed forces qualifying test: Evidence from school entry laws. *Journal of Human Resources*, 41(2), 294–318.
- Cascio, E., & Schanzenbach, D. W. (2007). First in the class? *Age and the education production function*. Working Paper 13663. NBER.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27(5), 703–722.
- Currie, J. (2001). Early childhood interventions. *Journal of Economics Perspectives*, 15(2), 213–238.
- Currie, J., & Thomas, D. (1995). Does Head Start make a difference? *American Economic Review*, 85(3), 341–364.
- Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, 25, 43–62.
- Dhuey, E. (2007). Who benefits from kindergarten? *Evidence from the introduction of state subsidization*. Mimeo.
- Dobkin, C., & Ferreira, F. (2006). *Do school entry laws affect educational attainment and labor market outcomes?* Mimeo.
- Elder, T. E., & Lubotsky, D. H. (2009). Kindergarten entrance age and children's achievement: Impacts of state policies, family background, and peers. *Journal of Human Resources*, 44(3), 641–683.
- Fredriksson, P., & Öckert, B. (2006). Is early learning really more productive? *The effect of school starting age on school and labor market performance*. Technical report 2006:12. Uppsala: IFAU.
- Garces, E., Thomas, D., & Currie, J. (2002). Longer-term effects of head start. *American Economic Review*, 92, 999–1012.
- Gormley, W. T., & Gayer, T. (2005). Promoting school readiness in Oklahoma: An evaluation of Tulsa's pre-K program. *Journal of Human Resources*, 60, 533–558.
- Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology*, 41(6), 872–884.
- Hansen, K., Heckman, J., & Mullen, K. (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics*, 121(1–2), 39–98.
- Heckman, J. J. (1999). *Policies to foster human capital*. Technical report 7288. NBER.
- Leuven, E., Lindahl, M., Oosterbeek, H., & Webbink, D. (2004). *New evidence on the effect of time in school on achievement*. HEW 0410001. EconWPA [<http://ideas.repec.org/p/wpa/wuwph/0410001.html>].
- Mayer, S. E., & Knutson, D. (1999). Does the timing of school affect how much children learn. In S. E. Mayer, & P. E. Peterson (Eds.), *Earning and learning: How school matters* (pp. 79–102). Brookings Institution and Russell Sage Foundation.
- Merens, A., & Hermans, B. (2009). *Emancipatiemonitor 2008*. Den Haag: Sociaal en Cultureel Planbureau.
- Puhani, P. A., & Weber, A. M. (2007). Does the early bird catch the worm? Instrumental variables estimates of early educational effects of age of school entry in Germany. *Empirical Economics*, 32, 359–386.
- Stipek, D. (2002). At what age should children enter kindergarten? A question for policy makers and parents. *Social Policy Report*, 16(2), 3–16.
- Strøm, B. (2004). *Student achievement and birthday effects*. Mimeo, Norwegian University of Science and Technology.