German Economic Review 12(4): 371–394

Why Young Boys Stumble: Early Tracking, Age and Gender Bias in the German School System

Hendrik Jürges
University of Wuppertal and DIW Berlin

Kerstin Schneider
University of Wuppertal and CESifo München

Abstract. *After primary school, German pupils are given a secondary school track recommendation. This recommendation and the actual track choice are strongly associated with later life outcomes. Using data from the German PISA 2000 extension study, we analyze the effect of relative age on track recommendations and actual choice. Younger pupils and boys are less likely to be recommended to and enrolled in the academic track (Gymnasium), the most attractive track in terms of later life outcomes. Flexible enrollment and grade retention partly offset these effects. We find no convincing evidence that postponing the recommendation by lengthening primary school by 2 years reduces the age or gender bias.*

JEL classification: *I*21, *I*28.

Keywords: Educational tracking; age and gender bias.

1. INTRODUCTION

One of the most debated features of the German schooling system is the early selection of students into different secondary school tracks. At the end of primary school (usually at the age of 10), German pupils are given a more or less binding recommendation (depending on the state) which type of secondary school they should attend. This recommendation, and the resulting decision (i.e., which track to take), is strongly associated with outcomes throughout a pupil's entire life, mainly in relation to labor market success (see e.g. Dustmann, 2004). Of the three main secondary school tracks [basic track (*Hauptschule*), intermediate track (*Realschule*) and academic track (*Gymnasium*)], the academic track is the most prestigious, and it is the only track that provides direct entry into tertiary education. Upon finishing the academic track successfully, children in Germany are awarded a general university-entrance diploma. While mobility between tracks is desirable and possible in theory, factual mobility between tracks is low, and the mobility that does exist is usually downward in nature.[1] Thus, the decision made at the end of primary school effectively limits the educational opportunities of children. This applies even though this initial disadvantage is reduced by the possibility of entering academic track schools after successful completion of the intermediate track (Mühlenweg and Puhani, 2010).

1. For instance, in our data (described below), 5.7% of students who did not attend the academic track in grade 5 attended the academic track in grade 9, whereas 17.5% of those who attended the academic track in grade 5 no longer did so in grade 9. In other words, we find three times more downward mobility than upward mobility.

A possible justification for school tracking is efficiency (Brunello and Giannini, 2004; Hallinan, 1994). Homogeneous classrooms provide a learning environment that is better adapted to the abilities of the individual pupil. Ability differences between the high- and low-achieving students in a class are smaller than in comprehensive school systems and allow more focused instruction without leaving the weakest pupils behind or having better students not be challenged by the curriculum. Overall, according to its proponents, tracking will benefit weak and strong pupils alike, leading to better aggregate educational outcomes.

However, the PISA studies showed that Germany's tracked school system was not as successful as had been assumed. On the contrary, Germany ended up in the bottom half of the 'PISA league table', while many of its neighboring countries performed better. In fact, similar results had already been obtained by the TIMSS study 5 years before the first PISA study (see e.g. Jürges and Schneider, 2004), but this went largely unnoticed by the wider public.

Opponents of tracking argue that creating a learning environment with peers of similar ability will harm low-ability students more than it benefits high-ability children. Aggregate performance (and school efficiency) might thus suffer from tracking. Using international data, Hanushek and Wößmann (2006) find that tracking after the primary school level increases educational inequality and reduces aggregate performance. Examining a Finnish comprehensive school reform, Pekkarinen *et al.* (2009) find that abandoning tracking had a small *positive* effect on cognitive test scores and no effect on the mean performance in the arithmetic or logical reasoning tests.

While this research suggests that tracking might not be a good idea in the first place, the present paper examines consequences of the German tracking system that would be undesirable even if tracking did increase aggregate performance. In the following, we analyze secondary school track recommendations made by teachers and the actual choices made by parents (and students). Teachers' recommendations are typically based on academic achievement in primary school and the teachers' appraisal of the individual pupil's ability. Our primary interest is to analyze the effect of gender and relative age on teacher recommendations. Since enrollment takes place only once per year, the age range of any given class is at least 1 year. Being a year older when the recommendation is made in fourth grade could increase one's chances of being sent to the academic track. The reason is that at this age, 1 year of difference is significant in terms of maturity and academic achievement (Crawford *et al.*, 2007). A similar reasoning applies to the chances of boys in the educational system. At the age of 10, boys are less mature than girls, display less academic achievement, and hence might receive academic track recommendations less often (Lehmann and Peek, 1997).

Both, gender and relative age are arguably determined by chance: the first is determined during conception and the second by a child's birthday in relation to an arbitrary enrollment cut-off date. Thus, if we find that age and gender influence track recommendations, the system is discriminatory in the sense that it creates inequality regarding access based on a random event. Teachers should recommend students based on their assessment of (innate) ability and prediction of future academic performance, and not on academic performance as early as at age 10. Crawford *et al.* (2007) show that the age effect in academic achievement is strongest at the age of 5 and decreases thereafter, but it is still significant at ages 16 and 18. Moreover, they find that the disadvantage of younger students in terms of being listed as having special educational needs is greatest at age 11, i.e., approximately at the age at which the tracking decision is made in Germany.

A number of recent economic studies have analyzed birthday effects on indicators of academic achievement, such as years of schooling (Angrist and Krueger, 1992), standardized test results (Bedard and Dhuey, 2006; Datar, 2006; Frederiksson and Öckert, 2006; Leuven *et al.*, 2004; Puhani and Weber, 2007; Strøm, 2004), grade retention (Eide and Showalter, 2001) and factual secondary track choice (Fertig and Kluve, 2005; Mühlenweg and Puhani, 2010; Puhani and Weber, 2007; Schneeweis and Zweimüller, 2009), as well as on labor market performance (Frederiksson and Öckert, 2006). Bedard and Dhuey (2006) present the most comprehensive study to date on the effects of relative age at school entry. Using data from TIMSS 1995 and 1999 to examine 19 different countries (but not Germany), they find significant age effects on test scores in all countries. More importantly, the size of the effect systematically varies across countries. The age effect is largest in those countries where compliance to the cut-off date rule is nearly perfect. They also find that grade retention and late enrollment reduce the age effect by about one-third.[2]

The evidence for age effects in Germany is somewhat inconclusive. Fertig and Kluve (2005), using retrospective survey data, do not find any significant effects for age at school entry on the likelihood of repeating a class and eventually attaining a leaving certificate. In contrast, using administrative data from the state of Hesse, Puhani and Weber (2007) find strong positive effects for school entry age on standardized test scores at the end of primary schools (using PIRLS data) and the likelihood of attending a more advanced school track. Using the same data, Mühlenweg and Puhani (2010) find that the effect of relative entry age on track attendance lasts until the end of lower secondary school (age 16).

It is not surprising that relative age effects – especially at young ages – are omnipresent.[3] The important question is whether such initial disadvantages persist throughout a child's schooling. In the German educational system, this is more a cause for concern than in other countries. Birthday effects may be long-lasting because children are separated early (at the age of 10) into different secondary school types. This separation is based on a recommendation given by the primary school. Clearly, the teacher's recommendation should not primarily reflect current ability and maturity but also be based on predictions of the future development of the child. Such predictions should certainly not be affected by the gender or the relative age of a child in the fourth grade. Moreover, while teacher recommendations used to be non-binding in most German states, we observe a general tendency in education policy of giving greater weight to teachers' recommendations.

Gender differences in student achievement, from kindergarten to college, have recently come to attention again in the relevant literature (cf. Bedard and Cho,

2. In addition, there might be *factual differences in ability* by season of birth that could trigger month of birth effects on educational achievement (e.g., due to nutritional status in *utero*). However, month of birth effects appears to exist independently of whether the cut-off date is in the summer or winter, with those born in months after the cut-off outperforming those born in months before the cut-off. This provides indirect evidence that relative age effects are more important than seasonal effects.
3. Relative age effects appear to be particularly pronounced in competitive environments such as sports (see, e.g., Barnsley *et al.*, 1985, 1992).

2010; Cho, 2007; Jacob, 2002; Lai, 2010; Lehmann and Peek, 1997; Sax, 2001). Education systems characterized by early tracking may perpetuate or even increase early gender differences in student achievement which are due to differences in the speed of cognitive development. On average, girls develop certain abilities more quickly: young girls learn to read more quickly, are more verbal and more able to sit still (see e.g. Sax, 2001 for an overview), whereas boys appear to learn faster in mathematics and science. Existing evidence from Germany with respect to track recommendations shows that girls are more often recommended to academic track schools (Lehmann and Peek, 1997). Part of the explanation is that – at the age of 10–12 – boys are less mature than girls and thus perform worse when secondary school track recommendations are made. However, Lehmann and Peek also show that when actual achievement (measured in standardized tests) is controlled for, the difference remains. One explanation is that girls better meet other (social) expectations of teachers in primary schools.[4]

Using data from the German PISA 2000 extension study (PISA-E), we estimate the effect of gender and relative age on academic track recommendations and follow students from primary school to the ninth grade to see whether the age effect diminishes over time. This could be true if, for instance, students do not adhere to their teachers' recommendations when they enter secondary school or change school types in the course of secondary school. We find that gender and relative age at the time of recommendation have a significant and sizable effect on teachers' recommendations to enter the academic track. Younger pupils and male pupils are recommended to the academic track less often. Since relative age at school entry is largely driven by institutional birth date regulations concerning school entry, this type of recommendation bias is arbitrary. We also find that early and late enrollment, non-compliance with the recommendation and grade retention partly offset this bias, thereby confirming the effectiveness of this built-in flexibility in reducing possible inequalities in education opportunities. Furthermore, the relative age effect diminishes as students get older but does not vanish, whereas the gender effect is surprisingly persistent until grade 9. Finally, we find no convincing evidence that postponing the recommendation by another 2 years reduces either bias in a significant way.

Our study contributes to the existing literature in several ways. First, it uses an alternative data source for Germany (PISA-E), hence covering all German states. Second, it looks at track recommendations at the end of primary school as the outcome variable. Third, it combines the literature on relative age effects with the literature on educational tracking in a novel way, leading to important policy conclusions for the practice of educational tracking in Germany. In contrast to Puhani and Weber (2007), we focus our interpretation of the age effect not on the *IV* age effect but on the policy relevant (reduced form) effect of the age bias that remains after all existing flexibility within the system has been accounted for. Moreover, we combine the discussion of the age bias with the debate on disadvantages experienced by boys in the educational tracking system, arguing

4. (Innate) gender differences in cognitive ability have long been studied in the literature on education and psychology. The consensus that seems to have emerged is that the average ability of both sexes is very similar, with slight advantages in mathematics for men and slight advantages in reading for women. Moreover, the variability of skills is larger among men; hence, more men than women are found in the upper tails of the joint distribution (Hedges and Nowell, 1995).

that the maturity effect is driving properties of the German early tracking system which are both undesirable and unintended.

2. INSTITUTIONAL BACKGROUND

In the following brief description of the German school system, we emphasize those aspects that are most relevant for understanding the importance of secondary school track recommendations in the context of Germany.[5] All children in Germany attend primary school, which covers grades 1–4, and in two states (Berlin and Brandenburg) grades 1–6. There is no formal exit examination at the end of primary schooling. If the primary school (teacher) considers a child suitable for a certain type of secondary school, the child will be admitted without any special admission procedure. If the primary school's recommendations conflict with the parents' wishes, however, the final decision about the child's future educational path lies either with the parents, the secondary school or the school supervisory authority, depending on state laws.

Each secondary school track (basic, intermediate, academic track) leads to a specific leaving certificate. The basic track provides its students with basic general education, the intermediate track provides a more extensive general education, and the academic track provides an in-depth general education. Depending on their academic performance, students can switch between school types. A fourth type of school is the *Gesamtschule*, or comprehensive school. The comprehensive school offers all lower secondary level leaving certificates, as well as providing upper secondary education. However, it does not exist in all states and usually plays a minor role. Less than 10% of all children in Germany attend a comprehensive school. Yet another type of secondary schooling, the orientation stage (*Orientierungsstufe*), will play a role in our analyses later on. After attending primary school for 4 years, students in the states of Niedersachsen and Bremen move on to a 2-year orientation stage. The orientation stage can be either organized as an independent school or be part of a primary or secondary school (actually, the orientation stage has been abolished recently in both states). The objective of the orientation stage has been to provide students a smoother transition between primary and secondary schools.

As in most other countries, age at school entry in Germany is defined relative to a specific cut-off date. Until recent years, this date was 30 June in west Germany and 31 May in the former GDR. These cut-off dates applied to all children in our sample. The rule is that children who turn 6 on or before 30 June are admitted to primary school in that school year, which starts in August or early September. Children who turn 6 after 30 June are admitted to primary school 1 year later (in recent years, many states in Germany have deferred the cut-off date by 3, sometimes even 6 months, admitting increasingly younger children to primary school).

If everyone complied with the cut-off date rule and if there was no grade retention, the oldest children in a class (born in July or June, respectively) would always be 11 months older than the youngest (born in June or May, respectively). In this paper, we estimate how this age difference affects the secondary school track

5. See Jonen and Eckardt (2006) for a detailed description of the German school system, and Potrafke (2010), for an analysis of determinants of state differences in education policies.

recommendation given by the primary school in fourth grade. We are also interested in how flexible enrollment and grade retention modify the age effect.

Before children are admitted to primary school, they have to pass some basic maturity test. Children who are old enough to enter school, but do not pass this test, are admitted to primary school one year later (*late enrollment*). Children who are born after the cut-off date (but before 31 December) may be admitted to school by parental request, provided they pass the maturity test (*early enrollment*). Moreover, children who do not fulfill minimal performance criteria are not promoted to the next grade. The decision not to promote a child lies entirely with the school, but children may repeat classes on their own (or their parents') initiative.

3. ESTIMATION STRATEGY

Estimating the gender effect is straightforward, as gender is arguably exogenous. The age effect is more problematic, because actual age at recommendation may be endogenous due to late or early enrollment and grade retention, and OLS estimates may be severely downward biased. Following the literature on age effects, we use assigned relative age as our instrument. As in Bedard and Dhuey (2006), we coded assigned relative age from 0 (months) for children born immediately before the cut-off date (those born in June in western Germany and those born in May in eastern Germany) to 11 (months) for those born immediately after the cut-off date.[6]

We compute two sets of estimates for two different subsamples. The first set contains the results of linear regressions (with heteroskedasticity-consistent standard errors) of academic track recommendations and academic track attendance on the *assigned relative age at recommendation*. These are *reduced form* regressions that show the net effect of month of birth created by the German education system. This effect is the most policy-relevant in terms of describing inequalities in educational access due to the way the school system is organized (this point has been made e.g. by Bedard and Dhuey, 2006). Finding no age effect in the reduced from regressions could either mean that there was no age effect in the first place or that the system was flexible enough to offset these age effects. The possibility of early and late enrollment (on the one hand) or grade retention and skipping (on the other hand) introduces some flexibility that can possibly counteract the 'pure' age effect on recommendation rates. Our interpretation of the reduced form coefficient can best be illustrated using the (mechanical) relationship between the reduced form and the instrumental variables estimator:

$$RF = IV \times [1 - p(NC)] \tag{1}$$

where RF denotes the reduced form coefficient, i.e., the slope coefficient from a regression of the track recommendation on assigned relative age. IV denotes the instrumental variable estimate of the effect of actual age on track recommendations using assigned age as an instrument. We interpret this parameter as the 'pure' age effect on academic track recommendations (see discussion below). $p(NC)$ is the proportion of non-compliers in the population. Non-compliers are all students who

6. The largest possible difference between the youngest and oldest entrants is 1 year minus 1 day. However, we have only information on month of birth. The age difference between those born in July of year t and June of year $t + 1$ is thus 11 months *on average*.

enroll a year too early or too late. The proportion of non-compliers thus also provides a measure of the *flexibility* of the enrollment laws. Clearly, the more flexible the enrollment regulation, the smaller the reduced form parameter is compared with the *IV* parameter. In other words: flexibility of enrollment reduces the 'pure' age effect.

To put our reduced form estimates in perspective, we also report the results of instrumental variables regressions of academic track recommendations on the *actual* age at recommendation using *assigned relative age* as an instruments. However, to be clear, we are not primarily interested in the *IV* estimates as such (in contrast to earlier studies using the same instrumental variables strategy), but in the *difference* between the reduced form regressions and the *IV* results. In the context of our study, the difference between the *IV* estimate and the reduced form regressions is more interesting, because it shows the effect of the flexibility of the enrollment policy on the chances of getting academic track recommendations and enrolling in the academic track.

Although we follow a procedure that is standard in the literature on relative age effects (e.g. Bedard and Dhuey, 2006; Mühlenweg and Puhani, 2010), the interpretation of the *IV* coefficient deserves some discussion. Since *IV* identifies the complier-specific causal effect in the presence of heterogeneous treatment effects, it is unclear whether it can be generalized to always-takers and never-takers.[7] The definition of always-takers and never-takers is arbitrary in our context because it is not clear, *a priori*, what the 'treatment' is. We prefer to think of children enrolling when relatively old, independent of month of birth, as always-takers (receiving the treatment of 'enrolling relatively old') and children enrolling relatively young, independent of month of birth, as never-takers. As discussed in detail in Appendix A, we have reason to believe that average treatment effect and complier-specific causal effect should not be too different.

We mainly use the *IV* estimate to gauge the beneficial effect of flexible enrollment and grade retention on recommendation rates. Thus we are interested in establishing a lower bound for the average treatment effect, which would also give us a lower bound for the effect of flexible enrollment and grade retention. Under the extreme assumption that always-takers have no larger returns from enrolling 'when relatively old' than compliers and that never-takers have zero returns, the average treatment effect should be at most 16% smaller than the complier-specific effect. More realistic scenarios would entail an even larger average treatment effect (cf. Appendix A).

We estimate reduced form and *IV* regressions, first on the full sample and then on a restricted sample that includes only those students who were born immediately before and after the threshold. In west Germany, this restricted sample includes those born in June and July, while in east Germany, this sample includes those born in May and June, respectively. The purpose of this exercise is to substantiate the claim that the instrument month of birth is in fact exogenous, i.e., uncorrelated with any relevant individual characteristics that are omitted or unobserved (Buckles and Hungerman, 2008). We argue that while there may be seasonal variation in fertility that is correlated with family characteristics, timing of

7. Note that the existing literature either neglects (Bedard and Dhuey, 2006) or dismisses this issue as unimportant (Mühlenweg and Puhani, 2010). We also tend to believe that it is of minor importance in a quantitative sense.

birth in a narrow band around the cut-off date is uncorrelated with any potential unobserved confounders. Moreover, any other potential seasonal variation of characteristics with month of birth (health, personality) should be negligible in our restricted sample. Finally, we also compute each model with and without the inclusion of several important individual background variables, thereby controlling for possible correlations between birth dates and observed background variables.

4. DATA DESCRIPTION

The data used in this study are drawn from the extension of the German PISA 2000 study (PISA-E; cf. Baumert *et al.*, 2003). In particular, we use the sample that contains information on children who are in the ninth grade in 2000, independent of their age. The total sample after deleting cases of item non-response contains data on 26,112 students from all 16 German states.

The data include detailed retrospective information on the school career of the children up to grade 9, such as age at school entry, whether classes were repeated (and which), track choice recommendations and actual track choices in grades 5, 7 and 9. The latter four variables are our main outcome variables. Table 1 shows the proportion of students receiving an academic track recommendation and taking the academic track in grades 5, 7 and 9, by state. Overall, more than 40% of the students in the sample have received an academic track recommendation. Actual enrollment rates are lower: 37% of grade 7 students are enrolled in the academic track, a figure that drops to 34.5% in grade 9. This illustrates the presence of downward mobility: during the course of lower secondary school, more students leave the academic track for non-academic tracks than vice versa. It is remarkable to find such a gap between recommendation and actual enrollment in grade 5. Obviously a sizable proportion of parents decide not to send their child to the academic track despite having received the recommendation.[8] In Table 1, we also find large differences in recommendation rates between states; hence, we always include state dummies in our regressions.

Table 2 describes the prevalence of flexible enrollment and grade retention in our sample for each primary school type. Early enrollment rates in Germany are fairly low: in our data, we find rates of about 3–5%. Late enrollment is more than twice as common, except in the two states with 6-year primary schools. The proportion of students recommended for the academic track differs substantially between the three enrollment groups. In the group of states with 4-year primary schools, the average recommendation rates for those enrolling early is 49.5%, compared with about 41.4% for children who enter school at the regular age, and 23.3% for those who enter school late. The bivariate relationship between enrollment and academic track recommendations suggests that students who enroll early (late) constitute a positive (negative) selection in terms of ability. The proportion of students who had to repeat a class in primary school is between 2%

8. Additional analyses show a clear education gradient. Less-educated parents are more likely not to send their children to Gymnasium although they have a recommendation. Conversely, better-educated parents are more likely to send their children to Gymnasium although they have not received a recommendation. The most likely explanations are that better-educated parents have different educational aspirations or have more confidence in their own assessment of their child's potential.

Table 1 Proportion of students receiving the academic track (AT) recommendation, and being enrolled in the academic track in grades 5, 7 and 9, by state and type of primary school (in per cent)

State/type of primary school	Recommendation	Grade 5	Grade 7	Grade 9	*N*
4-year primary schools					
Saarland	37.7	37.1	33.1	32.3	1,606
Rheinland-Pfalz	39.4	33.3	34.9	33.2	1,672
Nordrhein-Westfalen	34.8	33.7	31.2	28.7	1,871
Schleswig-Holstein	37.0	38.7	37.6	34.8	1,652
Hamburg	66.7	64.1	61.7	59.1	772
Mecklenburg-Vorpommern	35.5	28.6	33.0	28.0	2,259
Sachsen	39.6	30.1	31.0	28.8	2,211
Bayern	42.2	39.8	36.1	34.2	1,546
Baden-Württemberg	43.7	39.3	38.2	37.9	1,576
Hessen	36.8	28.4	34.2	28.6	1,927
Thüringen	37.5	30.7	30.8	28.3	2,368
Sachsen-Anhalt	51.8	48.4	46.7	46.5	1,314
Total (unweighted)	40.2	35.5	35.6	33.1	20,774
6-year primary schools					
Brandenburg	41.4		39.0	37.7	1,766
Berlin	53.0		58.6	55.3	810
4-year primary schools plus orientation stage					
Niedersachsen	37.1		40.0	37.5	1,505
Bremen	38.1		41.1	37.2	1,257
Total (unweighted)	40.4		37.0	34.5	26,112

Table 2 Description of enrolment choices and grade retention, by type of primary school

	4-year primary schools		6-year primary schools		4-year plus orientation	
Variable	Per cent in sample	. . . of which AT recommended (per cent)	Per cent in sample	. . . of which AT recommended (per cent)	Per cent in sample	. . . of which AT recommended (per cent)
Enrolment						
Early	3.4	49.5	3.1	63.7	5.5	38.6
Regular	88.3	41.4	92.5	45.1	81.0	40.3
Late	8.3	23.3	4.3	30.4	13.5	20.4
Repeated class						
No	96.1	41.5	97.8	45.8	92.0	40.2
Yes	3.9	7.1	2.2	10.7	8.0	7.3
Number of observations	20,774		2,576		2,762	

and 8%. Again, the low proportion of academic track recommendations among those who repeat a class in primary school is very low, suggesting that repeating a class indicates low ability.

5. BIRTHDAY EFFECTS ON SECONDARY TRACK CHOICE AND ATTENDANCE

Our estimation results for the effect of relative age on the probability of academic track recommendations and on the probability of actually taking the academic track are summarized in Tables 3–5. In these tables, we only report the parameters of interest, i.e., estimated relative age or gender effects obtained from different regressions. For instance, Table 3 contains the results of 32 different regressions: The four rows represent the four dependent variables (academic track recommendation and attendance in grades 5, 7 and 9) and the columns represent different estimators.

5.1. Four grade primary schools

We begin our analysis by estimating the net (reduced form) effect of assigned relative age on recommendation rates in the full sample (see columns denoted *RF*). Our results in Table 3 suggest that an 11-month difference in assigned relative age leads to a 10 percentage point difference in terms of receiving an academic track recommendation (note that relative age is coded so that the regression coefficients show the effect of the 11-month difference). This holds true regardless of whether covariates are controlled for or not. The control variables used in the regressions are parental education level, student sex, immigrant status, whether parents read daily to their child at preschool age, and state dummies. A ten percentage point difference is about 0.25 times the average probability of getting an academic track recommendation, and is roughly equal to the difference between having parents with an intermediate track leaving certificate and having parents with a basic track certificate or no leaving certificate. The net age effect is thus in fact sizable and – in our view – large enough to warrant a reassessment of the system of early tracking. The background characteristics controlled for are neither individually not jointly significantly correlated with the assigned relative age, which supports the conjecture that the month of birth (relative to the cut-off date) is exogenous.

The columns labeled *IV* show the results of instrumental variables regressions of the chances of getting an academic track recommendation on the actual relative age in the fourth grade, using assigned relative age as instrument. As explained in Section 3, the *IV* parameter approximately shows the relative age effect if everyone had entered school at the assigned age, i.e., if there had been no possibility of early or late enrollment, and if there had been no grade retention until grade 4. Given the selectivity of early and late enrollment and grade retention, it is not surprising that this counterfactual is much larger than the net effect of assigned relative age. Comparing the estimated effect size of about 17 percentage points to the 10 percentage points net effect suggests that grade retention and enrollment flexibility reduce the relative age disadvantage of younger children, i.e., those born shortly before the cut-off date, by about two-fifths. Again, from a policy perspective, the *RF* parameter is the relevant estimate, as it gives an estimate of the remaining age bias after system flexibility has reduced the age effect to some extent.

The second set of results in Table 3 uses the sample restricted to students born immediately before and after the threshold. As explained above, this sample restriction serves to substantiate the exogeneity assumption regarding month of birth because unobserved parental characteristics that influence student achieve-

Table 3 Reduced form and *IV* estimates of relative age effects on the probability of getting an academic track (AT) recommendation and being enrolled in the academic track in grades 5, 7 and 9 in states with 4-year primary schools (all states except Berlin, Brandenburg, Bremen and Niedersachsen)

	Full sample ($N = 20{,}774$)				Cut-off date sample ($N = 3{,}568$)			
	RF		*IV*		*RF*		*IV*	
Dependent variable	Without controls (1)	With controls (2)	Without controls (3)	With controls (4)	Without controls (5)	With controls (6)	Without controls (7)	With controls (8)
AT recommended	0.096**	0.099**	0.164**	0.170**	0.073**	0.082**	0.147**	0.166**
	(0.011)	(0.010)	(0.019)	(0.017)	(0.016)	(0.015)	(0.034)	(0.031)
AT grade 5	0.078**	0.082**	0.134**	0.140**	0.064**	0.072**	0.128**	0.147**
	(0.010)	(0.010)	(0.018)	(0.017)	(0.016)	(0.014)	(0.033)	(0.030)
AT grade 7	0.062**	0.065**	0.106**	0.112**	0.040*	0.048**	0.080*	0.098**
	(0.010)	(0.010)	(0.018)	(0.017)	(0.016)	(0.015)	(0.032)	(0.030)
AT grade 9	0.059**	0.062**	0.101**	0.107**	0.033*	0.042**	0.067*	0.085**
	(0.010)	(0.009)	(0.018)	(0.016)	(0.016)	(0.014)	(0.032)	(0.029)

Notes: *RF*, reduced form estimate (effect of assigned relative age); *IV*, instrumental variables estimate (effect of actual relative age instrumented by assigned relative age); control variables are as follows: parents' education level, student sex, immigrant status, whether parents read daily to child at preschool age; all regressions control for state.
$+p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$; heteroskedasticity-robust standard errors in parentheses.

Table 4 Reduced form and *IV* estimates of relative age effects on the probability of getting an academic track (AT) recommendation and being enrolled in the academic track in grades 7 and 9 in states with 6-year primary schools (Berlin, Brandenburg)

	Full sample ($N = 2{,}576$)				Cut-off date sample ($N = 432$)			
	RF		*IV*		*RF*		*IV*	
Dependent variable	Without controls (1)	With controls (2)	Without controls (3)	With controls (4)	Without controls (5)	With controls (6)	Without controls (7)	With controls (8)
AT recommended	0.076*	0.095**	0.108*	0.136**	0.103*	0.111*	0.158*	0.170*
	(0.032)	(0.030)	(0.046)	(0.044)	(0.050)	(0.046)	(0.078)	(0.073)
AT grade 7	0.058+	0.077**	0.083+	0.110*	0.086+	0.091+	0.132+	0.139+
	(0.032)	(0.030)	(0.046)	(0.043)	(0.049)	(0.046)	(0.077)	(0.073)
AT grade 9	0.049	0.069*	0.070	0.098*	0.081	0.086+	0.124	0.132+
	(0.032)	(0.029)	(0.046)	(0.043)	(0.049)	(0.046)	(0.077)	(0.072)

Notes: *RF*, reduced form estimate (effect of assigned relative age); *IV*, instrumental variables estimate (effect of actual relative age instrumented by assigned relative age); control variables are as follows: parents' education level, student sex, immigrant status, whether parents read daily to child at preschool age; all regressions control for state.

$^{+}p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$; heteroskedasticity-robust standard errors in parentheses.

Table 5 Reduced form and *IV* estimates of relative age effects on the probability of getting an academic track recommendation and being enrolled in the academic track (AT) in grades 7 and 9 in states with 4-year primary schools plus orientation stage (Niedersachsen, Bremen)

	Full sample ($N = 2{,}762$)				Cut-off date sample ($N = 488$)			
	RF		*IV*		*RF*		*IV*	
Dependent variable	Without controls (1)	With controls (2)	Without controls (3)	With controls (4)	Without controls (5)	With controls (6)	Without controls (7)	With controls (8)
AT recommended	0.098**	0.074**	0.306**	0.218**	0.035	0.037	0.142	0.145
	(0.029)	(0.026)	(0.100)	(0.082)	(0.044)	(0.040)	(0.185)	(0.158)
AT grade 7	0.089**	0.064*	0.278**	0.189*	0.014	0.005	0.056	0.018
	(0.029)	(0.027)	(0.101)	(0.083)	(0.044)	(0.040)	(0.183)	(0.156)
AT grade 9	0.073*	0.047+	0.227*	0.138+	0.022	0.019	0.088	0.073
	(0.029)	(0.026)	(0.098)	(0.080)	(0.043)	(0.039)	(0.181)	(0.154)

Notes: *RF*, reduced form estimate (effect of assigned relative age); *IV*, instrumental variables estimate (effect of actual relative age instrumented by assigned relative age); control variables are as follows: parents' education level, student sex, immigrant status, whether parents read daily to child at preschool age; all regressions control for state.

$^{+}p<0.10$, $^{*}p<0.05$, $^{**}p<0.01$; heteroskedasticity-robust standard errors in parentheses.

ment might also be related to seasonal variations in fertility (Buckles and Hungerman, 2008). Still, family planning should be imprecise enough to assume that being born a few weeks before or after the school entry cut-off date is a random event and hence uncorrelated with any omitted or unobserved family characteristics. Moreover, other potential covariation of student characteristics with month of birth, such as health or personality, should be negligible.

In the restricted sample, the net effect of assigned relative age is about 7–8 percentage points, which is slightly smaller than its effect in the full sample. However, due to the large standard error in the restricted sample, the effects are not statistically different from the effects obtained from the full sample. If there were no early and late enrollment or grade retention, the assigned relative age effect would be about 15–17 percentage points, as can be inferred from the *IV* estimations. Again, these findings are in the same range as those in the full sample, and they suggest that – as far as academic track recommendations are concerned – age disadvantages are effectively reduced by about 50% as a result of the remaining flexibility of the system.

We now continue by examining how the relative age effect diminishes throughout the school career up to grade 9. Table 3 also shows the effects of a 1-year age difference on the probability of being enrolled in the academic track in grade 5 (i.e., immediately following recommendation), grade 7 and grade 9. For instance, the full sample reduced form estimates without control variables are 7.8 percentage points in grade 5, 6.2 percentage points in grade 7 and 5.9 percentage points in grade 9. In other words, by grade 9, the initial disadvantage of younger students is reduced by more than one-third. This finding is robust across estimation methods (OLS vs. *IV*), samples (full sample vs. restricted sample) and specifications (with or without the inclusion of individual background control variables). Moreover, the effect of relative age is statistically different from zero, regardless of the specification chosen.

The mechanism by which the age effect diminishes is simple. As noted earlier, the proportion of students taking the academic track decreases over time. However, this is a selective process. The age gradient of academic track recommendations and enrollment becomes flatter because fewer older students take the academic track in spite of such recommendations. Moreover, because parents can always decide to send their children to a secondary school type less advanced than, but not more advanced than, the recommended track, our analysis suggests that the decisions of parents counteract the relative age effect. Moreover, relatively old students drop out of the academic track more frequently than younger students between grades 5 and 9, as the difference in maturity between the younger and older students decreases. This is illustrated in Figure 1, which shows, by assigned relative age, the difference between the proportion of students enrolled in the academic track in grades 5, 7 and 9, respectively, and the proportion of students who received a recommendation for the academic track. The difference can be interpreted as a net drop-out rate. For instance, among students with the lowest assigned relative age, the proportion enrolled in the academic track in grade 7 is 3.1 percentage points lower than the proportion who received a recommendation for the academic track (see the dashed line). Among the students with the highest assigned age, this difference is 6.4 percentage points. Thus, the initial advantage of the older students shrinks by more than 3 percentage points. Note that the difference is negative for all assigned relative ages, indicating that no age group benefits from upward mobility in the system; the effect of downgrading (downward mobility) dominates throughout.

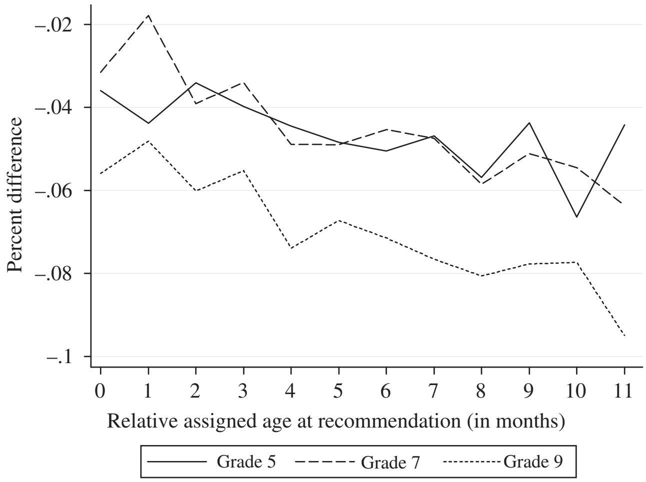

Figure 1 Difference between the proportion of students enrolled in academic track in grades 5, 7 and 9 and the proportion of students who received a recommendation for the academic track, by assigned relative age (sample restricted to states with 4-year primary schools)

5.2. 6-year primary schools

In the two German states with 6-year primary schools (Berlin and Brandenburg), students receive their recommendation when they are on average 2 years older than their peers in states with 4-year primary schools. Does delaying student age at recommendation by 2 years reduce the age bias? Our data do not allow us to distinguish whether students born in June in Berlin have entered school as the youngest group (if they lived in west Berlin) or as the oldest group (if they lived in east Berlin). Thus, we have excluded all Berlin students born in June from our sample.

The regression results for Berlin and Brandenburg are shown in Table 4. We basically ran the same regressions for this much smaller sample as for students in 4-year primary school states. First, we find statistically significant relative age effects on the probability of receiving an academic track recommendation across all samples, specifications and estimation methods. Second, although the estimated effects are somewhat smaller in the full sample than those for the 4-year primary schools, these differences are not statistically significant. Thus, judging from the data from Berlin and Brandenburg, 6-year primary schools are far from solving the problem of age-biased recommendations. We cannot even be sure whether they reduce the problem.

Third, with respect to actual academic track attendance in grades 7 and 9, the relative age bias gradually levels off and eventually becomes statistically insignificant when individual background is not controlled for. Although insignificant, the effects of relative age remain sizeable: we find an 11-month age advantage with respect to grade 9 academic track enrollment of 4.9 percentage points in the full sample and 8.1 percentage points in the cut-off point sample. When individual background is controlled for, the estimated reduced form effects are 6.9 and 8.6 percentage points, respectively, and they are both statistically different from zero. To summarize, 6-year primary schools are not significantly less biased toward older children than 4-year primary schools. This holds true in the short term – with respect to academic track recommendations – as well as in the medium term – with respect to actual enrollment.

5.3. 4-year primary schools with orientation stage

As explained above, some German states choose a third way of organizing the transition between primary and secondary schools by introducing a 2-year orientation stage (including grades 5 and 6) as another means to delay the tracking decision. The explicit aim of the orientation stage – abolished recently – was to improve the education system by providing a smoother transition between primary and secondary schools and by making sure that students receive the 'correct' secondary school track recommendation. The orientation stage combined characteristics of both primary and secondary schools. Students were taught together in most subjects, but there was also some internal tracking in key subjects such as German and mathematics.

The question is whether the orientation stage was successful in improving the allocation of students to secondary school types by reducing the relative age bias. Our results – derived from regression analyses – are shown in Table 5. As it turns out, these results are quantitatively similar to what we have obtained before. In the full sample, the estimated reduced form effect of having the highest rather than the lowest assigned relative age is 9.8 percentage points without control variables and 7.4 percentage points with control variables. This effect diminishes in secondary school but remains sizeable and statistically significant until grade 9.

The full sample instrumental variables estimates, which approximate the relative age effect in the absence of flexible enrollment and grade retention, are much larger than in states with 4- or 6-year primary schools. Controlling for individual characteristics our estimates suggest a difference of nearly 22 percentage points between the youngest and oldest students. This is related to the fact that early as well as late enrollment was more common in Niedersachsen and Bremen than in the other states. For instance, 45% of the students born in June in Niedersachsen and Bremen enrolled late, while the corresponding figure in the other 14 states is, on average, 20%. In essence, our results indicate that – in terms of the relative age bias – nearly nothing was gained by introducing an orientation stage, but that flexible enrollment had a very large bias reducing effect. Finally, we find considerably smaller and insignificant effects of relative age on academic track recommendations and enrollment in the restricted sample. However, because of large standard errors, the results are also generally not significantly different from the full sample results.

6. GENDER EFFECTS ON SECONDARY TRACK CHOICE AND ENROLLMENT

The results of our regression analysis for gender differences in recommendation rates and track attendance are shown in Table 6. Apparently as a consequence of differences in verbal and non-cognitive skills at age 10, gender differences are highly significant in all regressions.[9] In the sample of the states with a 4-year primary school girls are six percent more likely to get an academic track

9. We have also estimated models with gender and age effects interacted (both reduced form and *IV*). Interaction effects were quantitatively small and insignificant in all specifications. In other words, the age effect appears to be independent of the gender effect.

Table 6 Estimates of the effect of being male on the probability of getting an academic track (AT) recommendation and being enrolled in the academic track in grades 5, 7 and 9

	4-year primary school		6-year primary school		4 year plus two	
	$N = 20{,}744$		$N = 2{,}576$		$N = 2{,}762$	
Dependent variable	Without controls (1)	With controls (2)	Without controls (3)	With controls (4)	Without controls (5)	With controls (6)
AT recommended	-0.061^{**}	-0.079^{**}	-0.102^{**}	-0.106^{**}	-0.061^{**}	-0.100^{**}
	(0.007)	(0.006)	(0.019)	(0.019)	(0.018)	(0.020)
AT grade 5	-0.040^{**}	-0.056^{**}				
	(0.007)	(0.006)				
AT grade 7	-0.061^{**}	-0.075^{**}	-0.095^{**}	-0.099^{**}	-0.049^{**}	-0.085^{**}
	(0.007)	(0.006)	(0.019)	(0.019)	(0.019)	(0.020)
AT grade 9	-0.059^{**}	-0.073^{**}	-0.098^{**}	-0.102^{**}	-0.052^{**}	-0.083^{**}
	(0.006)	(0.006)	(0.019)	(0.018)	(0.018)	(0.020)

Notes: Control variables are as follows: actual relative age instrumented by assigned relative age, parental education level, immigrant status, whether parents read daily to child at preschool age; all regressions control for state.
$+p<0.10$, $p<0.05$, $^{**}p<0.01$; heteroskedasticity-robust standard errors in parentheses.

recommendation. Controlling for the covariates, this effect increases to almost 8 percentage points. This confirms earlier findings in the literature on Germany (Lehmann and Peek, 1997). As mentioned before, Lehmann and Peek also present some additional evidence that the gender difference remains when actual achievement is controlled for. This confirms the notion that even when controlling for cognitive skills, girls have better social (non-cognitive) skills and better meet other (social) expectations of primary school teachers. However, we believe that differences in the speed of developing cognitive and non-cognitive skills – differences which are rooted in differences in neurological development – should not affect educational chances. Observing that boys are at a disadvantage in regard to tracking decisions (due to these differences in neurological development) thus hints at another problem of the German school system.

In grade 5, the gender effect decreases by a third: more girls (particularly if they have an immigrant background) than boys are sent to tracks below the one recommended (while about the same proportion is sent to tracks above the one recommended). However, the difference in attendance rates increases again after grade 5 because the academic track drop-out rates of boys exceed those of girls. Thus, while the age effect decreases in lower secondary schooling, the gender effect remains high throughout lower secondary schooling. Developmental differences between girls and boys seem to be quite persistent at ages 10 through 15.[10] In light

10. If gender differences are truly persistent, one might argue that teachers' recommendations are less discriminatory as they seem at first sight, because teachers correctly anticipate future ability differences between boys and girls. The validity of this argument depends on the absence of discrimination when drop-out decisions are made.

of this, it is not surprising that delaying tracking for 2 years does not reduce the gender bias. The point estimates for the 6-year primary school are even higher.[11]

To summarize, the youngest boys have an almost 16% lower chance of getting an academic track recommendation than the oldest girls in the cohort. At grade 9, boys still experience this disadvantage (12%). And while the age effect decreases in lower secondary schooling, the gender effect is not reduced by grade 9. Hence delaying tracking by 2 years does not suffice to reduce the gender bias in tracking.

7. DISCUSSION

In this paper we study birthday and gender effects in the German school system. As in most other countries, school entry is subject to a cut-off date rule. Children who are born before a specific cut-off date are admitted to school in that year, while children born after the cut-off date are admitted 1 year later. Thus there is an 11-month difference in average assigned relative age between children born in the month before the cut-off date and those born in the month after the cut-off date. Relatively older children are more mature, perform better in school and have a higher level of social skills. Perhaps younger children also have problems asserting themselves in a group of older children. In addition, boys are outperformed by girls. Compared with girls, boys have significantly lower academic track recommendation and enrollment rates. Interestingly, this effect is persistent throughout a child's schooling, up to grade 9. Only in grade 5, when parents decide which track to enroll their child in, is the disadvantage of boys reduced. However, the disparity increases again after grade 5.

Although birthday effects have been documented in many countries, they do not raise too many concerns because one would generally believe that such birthday effects level off when children get older. However, in the German school system, there is more reason to worry than elsewhere. Birthday effects may be long-lasting because most children are separated into different types of schools on the basis of their scholastic achievement at the age of 10, i.e., when birthday effects are still relatively strong. Moreover, all the system's built-in flexibility with respect to entering the upper secondary school after grade 10, as well as its theoretical mobility between tracks in lower secondary schools, does not sufficiently compensate for the initial disadvantage of attending a lower track school.

Using data from the German PISA-E study, we study the effect of relative age and gender on a child's chances of getting a primary school recommendation for the academic track and of actually being enrolled in the academic track, the most academic and prestigious of Germany's secondary school tracks. We find that an 11-month difference in assigned relative age is associated with a 10 percentage point difference in receiving an academic track recommendation in grade 4, and a 6 percentage point difference in actually being enrolled in the academic track in grade 9. This is the net birthday effect observed in the German school system. The age effect would be about twice as large if there was no possibility of early and late enrollment, if the parents had no influence on secondary school choice, and if there was no grade retention. This 'pure' age effect is approximated by an

11. However, using cross-national data, Bedard and Cho (2010) find that the gender gap in math and science test scores (where boys outperform girls) in the eighth grade is smaller with less (distinct) tracking.

instrumental variable estimator using assigned relative age as an instrument for actual age, as it is common in the literature on age effects.

We substantiate our findings by performing the analyses on a sample restricted to students born in the 2 months immediately before and after the cut-off date. The overall evidence found in this paper is that the results from this restricted sample are not different from the full sample results. We interpret this as evidence that a selective timing of births does not drive the full sample results.

The birthday and gender effects also raise normative questions. Should assigned age or gender matter for the track choice recommendations, given that assigned age and gender are random and recommendations predetermine career opportunities? We do not believe it should. A solution often proposed is to abandon the current form of the tracking system altogether or track children at a later age, as it is common practice in other countries. As was discussed above, such a change might not only reduce educational inequality, but could also increase aggregate performance. Since some of the 16 German states have a 6-year primary school system, we had the opportunity to perform a (limited) test of the notion that late tracking reduces age and gender effects. However, we find practically no evidence that age and gender effects are smaller when recommendations are given in grade 6 rather than grade 4. One reason may be that tracking after sixth grade is still too early.

In practice, German parents have at least some limited influence on the secondary track taken, and we believe that parents should know about the age and gender effects described here, so that they can possibly counteract-biased recommendations by the primary school. Still, with early tracking some injustice generated by school entrance or exit laws probably has to be accepted as long as children are born during the entire year.

There are some tentative policy conclusions to be drawn from our analysis. Early tracking leads to a bias against relatively young children and boys, because academic performance and the level of social skills of older girls in grade 4 are higher than that of younger boys. Performance at grade 4, which is observed by the teachers, is a biased indicator of ability and future performance. Teachers seem to assign too much weight to current performance and too little weight to future performance. Since the differences in maturity level off over time, the tracking decision at age 10 is simply too early, taking place at an age at which the disadvantage of younger students in terms of being recorded as having special educational needs is greatest (Crawford *et al.*, 2007).

Our analysis shows that the built-in flexibility of the educational system helps to reduce the age bias but not the gender bias. What can be done? Delaying tracking is one option, but our analysis shows that 2 years might not suffice. Another idea is to sort children into classes by relative age and gender. This could easily be implemented in primary schools if there are several classes per cohort.[12] Furthermore, since parents' decisions tend to counteract the age bias, parents should be heard when making tracking decisions. Another important point is to raise teacher awareness. Teachers should be better aware of the disadvantaged groups and base their recommendations less exclusively on observed performance in grade 4. Teachers' guidelines for recommendations should mention potential age and gender bias so that teachers take these biases into account when making

12. We owe this suggestion to one of our reviewers.

recommendations.[13] To support the teachers' decision, a standardized cognitive abilities test could be used to substantiate the tracking decision by more reliably and objectively predicting future performance. One advantage of standardized tests is that age normalization can be applied, as is suggested in the context of the English school system (Crawford *et al.*, 2007; McNally, 2006). The shift of the cut-off date to 31 December recently introduced in some German states potentially increases rather than decreases the age bias, as children are getting even younger when the tracking decision is made. The unlucky birth draw will just be shifted from those born in June to those born in December.

Clearly, regardless of when the cut-off date is set, there will always be a youngest student in the class, and younger students will always be outperformed by older students on average. Nonetheless, it is important that educators are aware of this effect as well as the gender bias. Both are problematic from an equity and efficiency point of view. Particularly in Germany, with its early tracking system, the educational career path is largely determined at the age of 10, a time when age and gender effects are substantial.

APPENDIX A: AVERAGE TREATMENT EFFECT VS. COMPLIER-SPECIFIC TREATMENT EFFECT

In this appendix, we present some back-of-the-envelope calculations to show that the average treatment effect of age on academic track recommendations should not be very different from the complier-specific or local average treatment effect identified by *IV*. To ease the exposition, we concentrate on students from states with 4 year primary schools who are born either before or after the threshold (the cut-off date sample), and we also delete three students born before the threshold who enroll 1 year too early (at the age of 5) as well as 40 students born after the threshold who enroll late (at the age of 8). The endogenous treatment is now arbitrarily defined as enrolling at age 7 instead of age 6, and month of birth is used as an instrument. Table A1 shows the average recommendation rates and – in square brackets – the numbers of observations for each combination of treatment and instrument. Compliance types in each cell are shown in parentheses.

Never-takers are the group of students who always enroll at age 6 independent of their month of birth, always-takers are students who always enroll at age 7 independent of their month of birth and compliers are students who enroll at age 6 when born before the threshold and at age 7 when born after the threshold. Assuming that the instrument value is randomly allocated, we can estimate the proportion of compliers, always-takers and never-takers in the population from the column percentages in Table A1 (cf. Imbens and Jeffrey, 2009). Furthermore, the expected no-treatment outcomes for never-takers, $E(y0|n)$, and the expected treatment outcomes for always-takers, $E(y1|a)$, are known directly from the corresponding cells in Table A1. Together with the proportions of compliers, always-takers and never-takers, these can be used to compute the potential outcomes for compliers $E(y0|c)$ and $E(y1|c)$ – see Table A2.

Table A2 contains a number of relevant findings. First, there are more always-takers than never-takers (which reflects the fact that late enrollment is more

13. We have studied recommendation guidelines for two German states (Baden-Württemberg, Berlin), neither of which explicitly mentions relative age as a factor to be taken into account when formulating recommendations.

Table A1 Average outcomes (recommendation rates), by instrument and treatment value. Numbers of observations are shown in square brackets

	Before threshold ($Z = 0$)	After threshold ($Z = 1$)	Average
Age 6 ($D = 0$)	0.376 [1097] (never-taker/complier)	0.524 [317] (never-taker)	0.409
Age 7 ($D = 1$)	0.322 [456] (always-taker)	0.418 [1655] (always-taker/complier)	0.397
Average	0.361	0.435	

Table A2 Estimated proportions and potential outcomes for each compliance type

	Proportion	$E(y0)$	$E(y1)$	Difference
Always-takers	0.294		0.322	
Never-takers	0.161	0.524		
Compliers	0.546	0.333	0.470	0.137

common than early enrollment). Second, potential outcomes differ greatly between these compliance types. The estimated potential no-treatment outcomes are much larger for never-takers than for compliers (52.4% vs. 33.3%). This illustrates that students who enroll early are a positive selection of all students in terms of their academic achievement. The estimated potential treatment outcomes are much larger for compliers than for always-takers (47.0% vs. 32.2%), i.e., those who enroll late are a negative selection of all students in terms of their academic achievement. Third, the difference between treatment and no-treatment outcomes, i.e., the causal effect of age on recommendation rates, is known only for compliers. In our example, this difference amounts to 13.7 percentage points.

The issue is now, of course, whether the complier-specific causal effect is also representative for always-takers and never-takers; if so, we can interpret this effect as the average treatment effect for the entire population of students. We will never know for certain, but would suggest two ways of thinking about this. We are primarily interested in finding some 'lower bound' for the average treatment effect because the comparison of this lower bound with the reduced form parameter will also give an indication of the degree to which flexible enrollment and grade retention weaken the 'pure' age effect of recommendation rates.

One possibility is to assume that treatment effects are homogeneous and to compute potential no-treatment outcomes for always-takers (which would be 18.5%) and potential treatment-outcomes for never takers (which would be 66.1%) to see whether these become implausibly large or small. We do not think that this is the case. 18.5% is about the recommendation rate of students with less-educated parents (basic track), and 66.1% is about the recommendation rate of students with educated parents (university degree).

Alternatively, one can allow for heterogeneous effects and compute the average treatment effect under three different scenarios. Here, it is plausible to assume that always-takers have a higher return than compliers to receiving the treatment, and never-takers have a lower return than compliers to receiving the treatment; otherwise, they would not make the decisions they make. Further, it seems plausible that the treatment response is monotone, i.e., no group experiences disadvantages in terms of recommendation rates by enrolling late. (1) A *lower bound scenario* for the average treatment effect can be obtained if one assumes a treatment effect of 0% for never-takers and an effect for always-takers that is exactly the same as the complier-specific effect (i.e., 13.7 percentage points). In this scenario, the average treatment effect would amount to 11.5 percentage points. (2) An *upper bound scenario* uses the fact that the outcomes are proportions. The lowest possible value for the unknown $E(y0|a)$ is 0 and the largest possible value for the unknown $E(y1|n)$ is 1. This yields an upper bound of 24.6% for the average treatment effect. (3) In a *symmetrical scenario* where the absolute difference between the returns for always-takers and compliers and the absolute difference between the returns for never-takers and compliers are equal, the average treatment effect would be slightly larger than the complier-specific effect (because there are more always-takers than never-takers). In that case, the 'beneficial' effects of flexible enrollment and grade retention would be underestimated by our comparison of *IV* and reduced form coefficients.

As mentioned before, we are mostly interested in the lower bound scenario, which is mechanically driven by the proportion of never-takers in the population: the lower bound is equal to the *IV* estimate times 1 minus the proportion of never-takers. Estimating this proportion (on the basis of those born shortly after the threshold), we obtain 16% for states with 4-year primary schools, 13% for states with 6-year primary schools and 25% for states with 4-year primary schools plus orientation stage – the latter two numbers, however, are based on quite small samples. To summarize, a reasonable guess would be that *IV* estimates overestimate the average treatment effect by 16% at most.

ACKNOWLEDGEMENTS

We thank the editor, the three anonymous reviewers, and the participants at the 2008 SOLE meeting for their helpful comments.

Address for correspondence: Hendrik Jürges, Department of Economics, University of Wuppertal, Rainer-Gruenter-Str. 21 (FN.01), 42119 Wuppertal, Germany. Tel.: + 49 202 439 1385; fax: + 49 202 439 1384; e-mail: juerges@uni-wuppertal.de

REFERENCES

Angrist, J. and A. B. Krueger (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, 328–336.

Barnsley, R. H., A. H. Thompson and P. E. Barnsley (1985), 'Hockey Success and Birth-Date: The Relative Age Effect', *Journal of the Canadian Association for Health, Physical Education, and Recreation* 51, 23–28.

Barnsley, R. H., A. H. Thompson and P. Legault (1992), 'Family Planning: Football Style, the Relative Age Effect in Football', *International Review for the Sociology of Sport* 27, 77–88.

J. Baumert, C. Artelt and E. Klieme *et al.* (eds.), (2003), *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*, Leske + Budrich, Opladen, Germany.

Bedard, K. and I. Cho (2010), 'Early Gender Test Score Gaps Across OECD Countries', *Economics of Education Review* 29, 348–363.
Bedard, K. and E. Dhuey (2006), 'The Persistence of Early Childhood Maturity: International Evidence of Lung-Run Age Effects', *Quarterly Journal of Economics* 121, 1437–1472.
Brunello, G. and M. Giannini (2004), 'Stratified or Comprehensive? The Economic Efficiency of School Design', *Scottish Journal of Political Economy* 51, 173–193.
Buckles, K. and D. Hungerman (2008), 'Season of Birth and Later Outcomes: Old Questions', New Answers, NBER Working Papers 14573.
Cho, D. (2007), 'The Role of High School Performance in Explaining Women's Rising College Enrollment', *Economics of Education Review* 26, 450–462.
Crawford, C., L. Dearden and C. Meghir (2007), 'When are You Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England', CEE Discussion Paper 93.
Datar, A. (2006), 'Does Delaying Kindergarten Entrance Give Children a Head Start?', *Economics of Education Review* 25, 43–62.
Dustmann, C. (2004), 'Parental Background, Secondary School Track Choice, and Wages', *Oxford Economic Papers* 56, 209–230.
Eide, E. R. and M. H. Showalter (2001), 'The Effect of Grade Retention on Educational and Labor Market Outcomes', *Economics of Education Review* 20, 563–576.
Fertig, M. and J. Kluve (2005), 'The Effect of Age at School Entry on Educational Achievement in Germany', IZA Discussion Paper 1507.
Frederiksson, P. and B. Öckert (2006), 'Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance', IFAU Working Paper 2006:12.
Hallinan, M. T. (1994), 'Tracking. From Theory to Practice', *Sociology of Education* 67, 79–84.
Hanushek, E. and L. Wößmann (2006), 'Does Educational Tracking Affect Performance and Inequality? Differences-in-Differences Evidence across Countries', *Economic Journal* 116, C36–C76.
Hedges, L. V. and A. Nowell (1995), 'Sex Differences in Mental Test Scores, Variability and Numbers of High-Scoring Individuals', *Science* 269, 41–45.
Imbens, G. and W. Jeffrey (2009), 'Recent Developments in the Econometrics of Program Evaluation', *Journal of Economic Literature* 47, 5–86.
Jacob, B. (2002), 'Where the Boys Aren't: Non-Cognitive Skills, Returns to School and the Gender Gap in Higher Education', *Economics of Education Review* 21, 589–598.
Jonen, G. and T. Eckardt (2006), 'The Education System in the Federal Republic of Germany 2004', Secretariat of the Standing Conference of the Ministers of Education (KMK), Bonn.
Jürges, H. and K. Schneider (2004), 'International Differences in Student Achievement: An Economic Perspective', *German Economic Review* 5, 357–380.
Lai, F. (2010), 'Are Boys Left Behind? The Evolution of the Gender Achievement Gap in Beijing's Middle Schools', *Economics of Education Review* 29, 383–399.
Lehmann, R. H. and R. Peek (1997), 'Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Jahrgangsstufe an Hamburger Schulen. Bericht über die Untersuchung im September 1996', Hamburg, Behörde für Schule, Jugend und Berufsausbildung, Amt für Schule.
Leuven, E., M. Lindahl, H. Oosterbeek and D. Webbink (2004), 'New Evidence on the Effect of Time in School on Early Achievement', Mimeo.
McNally, S. (2006), 'Reforms to Schooling in the UK: A Review of Some Major Reforms and their Evaluation', *German Economic Review* 6, 287–296.
Mühlenweg, A. and P. A. Puhani (2010), 'Persistence of the School Entry Age Effect in a System of Flexible Tracking', *Journal of Human Resources* 45, 407–438.
Pekkarinen, T., R. Uusitalo and S. Kerr (2009), 'School tracking and development of cognitive skills', IFAU Working Paper 2009:6.
Potrafke, N. (2010), 'Public Expenditures on Education and Cultural Affairs in the West German States: Does Government Ideology Influence the Budget Composition?', *German Economic Review* 12, 124–145.

Puhani, P. A. and A. M. Weber (2007), 'Does the Early Bird Catch the Worm? Instrumental Variable Estimates of the Educational Effects of Age at School Entry in Germany', *Empirical Economics* 32, 359–386.

Sax, L. (2001), 'Reclaiming Kindergarten: Making Kindergarten Less Harmful to Boys', *Psychology of Men and Masculinity* 2, 3–12.

Schneeweis, N. and M. Zweimüller (2009), 'Early Tracking and the Misfortune of Being Young', Working Paper No. 0911, Johannes Kepler University of Linz.

Strøm, B. (2004), 'Student Achievement and Birthday Effects', Mimeo, Department of Economics, Norwegian University of Science and Technology.