Does Sharing the Same Class in School Improve Cognitive Abilities of Twins?

Dinand Webbink,[1] David Hay,[2] and Peter M. Visscher[3]

[1] *CPB Netherlands Bureau for Economic Policy Analysis, The Hague, the Netherlands*

[2] *Curtin University of Technology, Perth, Australia*

[3] *Queensland Institute of Medical Research, Brisbane, Australia*

This article analyzes the effect of classroom separation of twins on their cognitive abilities, measured at different ages in Dutch primary education. We use a large longitudinal school-based sample of twins and their classmates. The analysis tries to reduce the bias by unobserved factors due to the nonrandom assignment of twins by taking into account differences in school environment, previous test scores and variation in class assignment between years. We find that classroom separation matters for language in Grade 2. Nonseparated twins score higher on language, and the difference is larger for same-sex pairs. This finding is robust for various methods that take unobserved effects into account. In addition, there is some evidence for higher scores in arithmetic in Grade 2. For the higher grades we find no effect of classroom separation on cognitive ability. In the analysis of the effect of a separation of at least 3 years we find that separation increases language performance between Grade 6 and 8 for opposite-sex pairs.

A classic question for parents of twins is whether they should assign their children to the same class in school or to different classes. Sharing the same class has the advantage that twins can support each other and the presence of a co-twin can make them more confident in class. Twins are usually not used to being separated prior to primary school. On the other hand, assignment to different classes may stimulate the independent development of both twins and could prevent them being too competitive. It is not clear which alternative is better for the development of twins. This article investigates this question by comparing the cognitive performance of twin pairs sharing the same class with twin pairs in separate classes in the Netherlands. Dutch schools do not follow a general policy for the assignment of twins to classes. The Dutch Society for Parents of Multiples (Nederlandse Vereniging voor Ouders van Meerlingen; NVOM) advises parents to follow their own opinion, though generally NVOM believes separation stimulates the individualization of the twins (Geluk & Hol, 2001).

Two previous studies investigated the effect of classroom separation on twins' behavior and school achievement. Tully et al. (2004) used a sample of 878 same-sex twin pairs from the United Kingdom. A first assessment of the children took place at the age of 5 years and was followed by a second assessment 18 months later. The assessment included externalizing and internalizing problems, prosocial behavior, attention-deficit/hyperactivity disorder symptoms and learning effort, and was done by the teachers. In addition, children's IQ was assessed at the age of 5 and reading was assessed 18 months later. The difference between three groups was analyzed: pairs who were in the same class at both ages, pairs who were separated at both ages, and pairs who were in the same class at age 5, but separated by age 7. The major finding was that separated twins, both those separated early and late, have more internalizing problems than those not separated. In addition, twins separated later had lower reading scores than nonseparated pairs. Monozygotic (MZ) twins suffered more from separation than dizygotic (DZ) twins. The findings on the other variables were statistically not significant.

Van Leeuwen et al. (2005) investigated the short-term and long-term effects of school separation in a large sample of twins from the Netherlands Twin Registry. Short-term effects were studied at age 7 in twins separated at age 5 and long-term effects at age 12 in twins who had been separated or together most of the time in school. Mothers and teachers rated behavior problems at different ages. School performance was measured at age 12 using a national academic achievement test (CITO). Twins from separated pairs had more internalizing and externalizing problems than nonseparated twins at age 7 and age 12, as rated by both mothers and teachers. However, these effects could be explained by pre-existing differences. Only for the maternal rating of internalizing

Received 19 March, 2007; accepted 3 May, 2007.

Address for correspondence: Dr Dinand Webbink, CPB Netherlands Bureau for Economic Policy Analysis, PO Box 80510, 2508 GM, The Hague, the Netherlands. E-mail: h.d.webbink@cpb.nl

Table 1

Classroom Separation of Twins by Grade and Year (% in Same Class)

	Grade 2	Grade 4	Grade 6	Grade 8	All grades
1994	94	92	94	92	93
1996	73	81	81	85	79
1998	64	73	82	91	75
2000	59	66	77	85	70
2002	58	61	71	80	66

problems at age 7 the effects could be attributed to the separation. No differences were found in the cognitive test taken at age 12.

Both previous studies mainly concentrate on behavioral problems but also pay attention to cognitive achievements. Tully et al. (2004) analyze scores on a reading test at age 7 while controlling for IQ measured at age 5. Van Leeuwen et al. (2005) analyze test scores at age 12 without previous test scores for 843 twin pairs. Both studies show that there are pre-existing differences between separated and nonseparated twins. Tully et al. (2004) report that the separated early group contained a higher proportion of twins who had been referred to special education and their sample statistics also indicate a lower average family income for this group. Van Leeuwen et al. (2005) find that the probability of twins being separated at age 5 is positively related with the score on the scale for externalizing problems at age 3 and with socioeconomic status. This suggests that the assignment of twins is not random but the outcome of a decision by parents and teachers. It also raises concern about the impact of unobserved factors that determine this decision and which might be related to the performance of twins. The findings in the previous studies might be biased by these unobserved factors. For example, the peers in the classroom might differ for separated and nonseparated twins. If separated twins had peers with on average lower cognitive achievements, and these peers had a negative effect on the cognitive performance of these separated twins, a difference in the outcomes could be falsely attributed to the separation of twins.

This study focuses on differences in cognitive achievements of separated and nonseparated twins in Dutch primary education. We use a large longitudinal school-based sample which contains four measures of cognitive ability: IQ, language, arithmetic and CITO. The test for language and arithmetic are taken at ages 6, 8, 10 and 12 years.

Table 2

Means (Standard Deviations) for Dependent and Explanatory Variables

	Grade 2		Grade 4		Grade 6		Grade 8	
Co-twin in class	No	Yes	No	Yes	No	Yes	No	Yes
Language test	–1.31 (0.60)	–1.18 (0.53)	-0.20 (0.62)	–0.13 (0.59)	0.38 (0.60)	0.45 (0.52)	1.03 (0.64)	1.11 (0.59)
Arithmetic test	–1.19 (0.52)	–1.18 (0.54)	-0.46 (0.54)	–0.36 (0.55)	0.47 (0.38)	0.51 (0.36)	1.21 (0.36)	1.24 (0.38)
IQ test	n.a.	n.a.	100.7 (15.1)	100.9 (14.9)	99.6 (14.8)	100.4 (14.6)	101.1 (12.7)	100.6 (14.8)
CITO test	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	532.0 (11.3)	533.6 (10.2)
Share Girl (%)	47.3	52.7	46.0	51.6	49.4	52.5	53.8	53.5
Age at test (years)	5.85 (0.42)	5.83 (0.37)	7.97 (0.45)	7.95 (0.44)	9.99 (0.48)	9.97 (0.48)	12.07 (0.43)	11.95 (0.49)
Same sex pair (%)	71.5	66.6	75.4	68.4	68.7	68.2	61.5	68.1
Weight factor (%)								
1.0	53.6	55.8	57.6	53.9	55.7	52.8	49.4	50.0
1.25	13.9	22.4	11.3	23.2	15.2	23.6	15.4	27.5
1.9	27.9	16.7	26.2	18.1	25.0	18.2	33.3	18.1
Foreign country of birth (%)								
Father	38.3	26.8	36.3	26.2	35.2	26.7	41.0	26.7
Mother	37.7	21.6	35.8	23.5	32.2	24.1	37.8	23.5
Parents' education								
Father (1–5)	2.55	2.61	2.64	2.59	2.54	2.58	2.44	2.59
Mother (1–5)	2.49	2.52	2.53	2.51	2.40	2.49	2.33	2.44
Average class performance								
Language test	–1.18 (0.33)	–1.14 (0.31)	–0.17 (0.35)	–0.14 (0.33)	0.39 (0.30)	0.45 (0.25)	1.02 (0.32)	1.09 (0.28)
Arithmetic test	–1.06 (0.31)	–1.14 (0.35)	–0.47 (0.29)	–0.38 (0.32)	0.47 (0.19)	0.51 (0.17)	1.22 (0.23)	1.23 (0.22)
IQ test			101.6 (6.6)	100.2 (6.9)	100.4 (6.4)	100.2 (5.6)	100.3 (6.0)	100.4 (5.6)
Observations	562	1176	424	1168	264	1046	156	960

Note: Standard deviation in brackets.

Table 3

Probit Regression on Having a Co-Twin in your Class, Marginal Effect (Standard Errors)

	Marginal effect	Standard error
Same-sex pair	–0.023	(0.012)*
Girl	0.026	(0.011)**
Age at test (years)	0.000	(0.000)
Weigh factor		
Dutch lower educated	0.045	(0.019)**
Ethnic minority	–0.066	(0.022)**
Education father		
Level 2	0.046	(0.023)*
Level 3	0.047	(0.024)*
Level 4	0.053	(0.026)*
Level 5	0.002	0.028
Education mother		
Level 2	0.049	(0.022)**
Level 3	0.039	0.024
Level 4	–0.013	0.030
Level 5	0.018	0.029
Year of survey		
1994	0.000	0.000
1996	–0.234	(0.029)***
1998	–0.268	(0.027)***
2000	–0.339	(0.027)***
2002	–0.379	(0.027)***
Grade 2	0.000	0.000
Grade 4	0.061	(0.029)**
Grade 6	0.128	(0.045)***
Grade 8	0.186	(0.052)***
Observations	5704	

Note: ***/**/*significant at 1/5/10%-level

The IQ test is taken at age 8, 10 and 12 and the CITO test is taken at age 12. Hence, we can observe whether the findings on the separation of twins depend on age. In addition, the longitudinal character of the data enables us to study changes in cognitive ability between grades. The tests for languages and arithmetic are comparable between grades and designed to measure cognitive gains in these subjects.

The assignment to classes is measured in each wave of the study. In the analysis we use two definitions of class separation: at least 1 year separated, and separated in two subsequent waves. Another contribution of this study is that we try to reduce the bias by unobserved factors due to the nonrandom assignment of twins. We do this by taking into account differences in the performance of the classmates and previous test scores of the twins on tests designed to measure changes in achievements in language and arithmetic. In addition, we can identify small schools with only one class per grade level. These schools do not offer the opportunity to separate twins and we separately analyze the performance of pupils in these 'restricted schools'. Moreover, we use variation in class assignment between years in instrumental variable regressions.

Data

Data were available from the longitudinal PRIMA survey in the Netherlands, which aims to answer questions about educational strategies and performance in primary education in the Netherlands (Driessen, et al., 1994, 2004). We used the first five waves of the PRIMA survey including data on pupils, parents, teachers and schools. Each wave has approximately 60,000 pupils. The waves are: 1994, 1996, 1998, 2000 and 2002. The PRIMA project surveys a panel of approximately 600 schools, of which 180 schools are drawn for the over-sampling of pupils with a lower socioeconomic background. Within each school pupils in Grades 2, 4, 6 and 8 (average age: 6, 8, 10, 12 years) are tested in language and arithmetic. In addition, information on the social background is collected and teachers are asked about the behavior of the child in school. Twins were identified by matching on family name, date of birth, school and year of the survey (Webbink et al., 2006). For the analysis we use all twin pairs of which we have information on class sharing. This leaves us with 5756 twin observations and 2878 twin pairs.

In each wave of the study the assignment to classes has been measured. We use two definitions of class separation: at least 1 year separated, and separated in two subsequent waves. For the latter definition it is plausible that the separation had a duration of at least 3 years.

The cognitive measures we use in the analysis are an IQ score and three achievement scores; one for arithmetic, one for language and a nationwide achievement test (CITO). The last three measures are especially designed for the PRIMA survey. The IQ test focuses on nonverbal intelligence and has two components: 'composition of figures' and 'exclusion'. For the first component pupils are required to compose a figure, for instance a square, from several irregular segments. For the exclusion test, pupils are required to choose one figure that does not fit into a sequence of figures. These two nonverbal components have been chosen to measure intelligence unbiased by the socioeconomic background of the pupils. The IQ test is taken in Grade 4, 6 and 8. For each grade there is a different version of the test. For Grade 4 the composition of figures test has 17 items and the exclusion test has 20 items; for Grade 6 and 8 this is respectively 19 and 15 items. The score on the test is the number of items correctly answered. In all waves of the PRIMA survey the same IQ test has been used. Because of the difference in difficulty and number of items, the tests are not comparable between grades. Recently, a more extended test for

Table 4

Effect of at Least One Year in the Same Class (Random Effects Regression on Cognitive Ability in Grade 2, 4, 6 and 8)

	Grade 2	Grade 4	Grade 6	Grade 8
IQ				
Adjusted for				
Year		1.712 (1.079)	1.530 (1.335)	–0.439 (1.637)
Gender/same-sex pair/age		1.478 (1.072)	1.351 (1.315)	–0.685 (1.641)
Social background		0.989 (1.075)	1.027 (1.293)	–1.650 (1.679)
Class performance		1.493 (0.956)	1.014 (1.191)	–1.260 (1.518)
Observations		1393	1167	1035
Same-sex twins		2.169 (1.171)*	0.262 (1.439)	–1.068 (2.081)
Opposite-sex twins		0.494 (1.775)	3.408 (2.262)	–0.788 (2.287)
Language				
Year	0.134 (0.039)***	0.083 (0.044)*	0.068 (0.047)	0.071 (0.065)
Gender/same-sex pair/age	0.130 (0.039)***	0.071 (0.044)	0.052 (0.045)	0.050 (0.063)
Social background	0.064 (0.035)*	0.021 (0.039)	0.034 (0.041)	–0.015 (0.058)
Class performance	0.060 (0.031)**	0.022 (0.034)	0.021 (0.039)	–0.017 (0.053)
Observations	1600	1485	1247	1069
Same-sex twins	0.076 (0.037)**	0.021 (0.041)	-0.009 (0.047)	0.008 (0.072)
Opposite-sex twins	0.016 (0.061)	0.085 (0.068)	0.087 (0.070)	–0.004 (0.083)
Arithmetic				
Year	0.087 (0.035)**	0.017 (0.037)	0.027 (0.033)	0.059 (0.040)
Gender/same-sex pair/age	0.084 (0.035)**	0.020 (0.036)	0.027 (0.032)	0.036 (0.038)
Social background	0.040 (0.033)	–0.017 (0.034)	0.014 (0.031)	0.030 (0.038)
Class performance	0.066 (0.028)**	-0.002 (0.030)	0.015 (0.027)	0.012 (0.032)
Observations	1615	1489	1220	1039
Same-sex twins	0.083 (0.034)**	0.012 (0.037)	0.023 (0.034)	0.043 (0.040)
Opposite-sex twins	0.049 (0.051)	–0.014 (0.006)	0.006 (0.047)	–0.015 (0.058)
CITO test				
Year				1.726 (1.494)
Gender/same-sex pair/age				1.267 (1.437)
Social background				0.595 (1.407)
Class performance				–0.298 (1.210)
Observations				676
Same-sex twins				0.090 (1.573)
Opposite-sex twins				–1.762 (2.177)

Note: Estimates shown are the coefficient of a dummy for having a co-twin in the same class in regression models with a random effect for twin pair and dummies for year of survey. The first model only controls for year of the survey. In the next models potential confounders are introduced. The last model includes all controls. Standard errors in brackets. The IQ test was not taken in Grade 2. The CITO test is only taken in Grade 8. The most extended models are also separately estimated for same-sex pairs and opposite-sex pairs. ***/**/*significant at 1/5/10%-level

nonschool-related cognitive abilities, including the components 'composition of figures' and 'exclusion', has been judged favorably by the Committee On Test Affairs Netherlands (COTAN) of the Dutch professional association of psychologists (NIP). The tests for languages and arithmetic were developed by the CITO group. The language test for children in Grade 2, which is equivalent to infant school, measures the understanding of words and concepts. The arithmetic test for these children focuses on the sorting of objects. These tests can be taken in class. The tests for children in Grades 4, 6 and 8 all come from a system for following pupil achievements in primary education developed by the CITO group. From year to year the tests for the same grade levels are identical. The purpose of this is to compare achievement levels over time. In addition, the scores are comparable between grades. As the scales of the raw scores for language and arithmetic have no clear meaning, we transformed these scores for each test into wave-specific standardized scores, having mean zero and standard deviation one. IQ scores were normalized to a mean of 100 (*SD* 15) per grade. Since the start of the project several tests have been changed. In the PRIMA project these scores have been calibrated to the same scales as the

Table 5

Random Effects Regression of Having a Co-Twin in the Same Class on Cognitive Ability in Grade 2, 4, 6 and 8, Restricted Sample

	Grade 2	Grade 4	Grade 6	Grade 8
IQ		0.981 (1.025)	0.851 (1.257)	–1.086 (1.538)
Observations		1193	977	879
Same-sex twins		1.410 (1.270)	0.124 (1.507)	–0.849 (2.110)
Opposite-sex twins		0.028 (1.886)	3.008 (2.449)	–0.046 (2.326)
Language	0.050 (0.037)	0.023 (0.036)	0.032 (0.041)	–0.009 (0.052)
Observations	1104	1265	1045	907
Same-sex twins	0.059 (0.045)	0.006 (0.043)	–0.013 (0.050)	–0.006 (0.071)
Opposite-sex twins	0.016 (0.077)	0.089 (0.071)	0.115 (0.073)	0.039 (0.080)
Arithmetic	0.091 (0.034)***	-0.019 (0.032)	0.015 (0.028)	0.008 (0.032)
Observations	1120	1266	1020	878
Same-sex twins	0.101 (0.041)**	–0.009 (0.040)	0.024 (0.035)	0.033 (0.040)
Opposite-sex twins	0.100 (0.068)	–0.043 (0.064)	0.009 (0.048)	–0.009 (0.056)
CITO				–0.089 (1.164)
Observations				554
Same-sex twins				0.227 (1.512)
Opposite-sex twins				–1.508 (2.172)

Note: The restricted sample consists of twins in separate classes and twins in schools with only one class per specific grade.

Estimates shown are the coefficient of a dummy for having a co-twin in the same class in regression models with a random effect for twin pair and dummies for year of survey. All controls from the most extended model in Table 4 are included. Standard errors in brackets. The IQ test was not taken in Grade 2. The CITO test is only taken in grade 8. ***/**/*significant at 1/5/10%-level

tests that were previously used and are made comparable. It should be noted that over time comparability is hampered by other relevant differences between waves. In the first wave, tests were taken early in the school year. In the second wave, tests were taken halfway during the school year. In the first two waves tests were taken under the responsibility of an external examiner, while in the third wave the teacher of the class was responsible. As these differences may affect our findings we control for the year of the survey in all regressions.

The CITO test is a nationwide achievement test taken in Grade 8. The CITO consists of 240 multiple-choice items assessing four different intellectual skills: language, arithmetic, information processing and world orientation. Each performance scale contains 60 multiple-choice questions. These performance scales result in a standardized score of between 501 and 550 (Bartels et al., 2002). The CITO test is taken independently of the PRIMA project. Scores were obtained from the schools participating in the PRIMA project.

Statistical Methods

The main question we want to answer is whether having a co-twin in the same class has an impact on the cognitive performance of twins. We used a standard linear mixed model (in econometrics this is called a random effect regression) to estimate the difference in test scores between twins in the same class and twins in different classes, and included a random effect for twin pair. The following relationship has been estimated:

$$Y_{ic} = \alpha + \beta C + \gamma X + \eta_c + \varepsilon_{ic} \quad [1]$$

where Y is the score for cognitive ability of pupil *i* from twin pair *c*, C is a dummy for having a co-twin in the same class, X is a vector of covariates, η_c is random effect for twin pair and ε_{ic} is a random error term. The vector X includes year of the survey, gender, same-sex twin pair, age (at the time of taking the test), education levels and country of birth of the father and mother, and the so-called weight factor of the pupil according to the funding scheme for primary schools. The Dutch funding scheme for primary schools distinguishes several groups of disadvantaged pupils. The most important groups are Dutch pupils with lower educated parents and pupils with an ethnic minority background. Pupils not belonging to a disadvantaged group enter the funding scheme with a weight factor equal to unity. Dutch pupils with lower educated parents have a weight equal to 1.25 and pupils with an ethnic minority background have a weight factor of 1.9. Schools receive respectively 25% and 90% additional funding for these pupils. In addition, we control for the average test score of the classmates of the twins. We include a random effect for twin pair because the twins are not independently drawn. The indicator variable C for sharing the same class is our main variable. The regressions give a comparison of the performance of twins in the same class with twins in different classes. We do not have information on the zygosity of the twins. Hence, we cannot

Table 6

Two Stage Least Squares Regressions on Language and Arithmetic Using Year of the Survey as Instrument for Having a Co-Twin in the Same Class

Language	Grade 2			
	All		Same-sex twin	
	Language	Arithmetic	Language	Arithmetic
First stage				
1996	–0.229 (0.039)***	–0.248 (0.039)***	–0.312 (0.050)***	–0.318 (0.049)***
1998	–0.301 (0.038)***	–0.273 (0.041)***	–0.389 (0.048)***	–0.359 (0.051)***
2000	–0.354 (0.039)***	–0.335 (0.042)***	–0.407 (0.051)***	–0.389 (0.052)***
2002	–0.401 (0.039)***	–0.370 (0.041)***	–0.461 (0.049)***	–0.429 (0.051)***
F-value excluded instrument	31.4 $p = .000$	24.3 $p = .000$	30.3 $p = .000$	23.6 $p = .000$
Second stage				
Co-twin	0.257 (0.105)**	-0.017 (0.107)	0.263 (0.118)**	–0.056 (0.127)
Observations	1600	1615	1082	1101

Note: The first stage regression includes all covariates of the full model from the previous tables. The *F* value of the excluded instrument is a partial *F* test for all year dummies (conditional on the other covariates). The second stage regression includes all previous covariates and the predicted value of having a co-twin in the same class. Standard errors adjusted for clustering in twin pairs. ***/**/*significant at 1/5/10%-level.

distinguish between MZ and DZ pairs. We perform separate analysis for all twins, same-sex pairs and opposite-sex pairs.

A concern with the estimates of the random effect models is that the assignment of twins is not random and that unobserved factors that determine this decision are related with the performance of twins, cov $(C, \varepsilon_{ic}) \neq 0$. Such a covariance would violate a basic assumption for getting consistent estimates in equation [1]. For instance, if strongly motivated parents have a preference for the separation (or for the nonseparation) of their twins, this might bias the results in the case that these parents invest more time in helping and supporting their children at home, or if these parents invest more in selecting a high quality school. In that case, these unobserved factors will bias the results. It should be noted that it is not clear which alternative (separation or sharing) improves the cognitive performance of twins. Hence, it not clear how this bias might affect the estimate. We follow three approaches to reduce the bias by unobserved factors related to the assignment.

The first approach is based on restricted schools: schools with only one class per specific grade. In these schools parents do not have the opportunity to choose between the two alternatives. In that case, parents with a preference for separating their twins would have to assign their children to the same class. Hence, it seems likely that restricted schools contain a combination of parents with a preference for separation and parents with a preference for nonseparation. We repeat the estimations on a smaller sample that only includes twins in separate classes and twins in restricted schools. By reducing the sample we increase the share of parents with a preference for separation in the group of twins in the same class. It could be argued that parents that really want to separate their twins in school will choose large schools. This might be true but depends on the availability of large schools near the location of the parents. In addition, the number of classes at a specific grade level in school depends on enrolment. Variation in enrolment may induce variation in the number of classes, which may lead to unexpected restrictions for parents.

Second, we use variation in the assignment decisions between years in instrumental variable regressions. Table 1 shows that there is a clear time trend towards separation of twins, which is also observed within different grade levels. Whereas in 1994 approximately 93% of the twin pairs shared the same class, this drops linearly to 66% in 2002. This trend towards separation of twins is also noted by Van Leeuwen et al. (2005). To our knowledge there has not been a general policy change which might explain the time trend. If the time trend is not related to the unobserved factors of cognitive achievement we can use this trend as an instrument for class separation. In that case, the variation in classroom assignment induced by the time trend will not be correlated with the unobserved factors of cognitive achievement. We can isolate this variation in classroom assignment with two-stage least squares. In the first stage, classroom assignment is regressed on the time trend and the other covariates

$$C = \alpha + \beta T + \gamma X + \nu_{ic} \qquad [2]$$

where T is a vector of time dummies. In the second stage, the cognitive achievement is regressed on the predicted classroom assignment (based on the time trend) and the other covariates (see vector X in equation 1). This gives the instrumental variable estimator conditional on the covariates: $\beta_{IV} = \text{cov}(Y_{ic}, T)/\text{cov}(C, T)$.

Third, we exploit the longitudinal character of the data. We estimate models that control for the test score in the previous grade and analyze gains in test scores between grades.

Table 7

Random Effects Regression of Having a Co-Twin in the Same Class in Two Subsequent Waves on Cognitive Ability in Grade 2, 4, 6 and 8, Controlling for Previous Test Score

Regressions controlling for test scores in previous grade			
	Grade 4	Grade 6	Grade 8
Language test	0.048 (0.074)	0.010 (0.064)	–0.112 (0.068)*
N	426	518	475
Same-sex twins			–0.068 (0.091)
Opposite-sex twins			–0.212 (0.110)*
Arithmetic test	0.058 (0.055)	0.025 (0.039)	–0.000 (0.038)
N	444	502	448
Same-sex twins			
Opposite-sex twins			
Regressions on gains in test scores between grades			
	From Grade 2 to Grade 4	From Grade 4 to Grade 6	From Grade 6 to Grade 8
Language	0.053 (0.093)	0.023 (0.078)	–0.145 (0.075)*
N	426	518	475
Same-sex twins			–0.055 (0.102)
Opposite-sex twins			–0.248 (0.116)**
Arithmetic	0.033 (0.073)	0.058 (0.054)	–0.024 (0.043)
N	444	502	448

Note: Controls used are year, date of birth, gender, weight factor for financing of schools, education and country of birth father and mother, same-sex pair, average class performance in highest grade, random effect for twin pair, standard errors in brackets. ***/**/*significant at 1/5/10%-level

Results

Table 2 shows sample statistics for the dependent and explanatory variables in all four grades for separated twins and twins sharing the same class. The statistics are based on the cross-sectional data. Hence, separation or sharing the same class has a duration of at least 1 year.

We observe that separated twins score on average lower in language and arithmetic. The socioeconomic background, indicated by country of birth and education of parents and the weight factor for the financing of schools, of separated and nonseparated twins differs. Separated twins more often have parents born in a foreign country. This is also indicated by the higher proportion of 1.9 pupils (ethnic minorities) in the group of separated twins. Hence, Dutch parents seem to prefer their twins to share the same class. We also observe that the average test scores of the classmates of the separated twins are lower than the test scores of the class mates of the nonseparated twins. This is probably related to the fact that separated twins more often have parents born in a foreign country, and the fact that the free school choice in Dutch education leads to segregation based on socioeconomic factors.

To further investigate differences between separated and nonseparated twins we did a probit regression on 'having a co-twin in your class'. Table 3 shows the estimation results. As in Van Leeuwen et al. (2005), we find that the rate of separation increases after 1994. Hence, there is a time trend towards separation. We also observe that in higher grades it is less likely that twins are separated.

The results of the random effect regressions of the effect of at least 1-year separation are shown in Table 4. For the IQ test none of the estimates are statistically significant. For the language test we find that nonseparated twins have higher scores in Grade 2. The difference between separated and nonseparated twins reduces after adding controls. Although several models for the higher grades give statistically significant results, for the full model we only find a statistically significant difference in Grade 2. In the separate estimate for same-sex pairs we find a larger difference. The results for the arithmetic test show the same pattern. Twins sharing the same class have higher scores in Grade 2 and this difference is larger for same-sex pairs. In fact, for opposite-sex pairs we find no significant difference. For the higher grades we find no difference. We also find no difference for the CITO test which is only taken in Grade 8.

The estimates in Table 4 might be biased by unobserved factors related to the decision of parents and teachers about the separation of twins. To reduce this bias we focus on 'restricted schools' that have only one class per grade level. Table 5 shows the estimation results for the model using all controls. The results are quite similar to those in Table 4. Nonseparated twins have higher scores in language and arithmetic in Grade 2, and this especially holds for same-sex pairs. We do not find significant differences in higher grades.

The second approach is based on the variation in classroom assignment between years (Table 6). The first stage results show that the year of the survey has a clear effect on the outcome of the separation decision.

The *F* value for the excluded instruments clearly exceeds the critical value of 10 for weak instruments (Staiger & Stock, 1997). The second stage results show that nonseparated twins score significantly higher on the language test in Grade 2. For same-sex pairs we find the same results. For arithmetic we find no difference. In addition, for all higher grades we find no significant differences in test scores.

Our third approach to reduce the bias by unobserved factors exploits the longitudinal character of the data. We estimated linear mixed regressions on the language and arithmetic score, controlling for the test score in the previous survey. In addition, we estimated regression models on the gain in these test scores between grades. In these analyses we use the second definition of sharing the same class. Hence, we compare twins that were separated in two subsequent waves with twins that were sharing classes in two subsequent waves. The controls for the test scores of the classmates refer to the scores in the highest grade. We assume that the impact of the school environment in the previous grades is already picked up by the previous test score. We find that nonseparated twins have a lower score in languages in Grade 8 (Table 7). The separate analysis for same-sex and opposite-sex pairs shows that this finding comes from the opposite-sex pairs. The regressions on the gains in test scores yield similar results. Within the sample of opposite-sex pairs the effect of classroom separation is only found for boys, not for girls. The estimate in the regression on the gain in test scores for boys from opposite-sex pairs is –0.342 (standard error 0.174). For all other grade levels we find no significant effects in the separate samples of same-sex and opposite-sex pairs.

Conclusions and Discussion

The empirical findings in this article do not provide a clear answer for the classic question about class separation of twins. In the analysis of the effect of a separation of at least 1 year we only find an effect on cognitive ability in Grade 2. All three approaches show that nonseparated twins score higher on language. The difference is larger for same-sex pairs. In addition, there is some evidence for higher scores in arithmetic in Grade 2. For the higher grades, we find no effect of classroom separation on cognitive ability. In the analysis of the effect of a separation of at least 3 years, we find that separation increases language performance between Grade 6 and 8 for opposite-sex pairs. Hence, the presence of a co-twin seems to matter at the early stages of primary education for same-sex pairs. At the later stages of primary education we hardly find any differences. Only between Grade 6 and Grade 8 we find a difference in language skills for opposite-sex pairs.

As mentioned previously, the Dutch society for Parents of Multiples advises parents to follow their own opinion. Although we find some statistically significant effects of classroom separation, especially in Grade 2, we think that the findings in this article do not support a clear preference for separation or nonseparation. Hence, we think that our findings show that the advice of the Dutch Society for Parents and Multiples is not detrimental with respect to the cognitive ability of twins.

A limitation of our findings is that we cannot distinguish between DZ and MZ twins among the same-sex pairs. It is not clear how this would change our findings. Van Leeuwen et al. (2005) found no difference between DZ and MZ twins, whereas Tully et al. (2004) found that MZ twins suffered more from separation than DZ twins. In addition, our approach to reduce the bias by unobserved factors does not take into account the possibility that parents may select the school for their children on the basis of their prior belief about whether to separate or not, or on the school policy regarding twin separation.

Acknowledgments

PMV acknowledges the NHMRC for funding (grant 442915).

References

Bartels, M., Rietveld, M. J., Van Baal, G.C., & Boomsma, D.I. (2002). Heritability of educational achievements in 12-year-olds and the overlap with cognitive ability. *Twin Research, 5,* 544–553.

Driessen, G., Van Langen, A., & Vierke, H. (2004). *Basisrapportage PRIMA-cohortonderzoek, Vijfde meting 2002–2003* [Report on PRIMA-longitudinal research project, Survey 2002–2003]. Nijmegen: ITS.

Driessen, G., Van Langen, A., & Oudenhoven, D. (1994). *De toetsen voor het cohort Primair onderwijs, verantwoording* [The tests for the Cohort Primary Education, justification]. Nijmegen: ITS.

Geluk, A., & Hol, J. (2001). Samen of apart? Meerlingkinderen naar school, peuterspeelzaal of kinderopvang [Together or Apart? Multiples to school, playground or daycare; Brochure]. Bergen, the Netherlands: NVOM (Dutch Society for Parents of Multiples).

Staiger, D., & Stock, J. (1997). Instrumental variables regressions with weak instruments. *Econometrica, 65,* 557–586.

Tully, L. A., Moffit, T. E., Caspi, A., Taylor, A., Kierman, H., & Andreou, P. (2004). What effect does classroom separation have on twins' behavior, progress at school, and reading abilities? *Twin Research, 7,* 115–124.

Van Leeuwen, M., Van den Berg, S. M., Van Beijsterveldt, C. E. M., & Boomsma, D. I. (2005). Effects of twin separation in primary school. *Twin Research and Human Genetics, 8,* 384–391.

Webbink, D., Roeleveld, J., & Visscher, P.M. (2006). Identification of twin pairs from large population-based samples. *Twin Research and Human Genetics, 9,* 496–500.