Learning and Individual Differences 47 (2016) 156–166

Contents lists available at ScienceDirect

Learning and Individual Differences

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

Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children

Eveline L. de Zeeuw [a,b,*], Catharina E.M. van Beijsterveldt [a,b], Tina J. Glasner [a], Eco J.C. de Geus [a,b], Dorret I. Boomsma [a,b]

[a] Department of Biological Psychology, Vrije Universiteit, Amsterdam, The Netherlands
[b] EMGO + Institute for Health and Care Research, VU Medical Centre, Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 4 June 2015
Received in revised form 22 January 2016
Accepted 24 January 2016

Keywords:
Educational achievement
Heritability
Twins
Primary education

ABSTRACT

Even children attending the same primary school and taught by the same teacher differ greatly in their performance. In the Netherlands, performance at the end of primary school determines the enrollment in a particular level of secondary education. Identifying the impact of genes and the environment on individual differences in educational achievement between children is important. The Netherlands Twin Register has collected data on scores of tests used in primary school (ages 6 to 12) to monitor a child's educational progress in four domains, i.e. arithmetic, word reading, reading comprehension and spelling (1058 MZ and 1734 DZ twin pairs), and of a final test (2451 MZ and 4569 DZ twin pairs) in a large Dutch cohort. In general, individual differences in educational achievement were to a large extent due to genes and the influence of the family environment was negligible. Moreover, there is no evidence for gender differences in the underlying etiology.

1. Introduction

Children differ in their ability to learn the subject material that is taught at school. Some master basic skills, such as reading and arithmetic, and pick up knowledge about science, history and biology much faster than their peers. Low educational achievement is associated with continued low achievement, school dropout and delinquency (Moilanen, Shaw, & Maxwell, 2010). General cognitive ability is the most important predictor of educational achievement (Deary, Strand, Smith, & Fernandes, 2007) and explains about half of the variation (Frey & Detterman, 2004). Most research on educational achievement of children has focused on environmental factors, such as socioeconomic status (SES) of the parents and school characteristics, and on differences between groups of children, for example boys and girls (OECD, 2010). However, even children attending the same school and taught by the same teacher differ greatly in their performance at school. It may be less relevant to look at group differences when differences within a group are much larger. Causes for individual differences between children do not necessarily have to be the same as for average differences between groups. The main reason for mean differences in educational achievement between boys and girls might be environmental whereas the cause for differences in performance between individual children may be largely genetic in nature.

Genetic research can address questions about the causes of individual differences amongst children and disentangle the underlying etiology, i.e. the extent to which the differences in educational achievement between children are explained by their genes or by the environment (Boomsma, 2013; Plomin, Defries, McClearn, & McGuffin, 2008). One of the most often used designs in behavior genetics is the twin study, which is based on the difference in genetic relatedness between identical or monozygotic (MZ) and fraternal or dizygotic (DZ) twins. Monozygotic (MZ) twin pairs are genetically (nearly) identical while dizygotic (DZ) twin pairs share approximately 50% of their segregating genes. If the larger genetic resemblance of MZ twin pairs is mirrored in a larger resemblance for a phenotype the phenotype is being influenced by genetic effects. Genetic effects are the sum of the effects of all additive genetic variants with an influence on educational achievement. When the correlation between MZ twin pairs is higher than between DZ twin pairs, this constitutes evidence for the influence of genetic effects. Environmental effects often are distinguished into common or shared environmental and unique or non-shared environmental effects. Common environmental effects are influences that are shared between twins or siblings who grow up in the same family and enhance their similarity beyond the similarity due to shared genes. There are other effects that also make offspring from the same parents more similar, including the effects of assortative mating, the similarity between spouses, which in the classical twin design will also be detected as common environmental effects (Evans & Martin, 2000). When the correlation between DZ twin pairs is more than half the correlation between MZ twin pairs there is an indication for the influence of the common environment. Unique environmental effects are influences that are not

* Corresponding author at: Department of Biological Psychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.
E-mail address: el.de.zeeuw@vu.nl (E.L. de Zeeuw).

http://dx.doi.org/10.1016/j.lindif.2016.01.009
1041-6080/

shared between twins, and make children less similar. When the correlation between MZ twin pairs is not equal to unity the unique environment has an influence. The unique environmental effects also include measurement error.

Numerous studies have demonstrated that genetic effects have a substantial influence on differences between children in general cognitive ability. General cognitive ability is often seen as an aptitude while reading, mathematics and spelling are taught at school and perceived as the outcome of education. Hence, it seems reasonable to expect that heritability of educational achievement is lower than the heritability of general cognitive ability. However, a recent study has shown that the opposite was true for primary school children in the United Kingdom (UK). Literacy and numeracy were significantly more heritable than general cognitive ability at ages 7 and 9, but no longer at age 12 (Kovas et al., 2013). Heritability of general cognitive ability also increases in the Netherlands from low (20%) to moderate (50%) (Bartels, Rietveld, van Baal, & Boomsma, 2002) during primary school, to high (65%) in adolescence (van Soelen et al., 2011) while heritability of educational achievement is already high at the start of primary school (70%) (Calvin et al., 2012) and remains the same until the end of primary school (60%) (Bartels et al., 2002). It has been proposed that the equal opportunities in the relatively homogenous education environment provided in Western societies acts to reduce environmental variation, making differences in educational achievement between children to a greater extent due to genetic differences (Kovas et al., 2013).

Twin studies have mainly focused on the educational domain of reading and, more recently, mathematics, while less is known about the etiology of other educational domains, such as, science. Most studies have used teacher assessments or tests that had been administered by the researchers through the internet, telephone or during a home-visit while only a few used standardized tests administered at school. Even though teacher assessments are correlated with standardized tests they might be less reliable measures of educational achievement. Furthermore, teacher ratings may be subject to rater bias when twins are taught in the same classroom and their educational achievement is rated by the same person. Each teacher has his or her own perception on educational achievement, which can make children seem more similar when they are assessed by the same teacher (Kan, van Beijsterveldt, Bartels, & Boomsma, 2014; Simonoff et al., 1998).

Most twin studies on educational achievement have primarily included English speaking children from the USA and the UK. Studies from other countries with different educational systems and languages are scarce (Byrne et al., 2009; Chow, Ho, Wong, Waye, & Bishop, 2011; Samuelsson et al., 2008). The question addressed in this study is whether the same pattern of the relative contribution of genetic and environmental effects on the variation in educational achievement exists in the Netherlands. Calvin et al. (2012) found genetic effects to be an important cause of variation in achievement in the educational domains language (43–74%) and arithmetic (36–73%) at ages 8, 10 and 12 in Dutch primary school children. However, they used a population cohort without information on zygosity and estimated the resemblance between monozygotic and dizygotic twin pairs from the proportion of same-sex and opposite-sex twin pairs, which is a much less powerful design than when zygosity is known.

In the Dutch educational system, the majority of primary schools use a pupil monitoring system that includes standardized tests assessing educational achievement (Cito, 2014a; Vlug, 1997). Tests are available for all grades and all important educational domains. The tests are independent of teaching methods and can be used to monitor a child's educational development in comparison to peers and across grades and educational domains. Tests measuring arithmetic, word reading, reading comprehension and spelling are, according to an inventory amongst teachers, the most informative with regard to the educational development of children (Polderman et al., 2011). A standardized nation-wide educational achievement test is available for the last grade, measuring what a child has learned during all primary school years (Cito, 2002). Together, these data provide a unique opportunity to give an overview of the underlying etiology of differences in educational achievement on different educational domains across primary school grades.

Differences in average educational achievement between boys and girls could be due to differences in the etiology of educational achievement. Quantitative gender differences are present if one gender is affected to a greater extent by the same genetic or environmental effects. Qualitative gender differences exist when different genetic or environmental effects have an influence on boys and girls. Very few studies have examined differences between boys and girls in the etiology of educational achievement, probably due to small sample sizes. One large study utilizing teacher ratings to determine the level of educational achievement in language, reading, mathematics and science found no gender differences in the causes underlying differences between children (Kovas, Haworth, Dale, & Plomin, 2007). Gender differences were also absent in another study using a standardized test to measure the reading level of children (Harlaar, Hayiou-Thomas, & Plomin, 2005).

The aim of the current study is to identify the impact of genes and the environment on educational achievement, i.e. arithmetic, word reading, reading comprehension and spelling, in primary school in a large cohort of Dutch twins and to explore possible gender differences.

2. Methods

2.1. Participants

The Netherlands Twin Register (NTR) was established around 1987 by the Department of Biological Psychology at the Vrije Universiteit Amsterdam and registers approximately 40% of all multiple births in the Netherlands. The parents of twins participate in longitudinal data collection and receive a survey about the development of their children every 2 to 3 years up until the twins are 12 years old. Details about selection and response rates are described elsewhere (Bartels et al., 2007; Boomsma et al., 2002; Boomsma et al., 2006; van Beijsterveldt et al., 2013). In 1999 the NTR started asking parents for their consent to approach the teacher(s) of their children attending primary school. Each year the parents of twins aged approximately 7, 9 and 12 years have been approached with this request. After parental consent, a survey is sent to the primary school teacher(s) of the twins and their siblings with questions about behavioral and emotional problems, functioning at school and educational achievement. In the Netherlands, a widespread pupil monitoring system is developed with standardized tests to monitor a child's educational progress in primary school ('LeerlingVolgSysteem' in Dutch) (Cito, 2014a). Since 2008 teachers are systematically asked to submit copies (usually a print-out) of the results on the pupil monitoring tests administered during the primary school period. Teachers were asked to provide information on the results of the tests administered in the current and previous grades. Approximately one third of the teachers who returned the teacher survey also send in a student report with the pupil monitoring test results. A final educational achievement (EA) test is administered at the end of the last grade of primary school (Cito, 2002). Because the results of the EA test become available at the end of the last school year, both the parents as well as the children themselves instead of the teachers are asked to report the scores on this EA test. Results on the EA test at the end of the primary school period have already been collected since 2000.

Data on at least one of the pupil monitoring tests (cohorts 1995–2006) or on the EA test (cohorts 1987–2001) were available for 16,234 children. We excluded children who had a disease or handicap that interfered severely with daily functioning ($N = 90$) or attended specialized education ($N = 79$), in the Dutch education system special schools are available for children who need extra care due to learning problems, physical and/or mental disabilities or a behavioral disorder, resulting in data for 7228 complete and 1609 incomplete twin pairs. The main reason for incomplete data is that twins attend different classes or schools and only one of the teachers returned the results on the pupil monitoring tests. Approximately 45% of the twin pairs attend different classes in a given grade of primary school, with only a very small proportion of these twin pairs going to different schools.

This study included data of 2818 twin pairs of opposite sex. For same-sex twin pairs, the determination of zygosity status was based on blood group or DNA polymorphisms (N = 1363) or on parental report of items on resemblance in appearance and confusion of the twins by parents and others (N = 4586). This last method can establish zygosity with an accuracy of approximately 93% (Rietveld et al., 2000). Twin pairs for which zygosity was unavailable were excluded from the analyses (N = 70). Results for at least one of the pupil monitoring tests were available for approximately one third of these twin pairs (496 MZm, 456 DZm, 541 MZf, 418 DZf and 825 DOS) and data on the EA test were available for the majority of the twin pairs (1113 MZm, 1132 DZm, 1338 MZf, 1149 DZf and 2288 DOS).

2.2. Measurements

The pupil monitoring system consists of age appropriate tests, independent of teaching methods, which are administered at three fixed time points (i.e. beginning, half way and end of a school year) to assess the educational achievement of a child in all important educational domains (Cito, 2014a). Tests with increasing difficulty are available for both the different time points in each grade and the different grades of primary school (grade 1: 6–7 years; grade 2: 7–8 years; grade 3: 8–9 years; 4: 9–10 years; grade 5: 10–11 years; grade 6: 11–12 years). The results on the tests are entered into a database which enables teachers to compile student reports with the results of the different tests across all time points and grades. In this study the results on the tests administered halfway through the school year have been used in the analyses.

Each test consists of a number of items that can provide a correct response and the number of correct responses is converted in an ability score. Conversion of test scores is done based on item-response theory (IRT) to ensure that the development in ability scores is on a single scale (Vlug, 1997). This results in ability scores that can be compared within and between grades to monitor a child's development over time and in comparison to its peers. Some of the pupil monitoring tests have been updated with new questions which resulted in different versions of the tests administered across children. An inventory amongst teachers indicated that the tests measuring arithmetic, reading, reading comprehension and spelling were the most informative with regard to a child's educational achievement (Polderman et al., 2011).

2.2.1. Arithmetic

The arithmetic test (grades 1 to 6) consists of one part in which children have to solve simple math problems within a short time period and a second part with more complex math problems without a time limit. The test assesses general knowledge of mathematics and arithmetic and comprises written computational problems of addition, subtraction, multiplication and division and problems on the notion of measurements, time and money, and knowledge about fractions, ratios and percentages.

2.2.2. Word Reading

The word reading test (grades 1 to 6) measures word decoding skills by counting the total number of individual words a child can correctly read aloud in 1 min. The test consists of three levels of increasing difficulty and complexity. The first level includes words that are pronounced exactly as they are spelled, the second level includes also other monosyllabic words and the third level includes two or more syllabic words. This study used the most difficult level of the test which is in general not yet administered in the first grade.

2.2.3. Reading comprehension

The reading comprehension test (grades 3 to 6) includes a large variety of different text types and genres with two different types of multiple choice questions. The test consists of a first part in which a child has to read a number of short texts and answer questions related to the text and a second part with parts of the text left blank that need to be filled out. The test tries to assess different components of reading processing by questions regarding both the facts and events described in the texts as well as by questions about the purpose of the writer and the intended readership of the texts.

2.2.4. Spelling

The spelling test (grades 1 to 6) measures both active, writing down the words, and passive, recognizing spelling errors, spelling. Active spelling is measured with a dictation by the teacher where a sentence is read aloud and a child has to write down a specific word from this sentence. Passive spelling is measured with multiple choice questions were a student has to select the answer in which the word that is printed in bold is spelled incorrectly.

2.2.5. Educational achievement

At the end of primary school a nationwide educational achievement (EA) test is administered over a period of 3 days (Cito, 2002). This final test measures what a child has learned in 6 years of primary education. The results of this test determine, along with the advice of the teacher, which level of secondary education is suitable for a child. The EA test consists of multiple choice items in four different educational domains, namely Arithmetic, Language, Study Skills and Science and Social Studies. The first three test scales are combined into a Total Score, which is standardized on a scale from 501 to 550. The Arithmetic scale includes items on numbers and operations, ratios, fractions and percentages, and measurements, geometry, time and money. The Language scale includes items on writing, spelling, reading comprehension and vocabulary. The Study Skills scale includes items on handling of study texts, handling of information, reading diagrams, tables and graphs and map reading. The Science and Social Studies scale includes items on geography, history and biology. All scores on the scales are standardized to percentile scores to correct for differences in the number of items across the years. The internal reliability of the EA test is good ($\alpha = .95$) as is the simulated test–retest reliability ($r = .96$) (Herman, Ronald, & Anja, 2011). The score on the test has been shown to correlate moderately with general cognitive ability ($r = .63$) (Bartels et al., 2002).

2.3. Statistical analyses

Twin correlations, separately for gender, were estimated using maximum likelihood estimation in the R (R Core Team, 2014) package OpenMx Version 2.3.1 (Boker et al., 2012; Boker et al., 2011). A model that freely estimated all parameters, i.e. means, variances and covariances, separately for the different zygosity-by-gender groups (MZm, DZm, MZf, DZf and DOS), was fitted to the data (saturated model). In this model, it was tested whether constraining the means and variances to be equal across gender led to a significant deterioration of the model fit (Purcell, 2002). Questions have been updated for the tests for arithmetic, reading comprehension and spelling, resulting in different versions of the tests administered across children. To correct for differences across versions, means and variances were estimated separately for the two versions.

To gain further insight into the causes of individual differences in educational achievement of children in primary school, univariate genetic models were fitted to the data for each educational domain and grade. The variation in educational achievement was decomposed into variance due to additive genetic effects (A), to shared or common environmental effects (C) and to non-shared unique environmental effects (E) (Posthuma et al., 2003). Additive genetic effects are the sum of the additive effects of all genetic variants influencing educational

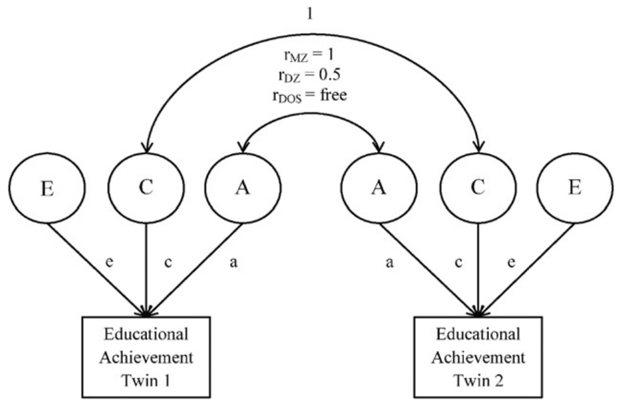

Fig. 1. Path diagram representing the twin model.

achievement. Shared or common environmental variance results from environmental effects that are shared by both members of a twin pair. Non-shared or unique environmental variance results from environmental effects that are not shared by a twin pair and also includes measurement error. The variance components A, C and E were estimated separately for boys and girls. The variance components are expected to correlate differently for MZ and DZ twin pairs due to the difference in genetic resemblance (Fig. 1). Since MZ twin pairs share (nearly) all their genes the correlation between the genetic effects of MZ twin pairs is fixed to 1.0. DZ twin pairs share approximately 50% of their segregating genes and therefore the correlation between the genetic effects of DZ twin pairs is fixed to 0.5. The correlation between the genetic effects is estimated freely for DOS twin pairs as different genetic effects could have an influence on educational achievement in boys and girls. For both MZ, DZ and DOS twin pairs the correlation between the common environmental effects is fixed to 1.0.

In the subsequent models, the influence of the gender of the student on the variance components was tested in two ways. First, we tested whether the same genetic effects are important in boys and girls (qualitative gender differences) by fixing the correlation between the genetic effects of DOS twin pairs to be equal to the correlation for DZ twin pairs. Qualitative gender differences will result in a lower genetic correlation between DOS twin pairs. Second, we tested whether the genetic effects had an influence to the same extent in boys and girls (quantitative gender differences) by fitting a model, which incorporates total variance differences, but does not allow the relative contribution of the variance components to be different between boys and girls. Quantitative gender differences will result in unequal variance components between boys and girls. Finally, the significance of the common environmental effects was tested by dropping them from the model. The difference in model fit between the nested models was assessed with a log-likelihood ratio test (LRT) which calculates the difference in −2log-likelihood (−2LL) between the models and evaluates the chi square-statistic using the difference in degrees of freedom between the models. A p-value smaller than 0.01 was considered significant. Constraints were kept, when a more restrictive model did not significantly decrease the goodness of fit, as a more parsimonious model is preferred.

3. Results

Table 1 gives the means and standard deviations of scores on the pupil monitoring tests across all grades and the educational

Table 1
Estimated means and standard deviations for the pupil monitoring tests on arithmetic, reading, reading comprehension, spelling across grades and the educational achievement test administered in the last grade of primary school.

	Version 1				Version 2					
	Boys		Girls		Boys		Girls			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	X^2 (df)	p
Arithmetic										
Grade 1 (ages 6–7)	429	53.7 (12.6)	418	53.0 (12.7)	1097	37.6 (15.4)	1093	34.3 (14.4)	26.2 (2)	<.001
Grade 2 (ages 7–8)	491	74.4 (11.5)	535	69.8 (11.1)	857	57.5 (14.5)	854	51.7 (13.3)	99.4 (2)	<.001
Grade 3 (ages 8–9)	536	90.7 (10.9)	555	86.3 (11.1)	800	79.1 (11.8)	805	73.0 (14.3)	113.1 (2)	<.001
Grade 4 (ages 9–10)	560	100.0 (9.0)	593	96.4 (9.7)	439	91.3 (11.5)	442	85.8 (11.6)	77.9 (2)	<.001
Grade 5 (ages 10–11)	515	111.5 (8.0)	530	108.3 (8.6)	434	104.4 (10.4)	410	100.3 (11.2)	62.0 (2)	<.001
Grade 6 (ages 11–12)	355	123.1 (7.8)	399	120.5 (7.9)	116	117.9 (11.4)	111	112.7 (9.6)	28.6 (2)	<.001
Reading										
Grade 2 (ages 7–8)	852	41.5 (18.9)	883	42.5 (19.1)					1.1 (1)	.292
Grade 3 (ages 8–9)	900	63.0 (18.3)	910	63.3 (18.9)					0.1 (1)	.733
Grade 4 (ages 9–10)	683	74.1 (17.4)	693	74.3 (16.9)					0.0 (1)	.853
Grade 5 (ages 10–11)	598	84.0 (16.6)	605	85.6 (15.5)					2.4 (1)	.119
Reading comprehension										
Grade 3 (ages 8–9)	693	30.4 (14.2)	717	33.6 (13.5)	704	29.7 (14.0)	690	31.8 (13.3)	23.4 (2)	<.001
Grade 4 (ages 9–10)	751	41.5 (14.1)	785	42.3 (13.7)	325	36.8 (13.3)	348	36.4 (12.2)	1.0 (2)	.593
Grade 5 (ages 10–11)	721	50.2 (14.8)	728	52.5 (14.9)	327	49.2 (13.6)	346	49.3 (15.7)	8.2 (2)	.017
Grade 6 (ages 11–12)	554	61.5 (14.9)	586	64.7 (15.0)	164	64.5 (17.2)	186	63.8 (17.7)	11.1 (2)	.004
Spelling										
Grade 1 (ages 6–7)	693	109.2 (8.8)	634	110.4 (8.7)	889	108.9 (7.4)	961	109.7 (7.1)	11.8 (2)	.003
Grade 2 (ages 7–8)	764	119.0 (7.5)	750	120.2 (7.3)	690	120.3 (6.8)	699	121.5 (6.6)	20.4 (2)	<.001
Grade 3 (ages 8–9)	738	127.0 (7.0)	751	128.0 (6.5)	674	128.0 (7.1)	672	129.2 (7.0)	15.2 (2)	<.001
Grade 4 (ages 9–10)	775	136.0 (7.2)	813	136.8 (7.1)	304	134.3 (7.6)	316	135.6 (7.2)	8.3 (2)	.016
Grade 5 (ages 10–11)	723	142.1 (6.5)	761	143.1 (6.5)	302	139.6 (6.3)	313	140.4 (5.8)	9.3 (2)	.010
Grade 6 (ages 11–12)	561	149.9 (5.9)	587	150.0 (5.9)	143	145.7 (9.1)	155	145.5 (8.0)	0.1 (2)	.957
Educational achievement										
Arithmetic	4330	64.8 (26.2)	4678	53.0 (28.2)					382.0 (1)	<.001
Language	4327	57.1 (27.5)	4681	60.7 (27.0)					39.6 (1)	<.001
Study skills	4322	62.1 (27.2)	4677	59.7 (28.0)					16.1 (1)	<.001
Science and social studies	3829	65.8 (25.9)	4133	51.5 (27.4)					537.4 (1)	<.001
Total score	6137	538.2 (8.5)	6794	537.1 (8.7)					48.4 (1)	<.001

achievement test, for boys and girls separately and for the different versions of the tests. There were significant gender differences for arithmetic in all grades and for reading comprehension and spelling in some grades. Boys were on average better at arithmetic and girls performed on average better on the reading comprehension and spelling tests. On average there were no gender differences for the word reading tests. Gender differences were also present for the total score on the educational achievement test and for all subscales. Boys scored on average higher on arithmetic, study skills and science and social studies while girls obtained on average better results for language.

Twin correlations were estimated in the five zygosity-by-gender groups (Table 2). All MZ correlations were higher than DZ correlations, suggesting additive genetic effects. Sometimes DZ correlations were larger than half the MZ correlations, suggesting common environmental effects. The genetic model fitting results with the standardized estimates and their 95% confidence intervals are reported for arithmetic, word reading, reading comprehension and spelling across grades and an educational achievement test administered in the last grade of primary school (Table 3). In bold the results for the most parsimonious model that did not significantly deteriorate the goodness of fit. The full models showed small differences between boys and girls in the estimates of the variance components, but these were, in general, not significant. The relative contribution of the additive genetic, common environmental and unique environmental effects are displayed for the models estimating all variance components equal for boys and girls while allowing total variance differences (Fig. 2).

Genetic effects were the most important contributor to individual differences in educational achievement in arithmetic (60–74%), word reading (72–82%) and reading comprehension (54–64%) and for most grades in spelling (33–70%). Common environmental effects had a negligible influence on arithmetic (0–8%), word reading (0–7%) and reading comprehension (1–12%) and a slightly larger influence on spelling (0–29%). Unique environmental effects explained the remaining variance in arithmetic (26–34%), word reading (11–29%), reading comprehension (32–35%) and spelling (30–39%). Genes were also the largest contributor to the variation in the educational achievement test (74%). The heritability differed somewhat between the educational domains measured with this test, i.e. arithmetic (68%), language (67%), study skills (60%) and science and social studies (56%). The common environmental effects were also small for the total score (8%), arithmetic (5%), language (10%), study skills (6%) and science and social studies (21%). Unique environmental effects explained the remaining variance (18–34%).

4. Discussion

The current study presents the etiology of achievement in several educational domains across primary school grades 1 to 6, approximately corresponding to ages 6 to 12, in a large, representative cohort from The Netherlands. The influence of the genetic and environmental effects was systematically examined for the educational domains of arithmetic, word reading, reading comprehension and spelling. The extent to which genes influenced differences in educational achievement was large and relatively stable across all grades for arithmetic (60–74%), word reading (72–82%) and reading comprehension (54–64%). Similar heritability estimates were obtained at all ages despite large differences in content across grades. The influence of the common environment was negligible for these educational domains. The heritability of spelling was somewhat small in the first grade (33%), but increased at later ages (58–70%). In contrast to later grades, the common environment had an influence on spelling in the first grade of primary school. Studies in preschool children report a much larger influence of the common environment, shared by all children in a family, on, for example, reading (Byrne et al., 2009; Oliver, Dale &

Table 2
Twin correlations for the pupil monitoring tests on arithmetic, reading, reading comprehension, spelling and the educational achievement test.

	N	MZm	N	DZm	N	MZf	N	DZf	N	DOS
Arithmetic										
Grade 1 (ages 6–7)	341	.664	292	.449	348	.721	271	.326	568	.386
Grade 2 (ages 7–8)	302	.667	271	.428	326	.660	253	.320	510	.318
Grade 3 (ages 8–9)	301	.659	269	.397	319	.733	248	.272	509	.332
Grade 4 (ages 9–10)	224	.704	201	.447	243	.690	194	.405	378	.296
Grade 5 (ages 10–11)	212	.727	190	.320	215	.696	176	.396	356	.399
Grade 6 (ages 11–12)	119	.645	93	.544	124	.735	96	.388	179	.294
Reading										
Grade 2 (ages 7–8)	194	.822	175	.434	215	.864	161	.459	328	.488
Grade 3 (ages 8–9)	199	.733	182	.382	215	.748	174	.475	369	.386
Grade 4 (ages 9–10)	147	.774	150	.520	156	.789	136	.298	276	.327
Grade 5 (ages 10–11)	145	.849	125	.545	144	.784	118	.432	224	.258
Reading comprehension										
Grade 3 (ages 8–9)	305	.667	285	.407	327	.665	261	.300	535	.326
Grade 4 (ages 9–10)	232	.710	219	.513	259	.651	215	.309	411	.333
Grade 5 (ages 10–11)	228	.649	215	.417	253	.717	202	.287	390	.375
Grade 6 (ages 11–12)	167	.649	147	.484	179	.648	147	.390	275	.288
Spelling										
Grade 1 (ages 6–7)	344	.623	316	.399	365	.567	281	.399	594	.524
Grade 2 (ages 7–8)	320	.648	283	.405	345	.669	262	.336	544	.356
Grade 3 (ages 8–9)	311	.689	285	.329	338	.709	255	.272	533	.294
Grade 4 (ages 9–10)	234	.699	225	.446	263	.605	215	.270	412	.303
Grade 5 (ages 10–11)	224	.704	211	.306	255	.675	199	.596	393	.334
Grade 6 (ages 11–12)	166	.721	141	.483	181	.687	144	.374	259	.201
Educational achievement										
Arithmetic	757	.700	787	.374	907	.757	765	.408	1618	.369
Language	757	.765	787	.457	908	.782	765	.423	1617	.449
Info skills	755	.633	786	.375	908	.697	765	.350	1618	.359
Science and social studies	668	.745	695	.515	802	.785	677	.455	1448	.472
Total score	1112	.804	1129	.468	1337	.830	1149	.432	2280	.437

MZm = monozygotic male twin pairs; MZf = monozygotic female twin pairs; DZm = dizygotic male twin pairs; DZf = dizygotic female. twin pairs; DOS= dizygotic of opposite sex twin pairs; N = number of (in)complete twin pairs.

Table 3
Genetic modeling results for arithmetic, reading, reading comprehension, spelling and educational achievement.

		ep	−2ll	df	model	χ^2	Δdf	p	a^2	c^2	e^2
Arithmetic											
Grade 1 (ages 6–7)											
0 Saturated		50	23,690.61	2915	–	–	–	–	–		
1 ACE	Boys	12	23,722.67	2948	0	32.06	33	.514	.35 [.12–.62]	.29 [.04–.49]	.36 [.30–.43]
	Girls								.71 [.65–.76]	.00 [.00–.17]	.29 [.24–.35]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	23,723.73	2949	1	1.06	1	.303	.38 [.15–.63]	.27 [.03–.47]	.35 [.29–.42]
	Girls								.65 [.52–.73]	.06 [.00–.17]	.29 [.24–.35]
3 ACE: Boys = Girls		9	23,728.15	2951	2	4.42	2	.110	.60 [.46–.71]	.08 [.00–.20]	.32 [.28–.37]
4 AE: Drop C		**8**	**23,729.56**	**2952**	**3**	**1.41**	**1**	**.234**	**.68 [.64–.72]**	–	**.32 [.28–.36]**
Grade 2 (ages 7–8)											
0 Saturated		50	20,784.12	2616	–	–	–	–			
1 ACE	Boys	12	20,821.89	2649	0	37.76	33	.261	.42 [.15–.70]	.24 [.00–.48]	.34 [.28–.42]
	Girls								.67 [.60–.73]	.00 [.00–.21]	.33 [.27–.40]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	20,821.91	2650	1	.02	1	.896	.42 [.15–.69]	.24 [.00–.47]	.34 [.28–.42]
	Girls								.66 [.55–.72]	.01 [.00–.09]	.33 [.27–.40]
3 ACE: Boys = Girls		9	20,825.79	2652	2	3.89	2	.143	.66 [.61–.71]	.00 [.00–.13]	.34 [.29–.39]
4 AE: Drop C		**8**	**20,825.79**	**2653**	**3**	**.00**	**1**	**1.000**	**.66 [.61–.71]**	–	**.34 [.29–.39]**
Grade 3 (ages 8–9)											
0 Saturated		50	20,201.25	2584	–	–	–	–			
1 ACE	Boys	12	20,246.89	2617	0	45.63	33	.071	.57 [.30–.71]	.07 [.00–.31]	.36 [.29–.44]
	Girls								.74 [.68–.79]	.00 [.00–.13]	.26 [.21–.32]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	20,246.89	2618	1	.00	1	.949	.57 [.32–.70]	.07 [.00–.29]	.36 [.30–.44]
	Girls								.74 [.68–.79]	.00 [.00–.12]	.26 [.21–.32]
3 ACE: Boys = Girls		9	20,251.48	2620	2	4.59	2	.101	.69 [.64–.73]	.00 [.00–.12]	.31 [.27–.36]
4 AE: Drop C		**8**	**20,251.48**	**2621**	**3**	**.00**	**1**	**1.000**	**.69 [.64–.73]**		**.31 [.27–.36]**
Grade 4 (ages 9–10)											
0 Saturated		50	14,586.79	1943	–	–	–	–			
1 ACE	Boys	12	14,641.28	1976	0	54.49	33	.011	.49 [.20–.75]	.20 [.00–.46]	.31 [.24–.39]
	Girls								.56 [.26–.73]	.10 [.00–.37]	.33 [.27–.42]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	14,641.66	1977	1	.39	1	.534	.49 [.20–.75]	.21 [.00–.46]	.31 [.24–.39]
	Girls								.67 [.30–.73]	.00 [.00–.34]	.33 [.27–.41]
3 ACE: Boys = Girls		9	14,643.19	1979	2	1.53	2	.466	.61 [.42–.73]	.07 [.00–.24]	.32 [.27–.38]
4 AE: Drop C		**8**	**14,643.87**	**1980**	**3**	**.678**	**1**	**.410**	**.68 [.63–.73]**	–	**.32 [.27–.37]**
Grade 5 (ages 10–11)											
0 Saturated		50	13,128.06	1802	–	–	–	–			
1 ACE	Boys	12	13,187.57	1835	0	59.51	33	.003	.75 [.68–.80]	.00 [.00–.14]	.25 [.20–.32]
	Girls								.73 [.66–.79]	.00 [.00–.22]	.27 [.21–.34]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	13,187.78	1836	1	.21	1	.645	.75 [.69–.81]	.00 [.00–.14]	.25 [.19–.32]
	Girls								.74 [.66–.79]	.00 [.00–.21]	.26 [.21–.34]
3 ACE: Boys = Girls		9	13,188.50	1838	2	.71	2	.700	.74 [.70–.78]	.00 [.00–.13]	.26 [.22–.30]
4 AE: Drop C		**8**	**13,188.50**	**1839**	**3**	**.00**	**1**	**1.000**	**.74 [.70–.78]**	–	**.26 [.22–.30]**
Grade 6 (ages 11–12)											
0 Saturated		50	6687.21	941	–	–	–	–	–		
1 ACE	Boys	12	6730.17	958	0	60.54	33	.002	.49 [.06–.75]	.17 [.00–.53]	.34 [.25–.48]
	Girls								.75 [.66–.81]	.00 [.00–.30]	.25 [.19–.34]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	6730.17	959	1	.00	1	.999	.49 [.07–.75]	.17 [.00–.52]	.34 [.25–.47]
	Girls								.75 [.44–.81]	.00 [.00–.29]	.25 [.19–.34]
3 ACE: Boys = Girls		9	6732.88	961	2	2.71	2	.258	.71 [.63–.77]	.00 [.00–.19]	.29 [.23–.37]
4 AE: Drop C		**8**	**6732.88**	**962**	**3**	**.00**	**1**	**1.000**	**.71 [.63–.77]**	–	**.29 [.23–.37]**
Reading											
Grade 2 (ages 7–8)											
0 Saturated		25	14,506.08	1689	–	–	–	–			
1 ACE	Boys	9	14,515.22	1705	0	9.14	16	.908	.84 [.57–.88]	.00 [.00–.26]	.16 [.12–.21]
	Girls								.83 [.57–.89]	.03 [.00–.29]	.14 [.11–.18]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	14,515.93	1706	0	.72	1	.396	.78 [.56–.87]	.06 [.00–.27]	.16 [.12–.21]
	Girls								.78 [.56–.88]	.08 [.00–.30]	.14 [.11–.18]
3 ACE: Boys = Girls		6	14,516.31	1708	1	.38	2	.827	.78 [.64–.87]	.07 [.00–.21]	.15 [.12–.18]
4 AE: Drop C		**5**	**14,517.08**	**1709**	**3**	**.76**	**1**	**.383**	**.85 [.82–.88]**	–	**.15 [.12–.18]**
Grade 3 (ages 8–9)											
0 Saturated		25	15,150.08	1755	–	–	–	–			
1 ACE	Boys	9	15,164.53	1771	0	14.45	16	.565	.70 [.37–.81]	.06 [.00–.36]	.25 [.19–.32]
	Girls								.66 [.37–.82]	.11 [.00–.38]	.22 [.17–.29]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	15,164.63	1772	1	.10	1	.750	.75 [.38–.81]	.00 [.00–.35]	.25 [.19–.32]
	Girls								.66 [.37–.82]	.12 [.00–.39]	.22 [.17–.29]
3 ACE: Boys = Girls		6	15,165.30	1774	2	.67	2	.714	.73 [.56–.80]	.03 [.00–.19]	.24 [.20–.28]
4 AE: Drop C		**5**	**15,165.42**	**1775**	**3**	**.12**	**1**	**.734**	**.77 [.72–.80]**	–	**.23 [.20–.28]**
Grade 4 (ages 9–10)											
0 Saturated		25	11,305.09	1332	–	–	–	–			
1 ACE	Boys	9	11,326.66	1348	0	21.56	16	.158	.47 [.19–.81]	.31 [.00–.56]	.22 [.16–.30]

(continued on next page)

Table 3 (*continued*)

		ep	−2ll	df	model	χ^2	Δdf	p	a^2	c^2	e^2
	Girls								.80 [.72–.85]	.00 [.00–.15]	.20 [.15–.28]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	11,326.67	1349	1	.01	1	.935	.48 [.21–.78]	.30 [.01–.55]	.22 [.16–.30]
	Girls								.80 [.71–.85]	.00 [.00–.05]	.20 [.15–.28]
3 ACE: Boys = Girls		6	11,331.58	1351	2	4.91	2	.086	.79 [.74–.83]	.00 [.00–.11]	.21 [.17–.26]
4 AE: Drop C		**5**	**11,331.58**	**1352**	**3**	**.00**	**1**	**1.000**	**.79 [.74–.83]**	–	**.21 [.17–.26]**
Grade 5 (ages 10–11)											
0 Saturated		25	9654.21	1157	–	–	–	–			
1 ACE	Boys	9	9671.31	1173	0	17.10	16	.379	.55 [.28–.87]	.29 [.00–.54]	.16 [.12–.23]
	Girls								.75 [.40–.85]	.05 [.00–.38]	.20 [.15–.28]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	9671.34	1174	1	.03	1	.861	.55 [.28–.85]	.28 [.00–.54]	.16 [.12–.23]
	Girls								.78 [.51–.85]	.02 [.00–.28]	.20 [.15–.28]
3 ACE: Boys = Girls		6	9680.20	1176	2	8.86	2	.012	.82 [.67–.85]	.00 [.00–.13]	.18 [.15–.23]
4 AE: Drop C		**5**	**9680.20**	**1177**	**3**	**.00**	**1**	**1.000**	**.82 [.77–.85]**	–	**.18 [.15–.23]**
Reading comprehension											
Grade 3 (ages 8–9)											
0 Saturated		50	21,744.90	2690	–	–	–	–			
1 ACE	Boys	12	21,785.57	2723	0	40.68	33	.168	.49 [.23–.70]	.16 [.00–.38]	.36 [.29–.43]
	Girls								.66 [.59–.72]	.00 [.00–.22]	.34 [.28–.41]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	21,785.61	2724	1	.04	1	.843	.49 [.24–.70]	.15 [.00–.38]	.35 [.29–.43]
	Girls								.65 [.45–.72]	.01 [.00–.14]	.34 [.28–.41]
3 ACE: Boys = Girls		9	21,787.20	2726	2	1.58	2	.453	.64 [.48–.70]	.01 [.00–.15]	.35 [.30–.40]
4 AE: Drop C		**8**	**21,787.22**	**2727**	**3**	**.03**	**1**	**.874**	**.65 [.61–.70]**	–	**.35 [.30–.39]**
Grade 4 (ages 9–10)											
0 Saturated		50	17,111.12	2122	–	–	–	–			
1 ACE	Boys	12	17,164.98	2155	0	53.86	33	.012	.49 [.25–.75]	.23 [.00–.44]	.28 [.23–.35]
	Girls								.63 [.31–.70]	.00 [.00–.29]	.37 [.30–.45]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	17,165.00	2156	1	.02	1	.899	.49 [.25–.75]	.23 [.00–.44]	.28 [.23–.35]
	Girls								.61 [.42–.70]	.02 [.00–.21]	.37 [.30–.45]
3 ACE: Boys = Girls		9	17,171.34	2158	2	6.34	2	.042	.59 [.43–.72]	.08 [.00–.23]	.32 [.28–.38]
4 AE: Drop C		**8**	**17,172.47**	**2159**	**3**	**1.13**	**1**	**.288**	**.68 [.63–.73]**	–	**.32 [.27–.37]**
Grade 5 (ages 10–11)											
0 Saturated		50	16,834.20	2045	–	–	–	–			
1 ACE	Boys	12	16,922.95	2078	0	88.75	33	<.001	.55 [.27–.73]	.11 [.00–.36]	.33 [.27–.42]
	Girls								.65 [.56–.71]	.00 [.00–.20]	.35 [.29–.44]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	16,923.51	2079	1	.57	1	.452	.56 [.29–.73]	.11 [.00–.35]	.33 [.27–.41]
	Girls								.62 [.40–.71]	.03 [.00–.21]	.35 [.29–.43]
3 ACE: Boys = Girls		9	16,924.54	2081	2	1.03	2	.597	.60 [.42–.71]	.06 [.00–.21]	.34 [.30–.40]
4 AE: Drop C		**8**	**16,925.04**	**2082**	**3**	**.50**	**1**	**.480**	**.66 [.61–.71]**	–	**.34 [.29–.39]**
Grade 6 (ages 11–12)											
0 Saturated		50	12,024.86	1427	-	-	-	-			
1 ACE	Boys	12	12,060.09	1460	0	35.23	33	.363	.48 [.15–.74]	.19 [.00–.47]	.33 [.25–.43]
	Girls								.45 [.09–.71]	.18 [.00–.49]	.37 [.29–.47]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	12,060.88	1461	1	.79	1	.375	.46 [.14–.74]	.20 [.00–.48]	.33 [.26–.43]
	Girls								.62 [.07–.71]	.02 [.00–.51]	.36 [.28–.46]
3 ACE: Boys = Girls		9	12,061.51	1463	2	.36	2	.728	.54 [.32–.70]	.12 [.00–.30]	.35 [.29–.42]
4 AE: Drop C		**8**	**12,062.79**	**1464**	**3**	**1.27**	**1**	**.259**	**.66 [.59–.71]**	–	**.34 [.29–.41]**
Spelling											
Grade 1 (ages 6–7)											
0 Saturated		50	21,082.01	3049	-	-	-	-			
1 ACE	Boys	12	21,122.46	3082	0	40.45	33	.174	.47 [.22–.69]	.17 [.00–.38]	.36 [.30–.44]
	Girls								.30 [.04–.62]	.29 [.00–.52]	.41 [.34–.49]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	21,124.26	3083	1	1.80	1	.180	.37 [.17–.53]	.26 [.13–.42]	.37 [.30–.45]
	Girls								.26 [.05–.46]	.34 [.16–.52]	.40 [.34–.48]
3 ACE: Boys = Girls		**9**	**21,124.50**	**3085**	**2**	**.23**	**2**	**.890**	**.33 [.18–.47]**	**.28 [.16–.40]**	**.39 [.34–.44]**
4 AE: Drop C		8	21,143.02	3086	3	18.52	1	<.001	.64 [.60–.68]	–	.36 [.32–.40]
Grade 2 (ages 7–8)											
0 Saturated		50	18,716.77	2783	-	-	-	-			
1 ACE	Boys	12	18,749.77	2816	0	33.00	33	.467	.54 [.29–.70]	.09 [.00–.31]	.36 [.30–.44]
	Girls								.66 [.38–.72]	.01 [.00–.27]	.34 [.28–.40]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	18,749.93	2817	1	.16	1	.685	.54 [.29–.70]	.10 [.00–.31]	.36 [.30–.44]
	Girls								.61 [.37–.71]	.05 [.00–.27]	.34 [.28–.40]
3 ACE: Boys = Girls		9	18,750.27	2819	2	.34	2	.845	.59 [.44–.69]	.06 [.00–.19]	.35 [.31–.40]
4 AE: Drop C		**8**	**18,750.99**	**2820**	**3**	**.72**	**1**	**.396**	**.66 [.61–.70]**	–	**.34 [.30–.39]**
Grade 3 (ages 8–9)											
0 Saturated		50	18,139.31	2720	-	-	-	-			
1 ACE	Boys	12	18,160.87	2753	0	21.56	33	.937	.70 [.47–.76]	.00 [.00–.21]	.30 [.24–.36]
	Girls								.70 [.54–.75]	.00 [.00–.15]	.30 [.25–.37]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	18,161.55	2754	1	.68	1	.408	.66 [.45–.75]	.04 [.00–.23]	.30 [.24–.37]
	Girls								.68 [.52–.75]	.01 [.00–.16]	.30 [.25–.37]
3 ACE: Boys = Girls		9	18,163.11	2756	2	1.56	2	.459	.70 [.62–.74]	.00 [.00–.06]	.30 [.26–.35]
4 AE: Drop C		**8**	**18,163.11**	**2757**	**3**	**.00**	**1**	**1.000**	**.70 [.65–.74]**	–	**.30 [.26–.35]**

Table 3 (*continued*)

		ep	−2ll	df	model	χ^2	Δdf	p	a^2	c^2	e^2
Grade 4 (ages 9–10)											
0 Saturated		50	14,445.63	2124	–	–	–	–			
1 ACE	Boys	12	14,478.83	2157	0	33.19	33	.458	.46 [.18–.73]	.22 [.00–.47]	.32 [.26–.40]
	Girls								.61 [.43–.69]	.00 [.00–.19]	.39 [.31–.48]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	14,478.91	2158	1	.09	1	.767	.47 [.19–.73]	.21 [.00–.46]	.32 [.26–.40].
	Girls								.61 [.46–.69]	.01 [.00–.12]	39 [.31–.48]
3 ACE: Boys = Girls		9	14,482.87	2160	2	3.95	2	.139	.64 [.46–.70]	.01 [.00–.16]	.35 [.30–.41]
4 AE: Drop C		**8**	**14,482.87**	**2161**	**3**	**.00**	**1**	**.945**	**.65 [.59–.70]**	–	**.35 [.30–.41]**
Grade 5 (ages 10–11)											
0 Saturated		50	13,199.57	2022	-	-	-	-			
1 ACE	Boys	12	13,235.51	2055	0	35.94	33	.333	.68 [.60–.74]	.00 [.00–.12]	.32 [.26–.40]
	Girls								.25 [.03–.51]	.44 [.20–.63]	.31 [.25–.38]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	13,235.79	2056	1	.28	1	.594	.63 [.49–.73]	.05 [.00–.16]	.32 [.26–.40]
	Girls								.26 [.04–.51]	.43 [.20–.62]	.31 [.25–.38]
3 ACE: Boys = Girls		9	13,245.36	2058	1	9.57	2	.008	.58 [.42–.73]	.11 [.00–.25]	.31 [.27–.37]
4 AE: Boys, Drop C	**Boys**	**10**	**13,239.12**	**2057**	**2**	**3.33**	**1**	**.068**	**.69 [.61–.75]**	–	**.31 [.25–.39]**
	Girls								**.38 [.19–.61]**	**.32 [.10–.50]**	**.29 [.24–.36]**
5 AE: Girls, Drop C	Boys	9	13,246.87	2058	4	7.75	1	.005	.68 [.60–.74]	–	.32 [.26–.40]
	Girls								.71 [.64–.76]	–	.29 [.24–.36]
Grade 6 (ages 11–12)											
0 Saturated		50	9115.90	1383	-	-	-	-			
1 ACE	Boys	12	9159.94	1416	0	44.04	33	.095	.46 [.15–.75]	.23 [.00–.51]	.30 [.24–.39]
	Girls								.57 [.23–.73]	.09 [.00–.39]	.34 [.27–.44]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	11	9160.14	1417	1	.19	1	.659	.48 [.15–.75]	.22 [.00–.51]	.30 [.24–.39]
	Girls								.65 [.30–.73]	.01 [.00–.33]	.34 [.27–.43]
3 ACE: Boys = Girls		9	9164.10	1419	2	3.96	2	.138	.66 [.44–.73]	.02 [.00–.21]	.32 [.27–.39]
4 AE: Drop C		**8**	**9164.13**	**1420**	**3**	**.03**	**1**	**.856**	**.68 [.62–.73]**	–	**.32 [.27–.38]**
Educational achievement											
Arithmetic											
0 Saturated		25	83,434.16	8978	–	–	–	–	–	–	–
1 ACE	Boys	9	83,448.10	8994	0	13.94	16	.603	.66 [.52–.74]	.05 [.00–.18]	.29 [.26–.33]
	Girls								.68 [.56–.77]	.07 [.00–.19]	.25 [.22–.27]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	83,448.58	8995	1	.48	1	.490	.71 [.52–.74]	.01 [.00–.17]	.29 [.26–.32]
	Girls								.6875 [.62–.77]	.08 [.00–.19]	.25 [.22–.27]
3 ACE: Boys = Girls		6	83,453.08	8997	2	4.50	2	.105	.68 [.61–.75]	.05 [.00–.12]	.27 [.25–.29]
3 AE: Drop C		**5**	**83,455.10**	**8998**	**3**	**2.02**	**1**	**.155**	**.74 [.72–.75]**	–	**.26 [.25–.28]**
Language											
0 Saturated		25	83,046.16	8978	–	–	–	–	–	–	–
1 ACE	Boys	9	83,062.47	8994	0	16.31	16	.431	.62 [.50–.75]	.15 [.02–.26]	.23 [.21–.26]
	Girls								.72 [.60–.80]	.06 [.00–.18]	.22 [.20–.24]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	83,062.67	8995	1	.20	1	.651	.62 [.50–.72]	.15 [.05–.26]	.23 [.21–.26]
	Girls								.70 [.61–.76]	.09 [.03–.17]	.22 [.20–.24]
3 ACE: Boys = Girls		**6**	**83,064.06**	**8997**	**2**	**1.39**	**2**	**.499**	**.67 [.61–.74]**	**.10 [.04–.16]**	**.22 [.21–.24]**
4 AE: Drop C		5	83,074.17	8998	2	10.11	1	.001	.78 [.76–.79]	-	.22 [.21–.24]
Study skills											
0 Saturated		25	83,903.97	8969	–	–	–	–	–	–	–
1 ACE	Boys	9	83,921.75	8985	0	17.78	16	.337	.51 [.37–.66]	.12 [.00–.25]	.37 [.33–.41]
	Girls								.69 [.55–.72]	.00 [.00–.13]	.31 [.28–.34]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	83,922.12	8986	1	.37	1	.541	.51 [.37–.65]	.12 [.00–.24]	.37 [.33–.41]
	Girls								.66 [.57–.71]	.03 [.00–.11]	.31 [.28–.34]
3 ACE: Boys = Girls		6	83,928.85	8988	2	6.73	2	.035	.60 [.53–.68]	.06 [.00–.13]	.33 [.31–.36]
4 AE: Drop C		**5**	**83,931.80**	**8989**	**3**	**2.95**	**1**	**.086**	**.67 [.65–.69]**	–	**.33 [.31–.35]**
Science and social studies											
0 Saturated		25	73,103.83	7934	-	-	-	-	-	-	-
1 ACE	Boys	9	73,023.23	7950	0	9.41	16	.896	.48 [.36–.61]	.27 [.14–.37]	.25 [.23–.29]
	Girls								.66 [.54–.79]	.13 [.00–.24]	.21 [.19–.23]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	73,023.25	7951	1	.02	1	.900	.48 [.36–.60]	.27 [.15–.37]	.25 [.23–.28]
	Girls								.65 [.57–.72]	.14 [.07–.21]	.21 [.19–.23]
3 ACE: Boys = Girls		**6**	**73,030.43**	**7953**	**2**	**7.18**	**2**	**.028**	**.56 [.50–.64]**	**.21 [.14–.27]**	**.23 [.21–.25]**
4 AE: Drop C		5	73,065.31	7954	3	34.88	1	<.001	.78 [.76–.79]	-	.22 [.21–.24]
Total score											
0 Saturated		25	88,844.71	12,878	–	–	–	–	–	–	–
1 ACE	Boys	9	88,870.41	12,894	0	25.70	16	.058	.71 [.61–.81]	.11 [.01–.20]	.18 [.17–.20]
	Girls								.77 [.67–.84]	.06 [.00–.15]	.18 [.16–.19]
2 ACE: $rA_{DOS} = rA_{DZ}$	Boys	8	88,870.60	12,895	1	.19	1	.663	.71 [.61–.81]	.11 [.01–.20]	.19 [.17–.20]
	Girls								.79 [.69–.83]	.04 [.00–.13]	.18 [.16–.19]
3 ACE: Boys = Girls		**6**	**88,871.41**	**12,897**	**2**	**.81**	**2**	**.667**	**.74 [.69–.80]**	**.08 [.02–.13]**	**.18 [.17–.19]**
4 AE: Drop C		5	88,879.25	12,898	3	7.84	1	.005	.82 [.81–.83]	-	.18 [.17–.19]

Note: In bold, the results for the most parsimonious model that did not significantly ($p > 0.01$) deteriorate the goodness of fit.

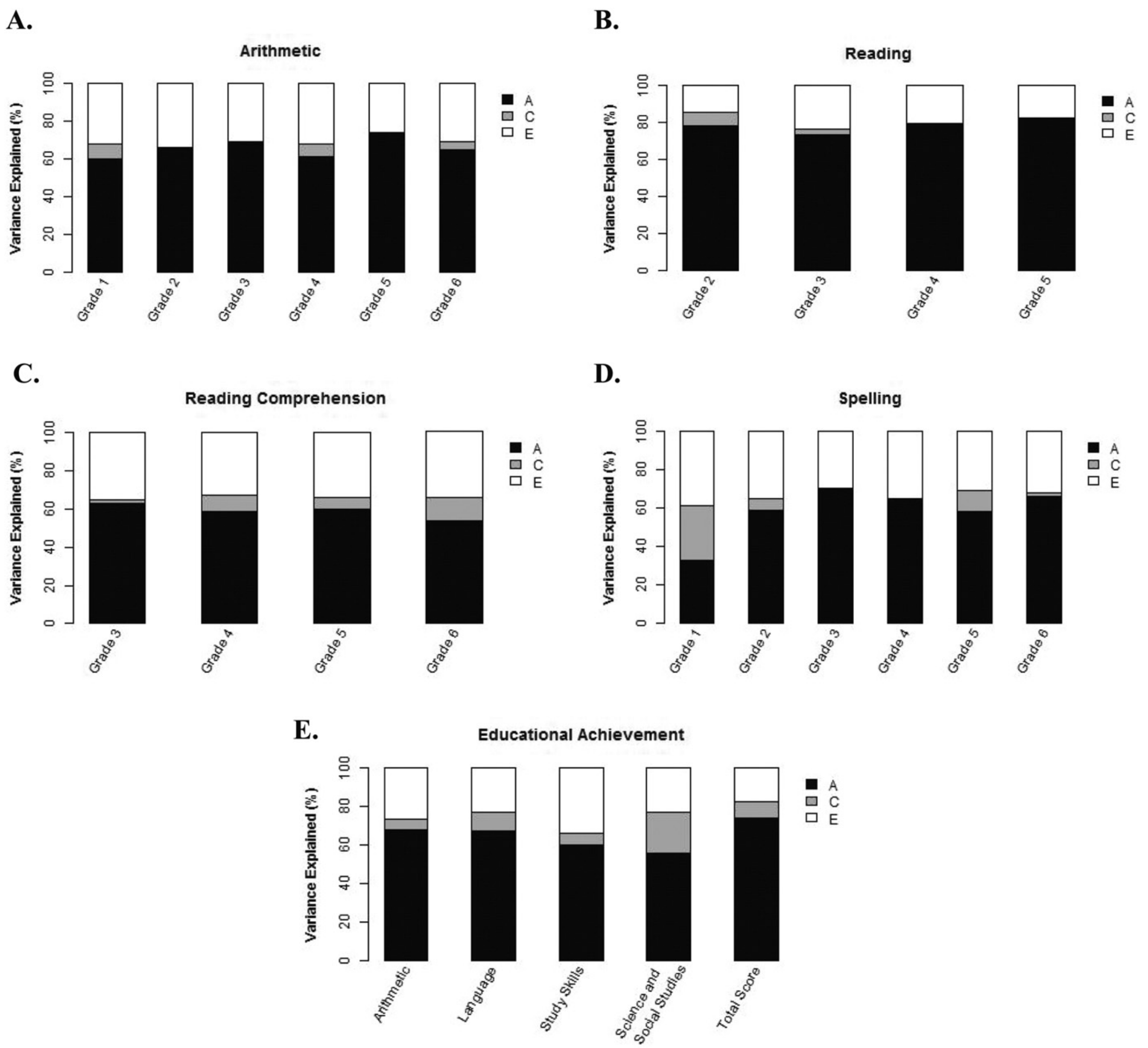

Fig. 2. The relative contribution of the genetic, common environmental and unique environmental effects for arithmetic (A), reading (B), reading comprehension (C), spelling (D) and educational achievement (E).

Plomin, 2005), than has been found for school children. Lessons in the first grade focus mainly on reading and arithmetic while spelling becomes more important in later grades when children have learned how to read words. It could be that, in the first grade, the common home environment still influences spelling while this is no longer the case for reading and arithmetic.

Heritability of a standardized educational achievement test in the last grade was high (74%) while the influence of the common environment was low. There was a noteworthy difference with regard to the etiology of the specific domains included in the educational achievement test. Science and social studies was less heritable with a larger influence of the common environment compared to arithmetic and language. A lower heritability is also observed in earlier research on science performance (Haworth, Dale, & Plomin, 2009; Haworth, Kovas, Dale, & Plomin, 2008). A possible explanation for this difference in heritability between the core educational domains, i.e. language and reading, and science and social studies is that there is more heterogeneity in the curriculum and lessons on topics related to geography, history and biology while the methods for teaching children to read and calculate are more standardized. The larger heterogeneity in education will increase differences in the environment and, as a result, individual differences between children will, relatively, to a lesser extent be explained by genes (Heath et al., 1985).

Traditional mean gender differences in educational achievement were observed with boys scoring, on average, better on numeracy and girls performing, on average, better at some of the literacy subjects (Cito, 2014b; OECD, 2010). However, there was no consistent indication for the presence of quantitative gender differences, meaning that the extent to which genes and the environment influence educational achievement is similar across gender. Qualitative gender differences were also not present which means that the genes that have an influence on educational achievement are the same for boys and girls. Taken together, this rules out a difference in the genetic architecture of educational achievement as an explanation for the mean differences between boys and girls. This is in line with findings from a study of teacher ratings measuring educational achievement in multiple educational domains (Kovas et al., 2007). The absence of gender differences in genes with an effect on educational achievement in children has implications for the attempts to find genetic variants with genome-wide association studies since it will be possible to combine the data on educational achievement in boys and girls resulting in larger sample sizes.

The twin method assumes that MZ twins are more similar than DZ twins because of their larger genetic resemblance and not because MZ twins are treated more alike than DZ twins, by for example teachers, or otherwise experience more similar environments, such as sharing the same classroom or friends. The equal environment assumption is potentially violated when characteristics of the environment are more often shared between MZ twins than DZ twins. This would mean that the difference between the similarity for MZ and DZ twin pairs is no longer due to differences in genetic similarity, but to a combination of genetic and environmental effects, and the impact of genes will be overestimated. However, this is only the case when the similarity in the environment relates to a resemblance in a phenotype. Research has shown that the equal environment assumption holds for, amongst others, general cognitive ability and educational achievement (Evans & Martin, 2000; Loehlin, 1989).

To be able to generalize the outcome of twin studies to the general population, the twin method further assumes that twins are representative of the general population for the phenotype of interest. Twins do differ from singletons in striking ways with regard to birth conditions. Twins are born, on average, 3–4 weeks prematurely and have approximately 1 kg lower birth weights (Martin et al., 2010). These differences dissipate fairly early on, however, and, already in childhood, twins and singletons have very similar scores for body size (Estourgie-van Burk, Bartels, Boomsma, & Delemarre-van de Waal, 2010) and, more relevantly, for general cognitive ability (Webbink, Posthuma, Boomsma, de Geus, & Visscher, 2008) and educational achievement (Cohen, van Goozen, Orlebeke, Buitelaar, & Cohen-Kettenis, 2002; de Zeeuw, van Beijsterveldt, De Geus, & Boomsma, 2012).

Heritability estimates for educational achievement in the Netherlands are comparable to other countries, especially the UK (Kovas et al., 2013). Apparently genes account for most individual differences across these populations. The relative contribution of genes and the environment to the variation in educational achievement can differ when either the genetic effects are different or the environmental effects differ, for example, due to differences in SES, national curriculum or educational opportunities. Differences across countries may lead to a relatively larger or smaller role of the environment. The educational system in The Netherlands is similar to the one implemented in the UK, as both countries have a national curriculum, while the educational system in, for example, the USA is more decentralized. A national curriculum likely restricts the variation in school environments leading to an increase in the relative contribution of genes to the variation in educational achievement. It would be of interest to systematically investigate whether these differences between countries in educational system are reflected in differences in heritability.

Although the individual differences between children in their achievement across different educational domains are for a large part due to innate differences, one must keep in mind that this heritability does not equal determinism. First of all, the variance between children may be heritable, but the mean can be positively influenced by a school environment of good quality. For example, when increasing instruction time for a certain educational domain there will be an increase in the average ability in this domain. Furthermore, some environments can affect some children more than others (gene–environment interaction). For example, children with a certain genetic make-up may respond better to a specific tutoring program than other children. In our view, the challenge for teachers is to make sure that children, who have more difficulty at school, will learn how to read, write and perform calculations, while the children that learn easily are still sufficiently challenged. Whole-classroom teaching might not be the best method to achieve this goal and a more personalized approach to education will be necessary. Unfortunately, the increasing number of children per teacher and the demand on teachers with regard to administrative duties might preclude teachers from customizing their lessons to the needs of each child. Advances in technology might perhaps in the future be the answer to providing each child with individual lessons targeting their specific needs.

Acknowledgments

We are grateful to the twin families and the teachers for their participation. We acknowledge the research program 'Innovative learning materials and methods' funded by the Netherlands Initiative Brain and Cognition, a part of the Organization for Scientific Research (NWO) under grant 056-31-001; 'Twin-Family Database for Behavior Genetics and Genomics Studies' (NWO 480-04-004); 'Genetics of Mental Illness' European Research Council (ERC-230374); 'Genetic Influences on Stability and Change in Psychopathology from Childhood to Young Adulthood' (NWO/ZonMW 91210020); 'Consortium on Individual Development (CID)' which is funded through the Gravitation program of the Dutch Ministry of Education, Culture and Science and the Netherlands Organization for Scientific Research (NWO 0240-001-003).

References

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

Bartels, M., van Beijsterveldt, C. E., Derks, E. M., Stroet, T. M., Polderman, T. J., Hudziak, J. J., et al. (2007). Young Netherlands Twin Register (Y-NTR): A longitudinal multiple informant study of problem behavior. *Twin Research and Human Genetics*, *10*, 3–11.

van Beijsterveldt, C. E., Groen-Blokhuis, M., Hottenga, J. J., Franic, S., Hudziak, J. J., Lamb, D., et al. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, *16*, 252–267.

Boker, S. M., Neale, M. C., Maes, H. H. M., Wilde, M. J., Spiegel, M., Brick, T. R., et al. (2011). An open source extended structural equation modeling framework. *Psychometrika*, *76*(2), 306–317.

Boker, S. M., Neale, M. C., Maes, H. H. M., Wilde, M. J., Spiegel, M., Brick, T. R., et al. (2012). *OpenMx 1.2 User guide*.

Boomsma, D. I. (2013). Twin, association and current 'omics' studies. *The Journal of Maternal–Fetal and Neonatal Medicine*, *26*, 9–12.

Boomsma, D. I., de Geus, E. J., Vink, J. M., Stubbe, J. H., Distel, M. A., Hottenga, J. J., et al. (2006). Netherlands Twin Register: From twins to twin families. *Twin Research and Human Genetics*, *9*, 849–857.

Boomsma, D. I., Vink, J. M., van Beijsterveldt, T. C., de Geus, E. J., Beem, A. L., Mulder, E. J., et al. (2002). Netherlands Twin Register: A focus on longitudinal research. *Twin Research*, *5*, 401–406.

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., et al. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to Grade 2. *Journal of Neurolinguistics*, *22*, 219–236.

Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., et al. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics*, *42*, 699–710.

Chow, B. W. -Y., Ho, C. S. -H., Wong, S. W. -L., Waye, M. M., & Bishop, D. V. (2011). Genetic and environmental influences on Chinese language and reading abilities. *PLoS ONE*, *6*, e16640.

Cito (2002). *Eindtoets Basisonderwijs*. The Netherlands, Cito: Arnhem.

Cito (2014a). *Cito Leerling Volgsysteem*. The Netherlands, Cito: Arnhem.

Cito (2014b). *Terugblik en resultaten 2014 — Eindtoets Basisonderwijs*. the Netherlands, Cito.: Arnhem.

Cohen, C. C., van Goozen, S. H., Orlebeke, J. F., Buitelaar, J. K., & Cohen-Kettenis, P. T. (2002). A comparison of educational achievement in a national sample of Dutch female twins and their matched singleton controls. *Twin Research*, *5*, 273–276.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*, 13–21.

Estourgie-van Burk, G. F., Bartels, M., Boomsma, D. I., & Delemarre-van de Waal, H. (2010). Body size of twins compared with siblings and the general population: From birth to late adolescence. *The Journal of Pediatrics*, *156*, 586–591.

Evans, D. M., & Martin, N. G. (2000). The validity of twin studies. *Genescreen*, *1*, 77–79.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, *15*, 373–378.

Harlaar, N., Hayiou-Thomas, M. E., & Plomin, R. (2005). Reading and general cognitive ability: A multivariate analysis of 7-year-old twins. *Scientific Studies of Reading*, *9*, 197–218.

Haworth, C., Dale, P. S., & Plomin, R. (2009). The etiology of science performance: Decreasing heritability and increasing importance of the shared environment from 9 to 12 years of age. *Child Development*, *80*, 662–673.
Haworth, C., Kovas, Y., Dale, P. S., & Plomin, R. (2008). Science in elementary school: Generalist genes and school environments. *Intelligence*, *36*, 694–701.
Heath, A. C., Berg, K., Eaves, L. J., Solaas, M. H., Corey, L. A., Sundet, J., et al. (1985). Education policy and the heritability of educational attainment. *Nature*, *314*, 734–736.
Herman, v. B., Ronald, E., & Anja, D. W. (2011). *Wetenschappelijke verantwoording van de Eindtoets Basisonderwijs 2010.* Cito: Arnhem.
Kan, K. J., van Beijsterveldt, C. E., Bartels, M., & Boomsma, D. I. (2014). Assessing genetic influences on behavior: Informant and context dependency as illustrated by the analysis of attention problems. *Behavior Genetics*, *44*(4), 1–11.
Kovas, Y., Haworth, C. M., Dale, P. S., & Plomin, R. (2007). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, *72*(3), 1–144.
Kovas, Y., Voronin, I., Kaydalov, A., Malykh, S. B., Dale, P. S., & Plomin, R. (2013). Literacy and numeracy are more heritable than intelligence in primary school. *Psychological Science*, *24*, 2048–2056.
Loehlin, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist*, *44*, 1285.
Martin, J. A., Hamilton, B. E., Sutton, P. D., Ventura, S. J., Mathews, T. J., Kirmeyer, S., et al. (2010). Births: Final data for 2007. *National Vital Statistics Reports*, *58*, 1–85.
Moilanen, K. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing, and academic competence from middle childhood to early adolescence. *Development and Psychopathology*, *22*, 635–653.
OECD (2010). *PISA 2009 Results: What students know and can do — Student performance in reading, mathematics and science (Volume I) OECD.*
Oliver, B. R., Dale, P. S., & Plomin, R. (2005). Predicting literacy at age 7 from preliteracy at age 4: A longitudinal genetic analysis. *Psychological Science*, *16*, 861–865.
Plomin, R., Defries, J. C., McClearn, G. E., & McGuffin, P. S. (2008). *Behavioral genetics* (5th ed.). New York: Worth Publishers.
Polderman, T. J., Huizink, A. C., Verhulst, F. C., van Beijsterveldt, C. E., Boomsma, D. I., & Bartels, M. (2011). A genetic study on attention problems and academic skills: Results of a longitudinal study in twins. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*, 22–34.
Posthuma, D., Beem, A. L., De Geus, E. J., Van Baal, G. C., von Hjelmborg, J. B., Iachine, I., et al. (2003). Theory and practice in quantitative genetics. *Twin Research*, *6*, 361–376.
Purcell, S. (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research*, *5*, 554–571.
R Core Team (2014). *R: A language and environment for statistical computing.* Vieanna, Austria: R Foundation for Statistical Computing.
Rietveld, M. J., van, D. V., Bongers, I. L., Stroet, T. M., Slagboom, P. E., & Boomsma, D. I. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research*, *3*, 134–141.
Samuelsson, S., Byrne, B., Olson, R. K., Hulslander, J., Wadsworth, S., Corley, R., et al. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral–genetic analysis. *Learning and Individual Differences*, *18*, 289–295.
Simonoff, E., Pickles, A., Hervas, A., Silberg, J. L., Rutter, M., & Eaves, L. (1998). Genetic influences on childhood hyperactivity: Contrast effects imply parental rating bias, not sibling interaction. *Psychological Medicine*, *28*, 825–837.
van Soelen, I. L., Brouwer, R. M., Leeuwen, M. v., Kahn, R. S., Hulshoff Pol, H. E., & Boomsma, D. I. (2011). Heritability of verbal and performance intelligence in a pediatric longitudinal sample. *Twin Research and Human Genetics*, *14*, 119–128.
Vlug, K. F. (1997). Because every pupil counts: The success of the pupil monitoring system in The Netherlands. *Education and Information Technologies*, *2*, 287–306.
Webbink, D., Posthuma, D., Boomsma, D. I., de Geus, E. J., & Visscher, P. M. (2008). Do twins have lower cognitive ability than singletons? *Intelligence*, *36*, 539–547.
de Zeeuw, E. L., van Beijsterveldt, C. E., De Geus, E. J., & Boomsma, D. I. (2012). Twin specific risk factors in primary school achievements. *Twin Research and Human Genetics*, *15*, 107–115.