

## From Learning to Read to Reading to Learn: Substantial and Stable Genetic Influence

Nicole Harlaar  
*King's College of London*

Philip S. Dale  
*University of New Mexico*

Robert Plomin  
*King's College of London*

Little is known about the underlying causes and developmental patterns of stability and change in early reading abilities. In a longitudinal study of twins ( $n = 4,291$  pairs), individual differences in reading achievement assessed by teachers using U.K. National Curriculum (NC) criteria showed substantial heritabilities at ages 7, 9, and 10 years (.57–.67) and modest shared environmental influences (.10–.17). Stability in NC scores was primarily mediated genetically. There was also evidence for age-specific genetic influences at each age. Genetic influences on reading are substantial and stable during the elementary school years despite the shift from “learning to read” to “reading to learn.”

Reading acquisition is a cumulative process, beginning in early childhood and extending through the school years. In the early stages of learning to read, children are expected to become aware of the systematic correspondences between letters and the sounds of words, to learn the exceptions to these correspondences, and to be able to learn to recognize words quickly and accurately. Beyond this formative period in reading development, reading instruction and curricula tend to focus more strongly on honing reading comprehension skills. For example, they must learn to make links across different parts of text, identify the main themes, and summarize and evaluate what they have read (Biancarosa & Snow, 2004). As such, there is a gradual shift from “learning to read” to “reading to learn” (Chall, 1983).

Notwithstanding these important normative changes, there are huge within-age individual differences in reading ability, and these individual differences tend to remain relatively stable through childhood. Studies examining continuous test scores have shown that age-to-age correlations between diverse aspects of reading are typically of large effect size (.60 or greater; e.g., de Jong & van der Leij,

2002; Leppänen, Niemi, Aunola, & Nurmi, 2006; Parrila, Aunola, Leskinen, Nurmi, & Kirby, 2005; Scarborough, 1998). Reading disabilities in early childhood also show a strong degree of stability, despite remedial efforts that are usually made to strengthen the skills of lower achievers (Scarborough, 1998). Taken together, this evidence suggests that early reading delays and difficulties are associated with persistent reading difficulties, whereas early success means that children often remain near the top of the class for reading.

Understanding the origins of individual differences in reading ability has been a research target for both molecular and quantitative genetic studies for more than three decades (Pennington & Olson, 2005). Using quantitative genetic methodologies, twin and adoption studies have consistently yielded evidence that individual differences in reading abilities are substantially heritable. That is, within the populations studied, additive genetic influences largely account for why children differ in their reading abilities.

In contrast, little is known about the genetic and environmental etiology of individual differences in the stability of early to later reading skills. Only three genetically informative studies have examined reading longitudinally. In the Colorado Adoption Project, word recognition was examined at ages 7, 12, and 16 in a sample of adoptive and nonadoptive sibling pairs (Wadsworth, Corley, Hewitt, & DeFries, 2001; Wadsworth, Corley, Plomin, Hewitt, & DeFries, 2006). In the International Longitudinal Twin Study

---

We gratefully acknowledge the ongoing contribution of the parents and children in the Twin Early Development Study (TEDS). TEDS is supported by a program Grant (G0500079) from the U.K. Medical Research Council and a Grant (HD-44454) from the U.S. National Institute of Child Health and Human Development.

Correspondence concerning this article should be addressed to Nicole Harlaar, MRC Social, Genetic, & Developmental Psychiatry (SGDP) Centre, Box Number P083, Institute of Psychiatry, King's College London, London, SE5 8AF, U.K. Electronic mail may be sent to n.harlaar@iop.kcl.ac.uk.

(Byrne et al., 2006) and the Western Reserve Reading Project (Petrill, Deater-Deckard, Thompson, Schatschneider, & DeThorne, 2006), twins were tested at kindergarten or first grade and again 1 year later on measures of reading and language-related skills (e.g., phonological awareness, rapid automatized naming, and spelling).

Across these studies, three broad trends have emerged. First, the phenotypic correlation between early and later reading performance is principally mediated by additive genetic factors. That is, if we think of the estimated phenotypic correlation between two waves of assessment as the sum of genetic and environmental factors, the available evidence indicates that genes are the primary factor contributing to this correlation. Second, substantially the *same* genetic influences affect reading performance across assessment occasions. For example, genetic correlations between measurement occasions were of large effect size in each study, ranging from approximately 0.56 (Byrne et al., 2006, for spelling; estimate based on path coefficients), to 1.00 in the Colorado Adoption Project (Wadsworth et al., 2006). Third, environmental factors that contribute to familial resemblance—*shared* environmental factors—generally make a smaller contribution to the covariance between early and later reading, but the actual factors themselves are correlated across measurement occasions (approximately 0.18–1.00). These estimates must be viewed with some caution because twin studies are generally underpowered to detect significant shared environmental influences (Hopper, 2000), but they nevertheless hint at the possibility that stable environmental influences constitute a second source of stability in reading development.

A further question of interest in each of these studies was whether new genetic or shared environmental influences are expressed through reading development in the school years. That is, is there evidence that new sources of genetic or environmental influences come “online” during development? Although the genetic makeup of an individual is invariant, the degree to which certain genes are expressed may vary over time. Similarly, it is clear that the effects of environmental influences may wax and wane over time, and that new environmental effects may also influence reading development. Results across studies yield a rather confused picture. In the Colorado Adoption Project, there was no evidence for reliable “new” sources of genetic or shared environmental influences at ages 12 or 16—all genetic influences at these ages reflected the effects of genetic influences at age 7. However, in the International Longitudinal Twin Study, reliable new

genetic sources emerged for word reading, with a similar (but nonsignificant) trend for spelling, but there was no evidence for a reliable second source of shared environmental influences; in the Western Reserve Reading Project, there was evidence for independent genetic influences and, for word knowledge and letter knowledge, independent shared environmental effects.

In summary, there is emerging evidence that genetic factors primarily contribute to the stability of reading achievement, and there is continuity in both genetic and shared environmental influences affecting reading through development. These studies provide an important step toward characterizing the stability of reading development. However, there is a gap in our knowledge about the trends specifically involved in the elementary school years, when the transition from “learning to read” to “reading to learn” is particularly salient in the school curricula. The Colorado Adoption Project covered this period by assessing children at age 7 and then at age 12, but this period—almost half the child’s life span—involves many changes. Arguably, more frequent assessments are needed. Conversely, the International Longitudinal Twin Study and the Western Reserve Reading Project assessed children’s reading skills over a 1-year period, from kindergarten to first grade or first grade to second grade, but these studies have not yet covered the later elementary school years.

Against this background, the purpose of the present study is to investigate the developmental etiology of reading achievement from ages 7–9 and 10 for a sample of 4,291 twin pairs. Our first aim was to examine the extent to which phenotypic (observed) variance in National Curriculum (NC) scores at each age is due to genetic and environmental influences. We have previously reported evidence for high heritability at 7 years for both teacher assessments (.49–.72), whereas shared environmental influences were modest (.04–.37 and .09–.44, respectively; Harlaar, Dale, & Plomin, 2005). These estimates are similar to studies with similar-age (e.g., Byrne et al., 2006; Hohnen & Stevenson, 1999; Petrill et al., 2006), older (e.g., Bates et al., 2004), and mixed-age (e.g., Gayán & Olson, 2003) twin samples. Thus, we predicted that heritability would be substantial and shared environment would be modest for individual differences at ages 9 and 10. Our second aim was to examine the relative contributions of genetic and environmental influences to associations between reading achievement across ages 7–10. On the basis of previous research, we predicted that phenotypic stability is primarily due to genetic influences. However, given the rapid development of

reading skills in early elementary school, we also leave open the possibility that independent genetic influences may emerge over time. Because single-method correlations (e.g., based on NC scores only) may inflate true correlations between variables (Bank, Dishion, Skinner, & Patterson, 1990), we conducted parallel analyses with a test of reading comprehension at age 10.

## Method

### *Sample*

Participants were from the Twins Early Development Study (TEDS), a longitudinal study of twins ascertained from population records of live twin births in England and Wales (Oliver & Plomin, in press). Zygosity was determined for same-sex twins using polymorphic DNA markers (Freeman et al., 2003) or a questionnaire of twin similarity, completed by parents at ages 2, 3, and 4 years (Price et al., 2000). For this study, we assessed families of twins born between January 1994 and August 1995 at ages 7, 9, and 10 years. Each child received a t-shirt at age 7, and a bookstore voucher at ages 9 and 10.

Children were being taught according to the U.K. National Curriculum (NC), a statutory school curriculum followed by all state-supported schools in the United Kingdom (Department for Education and Employment, 2000). In the elementary school years, the NC is split into two levels or *Key Stages*. Children from Kindergarten to age 7 follow Key Stage 1. Reading instruction at this level primarily focuses on developing word-level reading processes such as phonological awareness, word recognition, and grammatical awareness. Additionally, children are encouraged to use their background knowledge and knowledge of print conventions to derive meaning from text. Between ages 8 and 11, children move up to Key Stage 2. At this stage, reading instruction continues to emphasize word-level reading processes, but greater weight is placed on developing skills needed to understand text and read for information—for example, making inferences, skimming for the gist of the content, and distinguishing between fact and opinion.

Twins were excluded if teacher assessments (described below) were returned more than 90 days after they were originally sent, in order to minimize variation in the length of teacher contact time with children at the time of assessment. We also excluded pairs in which one or both twins had a neurological condition or specific medical syndrome (e.g., cystic fibrosis). Children whose first language in the home

was not English were not excluded because English as a second language was not associated with either teacher or test assessments. The final sample available for study consisted of 4,291 twin pairs: 1,561 monozygotic (MZ) twin pairs (718 male pairs, 843 female pairs), 1,407 same-sex dizygotic (DZ) pairs (682 male pairs, 725 female pairs), and 1,323 twins in opposite-sex dizygotic (DZO) pairs. The ethnicity of the sample was representative of the population in England and Wales. For example, 92% of the sample was White, compared with 91% in the general population (Sattar, Glen, & Diaper, 2005).

### *Measures and Procedure*

Reading achievement was assessed by teacher assessments at ages 7, 9, and 10. A web-based test was used to assess reading comprehension at age 10. At each of the three time occasions, informed parental consent was obtained. Parental consent was also obtained to contact teachers of twins, and informed consent was obtained from teachers. Of families who consented for teachers to be contacted, 80.3–87.0% of teachers consented and participated in the NC assessments.

### *Teacher Assessments*

Teachers assessed children's reading at ages 7, 9, and 10 years using a rating scale of general reading achievement that referenced U.K. NC achievement goals for literacy. The assessment at age 7 was based on Key Stage 1 criteria (shown in Appendix A1). Children can be awarded achievement levels ranging from 0 to 4. The assessments at ages 9 and 10 were based on Key Stage 2 (shown in Appendix A2). Children can be awarded achievement levels ranging from 1 to 5. There was a significant positive correlation between age at the time of assessment and NC scores: older children had higher NC scores (7-year NC:  $r = .08$ , confidence interval [CI]: .03–.09; 9-year NC:  $r = .08$ ; CI: .05–.11; 10-year NC:  $r = .12$ ; CI: .09–.15). Consequently, all analyses were based on age-adjusted NC scores.

Face validity of NC scores is provided by the distribution of scores. At all three waves of assessment, scores were normally distributed, with minimal skew or kurtosis. In support of the validity of the NC teacher assessments at Key Stage 1 and Key Stage 2, there is evidence for substantial agreement between these assessments and scores on group-administered NC reading tests. For example, cross-tabulation analyses on a nationwide sample of 600,000 children have shown that agreement be-

tween NC teacher assessments of reading and scores on group-administered NC reading tests at Key Stage 1 is good (Cohen's  $\kappa = .80$ ; Dale, Harlaar, & Plomin, 2005). Similarly high levels of agreement have been reported for Key Stage 2 teacher and test assessments (Reeves, Boyle, & Christie, 2001).

At each wave of assessment, teacher assessments were obtained by postal questionnaire during the spring semester. Twins in the same classroom were assessed by the same teacher, whereas twins in different classrooms were assessed by different teachers. Most primary schools in the United Kingdom have no formal policy about educating twins in the same classroom or in different classrooms (British Broadcasting Corporation, 2001). Placement in the same or different classrooms in our sample is not systematically related to zygosity at age 7,  $\chi^2_{(1)} = 1.26$ ;  $p = .26$ , age 9,  $\chi^2_{(1)} = 7.98$ ;  $p = .37$ , or age 10,  $\chi^2_{(1)} = .154$ ;  $p = .69$ . Twins were more likely to be rated by the same teacher at ages 7 and 9, but by different teachers at age 10 (64.7% at age 7, 58.5% at age 9, and 21% at age 10).

#### *Web-based Assessment*

At age 10, participants completed a web-based test battery at home that included adaptation of the reading comprehension subtest of the Peabody Individual Achievement Test (Markwardt, 1997) (hereafter referred to as PIAT<sub>rc</sub>). To maximize sample participation, the test battery was designed with minimal software requirement, and vouchers were given to remunerate families for the cost of Internet use. Families without a computer were given the opportunity to complete the assessments at the Institute of Psychiatry. Parents were reimbursed for the cost of Internet use during testing.

Families who consented to participate in the web-based testing received a secure username and password providing access to the test pages (at [www.teds.ac.uk](http://www.teds.ac.uk)). The PIAT<sub>rc</sub> assesses literal comprehension of sentences. Sentence items were presented visually and with oral instructions given by the computer using digitized speech. The children responded with the mouse, moving the pointer to the desired location and clicking on it. All the children started with the same items, but an adaptive algorithm modified item order and test discontinuation in order to tailor difficulty to the ability level of each participant. Participants could attempt each item only once. Parents were asked to supervise the testing by ensuring that each twin completed the test separately and did not observe their co-twin's test. The web-based adaptation of the PIAT<sub>rc</sub> contained

the same practice items, test items, and instructions as the original published test. Credit (automatic score of 1) was given for all items that were skipped due to upward branching. PIAT<sub>rc</sub> total scores were derived by summing correct and credited scores.

The reliability of the PIAT<sub>rc</sub> is indicated in two ways. First, the correlation between in-person testing of the standard PIAT<sub>rc</sub> test (without adaptive algorithms) and our web-based version is .89 ( $n = 30$ ). Second, the PIAT<sub>rc</sub> shows good internal consistency (Cronbach's  $\alpha = .95$ ). We should also note that test error due to unreliable scoring (e.g., clerical error or improper application of standard scoring criteria) is eliminated in the web-based PIAT<sub>rc</sub> because scoring does not require human transcription and data entry.

#### *Sample Representativeness*

We try to minimize the effects of attrition by emphasizing that families may skip individual assessment waves at any time, without dropping out of the study completely. As a result, overall rates of attrition in TEDS are low. However, response rates to NC assessments were variable—79.6%, 57.7%, and 59.9%, for NC assessments at ages 7, 9, and 10 years, respectively, and 59.8% for the PIAT<sub>rc</sub> at 10 years.

Variability in response rates can be problematic for behavioral genetic studies if the sample units that provide valid information are systematically different from those that are missing from the study—for example, if children with poor reading abilities are more likely to drop out of the study (Taylor, 2004). To assess this possibility, we used a linear regression analysis to estimate the extent completeness of participation predicts NC scores. Completeness of participation in NC assessments was assessed on a 3-point scale ( $n$  refers to families): (1) NC scores available for one wave of assessment only ( $n = 2,050$ ); (2) NC scores available for two assessment waves (ages 7 and 9, 7 and 10, or 9 and 10;  $n = 2,614$ ); and (3) NC scores available at all assessment waves (ages 7, 9, and 10;  $n = 5,515$ ). Because pairs of twins are nested within the same families, we used the Huber–White variance estimator in STATA (StataCorp, 2005) to correct for the nonindependence of data. Completeness of participation was positively associated with NC scores at age 7 ( $\beta = .11$ ,  $CI = .09-.13$ ), age 9 ( $\beta = .17$ ,  $CI = .13-.21$ ), and age 10 ( $\beta = .12$ ,  $CI = .08-.15$ ). That is, at each age, children with higher NC scores were also more likely to participate at each age. We should note, however, that the large sample size means that even minor and substantively trivial associations are significant, and thus it is more appropriate to consider effect size as

an indicator of “significance.” The effect sizes were completeness of participation accounted for no more than 2% of the variance in scores at each age—a negligible effect size (Cohen, 1988).

We also examined the relationship between participation and maternal educational qualifications, which may be regarded as a proxy for socioeconomic status (SES). Maternal educational qualifications were coded on a 7-point scale, from “no educational qualifications” through “postgraduate qualifications.” There was a significant correlation between completeness of participation and maternal education (Spearman’s  $r = .09$ ): twins participating at all three ages tended to have mothers who were more likely to have completed mandatory education, compared with mothers of twins who did not participate at each assessment wave. Similar results were obtained for the PIAT<sub>rc</sub>, which was coded as either completed (“1”) or missing (“0”) for each family. Participation in the PIAT<sub>rc</sub> assessment was significantly correlated with maternal education (point biserial  $r = .15$ ). By conventional standards (Cohen, 1988), correlations of .09 and .15 can be regarded as being of small effect size.

Overall, these findings suggest that our sample is skewed toward families with higher maternal education, although the effect sizes of the associations are marginal. This finding suggests that some caution is needed in generalizing our findings to other samples. However, we can probably assume that children with reading difficulties are not significantly more likely to drop out of the study.

### Analyses

Our genetic analyses were based on standard quantitative genetic principles for twin data. When longitudinal data are available, the phenotypic (observed) variance within each variable at each age and the covariance between variables across ages can be attributed to the combined influences of additive genetic variance and environmental variance. Additive genetic (A) variance reflects variation in genotypes transmitted from parents to offspring. Environmental variance is divided into two parts: shared environmental (C), reflecting variation in nongenetic influences that affect all persons within a family to the same degree (e.g., family SES), and nonshared environmental (E) (e.g., differential educational experiences), reflecting variation in environment influences that cause individual family members to differ from one another. A, C, and E components can be estimated from twin variances and covariances based on the assumption that fa-

miliar resemblance due to genetic influences varies as a function of zygosity (MZ twins are genetically identical, whereas DZ twins share on average 50% of their segregating genes), whereas familial resemblance due to shared environmental influences does not (environmental influences that contribute to familial resemblance are assumed to affect MZ and DZ twins equally) (Plomin, DeFries, McClearn, & McGuffin, 2001).

We used a Cholesky decomposition framework to examine the relative contributions of A, C, and E factors to the longitudinal relationships between NC reading scores at ages 7, 9, and 10, and the longitudinal relationships between 7-year NC scores and PIAT<sub>rc</sub> scores. For the purpose of these analyses, the Cholesky decomposition models were compared with variance–covariance matrices derived from raw age- and sex-corrected (McGue & Bouchard, 1984) data. Model parameters and 95% CIs were estimated by full-information maximum likelihood estimation using the computer program *Mx* (Neale, Boker, Xie, & Maes, 2002); this method allows missing data to be taken into account.

## Results

### *Descriptive Statistics*

Table 1 presents the means and standard deviations for 7-, 9-, and 10-year NC scores. Two comparisons are of interest. First, we examined the extent to which NC scores at each age differ as a function of sex and zygosity. A  $2 \times 2$  (sex  $\times$  gender) multivariate analysis of variance on NC scores revealed a main effect for sex, with boys scoring significantly lower at each age, on average, compared with girls,  $F(3, 3204) = 21.51, p < .01$ . Examination of effect sizes shows that gender accounted for no more than 2% of the variance at each age as estimated by partial eta squared ( $\eta^2$ ). With regard to zygosity, DZ twins scored somewhat higher at each age, on average, compared with MZ twins, but the main effect for zygosity was not significant,  $F(3, 3204) = 1.05, p = .37$ , and the effect size of zygosity on NC scores at each age was negligible ( $\eta^2 = .00$  at each age).

The second comparison of interest was between NC scores for twins in pairs assessed by the same teacher and NC scores for twins in pairs assessed by different teachers. We compared twin covariances for NC scores by teacher status (whether twins were assessed by the same teacher or by different teachers) separately at each age using *Mx*. In this analysis, variances and covariances for NC scores at ages 7, 9, and 10 years were allowed to differ by teacher status

Table 1  
National Curriculum (NC) Scores by Sex and Zygosity at Ages 7, 9, and 10 Years

	7-year NC			9-year NC			10-year NC		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
By zygosity									
MZ	2.12	.70	1,237	3.11	.76	899	3.49	.82	921
DZ	2.18	.69	2,179	3.14	.75	1,579	3.56	.66	1,651
By sex									
Males	2.08	.71	1,651	3.03	.78	1,179	3.45	.83	1,218
Females	2.24	.66	1,765	3.22	.71	1,299	3.62	.78	1,354
By sex and zygosity									
MZM	2.03	.72	576	2.98	.80	411	3.38	.84	406
DZM	2.09	.71	556	3.05	.77	375	3.52	.83	394
MZF	2.21	.67	661	3.22	.71	488	2.59	.78	515
DZF	2.26	.66	568	3.24	.73	429	3.60	.80	437
DZOm	2.12	.69	519	3.07	.78	393	3.46	.83	418
DZO <sub>f</sub>	2.27	.65	536	3.21	.71	382	3.68	.75	402

Note. Different = twins in pairs assessed by different teachers; DZF = dizygotic females; DZM = dizygotic males; DZO<sub>f</sub> = female twins in opposite-sex DZ pairs; DZO<sub>m</sub> = male twins in opposite-sex DZ pairs; MZF = monozygotic females; MZM = monozygotic males; *n* = number of pairs; Same = twins in pairs assessed by the same teacher.

and then were systematically equated. Equating the covariances resulted in a significant deterioration in model fit at each age, with the results indicating that twins assessed by the same teacher showed greater covariance (i.e., resemblance) than twins assessed by different teachers (details available on request from the first author). Further analyses indicated that this was the case for both MZ and DZ twins. This finding of greater resemblance in twins assessed by the same teacher is consistent with two interpretations. First, teacher ratings may include individual rater response tendencies (e.g., stereotyping, idiosyncratic response styles). To the extent that rater tendencies are shared across twins, twins assessed by the same teacher may be expected to show greater resemblance than twins assessed by different teachers. A second, not mutually exclusive, interpretation is that the greater resemblance reflects a real effect of being in the same classroom. Because our data do not allow us to disentangle these possibilities, and in order to maximize the power of our analyses (especially given the uneven proportions of children assessed by the same teacher and children assessed by different teachers at ages 7 and 10), we only present findings for the whole sample in this paper.

### How Stable is Reading From 7 to 10 Years?

NC scores across 7, 9, and 10 years correlated significantly across ages: .62 (95% CIs: .60–.63) from

7 to 9 years, .59 (.57–.61) from 7 to 10 years, and .63 (.61–.64) from 9 to 10 years. These correlations each represent a large effect size, according to conventional standards (Cohen, 1988).

### Twin Correlations

In a first stage of analysis, we calculated within-age, cross-twin correlations, and cross-age, cross-twin correlations between NC scores at ages 7, 9, and 10 years. Within-age, cross-twin correlations (e.g., the correlation between 7-year NC scores in Twin 1 and 7-year NC scores in Twin 2) can be used to decompose the *variance* of NC scores at a particular age into genetic and environmental influences. These correlations are presented in Table 2. If we focus only on zygosity differences in these correlations, we see consistent evidence for genetic influence as indicated by the higher MZ correlations ( $r_{MZ} = .73-.79$ ) than DZ correlations ( $r_{DZ} = .43-.48$ ). There is also evidence for shared environmental influences because the DZ correlations are more than half the MZ correlations.

Comparing twin correlations by sex as well as by zygosity provides information about possible sex differences in etiology. Specifically, sex-specific genetic or shared environmental influences (*qualitative* sex differences) are suggested when same-sex DZ twin pairs resemble each other to a greater extent than opposite-sex DZ twin pairs, as genetic or environmental influences on NC scores that are specific

Table 2  
 Within-Age and Cross-Age Twin Correlations for 7-, 9-, and 10-Year NC Scores

	7 years		9 years		10 years	
	$r_{ICC}$	CI	$r_{ICC}$	CI	$r_{ICC}$	CI
(A). Within-age twin correlations ( $r_{ICC}$ ) for 7-, 9-, and 10-year NC scores						
By zygosity						
MZ	.79	.77–.81	.75	.72–.78	.75	.71–.77
DZ	.47	.42–.51	.43	.37–.48	.48	.42–.53
By sex						
MZM	.77	.74–.80	.73	.68–.77	.73	.69–.77
DZM	.46	.41–.51	.43	.34–.50	.53	.45–.60
MZF	.78	.75–.80	.77	.73–.80	.74	.70–.78
DZF	.47	.42–.52	.41	.33–.48	.45	.38–.53
DZO	.41	.37–.46	.43	.37–.48	.41	.35–.47
	7–9 years		7–10 years		9–10 years	
	$r_{CA}$	CI	$r_{CA}$	CI	$r_{CA}$	CI
(B). Cross-age twin correlations ( $r_{CA}$ ) for 7-, 9-, and 10-year NC scores						
$r_p$	.61	.58–.64	.59	.56–.65	.62	.59–.65
By zygosity						
MZ	.56	.51–.61	.54	.49–.59	.55	.50–.60
DZ	.32	.27–.37	.32	.28–.37	.31	.26–.36
By sex and zygosity						
MZM	.60	.53–.67	.57	.50–.64	.60	.52–.67
DZM	.36	.27–.45	.36	.27–.45	.31	.20–.41
MZF	.52	.45–.59	.52	.45–.59	.51	.43–.58
DZF	.34	.24–.43	.30	.21–.40	.24	.13–.34
DZO	.26	.21–.35	.31	.24–.38	.35	.28–.42

Note. CI = confidence interval; DZ = same-sex dizygotic twin pairs; DZF = DZ female twin pairs; DZM = DZ male twin pairs; DZO = opposite-sex DZ twin pairs; MZ = monozygotic twin pairs; MZF = MZ female twin pairs; MZM = MZ male twin pairs; NC = National Curriculum;  $r_p$  = phenotypic correlation.

to one sex may be expected to reduce within-pair similarity. Differences in the magnitude of genetic and environmental contributions to phenotypic variance (*quantitative* sex differences) are suggested when resemblance within male and female same-sex twin pairs differs as a function of zygosity. In fact, the within-age correlations are similar for male and female same-sex pairs as well as for same-sex and opposite-sex DZ pairs, indicating no quantitative or qualitative sex differences in etiology.

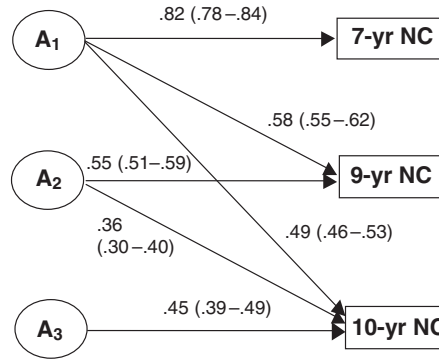
The bottom panel of Table 2 shows the cross-age twin correlations, as well as the phenotypic correlations across age. Cross-twin cross-trait correlations (e.g., the correlation between 7-year NC scores in Twin 1 and 9-year NC scores in Twin 2) can be used to decompose the *covariance*, or stability, between NC scores across ages into genetic and environmental influences. The degree to which MZ correlations exceed DZ correlations indexes the influence of genetic factors on twin similarity and stability. It

should be noted that phenotypic correlations across age create a ceiling for the cross-age correlations. The MZ cross-age correlations ( $r_{MZ} = .51–.60$ ) are greater than the DZ cross-age correlations ( $r_{DZ} = .24–.36$ ), but they are not twice the DZ cross-age correlations when CIs are taken into consideration. These differences suggest that genetic factors mediate the phenotypic stability from age to age. Again, there is no evidence for sex differences in etiology.

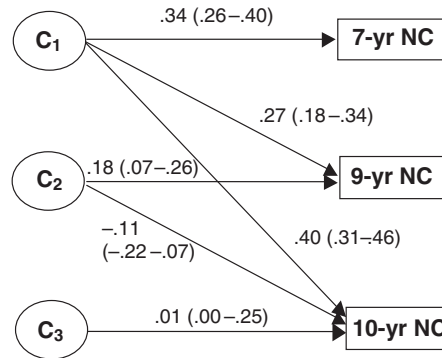
### Genetic and Environmental Influences on the Variance and Covariance of NC Scores?

Etiological patterns suggested by twin correlations can be tested more formally using structural equation models. In a second stage of analysis, we examined the data using a genetic Cholesky decomposition model. In this model, shown in Figure 1

1. Additive genetic influences



2. Shared environmental influences



3. Non-shared environmental influences

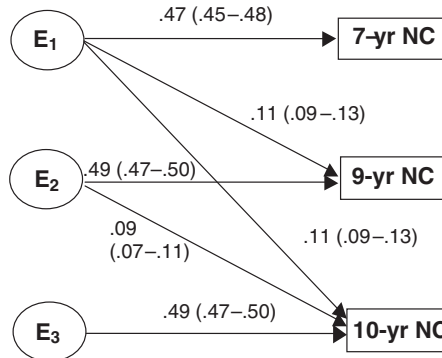


Figure 1. Standardized path coefficients from Cholesky decomposition model for National Curriculum (NC) scores at ages 7, 9, and 10 years. Within-age variances and cross-age covariances between NC scores at ages 7, 9, and 10 years are decomposed into latent additive genetic influences (A; upper panel), shared environmental influences (C; middle panel), and nonshared environmental influences (E; lower panel).

for one member of a twin pair, latent factors represent A, C, and E influences on NC scores at ages 7, 9, and 10 years. The first set of latent factors, A<sub>1</sub>, C<sub>1</sub>, and E<sub>1</sub>, indicate A, C, and E effects that contribute to the total phenotypic variance in 7-year NC scores, as well as A, C, and E effects that contribute to the *shared* variance between NC scores at ages 7, 9, and 10—that is, the stability of NC scores across all three ages. The second set of latent factors, A<sub>2</sub>, C<sub>2</sub>, and E<sub>2</sub>, represent A, C, and E effects that contribute to the covariance between 9- and 10-year NC scores inde-

pendent of A, C, and E effects that they share with 7-year scores. The final set of latent factors, A<sub>3</sub>, C<sub>3</sub>, and E<sub>3</sub>, represent residual A, C, and E effects on 10-year NC scores.

Before examining the relative contributions of genetic and environmental influences on the variance and covariance of NC scores, we tested whether there was evidence for quantitative and qualitative sex differences. The results (available from the first author) indicated that a model specifying no sex differences provided the best fit to the data, as sug-



gested by the twin correlations. Thus, we used the combined sample of boys and girls to test the aims of our study.

Our first aim was to examine the extent to which individual differences in reading experience are due to genetic (and environmental) influences. When the path coefficients in the trivariate Cholesky decomposition model are standardized, the proportion of the total phenotypic variance in each measure that is due to genetic influences (i.e., heritability) can be estimated by summing the squared path coefficients from the latent genetic factors loading on that measure. For example, the heritability of 10-year NC scores is 57%, which is the sum of 24% for  $A_1$ , 13% for  $A_2$ , and 20% for  $A_3$ . Similar calculations can be used to estimate the proportion of phenotypic variance in each measure due to shared environmental influences and nonshared environmental influences. These squared path coefficients and variance estimates are presented in Table 3 for clarity. Briefly, there is evidence for substantial heritability at each age, modest but significant shared environmental influence at each age, and moderate nonshared environmental influence.

Our second aim was to examine the extent to which genetic factors account for the stability of NC

scores from ages 7 to 9. An impression of the results can be obtained from the path coefficients in Figure 1. As described above, the first latent A factor,  $A_1$ , represents additive genetic influences on 7-year NC scores that also contribute to the stability of NC scores across all three ages. The factor loadings are significantly greater than zero, indicating that there is a significant degree of genetic covariance between NC scores at all three waves of assessment. The loadings of the second latent A factor,  $A_2$ , are also significantly greater than zero, indicating that variance in 9-year NC scores is partly due to genetic factors that are independent of genetic factors influencing 7-year NC scores, and that this second source of genetic influence contributes to variance in 10-year NC scores. Finally, the third source of genetic influence,  $A_3$ , loading only on 10-year NC scores, is also significant, indicating that there are genetic influences on individual differences in 10-year NC scores that were not previously expressed at age 7 or 9. The genetic effects suggest substantial genetic stability across all 3 years, and some additional stability specific to ages 9 and 10. In addition, some age-specific genetic influence emerged at age 10. In contrast, shared environment path coefficients (middle panel) are lower and suggest only one factor

Table 3

*Squared Cholesky Path Coefficients for Genetic, Shared Environmental, and Nonshared Environmental Components of Variance for National Curriculum (NC) Scores at Ages 7, 9, and 10*

	Additive genetic influences			
	$A_1$	$A_2$	$A_3$	$a^2$
7-year NC	.67 (.63–.67)			.67 (.62–.72)
9-year NC	.34 (.33–.34)	.31 (.23–.36)		.65 (.57–.73)
10-year NC	.24 (.19–.30)	.13 (.12–.19)	.20 (.12–.23)	.57 (.49–.65)
	Shared environmental influences			
	$C_1$	$C_2$	$C_3$	$c^2$
7-year NC	.11 (.11–.16)			.11 (.07–.16)
9-year NC	.07 (.04–.08)	.03 (.03–.09)		.10 (.03–.18)
10-year NC	.16 (.15–.24)	.01 (.00–.02)	.00 (.00–.07)	.17 (.11–.24)
	Nonshared environmental influences			
	$E_1$	$E_2$	$E_3$	$e^2$
7-year NC	.22 (.20–.23)			.22 (.21–.23)
9-year NC	.01 (.01–.02)	.24 (.22–.26)		.25 (.23–.27)
10-year NC	.01 (.01–.02)	.01 (.00–.02)	.24 (.22–.24)	.26 (.23–.28)

*Note.*  $A_1$ ,  $A_2$ ,  $A_3$  = latent additive genetic influences;  $C_1$ ,  $C_2$ ,  $C_3$  = latent shared environmental influences;  $E_1$ ,  $E_2$ ,  $E_3$  = latent nonshared environmental influences.

Total percentage of variance explained by A, C, and E in final column, estimated by summing the values in each row.

that is stable across all three ages. Nonshared environmental influences, including measurement error, are low and specific to each age (bottom panel).

### *Genetic and Environmental Overlap*

The contribution of genes, shared environmental influences, and nonshared environmental influences can be assessed more formally in terms of the extent to which each of these factors accounts for the stability between 7-, 9-, and 10-year NC scores, and the extent to which these genetic factors correlate across measures.

### **To What Extent Do Genetic/Environmental Factors Account for Stability?**

As described in the Introduction, the estimated phenotypic correlation between two waves of assessment reflects the combined effects of genetic and environmental factors. In the Cholesky decomposition model, the contribution of genetic factors to the phenotypic correlations between 7-, 9-, and 10-year NC scores can be estimated by summing the products of the path coefficients from the genetic factors that load on each pair of measures, and dividing the result by the phenotypic correlation between those measures. These results are presented in Table 4. The first row summarizes the phenotypic correlations. The next three rows indicate the proportion of these phenotypic correlations that can be ascribed to A, C, and E. For example, 77% of the phenotypic correlation of .62 between NC scores at 7 and 9 years is mediated genetically. This value can be derived from the path coefficients in Figure 1. The product of the A paths connecting 7- and 9-year NC

scores ( $.82 \times .58 = .48$ ) represents the genetic contribution to the phenotypic correlation. Dividing this genetic contribution (.48) by the phenotypic correlation (.62) indicates that 77% of the phenotypic correlation is mediated genetically. Genetic influences also substantially mediated the phenotypic correlation between NC scores at ages 7 and 10 (68%) and between NC scores at ages 9 and 10 (77%). The contribution of shared environmental influences to the phenotypic correlations are moderate (.14–.23) and the contribution of nonshared environment is modest (.08–.09).

### **To What Extent Are Genetic/Environmental Factors Correlated Across Measures?**

In addition to estimating genetic and environmental contributions to the phenotypic correlations, we derived the genetic and environmental correlations from the model. The genetic correlation indicates the extent to which individual differences on NC scores at two ages reflect the same genetic influences, whereas the environmental correlations indicate the extent to which individual differences reflect the same (shared or nonshared) environmental influences. Genetic and environmental correlations may take any value between  $-1$  and  $+1$ , and are independent of the extent to which two traits are each influenced by genetic and environmental influences (Plomin & DeFries, 1979; Posthuma et al., 2003).

As shown in the bottom panel of Table 4, the genetic correlations from year to year are substantial (.65–.80), indicating that more than two thirds of the genetic effects are in common from age to age, even from 7 to 10 years. The age-to-age shared environmental correlations are also substantial (.66–.96). Thus, even though shared environment contributes

Table 4

*Additive Genetic (A), Shared Environmental (C), and Nonshared Environmental (E) Contributions to the Stability of Reading Performance (with 95% Confidence Intervals)*

	7–9-year NC	7–10-year NC	9–10-year NC	7-year NC to 10-year PIAT <sub>rc</sub>
$r_p$	.62 (.60–.63)	.59 (.57–.61)	.63 (.61–.64)	.44 (.42–.46)
Percentage of $r_p$ due to				
A	.77 (.70–.86)	.68 (.60–.76)	.77 (.67–.87)	.68 (.52–.84)
C	.14 (.06–.22)	.23 (.15–.30)	.14 (.05–.22)	.27 (.08–.31)
E	.08 (.06–.11)	.09 (.06–.12)	.09 (.06–.12)	.05 (.01–.10)
Correlations				
A	.73 (.72–.74)	.65 (.60–.72)	.80 (.72–.80)	.56 (.44–.69)
C	.84 (.66–.96)	.96 (.96–1.00)	.66 (.33–1.00)	.77 (.41–1.00)
E	.22 (.21–.28)	.22 (.21–.29)	.22 (.21–.29)	.09 (.01–.16)

Note.  $r_p$  = phenotypic correlation.

only modestly to the variance of NC scores at each age, these are largely the same environmental factors across age, which explains the moderate contribution of shared environmental influences to the phenotypic age-to-age correlations. In contrast, nonshared environmental contributions to stability are much lower.

**What Factors Explain Stability in Reading Achievement as Assessed by Psychometric Tests?**

As described in the Introduction, the use of only one source of information (teachers) could contribute to the phenotypic and genetic stability seen in our analyses of NC scores. For example, it is possible that teachers' ratings are affected by personality characteristics of the children. For this reason, we applied a similar model to the analysis of NC assessments at 7 years and PIAT<sub>rc</sub> scores at 10 years. In addition to different methods of assessment (teacher ratings for NC assessments; web-based administration for the PIAT<sub>rc</sub>), the measures are also designed to assess different reading skills (reading achievement for NC assessments; reading comprehension for the PIAT<sub>rc</sub>).

The results are briefly summarized here. Individual differences in PIAT<sub>rc</sub> scores were significantly heritable (.41; CIs: .40–.50) and were also partly attributable to shared environmental influences (.25; CIs: .16–.32) and nonshared environmental influences and measurement error (.35; CIs: .32–.38). The phenotypic correlation between 7-year NC assessments and PIAT<sub>rc</sub> scores was .44. The key issue is to determine the extent to which genetic and environmental influences contribute to this developmental association. Figure 2 shows that the Cholesky results for the 7-year NC assessments and the PIAT<sub>rc</sub> at 10 years are similar to the results for NC scores (Figure 1) in providing evidence for substantial genetic stability. This is further confirmed in the right-most column of Table 3, which shows that the phenotypic correlation of .44 between the 7-year NC assessments and the PIAT<sub>rc</sub> is due primarily to genetic influences (68%), which is identical to the genetic mediation between 7- and 10-year NC scores (68%). The genetic correlation between the 7-year NC assessments and PIAT<sub>rc</sub> is .56, again very similar to the genetic correlation between the 7- and 10-year NC scores (.65).

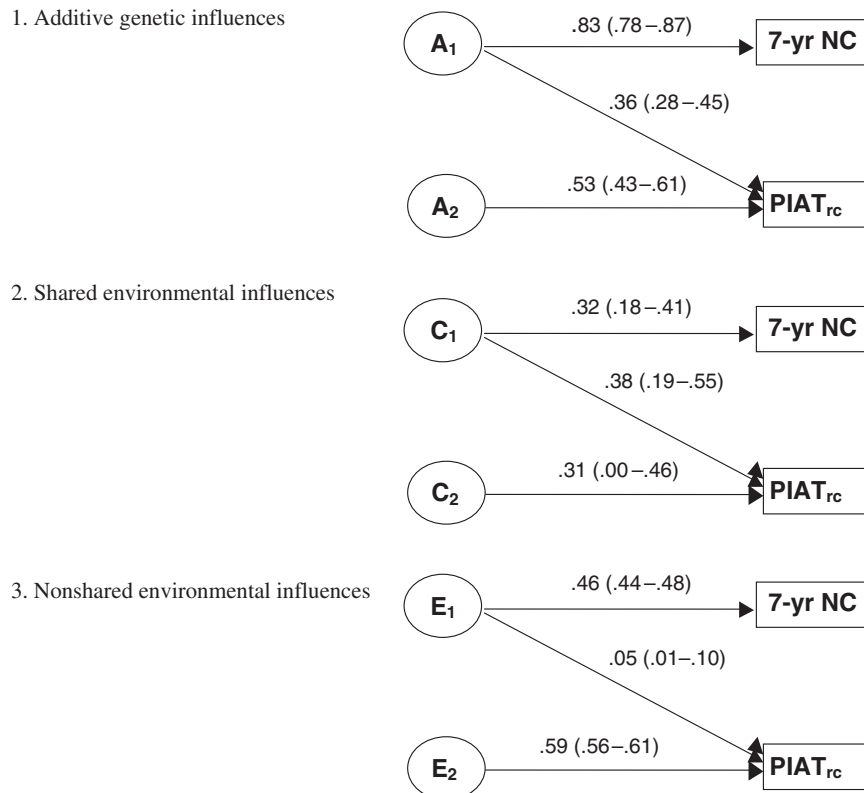


Figure 2. Standardized path coefficients (unsquared) from Cholesky decomposition of 7-year NC assessments and 10-year PIAT<sub>rc</sub> scores.

## Discussion

Teaching and assessment of reading achievement in the elementary school years shift from a primary focus on developing word-reading skills to a primary focus on reading comprehension. The results of this longitudinal study of reading from 7 to 10 years support the conclusion that genetic influence is substantial and stable during the elementary school years, across the developmental shift from “learning to read” to “reading to learn.”

### *Substantial Heritability of Individual Differences in NC Scores*

To date, the majority of genetic research has examined the etiology of individual variation in measures of reading achievement at only a single point in time and with samples spanning a wide age range. The results of the present study suggest that etiologies of NC scores of reading achievement at ages 7, 9, and 10 years are highly similar. Averaging across ages, genetic influences accounted for 63% of the variance in NC scores, shared environmental influences accounted for 13%, and nonshared environmental influences, including measurement error, accounted for the remaining 24%. These findings indicate that estimates of the proportion of variance in NC scores due to A, C, and E remained relatively constant at each age.

### *Substantial (but not Complete) Genetic and Shared environmental Overlap From 7 to 10 Years*

At a phenotypic level, NC scores were significantly correlated across ages 7, 9, and 10, despite the shift from Key Stage 1 to Key Stage 2 of the U.K. NC. In this context, a key finding of our study was that the stability could primarily be explained by genetic influences: about two thirds of the phenotypic correlations between ages 7 and 9, ages 7 and 10, and ages 9 and 10 were mediated genetically. We may interpret this genetic stability as indicating that the continuity of NC scores is largely due to genetic influences. That is, genetic influences partly explain why one child in our population sample has lower or higher NC scores than their peers, and genetic influences also partly explain why children’s reading performance is relatively stable across the elementary school years.

A related finding was that the age-to-age genetic correlations were considerable both for teacher assessments (e.g., .65 from 7 to 10 years) and for NC scores at age 7 and the PIAT<sub>rc</sub> at age 10. These

correlations may arise because some of the same genes that influence reading performance are expressed at significant levels through development, or because the effects of early-expressed genes (which result in, e.g., differences in neural networks) directly influence later reading performance, or some combination of these scenarios. Moreover, because the genetic correlation indexes the total complement of genetic influences correlated across development, we would predict that this is not just the case for one gene, but for many genes.

Despite the substantial genetic stability, age-to-age genetic correlations were not unity and our model-fitting results showed significant age-specific genetic influences at each age. Why should this finding emerge? One possibility is that these “new” genetic influences are concomitant with changes in cognitive and reading development. For example, the development of reading comprehension likely makes greater demands on inference making, anaphoric processing, and metacognitive knowledge and monitoring (Oakhill & Cain, 2003), which may, in turn, depend on partly different genes from those influencing the development of word recognition.

In contrast to genetic influences, shared and nonshared environmental influences made relatively smaller contributions to the phenotypic correlations across ages. Shared environmental influences nevertheless were almost perfectly correlated from ages 7 to 10, and there was little evidence for reliable age-specific environmental influences emerging at ages 9 or 10. This finding points to factors that contribute to individual differences in reading achievement and tend to remain relatively constant through childhood, for example, family or school SES, or the U.K. NC, which provides a common reading curricula for all children. Nonshared environmental influences, in contrast, showed little correlation across ages and were largely measure specific. This likely reflects, at least in part, measure-specific error. However, it is also possible that there are some real person-specific influences that have important but relatively ephemeral influences on reading achievement. These may include, for example, child-specific interactions between instructional practices and the child’s ability level (e.g., Connor, Morrison, & Petrella, 2004).

Overall, the findings from this study chime with those from previous twin studies, particularly with respect to genetic stability. Some points of differences are of note, however. Our finding of age-specific genetic influences is at odds with the Colorado Adoption Project, which found no reliable evidence for new genetic influences at ages 12 or 16. Similarly,

we found no evidence for age-specific shared environmental influences, whereas the Western Reserve Reading Project observed age-specific shared environmental influences for letter knowledge and word knowledge. It is important to note that estimates from quantitative genetic studies are always specific to the population under study, and thus the discrepancies may reflect differences in environmental or genetic variance across populations. However, mixed findings may also arise due to differences in measures, age of assessment, developmental shifts in reading strategies, statistical power, or some combination of these factors (and others). Some resolution of this issue may be achieved in a multitrait (e.g., word recognition, phonological awareness, comprehension monitoring) by multisource (e.g., psychometric tests, teacher assessments, experimental tasks) study, in which a diverse set of reading skills are assessed longitudinally.

#### *Genes and Environment Together*

In summary, this study adds to longitudinal genetic research characterizing the stability of reading development. Our results suggest that there is a genetic “core” to continuity in reading achievement from ages 7 to 9, and shared environmental influences comprise a second constant source of variability in reading development over this period. Given the importance of both genetic and environmental influences to the stability of reading development, a third putative source of stability, not directly examined here, is the interplay between genetic and environmental influences. For example, compared with children who find reading difficult, a child with genes that positively influence their ability to learn to read may be more likely to establish a habit of reading and to seek out more reading-enriched environments, an *evocative gene–environment correlation* (Petrill, Deater-Deckard, Schatschneider, & Davis, 2005). Moreover, this kind of interplay between genes and environments probably begins very early in childhood, before children even begin to read. For example, exposure and quality of preschool environmental experiences (e.g., shared book reading) may partly depend on genetic propensities of the parent (a *passive gene–environment correlation*) or their response to genetic characteristics of the child (a *reactive gene–environment correlation*), and in turn, this genetic contribution to individual differences in environmental experiences may influence subsequent reading achievement (Oliver, Dale, & Plomin, 2006). These scenarios suggest the possibility that the interplay between genes and environments set the

individual on a certain developmental “track” or trajectory, and as such is likely to contribute to stability in reading achievement. Research on the specific genes that confer risk for reading disability (Fisher & Francks, 2006), as well as specific environmental risk factors (Snowling & Hayiou-Thomas, 2006) is underway, and is likely to be particularly important for better understanding how genes and environments work together in reading development.

#### *Limitations*

In addition to the general methodological limitations of the twin design (Plomin et al., 2001), at least four caveats should be noted. First, our teacher assessments involve an overall rating of reading achievement that does not differentiate between specific reading skills, such as reading comprehension and word recognition. Although the Key Stage 2 criteria used to evaluate children’s reading achievement at ages 9 and 10 years place more weight on reading comprehension and higher order reading processes than Key Stage 1 criteria, it would be desirable to obtain specific measures of multiple reading skills at each of the three ages in order to clarify more precisely the interrelationships between lower and higher order reading skills across the years tested in this study.

A second difficulty concerns test validity. Although levels of agreement between NC teacher and test assessments of achievement are high (Dale et al., 2005; Reeves et al., 2001), there is evidence that teacher assessments may be influenced by nonreading characteristics of the children. For example, an analysis of Key Stage 2 NC scores showed that teacher assessments were more likely to be lower than test results when pupils had special educational needs (Reeves et al., 2001). These limitations do not apply to the PIAT<sub>rc</sub> but psychometric measures, especially when administered in nonstandard ways such as the Internet, may have their own problems.

A third and related problem is that we were not able to assess or control for individual rater response tendencies (e.g., stereotyping, idiosyncratic response styles). As described in the Method section, covariances for twins assessed by the same teacher were greater than for twins assessed by different teachers at each age, suggesting that twins with the same teacher were rated more similarly. Moreover, this was the case for both MZ and DZ twins. At a genetic level, a published assessment of same- and different-teacher ratings of NC scores at age 7 showed that MZ

and DZ twins were rated more similarly when they were assessed by the same teacher than when they were rated by different teachers (Harlaar et al., 2005). As a result, individual differences in NC scores in the same-teacher group were characterized by a lower nonshared environmental contribution and higher shared environmental contribution compared with the different-teacher group. However, as described in the Method section, interpretation of these findings is not wholly clear-cut. In a genetic analysis, higher shared environmental influences in twins assessed by the same teacher could reflect a real effect, in addition to (or instead of) rater response biases, because it is conceivable that the classroom context creates more shared environmental influences for twins in the same classroom. Our data do not make it possible to resolve this issue, however. Ultimately, steps should be taken to examine the extent of such biases, and where possible, to minimize them (Hoyt, 2000).

Finally, the present study is currently limited to three measurement points during the elementary school years. The development of reading skill, and particularly reading comprehension, clearly continues into middle and high school, however (e.g., Biancarosa & Snow, 2004). Moreover, there is some evidence that a significant minority of children develop "late-emerging" reading difficulties (Catts, Hogan, & Adlof, 2005; Leach, Scarborough, & Rescorla, 2003). These are children who respond well to early reading instruction, but who begin to show reading problems around the fourth grade, a phenomenon sometimes referred to as the "fourth-grade slump" (Chall, 1983; Snow, Burns, & Griffin, 1998). Further longitudinal research on individual differences in reading achievement in middle and high school is needed to elucidate genetic and environmental contributions to stability and change to reading development.

### Conclusions

We conclude that reading achievement rated by teachers during the elementary school years is highly heritable and shows substantial genetic stability across these years. Unravelling the complex processes in reading development is likely to be greatly facilitated if the specific genes that contribute to the substantial heritability and genetic stability of reading achievement can be found, in conjunction with detailed analyses of the interplay between genetic and environmental influences.

### References

- Bank, L., Dishion, T. J., Skinner, M. L., & Patterson, G. R. (1990). Method variance in structural equation modeling: Living with "glop". In G. R. Patterson (Ed.), *Depression and aggression in family interaction* (pp. 247–277). Hillsdale, NJ: Erlbaum.
- Bates, T. C., Castles, A., Coltheart, M., Gillespie, N., Wright, M., & Martin, G. (2004). Behaviour genetic analyses of reading and spelling: A component processes approach. *Australian Journal of Psychology, 56*, 115–126.
- Biancarosa, G., & Snow, C. E. (2004). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York*. Washington, DC: Alliance for Excellent Education.
- British Broadcasting Corporation. (2001, August 28). Twins in schools—Together or apart? Retrieved May 15, 2006, from <http://news.bbc.co.uk/1/hi/education/1480297.stm>
- Byrne, B., Samuelsson, S., Wadsworth, S., Hulstander, J., Corley, R., DeFries, J. C., et al. (2006). Longitudinal twin study of early literacy development: Preschool through Grade 1. *Reading and Writing*, May 20 [Epub ahead of print].
- Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chall, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child × instruction interactions. *Journal of Educational Psychology, 96*, 682–698.
- Dale, P. S., Harlaar, N., & Plomin, R. (2005). Telephone testing and teacher assessment of reading skills in 7-year-olds: I. Substantial correspondence for a sample of 5544 children and for extremes. *Reading and Writing, 18*, 385–400.
- De Jong, P. F., & van der Leij, A. (2002). Effects of phonological abilities and linguistic comprehension on the development of reading. *Scientific Studies of Reading, 6*, 51–77.
- Department for Education and Employment. (2000). *QCA key stage 1: Assessment and reporting arrangements*. London: HMSO.
- Fisher, S. E., & Francks, C. (2006). Genes, cognition and dyslexia: Learning to read the genome. *Trends in Cognitive Sciences, 10*, 250–257.
- Freeman, B., Smith, N., Curtis, C., Hockett, L., Mill, J., & Craig, I. W. (2003). DNA from buccal swabs recruited by mail: Evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behavior Genetics, 33*, 67–72.
- Gayán, J., & Olson, R. K. (2003). Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology, 84*, 97–123.

- Harlaar, N., Dale, P. S., & Plomin, R. (2005). Telephone testing and teacher assessment of reading skills in 7-year-olds: II. Strong genetic overlap. *Reading and Writing, 18*, 401–423.
- Hohnen, B., & Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology, 35*, 590–603.
- Hopper, J. L. (2000). Why 'common environmental effects' are so uncommon in the literature. In T. D. Spector, H. Snieder, & A. J. MacGregor (Eds.), *Advances on twin and sib-pair analysis* (pp. 151–165). London: Greenwich Medical Media.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods, 5*, 64–86.
- Leach, J. M., Scarborough, H. S., & Rescorla, L. (2003). Late-emerging reading disabilities. *Journal of Educational Psychology, 95*, 211–224.
- Leppänen, U., Niemi, P., Aunola, K., & Nurmi, J.-K. (2006). Development of reading and spelling Finnish from preschool to Grade 1 and Grade 2. *Scientific Studies of Reading, 10*, 3–30.
- Markwardt, F. C. Jr. (1997). *Peabody individual achievement test-revised (normative update) manual*. Circle Pines, MN: American Guidance Service.
- McGue, M., & Bouchard, T. J., Jr. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14*, 325–343.
- Neale, M. C., Boker, S. M., Xie, G., & Maes, H. M. (2002). *Mx: Statistical modeling* (6th ed.). Richmond, VA: Department of Psychiatry.
- Oakhill, J., & Cain, K. (2003). The development of comprehension skills. In T. Nunes & P. Bryant (Eds.), *Handbook of children's literacy* (pp. 155–180). Dordrecht: Kluwer Academic Publishers.
- Oliver, B., Dale, P. S., & Plomin, R. (2006). Predicting literacy at age 7 from preliteracy at age 4: A longitudinal genetic analysis. *Psychological Science, 16*, 861–865.
- Oliver, B., & Plomin, R. (in press). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research and Human Genetics*.
- Parrila, R., Aunola, K., Leskinen, E., Nurmi, J.-E., & Kirby, J. R. (2005). Development of individual differences in reading: Results from longitudinal studies in English and Finnish. *Journal of Educational Psychology, 97*, 299–319.
- Pennington, B. F., & Olson, R. K. (2005). Genetics of dyslexia. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 453–472). Oxford, UK: Blackwell Publishing.
- Petrill, S. A., Deater-Deckard, K., Schatschneider, C., & Davis, C. (2005). Measured environmental influences on early reading: Evidence from an adoption study. *Scientific Studies of Reading, 9*, 237–259.
- Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., & DeThorne, L. S. (2006). Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. *Reading & Writing*, May 25 [Epub ahead of print].
- Plomin, R., & DeFries, J. C. (1979). Multivariate behavioural genetic analysis of twin data on scholastic abilities. *Behavior Genetics, 9*, 505–517.
- Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: Worth Publishers.
- Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., von Helmborg, J. B., Iachine, I., et al. (2003). Theory and practice in quantitative genetics. *Twin Research, 6*, 361–376.
- Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*, 129–133.
- Reeves, D. J., Boyle, W. F., & Christie, T. (2001). The relationship between teach assessments and pupil attainments in standard test tasks at Key Stage 2, 1996–98. *British Educational Research Journal, 27*, 141–160.
- Sattar, G., Glen, F., & Diaper, A. (2005). Results from the General Household Survey, 2004. Retrieved August 8, 2006, from <http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=5756&Pos=&ColRank=1&Rank=272>
- Scarborough, H. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–110). Timonium, MD: York Press.
- Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Snowling, M. J., & Hayiou-Thomas, M. E. (2006). The Dyslexia spectrum: Continuities between reading, speech, and language impairments. *Topics in Language Disorders, 26*, 108–124.
- StataCorp. (2005). *Stata statistical software: Release 9.0*. College Station, TX: Stata Press.
- Taylor, A. (2004). The consequences of selective participation on behavioural–genetic findings: Evidence from simulated and real data. *Twin Research, 47*, 85–504.
- Wadsworth, S. J., Corley, R. P., Hewitt, J. K., & DeFries, J. C. (2001). Stability of genetic and environmental influences on reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Behavior Genetics, 31*, 353–359.
- Wadsworth, S. J., Corley, R. P., Plomin, R., Hewitt, J. K., & DeFries, J. C. (2006). Genetic and environmental influences on continuity and change in reading achievement in the Colorado Adoption Project. In A. Huston & M. Ripke (Eds.), *Developmental contexts in middle childhood: Bridges to adolescence and adulthood* (pp. 87–106). Cambridge, UK: Cambridge University Press.

### Appendix A: National Curriculum Scales for Reading Achievement

#### *A1. Key Stage 1 (KS1) Teacher Assessment Scale for Reading (Used at Age 7)*

- 0 Not yet functioning at Level 1
- 1 Pupils recognize familiar words in simple texts. They use their knowledge of letters and sound–symbol relationships in order to read words and to establish meaning when reading aloud. In these activities they sometimes require support. They express their response to poems, stories, and nonfiction by identifying aspects they like.
- 2 Pupils' reading of simple texts shows understanding and is generally accurate. They express opinions about major events or ideas in stories, poems, and nonfiction. They use more than one strategy, such as phonic, graphic, syntactic, and contextual, in reading unfamiliar words and establishing meaning.
- 3 Pupils read a range of texts fluently and accurately. They read independently, using strategies appropriately to establish meaning. In responding to fiction and nonfiction they show understanding of the main points of express preferences. They use their knowledge of the alphabet to locate books and find information.
- 4 Reading is substantially more advanced than most pupils at Level 3.

#### *A2. Key Stage 2 (KS2) Teacher Assessment Scale for Reading (used at ages 9 and 10)*

- 1 Not yet functioning at Level 2
- 2 Pupil's reading of simple texts shows understanding and is generally accurate. They express opinions about major events or ideas in stories, poems, and nonfiction. They use more than one strategy, such as phonic, graphic, syntactic, and contextual, in reading unfamiliar words and establishing meaning.
- 3 Pupils read a range of texts fluently and accurately. They read independently, using strategies appropriately to establish meaning. In responding to fiction and nonfiction they show understanding of the main points and express preferences. They use their knowledge of the alphabet to locate books and find information.
- 4 In responding to a range of texts, pupils show understanding of significant ideas, themes, events, and characters, beginning to use inference and deduction. They refer to the text when explaining their views. They locate and use ideas and information.
- 5 Reading is substantially more advanced than most pupils at Level 4.