Reading and Writing (2005) 18:401–423
DOI 10.1007/s11145-005-0271-1

Telephone testing and teacher assessment of reading skills in 7-year-olds: II. Strong genetic overlap

NICOLE HARLAAR[1], PHILIP S. DALE[2] and ROBERT PLOMIN[1]
[1]*MRC Social, Genetic, and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, King's College London, Box Number P083, SE5 8AF, Denmark Hill, London, UK;* [2]*Department of Communication Science and Disorders, University of Missouri, Columbia, USA*

Abstract. In a companion paper, word recognition skills assessed by telephone using the Test of Word Reading Efficiency (TOWRE) were found to correlate highly with National Curriculum (NC) teacher-assessed reading ability in 7-year-old twins. This study examined the genetic and environmental origins of this high correlation. TOWRE and NC scores were both highly heritable and the correlation between them was largely due to overlapping genetic effects. These findings were obtained both across the normal range of reading abilities and at the low extreme, defined by scores below a 13.4% cut-off on either measure. TOWRE and NC scores may provide promising phenotypes for further study of the aetiology of early reading abilities and disabilities.

Key words: Genetics, Individual Differences, Teacher Assessments, Twins, Word Recognition

In the preceding paper (Dale, Harlaar, & Plomin, 2005), we examined the relationship between a telephone-adaptation of a test of word recognition, the Test Of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), and teacher assessments of reading ability based on UK National Curriculum (NC) criteria (Department for Education and Employment, 2000). Replicating previous research on teacher judgements and test-assessed reading achievement in school-age children (Demarary & Elliott, 1998; Feinberg & Shapiro, 2003; Hecht & Greenfield, 2001, 2002; Triga, 2004), TOWRE and NC scores were found to be substantially correlated ($r = 0.69$) in a sample of 2772 pairs of 7-year-old twins. There was also good agreement between these measures for the identification of children in the lowest 11.6% of the distribution. As such, these measures appear to provide a useful assessment alternative in situations in which it is not feasible to administer standard psychometric tests of reading ability.

The apparent viability of using telephone and teacher assessments in large-scale research samples is relevant in the context of research examining the aetiology of reading abilities and disabilities. It is widely held that reading disabilities reflect a quantitative dimension of risk that is influenced by multiple genes of small to moderate effect size, known as quantitative trait loci (QTLs). Few of the specific QTLs for reading disabilities have yet been identified and the mechanisms by which the effects of genetic risk factors are mediated are unknown (Fisher & DeFries, 2002). However, current evidence and theory suggests that at least some of the genetic risk factors will contribute to variability in multiple aspects of reading ability and related cognitive abilities, and that elevated risk will also reflect interactions and correlations between genes and environmental risk factors (Fisher, in press; Plomin & Kovas, 2005). These complexities place a strong onus on identifying and employing suitable measures for genetic studies (Smith & Morris, 2005).

Evidence for the heritability of a measure is a prerequisite for genetic studies; that is, variation in scores must be at least partly due to genetic influences. If heritability is zero, for example, gene-finding methods will not succeed. Twin studies, which compare the resemblance of monozygotic (MZ) twin pairs and dizygotic (DZ) twin pairs living in the same family, have been widely used in the study of reading abilities and disabilities in order to estimate heritability (Pennington & Olson, 2005). Previous twin studies of school-age twins in the US and Australia have reported heritability estimates in excess of 0.50 across diverse measures of reading performance, including 'standard' tests of reading achievement, as well as experimental measures of orthographic coding and phonological awareness (Bates et al., 2004; Byrne et al., 2005; Foch & Plomin, 1980; Gayán & Olson, 2003; Reynolds et al., 1996). This finding can be interpreted as indicating that 50% or more of the measured variation in reading abilities within these samples is due to genetic influences. Moreover, results from a US study of 7–20-year-old twins attest to significant genetic influences on reading performance right across the spectrum of individual differences, including very able reading performance as well as performance levels characteristic of reading disabilities (Pennington & Olson, 2005).

Less consistent findings have been obtained in European twin samples. Heritability estimates for reading disabilities and normal variation in reading abilities were generally low or negligible in two UK studies, where twins were aged 13 years (Stevenson, Graham, Fredman, & McLoughlin, 1987) and 7–13 years (Bishop, 2001). A similar conclusion is suggested by twin correlations in a Swedish twin study for teachers' grades of reading ability in primary school (Husén, 1959). These findings imply that the main sources of variation in reading ability are non-genetic, reflecting shared environmental influences (environmental influences that contribute to familial resemblance, e.g., socio-economic status) or non-shared environmental influences (individual-specific environmental influences, e.g. differential learning experiences) and measurement error. In contrast, a fourth UK study showed that individual differences in literacy and phonological awareness in 6 and 7-year-old twins were substantially heritable (Hohnen & Stevenson, 1999). Similarly, we recently reported high heritabilities for normal variation in word reading and impaired word reading abilities, as assessed by the TOWRE at age 7 (Harlaar, Spinath, Dale & Plomin, 2005).

Direct comparisons across these studies are difficult because heritability is the proportion of phenotypic variance in a population due to genetic influences, and thus will depend on the degree of genetic and environmental variance within the population under study. For example, if the quality of reading instruction varies widely within a sample, then heritability will be lower than if all children are exposed to a more uniform environment (Bishop & Snowling, 2004). Furthermore, not all studies reported confidence intervals, or, where reported, confidence intervals were very wide, raising questions about the precision of point estimates. Nevertheless, the conflicting findings among these studies, which were all based on broadly representative samples of school-age children, are striking, and suggest that further exploration of the aetiology of reading abilities and disabilities is warranted.

The first aim of the present study was to estimate the relative contributions of genetic and environmental factors to NC scores for comparison with previous twin studies and our results for the TOWRE. To date, only one study has employed teacher assessments to assess reading abilities (Husén, 1959). The second and more central aim of the present study was to examine the nature of the association between TOWRE and NC scores. We used multivariate genetic analyses (Martin & Eaves, 1977) to estimate the extent to which these measures are influenced by the *same* genetic and environmental factors, as well as the extent to which the substantial phenotypic correlation between the two measures is mediated by genetic and environmental factors. We examined these issues at the level of individual differences across the whole ability range, as well as for the risk ('liability') for reading difficulties, as defined by low NC and TOWRE scores.

In considering these issues, we compared twin pairs in the same classroom, who were rated by the same teacher, to twin pairs in different classrooms with different teachers. When twins are assessed by the same teacher, rater response tendencies (e.g., stereotyping, idiosyncratic response styles) that influence ratings will be shared across co-twins. This was noted in the Swedish twin study, where twins in each pair were rated by the same teacher (Husén, 1959; p. 66): "We cannot exclude the possibility that a certain 'halo' has occurred when giving marks to the twins. The impression made on the teacher by one twin may have influenced him when giving marks to the other twin". If such biases are equal across zygosity, twins assessed by the same teacher will show greater resemblance than twins assessed by different teachers, resulting in higher shared environmental influences and lower non-shared environmental influences. An alternative explanation for higher shared environmental influences in twins assessed by the same teacher, however, is that they may reflect real effects that occur because the classroom context creates more shared environmental influence that contribute to variation in reading abilities. Twin correlations from a study of 12-year-old Dutch twins provided suggestive evidence that shared environmental influences made a greater contribution to individual differences in general academic achievement among twins taught by the same teacher compared to twins with different teachers (Bartels, Rietveld, Van Baal, & Boomsma, 2002). In the TEDS sample, NC assessments were obtained both for twins assessed by the same teacher as well as for twins assessed by different teachers, permitting a systematic assessment of whether components of variance differ for same- and different-teacher ratings of reading ability.

In summary, in Dale et al. (2005) we examined the phenotypic relationship between TOWRE and NC scores in 7-year-old twins. In the present analysis we examine the heritability of NC scores, the genetic and environmental relationship between NC and TOWRE scores, and possible heterogeneity that arises from twins having the same or different teachers.

Method

Sample

Full details of the sample are given in the companion paper. In brief, we studied 5544 children (2772 twin pairs) from the Twins Early Development Study (TEDS), a longitudinal study of twin pairs ascertained from population records of twin births in England and Wales (Trouton, Spinath, & Plomin, 2002). This sample consisted of twins born between January 1994 and August 1995 who consented to participate in child and teacher assessments when twins were 7-years-old. Zygosity was ascertained by parental ratings with an error rate of $\leq 5\%$, as validated by polymorphic DNA markers (Price et al., 2000). Unclear cases were resolved through DNA screening. Definitive zygosity assignment of a small group ($<5\%$) of same-sex twins awaits genotyping, and these twins were excluded from the analyses reported here. Twin pairs were also excluded if English was not the first language spoken in the home, untimely return of teacher assessments ($>$ 90 days after they were originally sent), or if either twin had a history of medical or perinatal complications commonly associated with problems in reading and cognitive development.

Analyses reported in this paper were based on same-sex twin pairs only. This sample included a total 1967 pairs, approximately evenly split between monozygotic (MZ) twins ($n = 1019$) and dizygotic (DZ) twins ($n = 948$). Twins in 1279 pairs (65%) had the same teacher, whereas twins in 688 pairs (35%) had different teachers. Of twins with different teachers, 660 pairs (96%) were also in different classrooms. Twins were placed in the same or different classes largely because of school policies on the separation of twins, which vary widely in the UK. The mean age of the twins when the TOWRE was administered was 7.06 years (SD = 0.21). The mean age of the twins when teacher assessments were returned was 7.10 years (SD = 0.26).

Measures and procedure

Telephone assessment

The Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999) was administered to each twin separately by telephone. The TOWRE is a measure of word recognition that comprises two sub-tests, sight-word Efficiency (SWE), which assesses fluency and accuracy in sight word reading, and phonemic decoding efficiency (PDE), which assesses phonemic decoding. Subtest item lists were mailed to families in a sealed envelope prior to the test sessions with separate instructions that the envelope should not be opened until the time of testing. Twins in each pair were tested within the same test session and by the same tester, who was blind to zygosity. Because the SWE and PDE subtests correlated substantially ($r = 0.83$; $P < 0.01$), composite TOWRE scores, derived by standardising and summing scores on the two subtests, were used in the current analyses.

Teacher assessments

Teachers' assessments were obtained by postal questionnaire and were based on UK National Curriculum (NC) Key Stage 1 criteria for reading attainment between 5 and 7 years of age (Department for Education and Employment, 2000). These criteria reference a statutory reading curriculum, the National Literacy Strategy (NLS), which defines literacy targets and instruction guidelines for teachers at state-funded primary schools in England and Wales (Department for Education and Employment, 1998; Stainthorp, 2002). The assessment of reading at Key Stage 1 requires teachers to rate children's reading ability on a 5-point scale ranging from far below average to far above average, based on their knowledge of the child's reading achievement over the academic year (see Appendix of Dale et al., 2005, for the rating scale). Because some evidence suggests that teacher assessments rise in accuracy through the school year (Glascoe, 2001), teacher assessments in the current study were obtained during the spring semester.

Analysis

Biometrical twin analyses were carried out for both individual differences in TOWRE and NC scores in the normal range (*individual differences analyses*) and for children with reading difficulties, as indicated by low TOWRE and NC scores (*liability threshold analyses*). The liability threshold analyses were based on the assumption that reading difficulties reflect a normally-distributed latent continuum of risk, or 'liability'. This liability is only expressed if an individual's liability is greater than a certain critical threshold value on the continuum (Falconer, 1965). For the purpose of the present analyses, reading difficulties were defined on the basis of scores below a threshold (cut-off) of 13.4% on the standardized distributions of TOWRE and NC scores. This cut-off corresponds to the percentage of children scoring 0 or 1 on the NC reading scale and an equivalent cut-off on the distribution of TOWRE scores. Because the evidence supporting the use of IQ in the classification of reading disabilities is limited (Vellutino, Fletcher, Snowling, & Scanlon, 2004), we did not reference low reading scores to either a minimum IQ level or a discrepancy between achievement and IQ test scores. All analyses were based on standardised TOWRE and NC scores adjusted for linear and quadratic effects of age and sex (McGue & Bouchard, 1984).

In a first stage of analysis, we calculated intraclass correlations to examine correlations between TOWRE and NC scores across the whole range of ability and tetrachoric correlations to examine correlations between the liabilities for low TOWRE and NC scores. Three types of intraclass and tetrachoric correlations were calculated: (1) within-trait, cross-twin correlations; (2) cross-trait, within-twin (phenotypic) correlations; and (3) cross-trait, cross-twin correlations. Cross-twin, within-trait correlations (e.g., the correlation between TOWRE scores in Twins 1 and 2) can be used to decompose the variance of a trait or liability into genetic and environmental influences, whereas cross-twin cross-trait correlations (e.g., the correlation between TOWRE scores in Twin 1 and NC scores in Twin 2) can be used to decompose the covariance between traits or liabilities, as indicated by the cross-trait, within-twin correlations, into genetic and environmental influences. Specifically, because MZ twins are genetically identical whereas DZ twins share, on average, 50% of their segregating genes, MZ correlations twice that of DZ correlations suggest that genetic factors are the primary cause of twin similarity and trait covariation. DZ correlations that are nearly equal to MZ correlations suggest that shared environmental influences are the primary cause of twin similarity and trait covariation. The extent to which MZ within-trait cross-twin correlations are less than unity is indicative of residual variance, which is attributed to non-shared (twin-specific) environmental influences and measurement error. Similarly, the extent to which MZ cross-trait within-twin correlations are less than the cross-trait, within-twin correlation implies the influence of non-shared environmental influences on trait covariation.

Aetiological patterns indicated by twin correlations can be tested more formally using structural equation models. Thus, in a second stage of analysis we tested the fit of a series of bivariate models (parameterised in terms of triangular decomposition matrices; Neale & Maes, 1999) using the structural equation modeling program *Mx* (Neale, Boker, Xie, & Maes, 2002). Models in the individual differences analyses were fit to variance–covariance matrices by maximum-likelihood estimation. Models in the liability threshold analyses were fit to polychoric correlation matrices using the method of asymptotic weighted least squares (Browne, 1984) in PRELIS (Jöreskog & Sörbom, 1995); the degrees of freedom were adjusted to take into account the use of correlation rather than covariance matrices (Neale & Maes, 1999).

Our baseline model is shown in Figure 1. In this model, variance in TOWRE and NC scores (or liabilities for low TOWRE and NC scores) and the covariance between them are modelled as a function of three latent factors, representing additive genetic (A), shared environmental (C), and non-shared environmental effects (E). The path coefficients *a*, *c*, and *e* are partial regressions that indicate the relative influence of the latent factors on measured TOWRE and NC scores. Additive genetic and shared environmental factors contribute to familial resemblance and thus are correlated across twins in each pair. Based on biometrical genetic principles, the correlation between additive genetic factors within a pair is specified to be 1.00 within MZ pairs and 0.5 within DZ twin pairs, whereas the correlation between shared environmental factors is specified to be 1.00 in both MZ and DZ twin pairs. Non-shared environmental

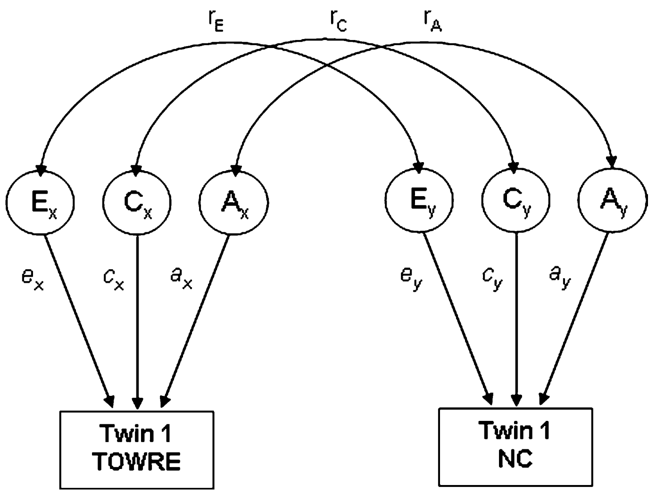

Figure 1. Bivariate genetic model for TOWRE and National Curriculum (NC) scores in one individual from a twin pair. Though not illustrated here, there are genetic and shared environmental correlations between the two members of a pair for both TOWRE and NC scores. Subscripts x and y denote TOWRE and NC scores, respectively. Paths are standardised regression coefficients and must be squared to equal the proportion of variance accounted for by latent additive (A), shared environmental (C), and non-shared environmental (E) factors. Correlations between the latent genetic, shared environmental, and non-shared environmental influences are denoted by $r_{A,}$ $r_{C,}$ and $r_{E.}$

influences contribute to differences between family members, reflecting the effects of individual-specific influences (e.g., differential educational experiences) as well as measurement error; these influences are therefore uncorrelated across twins in a pair (Evans, Gillespie, & Martin, 2002).

In order to test for differences between twins with the same teacher and twins with different teachers, we first allowed all parameters to vary across same- and different-teacher groups (the 'full' ACE model). We then successively constrained the A, C, and E parameters to be equal for both groups. Because the constrained models are nested within the full ACE model, we compared these models using the likelihood-ratio chi-square (χ^2) test. A significant change in chi-square, based on the difference of the degrees of freedom between two models, indicates that the model with fewer degrees of freedom should be adopted. The relative fit of the models was also assessed using the Akaike Information Criterion (AIC; Akaike, 1987) and the Bayesian Information Criterion (BIC; Raftery, 1995). Both criteria provide a quantitative index of the extent to which each model maximises correspondence between the observed and model-predicted variances and covariances, while minimising the number of parameters. Lower AIC and BIC values are associated with better-fitting models.

Results

Twin correlations

Table 1 summarises the intraclass and tetrachoric correlations for TOWRE and NC scores. The within-trait intraclass and tetrachoric correlations are presented in the left-hand side of the table. These correlations show that MZ twins were much more similar than DZ twins, although the MZ correlations were less than double the DZ twin correlations. It can therefore be inferred that both genetic and shared environmental influences contribute to the variance in TOWRE and TOWRE scores and liabilities for low TOWRE and NC scores. Correlations for twins with the same teacher were somewhat higher than for twins with different teachers. As might be expected, this pattern was more noticeable for NC scores than for scores on the TOWRE, which was administered completely independently of the teacher questionnaires. The difference between MZ and DZ intraclass NC correlations was comparable for twins with the same and twins with different teachers, indicating that the main effect of same- or different teacher status was to increase shared environmental influences contributing to twin resemblance and decrease residual variance due to non-shared environmental influences in twins with the same teacher. In contrast, the difference between the MZ and DZ tetrachoric correlations was greater for twins with the same teacher, suggesting that the liability for low NC scores was more heritable in twins with the same teacher compared to twins with different teachers.

The cross-trait intraclass and tetrachoric correlations are presented in the right-hand side of Table 1. The cross-trait, within-twin correlations show that TOWRE and NC scores were highly correlated phenotypically, both across the full range of abilities ($r = 0.68–0.74$) and for children with low TOWRE and NC scores ($r = 0.81–0.88$). The MZ cross-trait, cross-twin correlations exceeded, but were less than double, the corresponding DZ correlations. This pattern implicates both genetic and shared environmental influences on the covariance between TOWRE and NC scores (or liabilities). Because the MZ cross-trait, cross-twin correlations were almost as high as the cross-trait, within-twin correlations, it may be inferred that non-shared environmental influences on the covariance between TOWRE and NC scores (or liabilities) were negligible, however. The magnitude of the cross-trait, cross-twin correlations within zygosity group did not vary significantly as a function of whether twins had the same or different teachers, suggesting that same- or different-teacher status had little effect on genetic and environmental contributions to the covariance between TOWRE and NC scores.

Table 1. Intraclass and tetrachoric twin correlations for TOWRE and National Curriculum (NC) scores (95% confidence intervals in parentheses).

			Within-trait		*Cross-trait*	
			TOWRE scores	NC scores	Within-twin	Cross-twin
Intraclass correlations	*n* pairs					
MZ same teacher	664		0.85 (0.83–0.87)	0.83 (0.81–0.86)	0.69 (0.65–0.73)	0.64 (0.60–0.69)
MZ different teacher	355		0.84 (0.81–0.87)	0.71 (0.66–0.76)	0.74 (0.69–0.78)	0.67 (0.61–0.73)
DZ same teacher	615		0.53 (0.47–0.58)	0.51 (0.45–0.56)	0.68 (0.63–0.72)	0.43 (0.36–0.49)
DZ different teacher	333		0.44 (0.35–0.52)	0.34 (0.24–0.43)	0.74 (0.69–0.79)	0.35 (0.23–0.42)
Tetrachoric correlations	*n* probands					
	TOWRE	NC				
MZ same teacher	214	205	0.88 (0.82–0.94)	0.97 (0.94–0.99)	0.83 (0.76–90)	0.82 (0.74–0.90)
MZ different teacher	145	149	0.85 (0.77–0.94)	0.86 (0.77–0.94)	0.88 (0.80–0.96)	0.80 (0.70–0.91)
DZ same teacher	191	175	0.59 (0.46–0.72)	0.60 (0.47–0.73)	0.81 (0.72–0.89)	0.46 (0.30–0.61)
DZ different teacher	89	116	0.63 (0.45–0.80)	1 (0.44–0.78)	0.84 (0.75–0.93)	0.61 (0.44–0.78)

Note: MZ: Monozygotic; DZ: Dizygotic.

Model-fitting analyses

From the full ACE model that allowed model parameters to vary for twins with the same teachers and twins with different teachers, the proportion of the phenotypic variance or liability due to A, C, and E was estimated separating for TOWRE and NC scores by standardising and squaring path coefficients from the latent A, C, and E factors to TOWRE and NC scores. Figure 2 shows the contributions of A, C, and E to the variance in TOWRE and NC scores across the whole ability range. Additive genetic influences were substantial for both measures, accounting for between 63% and 74% of the variance in TOWRE and NC scores. Among twins assessed by the same teacher, shared environmental influences accounted for 22% of the variance in TOWRE scores and 18% of the variance in NC scores. Among twins assessed by different teachers, estimates of shared environmental influences were lower: 9% and 4%, respectively. The overlapping confidence intervals around the shared environmental estimates indicate that these estimates did not differ

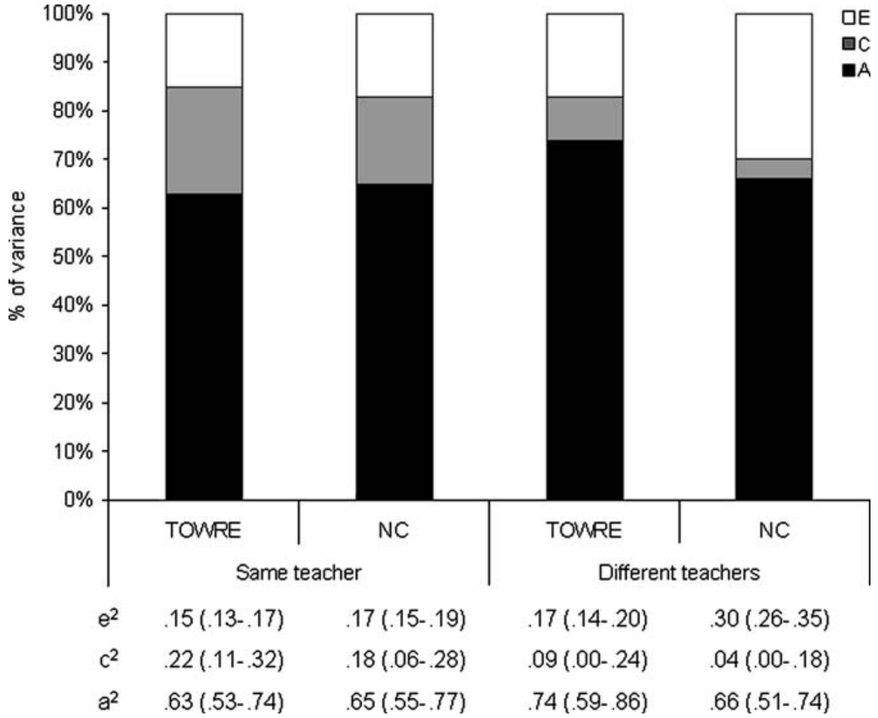

Figure 2. Variance in TOWRE and NC scores due to additive genetic (A), shared environmental (C), and non-shared environmental (E) influences for twins assessed by the same teacher and twins assessed by different teachers (95% CIs in parentheses).

significantly as a function of teacher status, however. Non-shared environmental influences accounted for 15% of the variance in TOWRE scores among twins with the same teacher and 17% of the variance in TOWRE scores among twins with different teachers. A comparable estimate was obtained for the contribution of non-shared environmental influences to NC scores among twins with the same teacher, but the proportion of variance in NC scores due to non-shared environmental influences among twins with different teachers was almost twice as great (30%), a significant difference.

Figure 3 shows the contributions of A, C, and E to the liabilities for low TOWRE and NC scores. For both children with the same teacher and children with different teachers, the liabilities for low TOWRE and NC scores were mainly due to genetic influences. Shared environmental influences were also comparable across groups and had moderate effect sizes, accounting for between 25% and 44% of the variance in liabilities. The point estimates for shared environmental influences were, in fact, slightly higher in twins with different teachers, compared to twins with the same teacher. However, mirroring the individual differences analyses, the confidence intervals around these estimates showed substantial overlap. Non-shared environmental influences accounted for between 3% and 14% of the liabilities. Overall, these findings suggest a departure from the results for the individual differences analyses in that estimates of the relative contributions of shared environmental and non-shared environmental influences to the liabilities for low TOWRE and NC scores were broadly similar for twins with the same teacher and twins with different teachers. We add the caveat, however, that the confidence intervals for all the point estimates were very wide, particularly for the relatively smaller group of twins with different teachers. This likely reflects the attenuation of power that results from the analysis of discrete scores and small samples (Neale, Eaves, & Kendler, 1994).

Two indices of the genetic and environmental contributions to the covariance between TOWRE and NC scores and liabilities were obtained. First, we estimated the genetic, shared environmental, and non-shared environmental correlations between TOWRE and NC scores and liabilities. These correlations index the extent to which individual differences or liabilities on these measures reflect the *same* genetic and environmental influences, and are independent of the extent to which two traits are each influenced by genetic and environmental influences. A genetic correlation of 1.00 between TOWRE and NC scores would indicate that genetic influences contributing to variance in TOWRE and NC scores were identical, whereas a genetic correlation of 0 would indicate that different gene loci or effects influence the two measures.

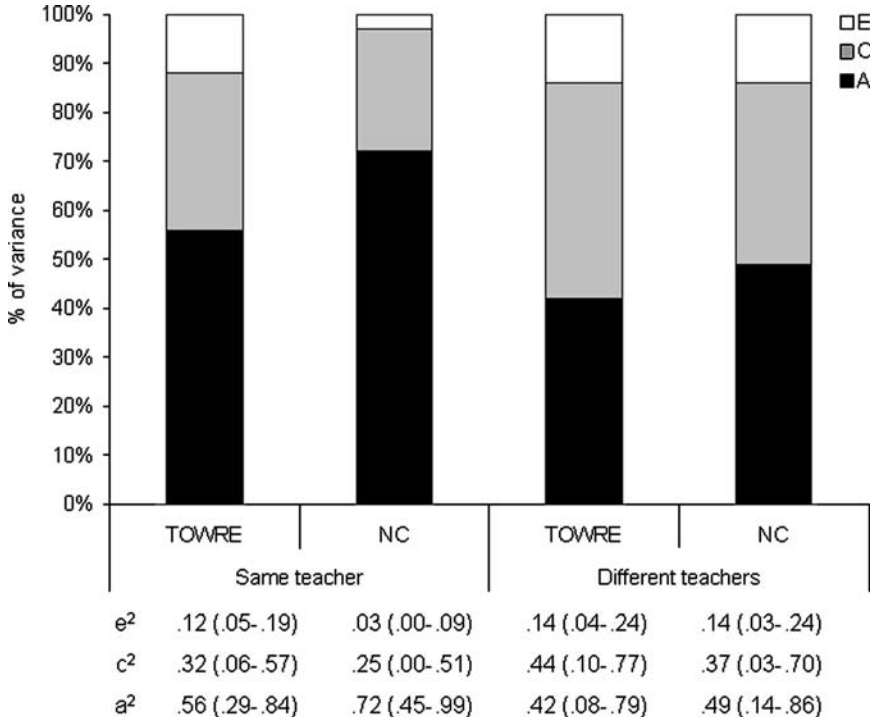

Figure 3. Liabilities for low TOWRE and NC scores due to additive genetic (A), shared environmental (C), and non-shared environmental (E) influences for twins assessed by the same teacher and twins assessed by different teachers (95% CIs in parentheses).

Table 2 shows the genetic, shared environmental, and non-shared environmental correlations from both the individual differences and liability threshold analyses. The genetic and shared environmental correlations were substantial, ranging from 0.65 to 1.00. This finding indicates that TOWRE and NC scores were largely influenced by the same genetic and shared environmental influences. Non-shared environmental influences correlated moderately at the level of individual differences analyses and more strongly at the low extreme. All correlations were slightly higher for twins with different teachers, but overall there were no significant differences as a function of same- or different teacher status.

Because the cross-trait within-twin (phenotypic) correlation between two traits partly reflects the relative contributions of genetic and environmental influences as well as the genetic correlation, we also estimated the extent to which the phenotypic correlation between TOWRE and NC scores was due to genetic, shared environmental, and non-shared environmental influences. The extent to which genetic influences contribute to

Table 2. Genetic, shared environmental, and non-shared environmental correlations (95% CIs in parentheses).

	Individual differences		Liability to low scores	
Correlation	Same teacher	Different teacher	Same teacher	Different teacher
A	0.74 (0.66–0.81)	0.84 (0.80–0.88)	0.99 (0.78–1.00)	1.00 (0.90–1.00)
C	0.89 (0.63–1.00)	1.00 (0.0–1.00)	0.65 (0.19–1.00)	0.83 (0.65–0.99)
E	0.28 (0.21–0.35)	0.32 (0.22–0.40)	0.81 (0.40–1.00)	0.80 (0.55–1.00)

the phenotypic correlation is the 'bivariate heritability', and is calculated as the product of the standardized *a* paths through the latent A factor (i.e. $a_X r_A a_Y$) divided by the phenotypic correlation (Plomin & DeFries, 1979). The relative contributions of shared and non-shared environmental influences to the phenotypic correlation can be estimated in a similar manner.

Table 3 shows the phenotypic correlations between TOWRE and NC scores and the decomposition of these correlations into genetic, shared environmental, and non-shared environmental components, both for the individual differences and liability threshold analyses. The findings closely mirror the patterns suggested by the cross-trait correlations shown in Table 1. Phenotypic correlations, reported here for same- and different-teacher groups overall, are similar to the zygosity-specific cross-trait within-twin correlations reported in Table 1 (0.68–0.88). Across both same- and different-teacher groups, these correlations were primarily mediated genetically. Specifically, the bivariate heritabilities (0.51–0.82) indicate that, in each case, approximately 50% of the phenotypic

Table 3. Decomposition of the cross-trait within-twin (phenotypic) correlations into proportions due to additive genetic, shared environmental, and non-shared environmental influences.

	Individual differences		Liability to low scores	
	Same teacher	Different teacher	Same teacher	Different teacher
Correlation (r_P)	0.68 (0.63–0.74)	0.73 (0.66–0.81)	0.84 (0.80–0.88)	0.88 (0.84–0.93)
% of r_P due to:				
A	0.68 (0.56–0.82)	0.82 (0.63–0.94)	0.75 (0.46–1.00)	0.51 (0.15–0.88)
C	0.25 (0.12–0.37)	0.08 (0.00–0.26)	0.20 (0.00–0.49)	0.37 (0.02–0.72)
E	0.06 (0.05–0.08)	0.10 (0.07–0.14)	0.05 (0.01–0.09)	0.12 (0.04–0.20)

correlation was due to genetic influences. Shared environmental influences had small to moderate effects on the phenotypic correlations, while non-shared environmental influences showed only small effects.

Our findings suggest that results were generally similar for same- and different-teacher groups, with the exception of some differences in the extent to which the variance and liabilities in TOWRE and NC scores were due to genetic and shared environmental influences. These impressions were clarified by further model-fitting analyses that tested for differences in parameter estimates between the same and different-teacher groups. Table 4 shows the results from these model-fitting comparisons, both for individual differences analyses (Models 1_{ID}–5_{ID}) and liability threshold analyses (Models 1_{LT}–5_{LT}). Models 1_{ID} and 1_{LT} refer to the baseline individual differences and liability threshold models, in which additive genetic, shared environmental, and non-shared environmental influences were estimated separately for twins with the same teacher and twins with different teachers. For both sets of analyses, there was no significant deterioration in model fit when additive genetic influences were equated for the same- and different-teacher groups (Models 2_{ID} and 2_{LT}), nor when shared environmental influences were equated for the same- and different-teacher groups (Models 3_{ID} and 3_{LT}). These findings converge with Figures 2 and 3 in showing that the relative magnitude of genetic and shared environmental influences on the variance in TOWRE and NC scores and the variance in liability for low TOWRE and NC scores do not vary significantly as a function of whether twins have the same teacher or different teachers. However, at the individual differences level of analysis, non-shared environmental influences could not be equated for the same- and different-teacher groups without a significant decrease in model fit (Model 4_{ID}), and a model in which all parameters (i.e., A, C, and E) were equated also fitted the data poorly (Model 5_{ID}). In contrast, the equivalent models for the liabilities for low TOWRE and NC scores (Models 4_{LT} and 5_{ID}) did not result in a significant deterioration in fit.

Overall, these findings indicate that non-shared environmental influences contributing to the variance in TOWRE and NC scores across the whole range of ability differed significantly for twins with the same and twins with different teachers, but same- or different teacher status did not seem to affect the relative magnitude of genetic and environmental contributions to the variance in liabilities for low TOWRE and NC scores. In the individual differences model comparison, the model equating shared environmental influences only for the same- and different-teacher groups (Model 3_{ID}) had the lowest AIC and BIC values and thus could be considered the most parsimonious model. In the liability thresholds

Table 4. Fit statistics from model-fitting analyses of TOWRE and NC scores.

		Model fit				Model comparison		
		χ^2	df	BIC	AIC	$\Delta\chi^2$	Δdf	*P*
Individual differences (ID)								
1_{ID}	ACE parameters vary for same and different teachers	21.41	22	−145.44	−22.59			
2_{ID}	CE parameters vary for same and different teachers (A equated)	26.65	25	−162.96	−23.35	5.24	3	0.16
3_{ID}	**AE parameters vary for same and different teachers (C equated)**	**24.86**	**25**	**−164.75**	**−25.14**	**3.45**	**3**	**0.33**
4_{ID}	AC parameters vary for same and different teachers (E equated)	72.05	25	−117.67	22.05	50.64	3	0.00
5_{ID}	ACE parameters for same and different teacher parameters for ACE	87.72	31	−147.39	25.72	66.31	9	0.00
Liability threshold (LT)								
1_{LT}	ACE parameters vary for same and different teachers	6.43	16	−114.92	−25.50			
2_{LT}	CE parameters vary for same and different teachers (A equated)	7.44	19	−136.66	−30.56	1.01	3	0.80
3_{LT}	AE parameters vary for same and different teachers (C equated)	7.10	19	−137.00	−30.90	0.68	3	0.88
4_{LT}	AC parameters vary for same and different teachers (E equated)	10.39	19	−133.71	−27.61	3.96	3	0.27
5_{LT}	**Equate same and different teacher parameters for ACE**	**15.96**	**25**	**−173.65**	**−34.04**	**9.53**	**9**	**0.39**

Note: A: additive genetic influences; C: shared environmental influences; E: non-shared environmental influences. df: degrees of freedom. BIC: Bayesian Information Criterion; AIC: Akaike's Information Criterion. Individual differences models 2_{ID} to 5_{ID} are compared with fit statistics for model 1_{ID}; Liability threshold models 2_{LT} to 5_{LT} are compared with fit statistics for model $1_{LT.}$ Boldface indicates the best-fitting model.

model comparison, the model equating all parameter estimates (Models 5_{LT}) could be considered most parsimonious.

Discussion

We used a genetically-informative design to examine genetic and environmental contributions to the covariance between TOWRE and teacher-rated NC scores of reading. Our analyses yielded three main findings.

Genetic and environmental influences on NC and TOWRE scores

The first finding was that individual differences in teacher assessments of 7-year reading ability, based on UK National Curriculum assessment criteria, were due to both additive genetic and shared environmental influences, with additive genetic influences generally being of greater magnitude. The heritability estimates mirror those for the TOWRE and were largely similar for twins with the same teacher and twins with different teachers. For both measures there were no significant differences in the parameter estimates for the individual differences and the liability threshold analyses. This finding partly reflects the wide confidence intervals around parameter estimates for the liability threshold analyses, but nevertheless suggests a conservative interpretation that the pattern of etiology was comparable across the normal range of ability and at the low tail of the distribution. The current study thus provides support from two different methods of assessment that individual differences in early reading ability and the liabilities for low reading ability are substantially heritable. The clear demonstration of genetic risk factors supports the rationality of using these measures in the TEDS sample for further research that explores the nature and mode of genetic risk factors affecting reading abilities and disabilities.

The finding that heritable effects were substantial and explained most of the phenotypic variation in reading ability contrasts with a previous Swedish study in which twin correlations for teacher-assessed primary school reading achievement suggested lower heritabilities, as well as two UK studies of twins ranging in age from 7 to 13 years (Bishop, 2001; Stevenson et al., 1987). However, our findings are consistent with a UK study of 6- and 7-year-old twins (Hohnen & Stevenson, 1999), as well as twin studies in the US and Australia. It is noteworthy that reading instruction in the UK school curriculum became progressively more structured and homogenous across schools in the 1990s following the introduction of major government initiatives to improve literacy levels in primary schools (Department for Education and Employment, 1998). These strategies may have served to reduce environmental variation in our study and in Hohnen & Stevenson (1999) compared to earlier studies or samples of older children, resulting, at least in part, in higher heritability estimates. This hypothesis is clearly speculative, but serves as a reminder of the importance of considering both environmental influences on reading ability, and possible secular (e.g., educational or societal) influences that increase or reduce environmental variation within the population.

Genetic influences on the relationship between NC and TOWRE scores

Our bivariate analyses yielded clear results on the nature of the association between TOWRE and NC scores. Our companion paper (Dale et al., 2005) reported that these measures correlated substantially at a phenotypic level. In the present study, we found that these measures are largely influenced by the same genetic and shared environmental influences and that phenotypic correlations between TOWRE and NC scores are primarily mediated genetically. Taken in conjunction with the findings of Dale et al. (2005), these results indicate that reading achievement in early elementary school, as assessed by teachers, and word recognition are closely linked both phenotypically and aetiologically.

Our data examined genetic and environmental influences at an aggregate level and provide no insight into causal or explanatory pathways. Nevertheless, it is interesting to speculate on why genetic and shared environmental influences on TOWRE and NC scores correlate so highly. Because the active processing of sentences and paragraphs cannot occur unless the reader can recognise individual words reliably and efficiently, it is likely that the word recognition abilities assessed by the TOWRE are involved both directly and indirectly in the skills assessed by teachers, at least on the face of the NC criteria. Thus, to the extent that the skills assessed by teachers draw on word recognition, some of the genetic and environmental influences that facilitate the skills assessed by the TOWRE may also facilitate reading achievement, as rated by teachers, at age 7. Concomitantly, the process of learning to read is likely to facilitate word recognition. It is also conceivable that the association between TOWRE and NC scores could reflect genetic and environmental influences on a common factor or set of processes. Our understanding of this relationship is likely to be facilitated by clarifying what underlying skills the measures are measuring at this age, coupled with identification of specific genetic and environmental risk factors associated with these measures.

Rater vs. classroom effects

A third aim of the present study was to examine the effects of having the same or different teachers. Due either to rater biases or a real effect of being exposed to classroom experiences that accentuate sibling resemblance, shared environmental influences on NC scores might be greater in twins with the same teacher, relative to twins with different teachers. Contrary to these expectations, we found that the relative contribution of shared environmental influences was generally comparable for children with the same and children with different teachers, both across the whole range of reading ability and for children with low TOWRE or NC scores. This was confirmed in our model-fitting comparisons, which indicated that shared environmental influences could be equated for same- and different-teacher groups. Thus, it appears that the higher twin correlations that emerge when twins are rated by the same teacher cannot be simply explained by correlated errors.

At the level of individual differences, same- and different-teacher groups did, however, differ with respect to estimates of non-shared environmental variance. When co-twins were rated by different teachers, estimates of non-shared environmental variance were higher. Having different teachers suggests that co-twins experience different classrooms, teaching styles, and learning experiences, all of which could contribute to individual differences in reading achievement. It is likely that at least some of the higher non-shared environmental variance for different teacher also reflects measurement variance as a result of having two raters. This variance may reflect rater biases (e.g., stereotyping, response styles) as well as measure unreliability. It is not possible to disentangle true non-shared environmental influences from measure unreliability in the current design. Greater purchase on this issue could be obtained by analysing latent traits of reading ability or collecting longitudinal data.

Limitations and conclusions

Some limitations need to be noted in considering these results. First, although our phenotypic findings provide evidence for the concurrent validity of these measures, probing the limits of the validity and reliability of the measures in the current sample is desirable. In particular, their diagnostic use in the accuracy of the classification of reading difficulties needs further investigation. A second and related limitation is that we have limited information about teachers themselves. Although teachers' assessments have been shown to correlate highly with their students' tested reading ability, they may be influenced by child characteristics (e.g., classroom behaviour) that do not uniquely influence reading attainment (Glascoe, 2001; Bennett, Gottesman, Rock, & Cerullo, 1993; Hecht & Greenfield, 2002). It would be valuable to examine the extent to which child characteristics explain the association between TOWRE and NC scores. Finally, conclusions drawn from this study may be specific to the age group studied (7-year-olds), to the time of assessment (Spring 2002 and Spring 2003), and to the UK. Genetic or environmental influences affecting reading abilities may change with age, across birth cohorts, or across societies.

Despite these caveats, the findings from our phenotypic and genetic analyses suggest that NC scores and our telephone adaptation of the TOWRE are expedient and valid alternatives to traditional methods of assessment that are sensitive to genetic variation in our sample. The strong phenotypic and genetic correlations suggest that an aggregate measure may provide a reliable phenotype for further genetic analyses. An alternative, but complementary approach, would be to use these measures in a multivariate design (cf. Marlow et al., 2003). We hope that such research may ultimately lead to a better understanding of the development of reading abilities and disabilities.

Acknowledgement

We gratefully acknowledge the contributions of the parents, children, teachers, and testers in the Twins Early Development Study (TEDS). TEDS is supported by program Grant G9424799 from the UK Medical Research Council.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

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

Bates, T. C., Castles, A., Coltheart, M., Gillespie, N., Wright, M., & Martin, N. G. (2004). Behavior genetic analyses of reading and spelling: A component processes approach. *Australian Journal of Psychology*, *56*, 115–126.

Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, *85*, 347–356.

Bishop, D. V. M. (2001). Genetic influences on language impairment and literacy problems in children: Same or different? *Journal of Child Psychology and Psychiatry*, *42*, 189–198.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different?. *Psychological Bulletin*, *130*, 858–806.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Statistical Psychology*, *37*, 62–83.

Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J. C., Willcutt, E. & Olson, R. K. (2005). Longitudinal twin study of early literacy development: Preschool and kindergarten phases. *Scientific Studies of Reading*, *9*, 219–235.

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: An Interdisciplinary Journal*, *18*, 385–400 (this issue).

Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgement of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*, 8–24.

Department for Education and Employment (1998). *The National Literacy Strategy: Framework for teaching*. London: HMSO.

Department for Education and Employment (2000). *QCA Key Stage 1: Assessment and reporting arrangements*. London: HMSO.

Evans, D. M., Gillespie, N. A., & Martin, N. G. (2002). Biometrical genetics. *Biological Psychology*, *61*, 33–51.

Falconer, D. (1965). The inheritance of liability to certain diseases estimated from the incidence among relatives. *Annals of Human Genetics*, *29*, 51–76.

Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, *18*, 52–65.

Fisher, S. E. (in press). Tangled webs: Tracing the connections between genes and cognition. *Cognition*.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: Genetic dissection of a complex cognitive trait. *Nature Reviews Neuroscience*, *3*, 767–782.

Foch, T. T., & Plomin, R. (1980). Specific cognitive abilities in 5- to 12-year-old twins. *Behavior Genetics*, *10*, 507–520.

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.

Glascoe, F. P. (2001). Can teacher's global ratings identify children with academic problems?. *Developmental and Behavioral Pediatrics*, *22*, 163–168.

Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on word recognition abilities and difficulties: A study of 7-year-old twins. *Journal of Psychology and Psychiatry*, *46*, 373–384.

Hecht, S. A., & Greenfield, D. B. (2001). Comparing the predictive validity of first grade teacher ratings and reading-related tests on third grade levels of reading skills in young children exposed to poverty. *School Psychology Review*, *30*, 50–69.

Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their students' reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing: An Interdisciplinary Journal*, *15*, 789–809.

Hohnen, B., & Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology*, *35*, 590–603.

Husén, T. (1959). *Psychological twin research*. Stockholm: Almqvist and Wiksell.

Jöreskog, K. G., & Sörbom, D. (1995). *PRELIS: A program for multivariate data screening and data summarization. A preprocessor for LISREL*. Chicago, IL: Scientific Software International.

Marlow, A. J., Fisher, S. E., Francks, C., MacPhie, I. L., Cherny, S. S., Richardson, A. J., Talcott, J. B., Stein, J. F., Monaco, A. P., & Cardon, L. R. (2003). Use of multivariate linkage analysis for dissection of a complex cognitive trait. *American Journal of Human Genetics*, *72*, 561–570.

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity*, *28*, 79–95.

McGue, M., & Bouchard, T. J. (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, Virginia: Department of Psychiatry.

Neale, M. C., Eaves, L. J., & Kendler, K. S. (1994). The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics*, *24*, 239–258.

Neale, M. C., & Maes, H. H. M. (1999). *Methodology for genetic studies of twins and families*. Dordrecht, Netherlands: Kluwer.

Pennington, B. F., & Olson, R. K. (2005). Genetics of dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 453–472). Oxford: Blackwell.

Plomin, R., & DeFries, J. C. (1979). Multivariate behavioral genetic analysis of twin data on scholastic abilities. *Behavior Genetics*, *9*, 505–517.

Plomin, R. & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131*, 592–617.

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.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–163.

Reynolds, M. T., Hewitt, J. K., Erickson, M. T., Silberg, J. L., Rutter, M., Simonoff, E., Meyer, J., & Eaves, L. J. (1996). The genetics of children's oral reading performance. *Journal of Child Psychology and Psychiatry*, *37*, 425–434.

Smith, S. D., & Morris, C. A. (2005). Planning studies of etiology. *Applied Psycholinguistics*, *26*, 97–110.

Stainthorp, R. (2002). Raising literacy standards for all? The National Literacy Strategy: An experiment in change in England and Wales. *Educational Psychology*, *22*, 473–487.

Stevenson, J., Graham, P., Fredman, G., & McLoughlin, V. (1987). A twin study of genetic influences on reading and spelling ability and disability. *Journal of Child Psychology and Psychiatry*, *28*, 229–247.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, Texas: Pro-ed.

Triga, A. (2004). An analysis of teachers' rating scales as sources of evidence for a standardised Greek reading test. *Journal of Research in Reading*, *27*, 311–320.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). The Twins Early Development Study (TEDS): A multivariate longitudinal genetic investigation of language, cognition, and problem behaviour in childhood. *Twin Research*, *5*, 444–448.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades?. *Journal of Child Psychology and Psychiatry*, *45*, 2–40.

Address for correspondence: Nicole Harlaar, MRC Social, Genetic & Developmental Psychiatry (SGDP) Centre, Box Number P083, Institute of Psychiatry, King's College London, Denmark Hill, London, SE5 8AF, UK

Phone: +44-207848 0600; Fax: +44-207848 0895; E-mail: n.harlaar@iop.kcl.ac.uk