Research Report

Predicting Literacy at Age 7 From Preliteracy at Age 4
A Longitudinal Genetic Analysis

Bonamy R. Oliver, Philip S. Dale, and Robert Plomin

MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, London, United Kingdom, and Department of Communication Science and Disorders, University of Missouri-Columbia

ABSTRACT—Early literacy experience and preliteracy knowledge have been shown to predict later literacy outcomes. Using a representative sample of 3,052 same-sex twin pairs (6,104 children) in the United Kingdom, we explored phenotypic and etiological interrelationships among early literacy experience, preliteracy knowledge, and school-based literacy outcomes (reading and writing). Both literacy experience and preliteracy knowledge at age 4 significantly and independently predicted literacy at age 7. Both measures also showed genetic influence that significantly predicted literacy at age 7, although genetic mediation was stronger for preliteracy knowledge than for early literacy experience. However, for both measures, shared environmental factors explained most of the association with literacy at age 7.

Literacy in the school years has been shown to be predicted both by early literacy-related experiences such as exposure to books, songs, and nursery rhymes (Burgess, Hecht, & Lonigan, 2002; Saracho, 2002) and by preliteracy skills such as phonological awareness and letter knowledge (Nathan, Stackhouse, Goulardis, & Snowling, 2004; Schneider, Roth, & Ennemoser, 2000; Treiman, Tincoff, Rodriguez, Mouzaki, & Francis, 1998). However, these predictions cannot be assumed to be causal. Although literacy experiences have traditionally been interpreted as purely environmental factors, there is a growing sense that at least some component of early literacy experience is child driven (Dale, Crain-Thoreson, & Robinson, 1995; Scarborough, 2001; Scarborough & Dobrich, 1994). More generally, advances in understanding the interplay between environmental and genetic factors have increased awareness of the fact that environments are often at least partly genetically influenced—that is, gene-environment correlation is often present (Plomin, 1994).

Another advance is the development of multivariate techniques that make it possible to estimate genetic and environmental contributions to predictions, such as those between preliteracy skills or experiences and literacy outcomes. There is some evidence that the association between preliteracy skills, such as phonological ability, and literacy outcomes is genetically mediated (Byrne et al., 2005; Cayan & Olson, 2003; Tunick & Pennington, 2002). However, possible genetic mediation of the link between early literacy experiences and literacy outcomes has not been examined. A better understanding of the etiological factors underlying phenotypic links between early literacy knowledge and experience and subsequent literacy outcomes will provide valuable clues about how literacy develops, and may identify appropriate foci for education.

To investigate the etiology of the relationship between early literacy experiences and knowledge and school-based literacy, we conducted a longitudinal genetic analysis using a large representative sample of twins. Our primary aim was to explore genetic and environmental mediation of the prediction of age 7 literacy measures from age 4 literacy experience and knowledge. These analyses also include the first genetic analysis of early literacy experience.

METHOD

Sample
The 'Twins' Early Development Study (TEDS), a longitudinal study of twins born in England and Wales in 1994, 1995, and 1996, formed the sampling frame for the current study. Data from TEDS families have been collected since the twins' first year of life, and collection is ongoing. The sample has been shown to be...
reasonably representative of the United Kingdom's population (Spinath, Ronald, Harlaar, Price, & Plomin, 2003; Trouton, Spinath, & Plomin, 2002).

For the present analyses, we selected 4,840 twin pairs for whom complete data on both twins were available both for the parental report obtained when the children were 4 years old and for the teacher report obtained during their second year of primary school. From this sample, we excluded 275 pairs in which at least one twin had a specific medical syndrome or was an extreme outlier for perinatal problems (e.g., extremely low birth weight) or for whom sex or zygosity information was not available, as well as 150 pairs for whom English was not the first language spoken at home. We also excluded the 1,363 opposite-sex twin pairs because there are methodological problems with running a sex-limited model for this type of genetic analysis (Loehlin, 1996). After exclusions, the sample for the current analyses consisted of 3,052 same-sex twin pairs (6,104 children), 3,242 girls and 2,862 boys in 1,565 monozygotic (MZ) and 1,487 dizygotic (DZ) pairs. The mean age of the twins at the time of the parental report was 4.0 years (range = 2.9–4.9). Their mean age when questionnaires were received back from the teachers was 7.2 years (range = 6.5–8.2). For individual analyses, extreme outliers (children scoring more than 3 standard deviations from the mean on the relevant measures) were excluded.

**Measures**

**Early Literacy Experience (ELE)**
The ELE measure was derived from three parent-report items collected from one parent when the twins were 4 years old: “Does your child read books or look at books with you?” “Does your child have any children's books?” and “Does your child have any children's tapes/records/CDs (for example, of nursery rhymes, stories)?” Each item was scored on a 5-point scale for the first-born twin and on a 5-point differential scale for the second-born twin. Factor analysis revealed that the covariance among these items as indexed by the principal component (general factor) accounted for 54.4% of the total variance of the items. Internal reliability was reasonable (α = .57) for a scale that aims to cover different, additive, and not necessarily correlated experiences.

**Preliteracy Knowledge (PK)**
The PK measure was derived from eight parent-report items collected at the same time as ELE items. There were three items on letter knowledge (“Can your child say the whole alphabet?” “If shown a letter, does your child name it?” and “If shown a letter, does your child know what sound it makes?”), three about word knowledge (“Can your child sound out words (for example c-at = cat)?” “Can your child read a word then tell you what it means?” and “Can your child tell a story back to you that s/he has read?”), and two on knowledge of rhyme (phonological awareness; “Ask your child: do ‘sip’, ‘tip’ and ‘lip’ sound the same” and “Ask your child: which word doesn't sound like the others ‘hall’, ‘shirt’, ‘ball’”). Factor analysis indicated that the covariance among the items accounted for 31.5% of their total variance. Internal reliability was again reasonable for this kind of scale (α = .62).

**Teacher Assessments of Reading and Writing**
Throughout their second school year, children in full-time education in the United Kingdom are assessed by their teachers in three domains of English. The criteria and tests for this assessment are part of the U.K. National Curriculum, developed by the Qualifications and Curriculum Authority and the National Foundation for Educational Research. Assessment involves two types of measurement, direct testing (DT) and teacher assessments (TA). Because DT scores were not available to us for the present study, we used TAs for reading and writing at Key Stage 1, designed for children ages 5 to 7 years. Along with the DT score, the TA ultimately determines the final score submitted to the Qualifications and Curriculum Authority to indicate a child's academic achievement at the end of the school year. The criteria for these assessments and evidence for their substantial validity are discussed elsewhere (Dale, Harlaar, & Plomin, in press; Oliver, Dale, & Plomin, 2004).

**Statistical Approach**
All measures were residualized for age and sex effects and standardized to a mean of 0 and standard deviation of 1, on the basis of the entire TEDS sample after exclusion of twins with major perinatal and medical problems (McGue & Bouchard, 1984).

Our main statistical approach was Cholesky decomposition, commonly used in longitudinal analyses. This method is similar to hierarchical multiple regression in its identification of independent variance explained by different predictors, but additionally allows parallel phenotypic and genetic analyses. The phenotypic and etiological results are depicted in path diagrams (Figs. 1 and 2), where boxes represent measured variables and circles represent latent factors: common phenotype (F), genetics (A), shared environment (G), and nonshared environment (E). The path coefficients between these latent factors and the measured variables are loadings (standardized partial regressions) that index the extent to which the measured variables covary. In each diagram, there are three latent factors, identified by subscript numbers: the first (i.e., F1, A1, C1, and E1) represent influences that the three observed variables have in common, the second (F2, A2, C2, and E2) loads only on the second and third variables, and the third (F3, A3, C3, and E3) represents influences specific to the final variable. Path coefficients to measured variables are squared to estimate variance explained. Because pairs of twins are nested within the same families, a
random twin from each pair was selected for the phenotypic analyses.

RESULTS

Phenotypic Analyses

Multivariate phenotypic Cholesky modeling (see Fig. 1) using the structural equation modeling software Mx (Neale & Maes, 1999) indicated that the ELE and PK scales were modestly correlated (.23) and significantly and independently predicted both TA Reading and TA Writing at 7 years. ELE correlated .19 (1.00 × .19) with both TA Reading and TA Writing, and PK correlated .29 (.23 × .19 + .97 × .25) with TA Reading and .26 (.23 × .19 + .97 × .22) with TA Writing. Furthermore, PK explained 5 to 6% (.25², .22²) of the variance in the literacy outcome measures independently of ELE.

Genetic Analyses

Multivariate genetic Cholesky analyses, also using Mx, were conducted to estimate genetic and environmental influences on ELE and PK and to investigate the extent to which the genetic and environmental variance in TA Reading and TA Writing is explained by these measures in concert and independently. The results are presented in Figure 2; Figure 2a shows the genetic pathways, Figure 2b the shared-environment pathways and Figure 2c the nonshared-environment pathways.

Univariate Results

As suggested by the difference between MZ and DZ correlations (.90 and .79 for ELE, .84 and .73 for PK), both ELE and PK are significantly heritable (Fig. 2a). The genetic variance (heritability) of F.I.F., estimated from the TA Reading model as the squared path coefficient leading to ELE, was .47², or .22 (95% confidence interval, CI = .19–.25); the estimated heritability of PK, calculated from the path diagram as the sum of the squared path coefficients leading to PK, was of the same magnitude, .21 (.07² + .45²; 95% CI = .17–.25). The heritability of TA Reading was estimated as .59 (.11² + .27² + .71²), which is in close agreement with our previous findings (Harlaar, Dale, & Plomin, in press). An identical model for TA Writing yielded the same estimated heritability for F.I.F. and PK (as expected). The heritability for TA Writing was estimated as .63 (.08² + .35² + .25²).
by Longitudinal good

The results of the present longitudinal study are consistent with previous studies which have found that exposure to literacy-relevant environments and preliteracy knowledge are correlated with literacy outcomes. Both preliteracy measures independently predicted TA Reading and TA Writing at age 7. The heritability estimate of .21 for PK in our study is consistent with the results of a previous study of 400 twin pairs, which estimated a heritability of .28 for preschool print knowledge using direct assessment of skills similar to those we measured using parent reports (Byrne et al., 2002).

Our study provides the first direct evidence that early literacy experience is at least to some modest extent influenced genetically, as proposed in previous literature (Dale et al., 1995; Scarborough, 2001; Scarborough & Dobrich, 1994) and also has—unsurprisingly—a large shared-environment component. However, the association between early literacy experience and literacy outcomes, at least at the age we examined, does not appear to be substantially mediated by these genetic factors, but rather is mediated by shared environmental influences. In contrast, and in line with a recent study of preliteracy knowledge and its links with later reading and spelling (Byrne et al., 2005), genetic influences on PK did substantially mediate its association with reading and writing in the present study. Nonshared environmental influences were significant for all measures but independent of each other. The large sample size should be considered in interpreting these effect sizes, which, although important conceptually, are small in magnitude.

The current study is limited in that the measures of ELE and PK were parental ratings and were not as detailed or comprehensive as desirable. However, to gain statistical power in a large-scale longitudinal study, one must accept the inevitable restrictions in data collection. We also acknowledge that both outcome measures were restricted to teachers’ ratings. However, the validity of these measures has been demonstrated elsewhere (Harlaar et al., in press; Oliver et al., 2004). In addition, and crucially, the measures we used are extremely relevant academically in the United Kingdom. Nevertheless, it will be important for future studies to use more objective reading and writing measures in exploring the phenotypic relationships between preliteracy and literacy outcomes, and their genetic and environmental mediation.

Phenotypic predictions from preschool-period variables to literacy outcomes have been used to support diverse theories of the emergence of literacy, as well as intervention practices. The present longitudinal genetic results suggest that these relationships are etiologically complex, so interpreting them will not be simple. Variation in preliteracy knowledge shows genetic links to later literacy outcomes, as are often found for early
cognitive predictors of later cognitive outcomes. In contrast, even though variation in literacy experience also reflects some genetic influences rather than being purely environmental, those genetic influences may not play much of a role in later literacy development. The contrast highlights the importance of investigating the etiology of noncognitive measures, such as literacy experience, and their predictive significance. Behavioral genetic methods can be powerful tools in this regard.

Acknowledgments—The Twins’ Early Development Study (TEDS) was supported by a program grant from the U.K. Medical Research Council (Programme Grant G0424799) and by the U.S. National Institute of Child Health and Human Development (HD04454). The authors thank all TEDS families for their continuing contribution to this study.

REFERENCES


(Received 12/6/04; Revision accepted 5/10/05; Final materials received 8/24/05)