Running head: GENETIC AND ENVIRONMENTAL INFLUENCES IN EARLY LITERACY

Genetic and Environmental Influences on Early Literacy

Brian Byrne[1], Richard K. Olson[2], Stefan Samuelsson[3], Sally Wadsworth[2], Robin Corley[2], John C. DeFries[2] and Erik Willcutt[2]

[1]University of New England, [2]University of Colorado, [3]Stavanger University and Linköping University

The authors acknowledge the Journal of Research in Reading, the United Kingdom Literacy Association and Blackwell Publishing. The article is published in the *Journal of Research in Reading, 29 (1),* pp. 33-49. The definitive version is available at www.blackwell-synergy.com'

Abstract

Prereading and early reading skills of preschool twin children in Australia, Scandinavia and the U.S. were explored in a genetically sensitive design (max. N = 627 preschool pairs and 422 kindergarten pairs). Analyses indicated a strong genetic influence on preschool phonological awareness, rapid naming, and verbal memory. Print awareness, vocabulary and grammar/morphology were subject primarily to shared environment effects. There were significant genetic and shared environment correlations among the preschool traits, as well as effects that were particular to each variable. Kindergarten reading, phonological awareness and rapid naming were primarily affected by genes, with spelling equally affected by genes and shared environment. Multivariate analyses revealed genetic and environmental overlap and independence among kindergarten variables. Longitudinal analyses showed genetic continuity as well as change in phonological awareness and rapid naming across the 2 years. Relations among the preschool variables of print awareness, phonological awareness and rapid naming and kindergarten reading were also explored in longitudinal analyses. Educational implications are discussed.

It has been known for a century that reading difficulties aggregate within families (Thomas, 1905). About 50 years ago evidence from limited twin studies that genetic transmission is in part responsible for this aggregation began to appear (Hallgren, 1950; Zerbin-Rubin, 1967), soon confirmed in methodologically more compelling studies (Decker & Vandenberg, 1985; DeFries, 1985; DeFries, Fulker, & LaBuda, 1987). Around 10 years ago a landmark article identifying a chromosomal locus associated with dyslexia was published (Cardon et al., 1994), identification of other loci followed (see Fisher & DeFries, 2002, for a summary), and within the last 2 years a Finnish group has identified an actual gene associated with the disorder (Taipale et al., 2003--although replication outside Finland has not yet been successful; Scerri et al., in press). Reading ability across the normal range and at the high end of the continuum is also affected by genes (Boada et al., 2002; Gayán & Olson, 2003). Genes are only part of the story, with the home and other aspects of the environment playing a role as well (Olson, Forsberg, & Wise, 1994), and concerted efforts are underway to identify which factors in the environment are particularly influential (Samuelsson et al., in press). Thus the amount and pace of research into the etiology of reading ability and disability is substantial and quickening.

In this article we describe our group's contribution to this research by summarizing published reports from a longitudinal twin study of early literacy development that we are conducting. We also take the opportunity to update some of our findings, made possible in part by an increased sample of twins in the project since those reports, and to include in our longitudinal analysis a variable, rapid naming, not hitherto covered.

Rationale

Our special contribution to the research into genetic and environmental influences on literacy has been to study the problem in a genetically informative, longitudinal design using twins recruited prior to the start of their school careers and followed for the first several years into school. Up until the beginning of this project in 1999, all studies of reading ability and disability conducted within a genetically sensitive design have involved children and adolescents aged 6 to 20 years (Hohnen & Stevenson, 1999; Olson & Byrne, 2005; Pennington & Olson, 2005). When studies are restricted to older children already experiencing reading success or failure, the confounding effects of reading levels on correlated abilities, such as phonological awareness, vocabulary, grammar and working memory, can cloud the interpretation of results, a point first brought home to the reading research community in the well-known article by Morais, Carey, Alegria, and Bertelson (1979). But if these associated skills are measured *prior to* formal (and much informal) reading instruction, their roles as causes or consequences of literacy development can be made clearer.

The longitudinal aspect of the project has allowed us to track changes in literacy and its related skills and insights as the children develop. The analyses will be extended to at least Grade 2 and, with further funding, to Grade 4, where children "read to learn" rather than just learn to read. By using a genetically sensitive design we can address questions of the continuity and change of etiological sources. For instance, should it turn out that phonological awareness is affected by genes at two or more phases (e.g., in both preschool and kindergarten), we can ask if the same or different genes are implicated across this developmental span. Similarly, should the twins' shared environment be playing a role in both phases, we can ask if new aspects of this source "kick in" in kindergarten or remain continuous with the earlier age.

Our sample comes from Australia, Scandinavia, and the USA. Thus the study is both cross-national and cross-linguistic, giving us a unique opportunity to integrate cultural and language differences into the picture of environmental factors that emerges from the project (see Sameulsson et al., in press, for examples of such differences from the preschool phase).

The twin children in our project are not selected because they bear a risk for dyslexia (as they would, for example, if we only selected those families in which one or both parents had reading difficulties). Thus our focus is on genetic and environmental influences that affect reading ability across its full range. Nevertheless, until and unless it becomes clear that etiological pathways to frank dyslexia are different from those that determine ability throughout the full range, we can assume that our findings can be generalized to very low as well as very high literacy levels, and those in between. At present, there is not sufficient evidence to support the conclusion that the etiology of dyslexia is qualitatively different from that of normal-range differences in reading (Pennington & Olson, 2005; Plomin and Kovacs, in press). Rather dyslexia can best be thought of as a particularly unfortunate "deal" from the deck of genetic and environmental factors that govern reading levels.

We present what follows in close to the usual format for research articles, though most of the techniques and results are already in published form (Byrne et al., 2002; Byrne et al., in press; Samuelsson et al., in press). We do so as the most convenient way to package this combination of review and new data.

Method

Materials

Test Selection at Preschool

The immediate goal of the preschool phase of the project was to identify insights and abilities in children that underpin subsequent literacy growth and to assess them in our twin sample. There is of course a large number of tests that we could have included in the preschool test battery. However, we needed to limit the number of these measures for two reasons. One is the power requirements of a behavior-genetic study; large numbers of participants are required, so we needed to balance the amount of data we collect on individual children against the number of children we test. The second reason is that preschool children generally have limited tolerance for extended testing sessions. To obtain a representative sample of twins in a longitudinal study, we needed to ensure that very few became frustrated and chose to leave the study. Given, then, the need to constrain the size of test battery, we measured candidate variables against the following criteria: (a) The variable is reliably associated with rates of reading growth, and with reading disability, in school-age children; (b) The variable is a known predictor of subsequent reading ability when administered before or at the beginning of school; (c) The variable discriminates between high risk and low risk preschoolers; (d) The variable is needed as a measure of general cognitive capacity. All measures we selected met one or more of these criteria.

Among the cognitive and linguistic variables we selected, some require little justification because their roles in literacy development are widely known and well documented. These include phonological awareness, letter knowledge and other aspects of print familiarity, rapid automatized naming, and measures of phonologically-based processes such as verbal short-term memory (see Byrne, 1998, Snowling, 2000, and Wolf & Bowers, 1999 for relevant reviews). Vocabulary and other aspects of “higher” language functions are also known to relate to literacy growth (Scarborough, 1998). We sought, too, to assess the children’s learning

potential, both within the domain of phonological awareness and more generally, prompted by evidence that how quickly children respond at a particular stage of literacy instruction serves as a good predictor of subsequent growth rates (Byrne, Fielding-Barnsley, & Ashley, 2000; Hindson et al., in press). We needed to include nonverbal measures to meet criterion (d) above.

Behavioral assessment in the form of parent, teacher, and tester ratings of attention and hyperactivity was justified by evidence of comorbidity between symptoms in the ADHD spectrum and reading disability, including common genetic sources (Pennington & Olson, 2005; Willcutt et al., 2002). Finally, home literacy factors such as parental reading habits and the child's willingness to engage in literacy activities were included as a means of specifying possible sources of family and outside environmental factors that our design identified. There is substantial evidence that such factors matter in literacy and language development (e.g., Sénéchal & LeFevre, 2002; Sénéchal, LeFevre, Thomas, & Daley, 1998).

Abbreviated details of the tests and other forms of assessment are as follows (for full details, see Byrne et al., 2002 and Samuelsson et al., in press).

Phonological awareness. These were mostly based on a preschool version of an established test, the Comprehensive Test of Phonological Processing (CTOPP--Wagner et al., 1999) supplied by C. Lonigan (personal communication, 2000). Tests included word, syllable and phoneme elision and blending, matching of phonemes, and rhyme. Also included was a specially designed phoneme awareness training routine that targets four phonemes, one per day, and teaches children to recognize words that begin, and end, with that phoneme (eg, *sun* and *sail*, *dress* and *bus*).

Print awareness. Letter knowledge (names and sounds); Clay's concepts of print; knowledge of common environmental print (words such as *Stop, Exit*); the

Word Identification subtest from the Woodcock Reading Mastery Test, as a screening measure.

Verbal fluency. Rapid naming of colours and objects from the CTOPP.

Language measures. Vocabulary (Hundred Pictures Naming Test (Fisher & Glennister, 1992); WPPSI Vocabulary); Productive morphology, a test based on Berko's methods, adapted by D. Bradley and further adapted for this project; Grammatic Closure from the Illinois Test of Psycholinguistic Abilities (ITPA,).

Verbal learning and memory. Gathercole's Nonword Repetition Test; WPPSI Sentence Memory; Story Recall and Sound Symbol Learning from the Wide Range Assessment of Memory and Learning (WRAML, Adams & Sheslow, 1990).

Visuospatial skills. WPPSI Block Design; visuospatial learning from the WRAML.

Reliabilities of these tests, in most cases calculated from our own data as Cronbach's alpha values, range from .46 to .92, with a mean of .80.

Parent/teacher measures. Home literacy environment was assessed by the Home Literacy Environment Questionnaire (Griffin & Morrison, 1997), and the Family Reading Survey (Whitehurst, 1992), with 19 questions interrogating book exposure, reading practices (both the child's and parents'), and so on; Disruptive Behaviour Rating Scale (DBRS—Barkley & Murphy, 1998), a measure designed to detect signs of inattention and hyperactivity (this test was also completed by preschool staff).

Tester assessments of behaviour. Because 4-yr-olds' commitment to task can be variable, each tester provided a rating of 1-3 of the degree of external distraction present for each test. They also provided an overall rating of attentiveness for each day (see below) and the full test schedule using the DBRS.

Test selection for subsequent assessment phases.

In the follow-up test cycles we focus on the literacy variables of letter knowledge, word and nonword reading, reading comprehension, and spelling. We include further measures of phonological awareness, vocabulary and grammar, and continue the theme of assessing learning potential, both specific to literacy and of a more general character. Here we nominate the tests given in the first follow-up year, referred to as Kindergarten. For details, see Byrne et al. (in press).

Word-level skills. Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), which includes both real and pseudoword items.

Spelling. Ten simple words and four nonwords, as used by Byrne and Fielding-Barnsley (1993).

Phonological awareness and rapid naming. Sound matching, blending and elision, and rapid color, letter, and digit naming from the CTOPP.

Print awareness. Letter sound knowledge.

Grammar. Test for the Reception of Grammar (TROG, Bishop, 1989).

Participants

At the time of writing, analyses are available for a maximum of 627 same-sex twin pairs recruited from the USA (355 pairs, 165 monozygotic (MZ), 190 dizygotic (DZ), mean age 58.8 months), Australia (150, 95, 55, 57.8), and Norway and Sweden, grouped as Scandinavia (122, 52, 70, 61.2). In this article we include the Scandinavian sample in the preschool analyses but exclude them from the kindergarten analyses because educational practices in Norway and Sweden make data pooling suspect: Children in those countries are typically not taught to read during the kindergarten year, in contrast to the situation in the USA and Australia. The maximum numbers of kindergarten pairs available are 213 MZ and 209 DZ.

Zygosity was determined by DNA analysis from cheek swab collection or, in a minority of cases, by selected items from the questionnaire by Nichols and Bilbro (1966).

Procedure

Children were tested in their homes and/or preschools and schools. The preschool battery was administered over five days, in sessions lasting around 45 minutes. The kindergarten follow-up was administered in a single, hour-long session. Except in Scandinavia, where this was not possible, each member of a twin pair was assessed by a different tester, at the same time. This avoids bias that might result from knowing the zygosity of the pair. (In the Scandinavian data the correlations within pairs tended to be somewhat higher, probably because of the absence of between-tester variance, but equally so for MZ and DZ pairs, suggesting a lack of bias towards judging MZ pairs to be more similar).

Results and Discussion

Data Reduction

Our assessment battery included tests that, on the face of it, overlap in content, such as word and phoneme blending and elision, word and nonword reading, and so on. For this reason, and because of the sheer number of measures, we sought to reduce the total number of variables, typically via factor analysis. We have in most instances formed latent traits using the multiple indicators yielded by the factor analyses and used these latent traits in modelling genetic, shared environment and unique environment effects. Latent traits model the common variance across a group of measure, producing “error free” estimates of the construct (Gayán & Olson, 2003). For the preschool results, we also checked if a more direct method of determining indices of the factors, namely a composite created by summing standard scores of the

contributing measures, gave us similar model results, and they did (Samuelsson et al., in press).

The logic behind the twin design and a description of the primary methods of analysis are provided elsewhere in the edition. In this report we provide information about analytic methods that are peculiar to our dataset.

Preschool Phase

Prereading Skills

Exploratory factor analysis of the cognitive and linguistic measures yielded four interpretable factors, accounting for 53.6% of total variance. The factors and their contributing individual measures were General Verbal Ability (the two vocabulary measures, WPPSI sentence memory, WRAML story memory, nonword repetition, productive morphology and grammar (ITPA)), Phonological Awareness (all of the measures listed in Methods under Phonological Awareness except for word blending, subject to a strong ceiling effect), Rapid Naming (the two rapid naming tests), and Print Awareness (all of the measures listed in Methods under Print Awareness except for the Woodcock Word Identification test, excluded because of floor effects). We elected to further divide General Verbal Ability into Verbal Memory (sentence memory, story memory, and nonword repetition), Vocabulary (HPNT and WPPSI vocabulary), and Grammar/Morphology (productive morphology and the ITPA subtest). See Samuelsson et al. (in press) for a justification of this subdivision.

We present the twin correlations and estimates of additive genetic (a^2), shared environment (c^2), and unique environment (e^2) influences on the four major latent traits, along with (a^2), (c^2), and (e^2) values for the subdivided latent traits, in Table 1. For this article we used combined data from the three country samples (USA, Australia, and Scandinavia) because the Australian and Scandinavian samples are as

yet too small to support independent genetic analyses. For the available individual country data and for comparisons of country means, see Samuelsson et al. (in press).

The four major latent traits all showed effects of both genes and environment, though the mix differed. For Phonological Awareness and Rapid Naming, genes dominated shared environment, shared environment was the primary influence for Print Awareness, and there were roughly equal effects for General Verbal Ability (confidence intervals for each source clearly overlap point estimates for the other source). The critical contrast between the patterns for Phonological Awareness and Print Awareness, both precursors to early literacy growth, reached significance; Phonological Awareness had higher genetic influence and lower shared environment influence than Print Awareness. (Note that within the Phonological Awareness variable, we were unable to detect a different pattern of influence for the phoneme training variable and scores based on the more "static" tests, even though there was a tendency for genes to be more influential in the static tests, contrary to our initial hypothesis that the phoneme learning routine would tap genetically driven factors more substantially. The training variable showed relatively high nonshared environment effects, suggesting lower reliability, a situation that may lead to underestimates of both genetic and shared environment effects.)

Although General Verbal Ability was affected both by genes and shared environment in about equal measure, the subcomponents did not all fit that pattern. Verbal Memory was more influenced by genes and the other components, Vocabulary and Grammar/Morphology, by shared environment. The contrast achieved significance in the case of Verbal Memory and Grammar/Morphology, and with planned larger samples we will determine if Vocabulary also contrasts with Verbal Memory. One interesting aspect of this result is that under criteria normally used in

factor analysis to group measures into factors (e.g., by inspection of the factor loadings--see Samuelsson et al., in press, for the details in our data) there were no compelling reasons to further subdivide the overall factor; nevertheless the subcomponents showed different patterns of genetic and environmental influence. Thus, latent traits capture what is common across constructs but there are still reliable sources of variance that are not common. Another aspect of interest is that these verbal components have all been shown to be related to reading ability, often in predictive studies with quite young children (e.g., Scarborough, 1998), suggesting a role for both genes and environment, in different mixes, among foundations of literacy growth. Of course, this conclusion is also warranted from the contrast among the four major latent traits we used for these analyses, for instance from the one between Phonological and Print Awareness.

Home Literacy Environment

Factor analysis of the two home literacy questionnaires yielded a four-factor solution accounting for 46.5% of the variance. Factor 1 was labelled Shared Book Reading, identified by questions about parent- and child- initiated book reading. Factor 2 was labelled Letter-Based Activities, with questions about alphabet games and writing and reading attempts by the child. We called the third factor Print Motivation, defined by questions about the child's lack of interest in reading sessions. Factor 4, Parent Reading Behavior, reflected amounts of reading by parents and of reading material available in the household.

These home literacy variables were related to certain between-country differences noted in the samples. For instance, Scandinavian parents' own reading behaviour was similar to that of the other countries but the amount of shared book reading and letter-based activities was substantially less. These differences, grounded

in cultural practices such as a tradition in Norway and Sweden not to subject children to formal or informal literacy instruction in the home during the preschool years, probably explain the lower levels of letter and print awareness noted in our data (see Samuelsson, in press, for details). We have not yet integrated the home literacy into our behaviour-genetic analyses to check, for instance, which of the four factors we identified best defines the shared environment effects noted above. However, although the correlations among these variables and the prereading skills of the children are mostly significant they are not particularly substantial. The highest correlation is between Shared Book Reading and Print Awareness, at .44. Thus even our "best" variable only accounts for 19% of the variance in print knowledge, itself subject to quite high shared environment effects (c^2 = .68). Thus we are a long way from identifying the major source(s) of the shared environment that drives print knowledge, and much the same goes for other preliterate skills.

Multivariate Analyses

Multivariate extensions of analyses from the basic twin design permit examination of the degree to which genetic and environmental influences are common to more than one measure. For example, we can ask if the same genes appear to be involved in two variables that are each genetically affected. Table 2 presents the genetic and shared environment correlations among the four major latent traits identified in our preschool data. It can be seen that for all four variables there is substantial genetic and shared environment overlap, indicating the presence of genetic and shared environment influences that are common to these traits. Note that these correlations are, to a degree, independent of the actual levels of genetic and shared environment influences on the individual traits. For instance, even though the

heritability of Print Awareness is relatively low (a^2 = .23) in comparison to Phonological Awareness (a^2 = .61), their genetic correlation is substantial at .68.

It is not known which of the chromosomal regions so far identified as contributing to reading ability (Fisher & DeFries, 2002), if any of them, are the sources of the genetic correlations. In addition, we do not yet have a clear picture of common environment sources for the shared environment correlations, though we hope that further exploration of our substantial database will shed light on this question.

It is important to note that the data in Table 2 does not mean that there are no independent genetic and environmental influences on the traits, since the genetic correlations are less than 1. Another analytic technique, Cholesky decomposition (Neale, Bokor, Xie, & Maes, 2002) has revealed significant independent effects as well (Sameulsson et al., in press; see below for a more complete example of a Cholesky decomposition). The Cholesky procedure is similar in principle to hierarchical regression in non-genetic studies, where the independent contribution of a predictor is assessed after the contributions of other predictors with which it shares variance are taken into account. Using this analysis, we found for instance that Phonological Awareness, Rapid Naming, and Print Awareness are each subject to independent genetic influences after that shared with General Verbal Ability is removed—see Table 8 in Samuelsson et al. That second genetic factor affecting Phonological Awareness also affects Print Awareness but not Rapid Naming. For shared environment, there is a second source which affects all three variables independently of the source shared with General Verbal Ability, but no other sources. The picture that emerges from these analyses will guide the search for genetic and

environmental factors, at least to the extent of identifying how many factors we need to be looking for.

Kindergarten Phase

Univariate Analyses

We subjected the kindergarten data to factor analysis and identified three factors, one that included the reading, spelling and phonological awareness measures, a second with the three rapid naming tests, and the third based on the TROG (see Methods for a description of all measures). For the behaviour-genetic analyses we subdivide Factor 1 into reading, spelling, and phonological awareness measures to check on possible differential patterns of influence and modelled them and rapid naming as latent traits. The TROG, as a single variable, was not treated as a latent trait. The results of those analyses are presented in Table 3. Reading, Phonological Awareness and Rapid Naming are substantially affected by genes, Spelling by genes and shared environment in equal measure, and the grammatical test largely by shared and nonshared environment. The a^2 value of .70 for Reading is close to that obtained for older school children (Gayán & Olson, 2003; Harlaar, Spinath, Dale, & Plomin, in press), showing that just about as early as children can be reliably assessed for isolated word and nonword reading they are as subject to genetic influence as they will remain throughout school. The findings for Phonological Awareness and Rapid Naming are in line with results from the preschool phase, as is the TROG, whose preschool counterpart is the IPTA and morphology, both less affected by genes than shared environment. See Byrne et al. (in press) for further details, including twin correlations for individual tests.

Multivariate Analyses

In Table 4 we present a previously unpublished Cholesky decomposition for the two literacy variables, Reading and Spelling, and the two variables traditionally considered in relation to them, Phonological Awareness and Rapid Naming (we will include the TROG in subsequent analyses when we incorporate reading comprehension, from the second follow-up phase). The order of entry of the variables, as in hierarchical regression, is guided by background empirical observations and current conceptual issues, in this case primarily by the specificity and covariation of genetic and environmental influences on reading and spelling skills after the effects shared with phonological awareness and rapid naming have been partialed out.

To illustrate how to interpret Table 4, consider the genetic path coefficients. The first genetic factor, A_1, has significant loadings on each of the four latent traits indicating a common source of genetic variation, including for Reading and Spelling. The loading of .61 for factor A_2 on Rapid Naming indicates that that variable is subject to a second source of genetic variance, and that source also affects Reading (loading of .34) though less so than factor A_1 (loading of .59). However, that second genetic factor is not also exerting influence over spelling that is independent from its influence on reading (loading of -.02). The loading of .51 of A_3 on Reading indicates a separate source of genetic variance for reading, and it also affects Spelling to a more modest extent. There is no independent genetic influence on Spelling, and thus all of the genetic variance in our spelling measure is shared with Phonological Awareness and Reading.

The pattern for shared environment indicates that a single source, C_1, determines all of the effect, with all other paths nonsignificant. Put another way, in contrast to the genetic pattern, where three independent sources of variation exist for these four latent variables, just one shared environment factor appears to be in operation. At this

stage we are unable to specify its nature, though data that we have collected on family literacy, SES, and so on, may help us to do so in future analyses.

The picture for nonshared (unique environment) is as follows: There appears to be modest effects of a single source, E_1, on all four variables, and a separate source for Rapid Naming but not for Reading and Spelling. The fact that E_1 influences all variables is interesting: Recall that all tests were given in a single session, and thus the individual child's attentional set may be at work, and because each child was assessed by a different tester we cannot dismiss tester effects. Other than that, we have no suggestions to offer as to what the source may be. Information that we have on factors that may differentiate within twin pairs, such as medical histories, birth order, birth weight, the school situation, and so on, may shed light on this. Note however that the physical variables just listed would be expected to show effects continuous across time, and we have not found continuity in nonshared environment effects in the longitudinal analyses (see below).

In summary, individual differences at kindergarten in phonological awareness and rapid naming appear to be driven by two genetic sources, one which they share and the other which primarily influences rapid naming. Both sources affect word and nonword reading efficiency, but only the shared path affects spelling accuracy. A third genetic factor influences reading independently of the other two. A single shared environment factor apparently influences all four of these literacy-related kindergarten measures, as does a unique environment factor. Rapid naming is also influenced by a separate unique environment source, but reading and spelling are not. Further analyses are required to clarify the nature of these environmental factors.

Longitudinal Analyses

The Cholesky decomposition model allows us to examine continuity and change in genetic and environmental factors affecting measures at two or more time points. The measures can be the same variable or different variables at the test occasions. In Table 5 we present such an analysis for the Phonological Awareness latent trait based on a total of 547 twin pairs at preschool and 422 at kindergarten. There is both genetic continuity and change. Path A1 is significant for both phases, and Path A2 is for kindergarten, indicating the new genes begin to affect this construct in kindergarten. (The values in Table 5 are very close to those in Table 5 of Byrne et al., in press, based on fewer twins, but Path A2 did not reach significance in the earlier report.) Shared environment is only influential at the preschool phase, showing that most of the reliable variance in phonological awareness is genetically driven by the time children reach the end of their first school year. This pattern of genetic change and continuity was also evident with rapid naming in the progression from preschool to school (Byrne et al., in press). Genes for the preschool measure were also in evidence in kindergarten, but an independent source came into play at the second time point. In assessing rapid naming we used color and object naming in preschool and colors, letters, and digits in kindergarten, and it is likely that the new genetic influence was associated with the new items because when only colors were analysed at both phases only a single genetic source was detectable (Byrne et al., in press).

To illustrate a longitudinal analysis with different variables, in Table 6 we present a Cholesky model of preschool Print Awareness, Phonological Awareness, and Rapid Naming and kindergarten Reading. This complements Table 6 in Byrne et al. (in press), which did not include Rapid Naming (and was based on fewer twins). A single genetic source influences all variables, a second source affects Phonological

Awareness but not Rapid Naming or kindergarten Reading, and a third source influences Rapid Naming but not, again, Reading. The fourth genetic path to Reading does not achieve significance on this sample size, but its value of .41 suggests that with a larger sample, currently being assembled, this might change. Thus reading ability at the end of the first year of school is affected by genes that also affect the two preschool foundations for literacy of letter and book awareness and sensitivity to phonological structure as well as verbal fluency as tapped by Rapid Naming. It is also possibly affected by a second, independent genetic source. (Byrne et al. also explored the independent genetic influences on reading from preschool Print Awareness and Phonological Awareness on Reading by reversing their order in two Cholesky models: Print had a significant independent genetic influence on reading, but Phonological Awareness did not.)

A single shared environment factor affects all variables, and a second source influences Phonological Awareness and Rapid Naming. There is one non-shared environment source affecting all variables and a second specific to Rapid Naming.

General Discussion

Levels of reading skill, from superior to deficient, are known to be affected by genetic variation in school children and adolescents, as well as by the kinds of environmental factors that twins share (family, school, for example). With this longitudinal project we have begun to trace those genetic and environmental factors as they operate from before formal schooling and on into the school years. We have shown that processes known to predict later literacy growth are already subject to substantial genetic influence at the preschool level—including phonological awareness, rapid naming, and verbal short-term memory. Other factors also show a degree of genetic influence, but more the effects of the home/preschool

environment—including vocabulary, print knowledge, and the higher-order language processes of morphological and syntactic control. There are both genetic and shared environment correlations among our preschool measures, indicating etiological overlap. But the overlap is not complete, with evidence also emerging of independence. For instance, phonological awareness and rapid naming are subject to genetic effects not shared with general verbal ability (Samuelsson et al., in press).

The fact that genetic effects emerge prior to schooling and therefore prior to the point at which reading levels are beginning to settle in children tells us that they are not secondary consequences of genetic determination of those reading levels, known to be substantial. This enhances the case for treating these processes as part of the causal chain in reading ability.

We have measured aspects of the home literacy environment, and, like others, have identified significant correlations between this environment and preschool abilities and insights. But we cannot be said to have accounted for substantial portions of the variance in those pre-literacy skills. This is a particular gap in our understanding of those variables that we have also shown to be largely determined by the overall early home and preschool environment, letter and print knowledge for example. Of course, the home environment may also be linked genetically to the twins' performance because parents both shape the environment and furnish their children's genes.

By the time children have completed their first school year, efficiency of word (and nonword) identification is largely heritable, as it is destined to remain throughout schooling. Shared environment effects are also at work. Spelling, too, shows effects of genes, and in this case the environment that the twins share plays a more substantial role. As was the case at preschool, phonological awareness and rapid

naming were significantly heritable, and our measure of grammar, the TROG, was more affected by shared environment. Among the literacy variables and those close to them (phonological awareness and rapid naming), there were genetic dependencies and independencies similar to those found in studies with older children (Compton et al., 2001): Rapid naming accounted for significant genetic influence on reading after controlling for phonological awareness, and reading was subject to a genetic effect independent of the genes that affect phonological awareness and rapid naming (as well as sharing genes with those traits), but spelling was not; once genetic influences on phonological awareness and reading had been factored out there was no independent genetic source for spelling. The presence of an independent genetic influence from rapid naming on reading but not spelling may be at least partly due to the speed demands in the word and nonword reading efficiency tasks, a demand that was not present for spelling.

Genetic and environmental continuity and change were evident in longitudinal analyses tracking the children from preschool to kindergarten. Phonological awareness and rapid naming each were subject to a genetic factor spanning both years and to a new one emerging in kindergarten (though the composition of the trait for rapid naming had changed across the two phases with the introduction of letters and digits at follow-up—see above). When kindergarten reading was considered in conjunction with the preschool measures of phonological and print awareness and rapid naming, a common genetic source for all except rapid naming emerged, and although phonological awareness and rapid naming were both subject to genetic influence separate from that shared with reading, that second source did not in turn influence reading in school. (We have not yet conducted these analyses with kindergarten spelling.) Thus reading is only affected by genes for preschool

phonological awareness that are shared with print knowledge, and not at all by genes for preschool rapid naming. We noted the former of these findings in Byrne et al. (in press), and return to it below. The latter is new for this report, and presents an apparent paradox because in the multivariate analysis restricted to the kindergarten variables (Table 4) we found a significant genetic relationship between rapid naming and reading (see Path A2). The paradox is probably solved, however, by the earlier observation that the genetics of rapid naming changes in kindergarten with the introduction of letters and digits to the test items. Apparently it is those genes that affect reading rather than the ones operative for the preschool version of the "same" test.

Returning to the situation with phonological and print awareness: We found phonological awareness to be highly heritable in preschool, significantly more so than print awareness, yet only the genes it shares with print awareness appear to affect kindergarten reading, and then quite substantially (a loading of .72 on Reading in Path A1, Table 6, can be squared to yield a heritability value of .52, the majority of the total heritability of reading). So even though the print variable is itself relatively modest in terms of genetic influence (a^2 = .23, see Table 1), the effects of those genes "amplify" later in school to drive a considerable part of the genetic effect on reading, and the more highly heritable phonological awareness exercises less genetic influence on reading, and no influence independent from that for print awareness.

The shared environment effect on reading in school is continuous with that of the preschool variables included in the longitudinal analysis (Table 6), and there does not appear to be a separate c^2 source emerging in school for reading. Later and larger samples may alter this picture, but it is clear that the environment provided by home and preschool carries over to reading in school, even if to a modest extent (loading of

.31 in Path C1, Table 6). We hope, in later analyses, to be able to better identify the full range of shared environment processes that drive these effects.

Implications

What are the implications for this research, ours and its predecessors with older children, for education? What messages should educators take away from the findings? First, consider the situation if the research had not been conducted. Imagine, that is, that we knew that reading problems ran in families but not that shared genes were part of the story. It is hard to be sure about this, but if Pinker (2003) is correct in his characterisation of Psychology still being largely under the sway of the Standard Social Science Model, with its emphasis on environmental determination of behaviour, the bias would be to look for environmental causes. When combined with evidence that children's reading levels tend to settle quite early in their school careers (Byrne, Freebody, & Gates, 1992; Freebody & Byrne, 1988; Juel, 1988) and that family practices do make a difference, parents would be natural targets for blame in cases of reading failure. In many, probably most, cases, this would be without justification.

Still, to many, identification of a genetic component in a disorder means that little, or nothing, can be done in its amelioration. But that, in principle, is wrong, as successful dietary intervention for the genetic metabolic disorder phenylketonuria shows. More directly, there is evidence that early and focused intervention for potential reading disorders in children with family histories of dyslexia can lead to grade-level performance in the early school years (Hindson et al., in press). So rather than leading to despair, research of this kind better serves as a stimulus to early and sustained intervention.

More generally, behaviour-genetic studies of our sort focus on human differences, not average levels of performance. If all the children in the research we have reviewed, including that with older children, had been reading at much higher levels such that none were hampered in school by inefficient reading, the pattern of differences could still be as heritable. So nothing in this research means that the search for better ways to teach reading is futile. On the contrary, the findings should act as a spur to continued research, to the adoption of the best evidence-based teaching practices, and to the early identification of children at risk for reading disability.

References

Adams, W., & Sheslow, D. (1990). *Wide range assessment of memory and learning*. Wilmington, Delaware: Jastak Associates.

Barkley, R., & Murphy, K. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook (2nd. ed.).* New York: Guilford Press.

Bishop, D. V. M., (1989). *Test for reception of grammar (TROG) 2nd edition.* Abingdon, UK: Medical Research Council.

Boada, R. , Willcutt, E.G., Tunick, R.A., Chhabildas, N.A., Olson, R.K., DeFries, J.C., & Pennington, B.F. (2002). A twin study of the etiology of high reading ability. *Reading and Writing: An Interdisciplinary Journal*, *15*, 683-707.

Byrne, B. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle.* Hove, UK: Psychology Press.

Byrne, B., Delaland, C., Fielding-Barnsley, R., Quain, P., Samuelsson, S., Hoien, T., Corley, R., DeFries, J. C., Wadsworth, S., Willcutt, E., Olson, R. K. (2002). Longitudinal twin study of early reading development in three countries: Preliminary results. *Annals of Dyslexia, 52*, 49-73.

Byrne, B. & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85*, 104-111.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of phoneme identity training after six years: Outcome level distinguished from rate of response. *Journal of Educational Psychology, 92,* 659-667.

Byrne, B., Freebody, P., & Gates, A. (1992). Longitudinal data on the relations of word-reading strategies to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly, 28,* 141-151.

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

Cardon, L.R., Smith, S.D., Fulker, D.W., Kimberling, W.J., Pennington, B.F., & DeFries, J.C. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science, 226,* 276-279.

Compton, D.L., Davis, C.J., DeFries, J.C., Gayan, J., & Olson, R.K. (2001). Genetic and environmental influences on reading and RAN: An overview of results from the Colorado Twin Study. In M. Wolf (Ed.), *Conference proceedings of the Dyslexia Research Foundation Conference in Extraordinary Brain Series: Time, fluency, and developmental dyslexia* (pp. 277-303). Baltimore MD: York Press.

Decker, S.N., & Vandenberg, S.G. (1985). Colorado twin study of reading disability. In D.B. Gray & F.J. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 123-135). Parkton, MD: York Press.

DeFries, J.C. (1985). Colorado Reading Project. In D.B. Gray & J.F. Kavanagh (Eds.), *Biobehavioral measures of dyslexia* (pp. 107-122). Parkton, MD: York Press.

DeFries, J.C., Fulker, D.W., & LaBuda, M.C. (1987). Evidence for a genetic aetiology in reading disability of twins. *Nature, 329,* 537-539.

Fisher, S.E., and DeFries, J.C. (2002). Developmental dyslexia: genetic dissection of a complex cognitive trait. *Nature Reviews Neuroscience, 3,* 767-780.

Fisher, J. P., & Glennister, J. M. (1992). *The hundred pictures naming test*. Hawthorn, Australia: Australian Council for Educational Research.

Freebody, P., & Byrne, B. (1988). Word-reading strategies in elementary school children: Relationship to comprehension, reading time, and phonemic awareness. *Reading Research Quarterly, 23,* 441-453.

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.

Griffin, E.A., & Morrison, F.J. (1997). The unique contribution of home literacy environment to differences in early literacy skills. *Early Child Development & Care, 127-128,* 233-243.

Hallgren, B. (1950). Specific dyslexia (congenital word-blindness): A clinical and genetic study. *Acta Psychiatrica et Neurologica Supplement, 65,* 1-287.

Harlaar, N., Spinath, F.M., Dale, P.S., & Plomin, R. (in press). Genetic influences on word recognition abilities and disabilities: A study of 7 year old twins. *Journal of Child Psychology and Psychiatry*.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (in press). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology.*

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

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447.

Morais, J., Content, A., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7,* 323-331.

Neale, M.C., Boker, S.M., Xie, G., Maes, H.H. (2002). *Mx: Statistical Modeling*. VCU Box 900126, Richmond, VA 23298: Department of Psychiatry. 6th Edition.

Nichols, R. C. & Bilbro, W. C., (1966). The diagnosis of twin zygosity. *Acta Genetica, 16,* 265-275).

Olson, R.K., Forsberg, H., & Wise B. (1994). Genes, environment, and the development of orthographic skills. In V.W. Berninger (Ed.), *The varieties of orthographic knowledge I: Theoretical and developmental issues* (pp. 27-71). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Olson, R. K., & Byrne, B. (2005). Genetic and environmental influences on reading and language ability and disability. In H. Catts, & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 173-200). Mahwah, NJ: Laurence Erlbaum Associates.

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: Blackwell Publishing.

Pinker, S. (2003). *The blank slate*

Samuelsson, S., Byrne, B., Quain, P., Corley, R., DeFries, J.C., Wadsworth, S., Willcutt, E., & Olson, R.K. (in press). Environmental and genetic influences on pre-reading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology.*

Scarborough, H. S. (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 disabilities: A view of the spectrum* (pp. 75-119). Timonium, MD: York Press.

Scerri, T. S., Fisher, S. E., Francks, C., MacPhie, I. L., Paracchini, S., Richardson, A. J., Stein, J. F., & Monaco, A. P., (in press). Putative functional alleles of DYX1C1 are not associated with dyslexia susceptibility in a large sample of sibling pairs from the UK. *Journal of Medical Genetics.*

Sénéchal, M., & LeFevre, J. (2002). Parental involvement in the development of children's reading skill: A five-year longitudional study. *Child Development, 73*, 445-460.

Sénéchal, M., LeFevre, J., Thomas, E. M., & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*, 96-116.

Snowling, M. J. (2000*). Dyslexia*. Oxford: Blackwell.

Thomas, C. J. (1905). Congenital word-blindness and its treatment. *Ophthalmoscope, 3*, 380-385.

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

Wagner, R.K., Torgesen, J.K., & Rashotte, C.A. (1999). The Comprehensive Test of Phonological Processes (CTOPP). Austin, Texas: PRO-ED.

Whitehurst, G. J. (1992). *Family reading survey*. Stony Brook: State University of New York.

Willcutt, E.G., Pennington, B.F., Smith, S.D., Cardon, L.R., Gayán, J., Knopik, V.S., Olson, R.K., & DeFries, J.C. (2002). Quantitative trait locus for reading disability on chromosome 6p is pleiotropic for attention deficit hyperactivity disorder. *American Journal of Medical Genetics (Neuropsychiatric Genetics), 114*, 260-268.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*, 415-438.

Zerdin-Rudin, E. (1967). Congenital word-blindness. *Bulletin of the Orton Society*, *17*. 47-56.

Author Note

Brian Byrne, School of Psychology, University of New England; Richard K. Olson, Sally Wadsworth, Robin Corley, John C. DeFries, and Erik Willcutt, Institute for Behavioral Genetics, University of Colorado; Stefan Samuelsson, Department of Behavioural Sciences, Linköping University.

This project is funded by the Australian Research Council (A79906201), National Institute for Child Health and Human Development (2 P50 HD27802 and RO1 HD38526), the Research Council of Norway (154715/330), the Swedish Research Council (345-2002-3701), and Stavanger University. We are grateful for support from the Australian Twin Registry and the Australian Multiple Birth Association. We also thank our testers: in Australia, Frances Attard, Fiona Black, Nicole Church, Marreta Coleman, Cara Newman; in Norway, Bjarte Furnes; in Sweden, Inger Fridolfsson; in the USA, Kim Corley, Rachael Cole, Barb Elliott, Kari Gilmore, Angela Villella, and Ingrid XXX.

Correspondence about this article can be directed to Brian Byrne, School of Psychology, University of New England, Armidale NSW 2351, Australia, electronic mail to bbyrne@une.edu.au

Table 1

Preschool Genetic (a^2), Shared Environment (c^2), and Non-shared Environment (e^2) Influences on Latent Traits (95% Confidence Intervals in Parentheses)

Variable	a^2	c^2	e^2
General verbal ability	.43(.30,.58)	.52(.37,.64)	.06(.03,.08)
Verbal memory	.57(.35,.79)	.29(.08,.48)	.14(.07,.21)
Vocabulary	.32(.06,.56)	.60(.38,.81)	.08(.01,.17)
Grammar/morphology	.29(.07,.53)	.59(.38,.77)	.12(.05,.19)
Phonological awareness	.61(.41,.83)	.30(.10,.48)	.09(.05,.14)
Rapid naming	.64(.40,.81)	.11(.00,.57)	.25(.18,.32)
Print awareness	.23(.12,.35)	.68(.56,.77)	.10(.06,.14)

Note: All estimates are significantly greater than 0 (p<.05) except the Rapid Naming (c^2)

Table 2

Genetic (Above Diagonal) and Shared Environment (Below Diagonal) Correlations for Preschool General Verbal Ability (GVN), Phonological Awareness (PA), Rapid Naming (RN), and Print Awareness (PRINT)

	Latent trait	1	2	3	4
1.	GVN	-	.62*	.43*	.56*
2.	PA	.94*	-	.35*	.68*
3.	RN	.57*	.78*	-	.35*
4.	PRINT	.67*	.86*	.74*	-

Note: *p<.05

Table 3

Model Fitting Estimates for Kindergarten Latent Traits (95% Confidence Intervals in Parentheses)*

Variable	a^2	c^2	e^2
Reading	.70*	.22 *	.07*
Spelling	.39*	.40*	.20*
Phonological awareness	.63*	.28	.10*
Rapid naming	.60*	.17	.23*
TROG	.21	.40*	.40*

Note: TROG = Test for Reception of Grammar. *This variable is a single score.

* indicates that component cannot be dropped without significant loss of fit, $p < .05$.

Table 4

Cholesky Model of Additive Genetic (A), Shared Environment (C), and Unique Environment (E) Factor Loadings on Kindergarten Variables (95% Confidence Intervals in Parentheses)

	Factor			
Variable	A1	A2	A3	A4
PA	.81 (.64, .94)			
RN	.34 (.14, .52)	.61 (.45, .72)		
READ	.59 (.44, .75)	.34 (.18, .50)	.51 (.34, .61)	
SPELL	.59 (.41, .76)	-.02 (-.15, .13)	.29 (.13, .41)	.10 (-.34, .34)
	C1	C2	C3	C4
PA	.51 (.24, .70)			
RN	.47 (.20, .65)	.17 (-.45, .45)		
READ	.42 (.09, .60)	-.17 (-.38, .38)	.00 (-.37, .37)	
SPELL	.63 (.38, .76)	.07 (-.32, .32)	.00 (-.30, .30)	.00 (-.28, .28)
	E1	E2	E3	E4
PA	.28 (.19, .40)			
RN	.27 (.09, .43)	.43 (.30, .54)		
READ	.26 (.17, .31)	.00 (-.09, .10)	-.05 (-.19, .19)	
SPELL	.32 (.18, .44)	-.01(-.17, .12)	-.07 (-.36, .36)	.23 (-.34, .34)

Note: PA = Phonological Awareness, RN = Rapid Naming, READ = Reading, SPELL = Spelling

Table 5

Cholesky Model of Additive Genetic (A), Shared Environment (C), and Non-shared Environment (E) Factor Loadings on Phonological Awareness Latent Variable at Preschool and Kindergarten (95% Confidence Intervals in Parentheses)

	Factor	
Variable	A_1	A_2
PA_1	.82 (.67,.97)	
PA_2	.65 (.45,.86)	.55 (.18,.73)
	C_1	C_2
PA_1	.56 (.24,.72)	
PA_2	.27 (-.15,.56)	.37 (-.57,.57)
	E_1	E_2
PA_1	.13(-.27,.27)	
PA_2	.25 (-.36,.36)	.00 (-.32,.32)

Note: $PA_{1,}$ PA_2 = Phonological awareness at preschool, kindergarten.

Table 6

Cholesky Model of Additive Genetic (A), Shared Environment (C), and Non-shared Environment (E) Factor Loadings for Preschool Print, Phonological Awareness and Rapid Naming and Kindergarten Reading (95% Confidence Intervals in Parentheses)

	Factor			
Variable	A1	A2	A3	A4
PRINT1	.47 (.34, .60)			
PA1	.61(.44, .83)	.50 (.22, .64)		
RN1	.41(.15, .66)	-.24 (-.55, .09)	.58 (-.01, .65)	
READ	.72 (.51, .91)	-.08 (-.47, .24)	-.09 (-.51, .23)	.41 (-.63, .63)
	C1	C2	C3	C3
PRINT1	.85 (.77, .90)			
PA1	.52 (.38, .64)	.28 (.06, .46)		
RN1	.24 (.09, .35)	.40 (.33, .59)	.09 (-.51, .51)	
READ	.31 (.16, .44)	-.04 (-.41, .34)	.36 (-.50, .50)	.00 (-.49, .49)
	E1	E2	E3	E3
PRINT1	.23 (.15, .29)			
PA1	.16 (.05, .27)	.03 (-.22, .22)		
RN1	.13 (-.02, .20)	.20 (-.52, .52)	.40 (.39, .52)	
READ	.15 (.11, .21)	-.20 (-.20, .26)	-.08 (-.26, .26)	.00 (-.23, .24)

Note. PRINT1 = Preschool Print Awareness, PA1 = Preschool Phonological Awareness, RN1 = Preschool Rapid Naming, = kindergarten Reading.