Published in final edited form as:

J Exp Child Psychol. 2013 July ; 115(3): 453–467. doi:10.1016/j.jecp.2013.03.008.

The Genetic and Environmental Etiologies of Individual Differences in Early Reading Growth in Australia, the United States, and Scandinavia

Micaela E. Christopher^{*,a,b}, Jacqueline Hulslander^b, Brian Byrne^{c,d}, Stefan Samuelsson^d, Janice M. Keenan^e, Bruce Pennington^e, John C. DeFries^{a,b}, Sally J. Wadsworth^b, Erik Willcutt^{a,b}, and Richard K. Olson^{*,a,b,d}

^aDepartment of Psychology and Neuroscience, University of Colorado Boulder, 345 UCB, Boulder, CO 80309-0345

^bInstitute of Behavioral Genetics, University of Colorado Boulder, 447 UCB, Boulder, CO, 80309-0447

^cSchool of Behavioural, Cognitive and Social Sciences, University of New England, Armidale NSW 2351, Australia

^dDepartment of Behavioral Sciences and Learning, Linköping University, 581 83 Linköping, Sweden

^eDepartment of Psychology, University of Denver, 2155 South Race, Denver, CO 80208

Abstract

This first cross-country twin study of individual differences in reading growth from post-kindergarten to post-2nd grade analyzed data from 487 twin pairs from the United States, 267 pairs from Australia, and 280 pairs from Scandinavia. Data from two reading measures were fit to biometric latent growth models. Individual differences for the reading measures at post-kindergarten in the U.S. and Australia were due primarily to genetic influences, and to both genetic and shared environmental influences in Scandinavia. In contrast, individual differences in growth generally had large genetic influences in all countries. These results suggest that genetic influences are largely responsible for individual differences in early reading development. In addition, the timing of the start of formal literacy instruction may affect the etiology of individual differences in early reading development, but have only limited influence on the etiology of individual differences in growth.

Keywords

Individual differences; Early reading development; Genetic; Twins; Cross-Linguistic

Understanding why children vary in their early reading development is the goal of much research. The focus of the present study is on the genetic and environmental etiologies of individual differences in early reading development, specifically rates of growth, within the first years of primary school. What sets this study apart from previous biometrical growth modeling research of early reading development is our inclusion of identical and fraternal twin data from Australia and Scandinavia (i.e., Norway and Sweden), in addition to the United States (U.S.). This allows us to address the question of how genetic and

*Corresponding authors. Phone 303-735-6595 (M. Christopher) or 303-492-8865 (R. Olson).

environmental influences on individual differences in early word reading and non-word reading growth vary cross-nationally and cross-linguistically.

Recent behavioral genetic research has begun exploring the genetic and environmental etiologies of individual differences in growth on reading in three different U.S. twin samples (Christopher et al., 2013; Hart et al., 2013; Logan et al., in press; Petrill et al., 2010). Rather than focusing on static time points, these biometric growth-curve studies have explored whether the genetic and environmental factors that drive differences as children start to read continue to affect subsequent reading growth across the early grades.

In our previous study (Christopher et al., 2013), we fit biometric growth models to reading and spelling data from twins in Colorado. The twins were tested starting at either the end of kindergarten (for two fluency- and accuracy-based word reading and non-word reading measures) or the end of first grade (for reading comprehension and spelling) and were tested through the end of fourth grade. The results showed that for all four measures, genetic factors were the main influences on individual differences at the end of first grade (the intercept), with small and nonsignificant shared environmental influences. Patterns of genetic and environmental influences for growth on the measures from first grade to fourth grade also showed strong and significant genetic influences. Spelling and reading comprehension also showed evidence of moderate, although not significant, shared environmental influences. We concluded that, for our measures and sample, variance in both reading and spelling ability at the end of first grade and subsequent growth was driven primarily by genetic influences. In addition, we found that, overall, there was very little variance in rates of growth on the measures compared to the much larger variance at intercept; thus, our twins appeared to grow on the reading and spelling measures at roughly the same rate.

Recent studies from an Ohio twin sample have used biometric growth modeling to test the etiology of individual differences in early reading development. The first, Petrill et al. (2010), found moderate genetic and strong shared environmental influences on three reading measures (letter identification, word reading, and decoding) at the first assessment wave (the intercept) when the twins ranged in age from 4.33 to 7.92 years ($M = 6.07$). Individual differences in linear growth on the three reading measures over the next two yearly assessment waves, however, were driven solely by shared environmental influences.

Logan et al. (in press) expanded upon Petrill et al. (2010) by including three additional yearly assessment waves up until the twins' mean age was 12.17-years, by including a reading comprehension measure, by fitting the data to nonlinear growth models, and by testing how the estimates of genetic and environmental variance at each wave changed as the children aged (i.e., by running multiple additional models that centered the intercept at each subsequent assessment wave). Individual differences for growth (both linear slope and quadratic growth) on word reading and non-word reading now had both genetic and shared environmental influences, although only shared environmental influences were significant for growth on reading comprehension. In addition, Logan et al. found that the estimates of genetic and shared environmental influences at the intercept varied depending on the wave on which the intercept was centered. While their main analyses set the intercept at the first assessment wave when the children were between four and seven years old, the later assessment waves had much smaller amounts of variance overall and recentering the intercept at later waves resulted in shared environmental estimates sharply declining and genetic estimates increasing compared to when the intercept was centered at the first wave. While overall variance in each assessment wave declined over time, these results suggest that the declines were largely due to decreases in shared environmental variance rather than decreases in genetic or nonshared environmental variance. Finally, Logan et al. found that

the genetic influences on growth rates were independent of the genetic influences on the first assessment wave, and there was significant negative overlap between the shared environmental influences at the first wave and subsequent rates of growth.

An additional study from Hart et al. (2013), using a racially and socioeconomically diverse sample of Florida twins, found that individual differences in oral reading fluency at the beginning of first grade are driven primarily by genetic influences, similar to the magnitudes of genetic influences found by Christopher et al. (2013). Variance in growth on the measure, both linear and quadratic, through the end of fifth grade was split between genetic and shared environmental influences in line with Logan et al. (in press). In addition, there was significant positive overlap between the genetic influences at the end of first grade and subsequent rates of growth, with genetic factors associated with higher oral reading fluency initially also associated with faster rates of growth. There was no significant overlap between shared environmental influences on individual differences at intercept and individual difference in growth rates.

To summarize, recent behavioral genetic studies exploring the etiology of individual differences in early reading growth have shown that growth on some early reading and reading-related measures is due primarily to genetic factors (i.e., fluent and accurate word reading and non-word reading in Christopher et al., 2013), with other measures showing evidence for small and non-significant shared environmental influences in addition to large genetic influences (i.e., reading comprehension and spelling in Christopher et al., 2013), both significant genetic and shared environmental influences (word reading and decoding in Logan et al., in press; oral reading fluency in Hart et al., 2013), or primarily shared environmental influences (reading comprehension in Logan et al., in press). The etiology of individual differences in reading measures, therefore, may vary depending on the measures and sample. The small number of studies overall, however, as well as the discrepant findings, suggests the need for additional analyses using independent samples.

In addition, all previous biometric growth models of early reading development were fitted to data from U.S. twins. It is possible that children learning to read in countries with different approaches to literacy instruction or different orthographies may show different genetic and environmental etiologies. The present study includes children from Scandinavia, Australia, and the U.S. (We note here that Norway and Sweden are combined together to increase sample size and are jointly referred to as “Scandinavia.”) Previous studies using the same dataset, the International Longitudinal Twin Study (ILTS; Byrne et al., 2009), argued that the initiation of formal literacy instruction, rather than general cognitive development or other developmental factors, leads to increasing genetic estimates and decreasing shared environmental estimates by the end of the first year of formal literacy instruction. Samuelsson et al. (2007) found that individual differences in post-kindergarten reading ability for Australian twins were largely due to genetic factors while U.S. twins showed evidence of modest shared environmental influences in addition to genetic influences. The Australian twins primarily live in New South Wales, a region of Australia that provides full-day kindergarten with a strong emphasis on literacy instruction. The U.S. twins, on the other hand, were recruited from Colorado, which has a mixture of full- and half-day kindergartens and heterogeneity in the amount of time dedicated to literacy instruction in kindergarten. Thus, the Australian twins received similar amounts of literacy instruction, leading to less environmental variance related to reading at the end of kindergarten than the U.S. twins. By the end of first grade, however, the U.S. twins had all received at least one year of formal literacy instruction; thus, shared environmental influences in the U.S. sample were smaller, no longer significant, and similar to the Australian sample (Byrne et al., 2007; 2009).

While Norway and Sweden do provide schooling for kindergarten-aged children, traditionally the focus in Scandinavian kindergartens has been on social development; literacy instruction is not emphasized until first grade (Lundberg, 1999)¹. Scandinavian kindergarteners, therefore, show lower genetic and higher shared environmental influences compared to U.S. and Australian children at the end of kindergarten (Samuelsson et al., 2007). How the etiology of variance in growth rates in Scandinavia compares to the U.S. and Australia remains an open question.

Of course, differences in children's early reading development in the U.S., Australia, and Scandinavia may not be due only to differences in early literacy education. Most obviously, the Scandinavian children are learning to read more transparent orthographies than English. Previous phenotypic research has shown that word reading accuracy is already close to ceiling within the first few months of formal reading instruction in orthographically transparent languages such as Italian and German (e.g., Cossu, Giuliotta, & Marshall, 1995; Landerl & Wimmer, 2008). To the extent that learning to read in a more transparent orthography affects reading development, differences in biometric estimates for growth rates between the countries may be found (see Ziegler & Goswami, 2005 for a review of the role of orthography in reading acquisition).

In order to best understand the etiology of individual differences in learning to read, it is important to analyze the etiologies of individual differences in early reading development in multiple countries and orthographies. By adding samples from Scandinavia and Australia, the present study expands upon the results of Christopher et al. (2013) by comparing each sample's etiology of individual differences in post-kindergarten word reading and non-word reading as well as individual differences in rates of growth through the end of second grade.² If the etiologies of word reading and non-word reading growth are similar once all children have started formal literacy education, that would argue against claims that societal differences, including orthography and approaches to literacy education, are the primary sources of individual differences in early word reading and non-word reading growth.

Methods

Participants

Participants in the current study are part of the ongoing International Longitudinal Twin Study (ILTS; Byrne et al., 2009) that includes twins from Australia, U.S., and Scandinavia. The twins were recruited from birth records in Colorado and Scandinavia and from the (volunteer) Australian Twin Registry in Australia. Zygosity was determined from DNA extracted from cheek swabs, or in a minority of cases from selected items from the Nichols and Bilbro (1966) questionnaire. All twins were learning to read in their first language (English for Australia and the U.S., Norwegian or Swedish in Scandinavia).

The U.S. sample at the end of kindergarten consisted of 224 monozygotic (MZ; i.e., identical) twin pairs and 263 same-sex dizygotic (DZ; i.e., fraternal) twin pairs for a total of 487 pairs. By post-2nd grade the sample consisted of 221 MZ and 261 DZ pairs for a total of 482 pairs. The Australian sample at the end of kindergarten consisted of 152 MZ twin pairs and 115 DZ twin pairs for a total of 267 pairs. By post-2nd grade, the Australian sample

¹Since 2006 Norwegian kindergarteners do receive formal literacy instruction in kindergarten. However, this change was instituted after our data collection.

²Christopher et al. (2013) included data from the end of fourth grade and estimated the intercept at the end of first grade. We only have complete fourth grade data for our U.S. sample. In order to make the countries as comparable as possible, the present study only includes data that was present in all of the countries. In addition, the intercept was moved to the end of kindergarten in the present study in order to estimate the genetic and environmental etiologies of individual differences in reading as children first started primary school.

included 120 MZ and 92 DZ twin pairs, a total 212 pairs. The Scandinavian sample at the end of kindergarten consisted of 138 MZ twin pairs and 142 DZ twin pairs for a total of 280 pairs. The post-2nd grade Scandinavian sample consisted of 122 MZ twin pairs and 127 DZ twin pairs, a total 249 pairs.

Mean levels of parental education were comparable across samples: U.S. mean level (standard deviation) was 14.32 years (2.21); Australia mean level was 13.46 years (1.78); Scandinavia mean level was 13.31 years (2.35). Mean ages in years for the U.S. sample (standard deviation, range) were 6.27 (.31, 5.50–7.08), 7.42 (.32, 6.58–8.67), and 8.45 (.31, 7.67–9.50) for the post-kindergarten, post-1st grade, and post-2nd grade waves, respectively. Mean ages in years at each wave for the Australian sample (standard deviation, range) were 6.07 (.35, 5.33–6.83), 7.00 (.35, 6.17–7.75), and 7.95 (.37, 7.25–8.67). For the Scandinavian sample, mean ages in years (standard deviation, range) at each wave were 6.75 (.29, 6.17–7.67), 7.74 (.32, 7.08–8.50), and 8.74 (.30, 8.17–9.50). All age differences were significant ($p < .01$), with the Australian sample youngest at each wave and the Scandinavian sample oldest at each wave.

After correcting for non-independence, significant sex effects were found in the Australian sample for post-1st and post-2nd non-word reading (post-1st $d = .29$; post-2nd $d = .34$; with boys scoring higher than girls at both time points), as well as in the Scandinavian sample for post-kindergarten and post-1st grade word reading and post-kindergarten non-word reading (post-kindergarten word reading $d = .31$; post-1st grade word reading $d = .23$; post-kindergarten non-word reading $d = .28$; girls scoring higher than boys at both time points). To control for the sex differences and possible age differences, age at the first wave (post-kindergarten) and sex were controlled for as definition variables in Mx (Neale, Boker, Xie, & Maes, 2003) in all analyses.

Procedure and Measures

The measures included in the present analyses are from larger test batteries that were administered in the ILTS either in the final three to four months of the school year (Australia) or in the summer after each school year (U.S. and Scandinavia). Testing at each time point was conducted in a single session lasting about 1 hour in the twins' homes or schools. Two testers separately assessed each twin at the same time, except for the Scandinavian sample where only one tester per pair was available. For all measures, raw scores based on total number correct were used.

Reading—The *Test of Word Reading Efficiency* (TOWRE; Torgesen, Wagner, & Rashotte, 1999), Form A was administered post-kindergarten, post-1st grade, and post-2nd grade. In the Sight Word Efficiency subtest (word reading) children read a list of difficulty-ordered words as quickly as possible, with the score being the number correctly read in 45 seconds. The Phonemic Decoding Efficiency subtest (non-word reading) is a list of pronounceable non-words administered in the same way. Test-retest reliability for English-speaking children aged 6- to 9-years-old was reported as .97 for word reading and .90 for non-word reading.

All tests were translated into Swedish and Norwegian for the Scandinavia sample. For the word reading test, the translations sought to maintain semantic, syllabic, orthographic, and phonemic similarities to the original English version. For example, the Swedish version used words that were familiar to young readers (as judged by the Swedish and Norwegian ILTS researchers), with 83% of the words having the same number of syllables as the English version and 76% of the words being the same semantically. For the non-word reading test, approximately 50% of the words used on the Swedish and Norwegian versions are exactly the same as the English version. For the remaining non-words, only minor changes (i.e.,

replacing a single letter) were made. These changes were necessary to avoid having stimuli that were actual Swedish or Norwegian words or to make the non-words follow common spelling rules.

Details of Analyses

To assess genetic, shared environmental, and non-shared environmental influences on initial performance and on subsequent growth on the reading measures, we fit biometric growth models (see McArdle, Prescott, Hamagami, & Horn, 1998; Reynolds et al., 2005) using the Mx software (Neale et al., 2003). As shown in Figure 1, biometric growth modeling is a combination of latent growth curve modeling (Byrne & Crombie, 2003; Loehlin, 1998; Meredith & Tisak, 1990; Preacher, Wichman, MacCallum, & Briggs, 2008) and the standard twin model that decomposes phenotypic variance into independent genetic variance (a^2), shared environmental variance (c^2), and non-shared environmental variance (e^2). The latent intercept factor represents ability level at Time 1, while the latent shape factor represents increases in scores over time. The loadings from intercept are fixed to 1 as intercept does not change, while the loadings from shape are scaled to the interval between time points. Unlike a linear growth model, where all growth loadings are fixed *a priori* to a linear function, we allowed the loading from shape to the third time point to be estimated in the model³. Finally, variance unique to each time point (including factors such as family disruptions, motivation, tester effects, and time-point-specific error) that is not captured by the latent factors is denoted by the small u 's under each time point. The correlations between unique variance for twin one and twin two are constrained to be equal within zygosity, but can vary across zygosity. Given that biometric growth curve models treat each twin pair as one observation, these correlations capture twin similarity unrelated to the latent growth model (see Christopher et al., 2013 for more details including comparing models with and without these correlations).

As in the twin model, the covariance between Twin 1 and Twin 2 on both intercept and shape is a combination of genetic influences (A) and shared environmental influences (C). The genetic and shared environmental intercept and shape correlations between Twin 1 and Twin 2 for MZ twins are set to 1 as they share 100% of both their genes and shared environment (see Figure 1). For DZ twin pairs, the correlations between Twin 1 and Twin 2 for intercept C and shape C are also set to 1, as they also share 100% of shared environment. However the twin correlations for both intercept A and shape A are set to .5 as DZ twins share 50% of their segregating genes on average. There are no correlations between twins for non-shared environment (E), as that is independent for each member of the twin pair. Squaring the resulting loadings from each A, C, and E factor onto intercept and shape estimates the proportion of variance accounted for by genetic, shared environmental, and non-shared environmental influences.

Results

Descriptive Statistics and Univariate Heritability Estimates

The means and standard deviations for the measures are presented in Table 1. While raw scores are used in the biometric growth modeling, standard scores (SS) for word reading and non-word reading are shown in the table to allow comparison of each sample to the U.S. standardizing populations for the measures. The standard score means show that the U.S.

³The decision to use non-linear growth curves was based on the fact that, as seen in Table 1, our subjects grew more between the first and second waves than between the second and third waves. For example, on the word reading test, the U.S. sample grew an average of 27.40 raw score points between post-kindergarten and post-1st grade, but only 14.04 raw score points between post-1st grade and post-2nd grade. Non-linear growth models also always resulted in significantly better model fit over linear growth models.

twins were relatively close to the U.S. standardizing population standard score mean of 100. Despite the fact that the Australian twins were significantly younger than the U.S. twins, the standard scores for the Australian twins were significantly higher than those of the U.S. twins on all measures at all waves ($p < .01$), suggesting that the Australian twins had received more literacy instruction. Conversely, although the Scandinavian twins were significantly older than the U.S. twins, the mean word reading and non-word reading standard score for the Scandinavian twins were significantly lower than the U.S. sample ($p < .01$), except on post-2nd grade non-word reading ($p = .08$).

The standard score standard deviations on word reading and non-word reading were generally less than the standardizing population average of 15. This was especially true at the first wave, potentially due to floor effects. Examination of the raw means and standard deviations also supports the existence of floor effects for all three samples, particularly in Scandinavia. This finding, as well as the low standard scores on word reading and non-word reading at post-kindergarten for the Scandinavian twins, partially reflects the lack of formal literacy instruction in Scandinavian kindergartens (Samuelsson et al., 2008).

Also shown in Table 1 are the results from the univariate analyses that estimate how much variance in each measure at each time point is due to genetic (a^2), shared environmental (c^2), and nonshared environmental (e^2) influences. In general, individual differences in the measures were driven by large genetic influences, with small and nonsignificant shared environmental influences. Significant shared environmental influences were only present at post-kindergarten word reading for the American and Scandinavian samples, as well as post-kindergarten non-word reading for the Scandinavian sample. Nonshared environmental influences, which include measurement error, were moderate and significant for all time points and measures. The a^2 , c^2 , and e^2 estimates are proportions of overall variance, with decreases in one leading to increases in the others. The decrease in the magnitudes of the shared environmental influences between post-kindergarten and post-1st grade support the idea that formal literacy instruction limits shared environmental variance, leading to lower shared environmental estimates and increased genetic estimates.

Biometric Growth Models

Table 2 presents the biometric growth model results for word reading and non-word reading. In order to capture nonlinear growth, all measures were fit to unspecified growth models, with the loading from shape onto the third time point estimated from the model (shown in Table 2). For all measures and countries, the 3rd loadings were less than 2, suggesting that participants gained more raw score points between post-kindergarten and post-1st grade than between post-1st grade and post-2nd grade.

As noted earlier, Norway and Sweden were combined together to increase power; thus, we standardized the word reading and non-word reading raw data relative to the first wave's mean and variance (within country). While using raw variables is standard practice in latent growth modeling, standardizing relative to the first wave does not change any of the psychometric properties of the raw variables and provides the exact same estimates for variance and patterns of growth as using raw variables (Ferrer, Balluerka, & Widaman, 2008). Although the variables were standardized relative to each country's first time point, Table 2 presents the raw variables results in order to aid in interpretation of the latent intercept and shape values. For example, mean shape for Australian word reading is 23.49. Multiplying 23.49 by the loading from shape to the third time point (1.57) estimates that the mean word reading score for Australian twins grew approximately 36.88 points from the end of kindergarten to the end of second grade.

In all three samples, individual differences in post-kindergarten word reading and non-word reading (intercept) had significant genetic influences, with estimates ranging from .30 (non-word reading in Scandinavia) to .85 (word reading in Australia). Significant shared environmental influences were found on post-kindergarten word reading in Colorado and Scandinavia (.29 and .56, respectively), as well as non-word reading in Scandinavia (.68). While the large confidence intervals limit our ability to assess significant differences between countries, the magnitudes of the estimates as well as the patterns of statistical significance suggest that post-kindergarten individual differences on word reading and non-word reading scores in the American and Australian samples were due primarily to genetic influences. In contrast, individual differences in the Scandinavian sample appeared to be jointly influenced by both genetic and shared environmental influences. Consistent with the univariate results shown in Table 1, these results suggest that starting formal literacy instruction may decrease environmental variance, resulting in larger genetic estimates and nonsignificant shared environmental estimates.

While biometric estimates on post-kindergarten intercept varied cross-country, the estimates for shape were largely similar. With the exception of word reading in Australia which did not have significant genetic or shared environmental estimates for shape, individual differences in growth through the end of second grade on word reading and non-word reading appeared to be largely due to genetic influences, with estimates ranging from .57 (non-word reading in Scandinavia) to .68 (non-word reading in the U.S.). Shared environmental influences were only significant for word reading and non-word reading in Scandinavia (.17 and .25, respectively). In addition, small and significant estimates of nonshared environmental influences were found in all three samples, and were sometimes of similar magnitude to the shared environmental estimates. Because time-point-specific measurement error is captured by the unique variance terms tied to each assessment point, the latent growth factor is thought to be largely devoid of measurement error, such as tester effects, present at individual testing sessions. It is possible, therefore, that these estimates are capturing important child-specific variance in growth. However, these significant nonshared environmental influences could also reflect measurement error common across all assessment waves, such as general unreliability in the measure.

As with the intercept estimates, the large confidence intervals limit our ability to assess differences between countries. Nevertheless, the magnitudes of the estimates in the countries, with the exception of word reading in Australia, were similar, with moderate to large genetic estimates and smaller, largely nonsignificant, shared environmental estimates. The etiology of individual differences in post-kindergarten word reading and non-word reading, therefore, appears to vary somewhat due to country differences, while the etiology of individual differences in growth was largely similar and due primarily to genetic influences.

To briefly summarize, the results of the biometric growth models for the two reading variables, word reading and non-word reading, are consistent with the idea that individual differences in reading ability at the end of kindergarten are affected by how much formal literacy instruction the children have received. In Australia and the U.S., the amount of environmental variance in reading performance has decreased via literacy instruction, leading to higher genetic estimates and lower shared environmental estimates. In Scandinavia, where children have not received consistent literacy instruction across their kindergarten classrooms, significant shared environmental influences were present. Individual differences in growth in reading through the end of second grade, however, were due primarily to genetic influences in all of the countries.

Estimating overlap of biometric influences between latent intercept and latent shape

We are able to explore the relation between intercept and shape for word reading and non-word reading by assessing the extent to which genetic and shared environmental influences present at intercept overlap with the biometric influences on shape (see Table 4). While all countries showed significant overlap for nonshared environmental influences on word reading and non-word reading, genetic and shared environmental influences on intercept overlapped with growth on word reading only in Scandinavia ($r_{\text{genetic}} = .76$, $r_{\text{shared environmental}} = -1.00$). Some of the genetic factors that made Scandinavian twins higher at the end of kindergarten on word reading, therefore, were related to faster growth rates through the end of second grade. Conversely, shared environmental influences that made children higher at the end of kindergarten on word reading were related to slower growth rates. A similar pattern was found for Scandinavian twins on non-word reading; the shared environmental correlation was negative and significant ($r_{\text{shared environmental}} = -1.00$), and the genetic correlation was positive, although not significant ($r_{\text{genetic}} = .67$).

The Australian and U.S. twins had nonsignificant shared environmental correlations for non-word reading intercept and shape. However, there was significant positive overlap between the genetic factors on intercept and shape in the Australian and U.S. samples ($r_{\text{genetic}} = .69$ and $.91$, respectively). Thus, the genetic influences on post-kindergarten non-word reading also affected how quickly the Australian and American children grew on non-word reading ability.

Discussion

The present study extends recent behavioral genetic results on the etiology of individual differences in early reading growth by including twin samples from outside the U.S. By using data from the ILTS (Byrne et al., 2009), one of only two twin studies that measure early reading ability in samples outside of the U.S. (the other being the Twins Early Development Study in England; Trouton, Spinath, & Plomin, 2002), we were able to examine cross-country (U.S., Scandinavia, and Australia) differences on the etiology of individual differences in early word reading and non-word reading growth. This extension allows us to explore two important potential sources of these etiological differences. First, all three countries differ in the cultural and instructional environments for reading. Second, our study is the first of its type to include a non-English sample, allowing us to assess differences that might occur with an orthography more transparent than English. Given the complexity of our results, we will briefly review our main findings before discussing their implications and how they expand upon previous research.

Etiology of Individual Differences in Early Reading and Subsequent Growth

Overall we found that the etiology of individual differences in post-kindergarten word reading and non-word reading appeared to vary somewhat due to country differences, while the etiology of individual differences in growth was largely similar and due primarily to genetic influences in all samples. Specifically, the Australian and U.S. samples showed large and significant genetic influences on word reading and non-word reading at the end of kindergarten, the intercept. Shared environmental influences on intercept, on the other hand, were much smaller in magnitude and, in the case of Australia, not significant. In contrast to the U.S. and Australian results, individual differences in the Scandinavian sample's intercept showed significant moderate genetic and shared environmental influences. We note here that previous research using the ILTS dataset found that pre-reading skills measured in preschool, such as letter name and sound knowledge, were predominantly influenced by shared environment in all of the countries (Samuelsson et al., 2005). In addition, the univariate results in Table 1 show that genetic estimates in the Scandinavian sample

increase, and shared environmental estimates decrease, by the end of first grade. Given that the Scandinavian sample did not start formal literacy instruction until first grade, and the Australian and U.S. twins both received at least some formal literacy instruction in kindergarten, the end-of-kindergarten intercept results suggest that environmental variance is largely minimized after a year of formal reading instruction, and children's reading ability in relation to their peers becomes more strongly tied to genetic factors.

Turning to the shape results, individual differences in rates of growth on the two reading measures showed moderate to large genetic influences. The genetic influences were significant for non-word reading in the three samples, and were significant for word reading in the U.S. and Scandinavia. Shared environmental influences for word reading and non-word reading were only significant in Scandinavia. With the exception of Australia on word reading, the shared environmental estimates were of smaller magnitude than on intercept. These results argue that the etiology of individual differences in early reading growth in societies that provide consistent and universal literacy education is largely due to genetic influences, especially after the first year of formal reading instruction.

Extending Previous Biometric Growth Curve Analyses

The current results expand upon previous biometric growth curve analyses of early reading that utilized data from U.S. twins only (Christopher et al., 2013; Hart et al., 2013; Logan et al., in press; Petrill et al., 2010). For example, our previous study found strong genetic influences on individual differences in early reading ability and subsequent growth (Christopher et al., 2013). The present results provide evidence from two additional samples to support the hypothesis that, within a year of consistent literacy instruction, variance in how quickly children's reading develops is generally more influenced by genetic than environmental factors.

We note that our estimates of genetic, shared environmental, and nonshared environmental influences are proportional and dependent upon the total amount of genetic and environmental variance in the sample. Increasing the overall environmental variance in a sample by encompassing a larger range of socioeconomic statuses or increased educational variance could result in different etiological patterns. For example, biometric growth curve analyses using a twin sample in Ohio have found much larger shared environmental estimates than our samples for both the intercept and growth on word reading and non-word reading when there was a much wider range of literacy instruction in their early test waves than in the samples used in the present study (Logan et al., in press; Petrill et al., 2010).

Limitations

Before discussing the larger implications of our results, it is important to highlight limitations that our study shares with other developmental studies of early readers as well as other behavioral genetic studies. First, as noted above, the genetic and environmental estimates are proportions of overall variance, and samples with more extreme environmental variance may show larger shared environmental influences. In addition, the overall variance in our measures at the first wave was also affected by the existence of floor effects. While floor effects are common and nearly impossible to avoid in studies of early readers (e.g., Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009), the fact that our measures were largely insensitive to differences amongst our lowest readers may somewhat limit our ability to assess individual differences in growth. The large confidence intervals, especially in Australia and Scandinavia, limited our ability to compare genetic and environmental estimates between countries. Consequently, we only described how the overall patterns compared in general.

In addition, our conclusions regarding the primary role of genetic influences on variance in early reading development are based both on the magnitude of the estimates as well as patterns of statistical significance. The large confidence intervals on the shared environmental estimates, likely due to the small amount of shape variance, indicate that we do not have the power to detect small to moderate shared environmental influences in the present samples.

Also, while not a limitation of the data but instead a feature of all behavioral genetic twin studies, it is important to note that the present study is concerned with differences in the etiology of *variance* within the samples, and not mean differences between the samples. It is not accurate to say, therefore, that our results show that the Scandinavian twins' average word reading and nonword reading scores were lower than the U.S. and Australian twins' scores at the end of kindergarten because of genetic and environmental factors.

Finally, it is important to make clear that we are limited in our ability to generalize our results to different languages and societal approaches to reading instruction. Additional research in this field is needed, especially in studies that include orthographically diverse samples. Inclusion of Danish subjects, for example, would be interesting as Danish has lexical and grammatical properties similar to Norwegian, but has a less transparent orthography.

Implications

The amount of variance in the measures affects the practical implications of our results. As shown in Table 2, with the exception of U.S. non-word reading, there was more variance on word reading and non-word reading intercepts than on rates of growth (approximately two to five times more), and the overall magnitude of shape variance was small. The small shape variances suggest that, within each of the samples, our twins were growing at fairly similar rates; thus, individual differences in rates of growth had less influence on children's reading levels at the end of second grade compared to individual differences in ability at the end of kindergarten.

The present study is an important addition to early reading research, as it is the first to explore the etiology of individual differences in early reading growth in samples from outside of the U.S. Although reading ability at the first wave (post-kindergarten) was affected by whether the pairs had started formal reading instruction, the etiological patterns of growth in the countries were similar. While all of the twins received literacy instruction from teachers in their schools, the small to moderate (and largely nonsignificant) shared environmental estimates for growth rates argues against claims that differences amongst educational approaches to teaching reading are a primary determiner of how quickly children learn to read. It is important to emphasize that, although the exact timing varied, all our pairs had consistent and universal access to literacy education. We are unable to generalize our results to children growing up in areas without this level of commitment to literacy education.

It is also important to note that we are limited in this study to the development of word reading and non-word reading. While Christopher et al. (2013) included reading comprehension and spelling measures as well, we were unable to include those measures in the present study as we only have complete data on these measures in Scandinavia and Australia at two time points (post-1st and post-2nd grades). Given that Christopher et al. (2013) found moderate genetic estimates with moderate, although nonsignificant, shared environmental estimates for reading comprehension and spelling growth from post-1st to post-4th grades in the U.S. sample, different aspects of reading may display different patterns of etiology for growth rates. Future studies aimed at early reading development

should include multiple measures of reading, and carefully distinguish the results with respect to different aspects of reading ability.

Conclusion

The goal of the present study was to begin to explore how the etiology of individual differences in early word reading and non-word reading growth rates varied in Australia, Scandinavia, and the U.S. While all of these countries provide their children with formal literacy instruction, the exact timing of the start of the instruction varies. The results were consistent with the idea that differences in when formal literacy instruction starts in schools changes the balance of genetic and shared environmental influences, but that the etiology of early reading growth through the end of second grade is largely similar. While additional studies using other types of languages and approaches to literacy instruction are needed, the current findings suggest that individual differences in early reading growth are largely driven by genetic influences.

Acknowledgments

Funding was provided by the National Institutes of Health, grant numbers P50 HD027802 for the Colorado Learning Disabilities Research Center, and R01 HD038526 for the Colorado component of the International Longitudinal Twin Study (ILTS). The Australian component of the ILTS was facilitated through access to the Australian Twin Registry, a national resource supported by an Enabling Grant (ID 628911) from the National Health & Medical Research Council. Funding was provided by the Australian Research Council (DP0663498 and DP0770805). The Scandinavian component of the ILTS was supported by the Research Council of Norway 154715/330, the Swedish Research Council grants 345-2002-3701, PDOKJ028/2006:1, and 2011-1905, and the Swedish Council for Working Life and Social Research (2011-0177). We thank the twins and their families who participated in our research.

References

- Byrne B, Coventry WL, Olson RK, Samuelsson S, Corley R, Willcutt EG, DeFries JC. Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to grade 2. *Journal of Neurolinguistics*. 2009; 22:219–236. [PubMed: 20161176]
- Byrne B, Fielding-Barnsley R. Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology*. 1993; 85:104–111.
- Byrne B, Samuelsson S, Wadsworth S, Hulslander J, Corley R, DeFries JC, Olson RK. Longitudinal twin study of early literacy development: Preschool through grade 1. *Reading and Writing*. 2007; 20:77–102.
- Byrne BM, Crambie G. Modeling and testing change: An introduction to the latent growth curve model. *Understanding Statistics*. 2003; 2:177–203.
- Catts HW, Petscher Y, Schatschneider C, Bridges MS, Mendoza K. Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*. 2009; 42:163–176. [PubMed: 19098274]
- Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington B, Olson RK. Modeling the etiology of individual differences in early reading development: Evidence for strong genetic influences. *Scientific Studies of Reading*. 2013 Advance online publication.
- Cossu G, Giuliotta M, Marshall JC. Acquisition of reading and written spelling in a transparent orthography: Two non-parallel processes? *Reading and Writing*. 1995; 7:9–22.
- Ferrer E, Balluerka N, Widaman KF. Factorial invariance and the specification of second-order latent growth models. *Methodology*. 2008; 4:22–36. [PubMed: 20046801]
- Hart SA, Logan JAR, Soden-Hensler B, Kershaw S, Taylor J, Schatschneider C. Exploring how nature and nurture affect the development of reading: An analysis of the Florida Twin Project on Reading. *Developmental Psychology*. 2013 Advance online publication.
- Jastak, S.; Wilkinson, GS. *The Wide Range Achievement Test-Revised: Administration manual*. Wilmington, DE: Jastak Associates, Inc.; 1984.

- Landerl K, Wimmer H. Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*. 2008; 100(1):150–161.
- Lieberman, I.; Rubin, H.; Duqués, S.; Carlisle, J. Linguistic abilities and spelling proficiency in kindergartners and adult poor spellers. In: Gray, D.; Kavanagh, J., editors. *Behavioral measures of dyslexia*. Parkton, MD: York Press; 1985. p. 163–176.
- Logan JAR, Hart SA, Cutting L, Deater-Deckard K, Schatschneider C, Thompson LA, Petrill SA. Reading development in young children: Genetic and environmental influences. *Child Development*. (in press).
- Loehlin, JC. *Latent variable models: An introduction to factor, path, and structural analysis*. 3rd ed.. Mahwah, NJ: Lawrence Erlbaum Associates; 1998.
- Lundberg, I. Literacy in Scandinavia. In: Wagner, DA.; Venezky, RL.; Street, BV., editors. *Literacy: An international handbook*. Oxford: Westview Press; 1999. p. 396–399.
- McArdle JJ, Prescott CA, Hamagami F, Horn JL. A contemporary method for developmental-genetic analyses of age changes in intellectual abilities. *Developmental Neuropsychology*. 1998; 14:69–114.
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika*. 1990; 55(1):107–122.
- Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. *Mx: Statistical modeling*. 6th ed.. Richmond, VA: Department of Psychiatry, Virginia Commonwealth University; 2003.
- Nichols RC, Bilbro WC Jr. The diagnosis of twin zygosity. *Acta Genetica et Statistica Medica*. 1966; 16:265–275. [PubMed: 4959152]
- Petrill SA, Hart SA, Harlaar N, Logan J, Justice LM, Schatschneider C, Cutting L. Genetic and environmental influences on the growth of early reading skills. *Journal of Child Psychology and Psychiatry*. 2010; 51:660–667. [PubMed: 20059623]
- Preacher, KJ.; Wichman, AL.; MacCallum, RC.; Briggs, NE. *Latent growth curve modeling*. Thousand Oaks, CA: Sage; 2008.
- Reynolds CA, Finkel D, McArdle JJ, Gatz M, Berg S, Pedersen NL. Quantitative genetic analysis of latent growth curve models of cognitive abilities in adulthood. *Developmental Psychology*. 2005; 41:3–16. [PubMed: 15656733]
- Samuelsson S, Byrne B, Quain P, Wadsworth S, Corley R, DeFries JC, Olson R. Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology*. 2005; 97(4):705–722.
- Samuelsson S, Olson R, Wadsworth S, Corley R, DeFries JC, Willcutt E, Byrne B. Genetic and environmental influences on prereading skills and early reading and spelling development in the United States, Australia, and Scandinavia. *Reading and Writing*. 2007; 20:51–75.
- Seymour PHK, Aro M, Erskine JM. Foundation literacy acquisition in European orthographies. *British Journal of Psychology*. 2003; 94:143–174. [PubMed: 12803812]
- Torgesen, JK.; Wagner, RK.; Rashotte, CA. *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-Ed.; 1999.
- Trouton A, Spinath FM, Plomin R. Twins early development study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Research and Human Genetics*. 2002; 5:444–448.
- Wimmer, H.; Landerl, K. How learning to spell German differs from learning to spell English. In: Perfetti, CA.; Rieben, L.; Fayol, M., editors. *Research, theory, and practice across languages*. Mahwah, NJ: Erlbaum; 1997. p. 81–96.
- Ziegler JC, Goswami U. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*. 2005; 131:3–29. [PubMed: 15631549]

Figure 1.

Biometric growth model. Not shown in the figure is that age at the 1st time point and sex are controlled for as definition variables in Mx. The correlations between the unique variances (u's) are constrained equal within zygosity but allowed to vary across zygosity.

Table 1

Descriptive and Univariate Statistics.

		SS		Raw		Univariate Estimates		
	n	Mean	SD	Mean	SD	a^2	c^2	e^2
Word Reading (TOWRE Sight)								
United States								
Post-KG	973	96.99	10.54	12.85	12.63	.68* [.54, .85]	.22* [.05, .36]	.10* [.08, .12]
Post-1 st	955	102.23	14.02	40.25	16.68	.84* [.66, .89]	.02 [.00, .20]	.14* [.11, .17]
Post-2 nd	964	103.04	14.28	54.29	14.30	.73* [.55, .84]	.08 [.00, .24]	.19* [.16, .24]
Australia								
Post-KG	522	104.10	10.37	20.70	13.58	.74* [.51, .90]	.13 [.00, .36]	.13* [.10, .16]
Post-1 st	502	110.22	12.83	43.78	16.26	.83* [.60, .87]	.00 [.00, .22]	.17* [.13, .22]
Post-2 nd	415	110.81	13.13	57.65	13.34	.71* [.40, .82]	.05 [.00, .34]	.24* [.18, .32]
Scandinavia								
Post-KG	560	81.99	10.89	6.36	10.72	.46* [.31, .65]	.43* [.23, .58]	.11* [.09, .15]
Post-1 st	576	84.97	13.41	25.58	14.62	.80* [.57, .87]	.03 [.00, .26]	.17* [.13, .22]
Post-2 nd	495	90.24	14.65	42.48	14.61	.78* [.52, .83]	.00 [.00, .24]	.22* [.17, .30]
Non-word Reading (TOWRE Decode)								
United States								
Post-KG	973	102.13	8.37	5.51	6.32	.62* [.43, .78]	.12 [.00, .30]	.26* [.21, .31]
Post-1 st	953	100.71	12.79	17.13	10.56	.82* [.70, .85]	.00 [.00, .12]	.18* [.15, .22]
Post-2 nd	964	100.22	13.42	23.94	11.58	.73* [.55, .84]	.07 [.00, .24]	.20* [.16, .25]

	n	SS		Raw		Univariate Estimates		
		Mean	SD	Mean	SD	a^2	c^2	e^2
Australia								
Post-KG	520	107.30	8.82	9.12	7.85	.83* [.54, .87]	.00 [.00, .28]	.17* [.13, .22]
Post-1 st	502	109.20	12.71	20.39	11.76	.60* [.31, .80]	.15 [.00, .42]	.26* [.20, .33]
Post-2 nd	414	109.86	14.59	29.24	12.47	.81* [.60, .86]	.00 [.00, .21]	.19* [.14, .25]
Scandinavia								
Post-KG	560	92.81	10.97	4.82	7.84	.42* [.26, .62]	.45* [.25, .60]	.13* [.10, .17]
Post-1 st	576	95.73	10.66	15.66	8.76	.67* [.43, .83]	.12 [.00, .34]	.22* [.17, .28]
Post-2 nd	495	98.65	11.14	23.18	9.50	.71* [.43, .80]	.02 [.00, .27]	.26* [.20, .36]

Note: KG = Kindergarten; a^2 = proportion of variance due to genetic influences; c^2 = proportion of variance due to shared environmental influences; e^2 = proportion of variance due to non-shared environmental influences. Sex and age at each time point have been controlled for as definitions variable in Mx.

* $p < .05$ determined using 95% confidence intervals.

Table 2

Estimates from the Biometric Growth Models of Word Reading and Non-word Reading KG-2nd: Means, Variance, and Proportions of Variance in Intercept and Change Over Time Due to Genetic and Environmental Influences.

	Raw Scores				3 rd Loading	r _{errors}	Latent Intercept			Latent Shape			
	Intercept		Shape				a ²	c ²	e ²	a ²	c ²	e ²	
	Mean	Var.	Mean	Var.									
Word Reading													
United States	12.72	132.40	27.79	41.47	1.49	MZ = .58 DZ = .23	.69* [.49, .93]	.29* [.06, .49]	.01* [.004, .03]	.62* [.19, .91]	.24 [.00, .65]	.13* [.07, .22]	
Australia	20.91	156.75	23.49	27.66	1.57	MZ = .48 DZ = .19	.85* [.55, .99]	.12 [.00, .42]	.03* [.01, .07]	.41 [.00, .92]	.43 [.00, .91]	.16* [.06, .37]	
Scandinavia	6.30	86.19	19.72	31.03	1.83	MZ = .66 DZ = .36	.40* [.18, .70]	.56* [.27, .76]	.04* [.01, .10]	.62* [.23, .85]	.17* [.003, .53]	.21* [.10, .36]	
Non-word Reading													
United States	5.40	25.96	11.94	25.56	1.54	MZ = .40 DZ = .11	.70* [.41, .97]	.24 [.00, .52]	.05* [.02, .11]	.68* [.39, .92]	.20 [.00, .47]	.12* [.06, .20]	
Australia	9.03	44.98	11.70	21.57	1.67	MZ = .39 DZ = .13	.73* [.34, .97]	.22 [.00, .60]	.05* [.02, .11]	.62* [.14, .91]	.20 [.00, .65]	.18* [.08, .32]	
Scandinavia	4.82	39.95	10.98	14.14	1.68	MZ = .57 DZ = .32	.30* [.09, .61]	.68* [.37, .87]	.03 [.01, .07]	.57* [.14, .84]	.25* [.02, .65]	.17* [.07, .34]	

Note: Word reading = TOWRE Sight Word Reading Efficiency subtest; Non-word Reading = TOWRE Phonemic Decoding Efficiency subtest; While raw scores are shown to aid interpretability, all models used raw scores normalized to Time 1; All models assume unspecified change and errors correlated across time-points and within-zygosity; 3rd Loading = the loading from latent shape factor onto 3rd time point; r_{errors} = correlation of time-point-specific errors were allowed to correlate within-zygosity; MZ = monozygotic; DZ = dizygotic; a² = proportion of variance due to genetic influences; c² = proportion of variance due to shared environmental influences; e² = proportion of variance due to non-shared environmental influences. Sex and age at the first time point have been controlled for as definitions variable in Mx.

* p < .05 determined using 95% confidence intervals, in brackets.

Table 4

Biometric Correlations Between Latent Intercept and Latent Shape.

		r_{genetic}	$r_{\text{shared environ.}}$	$r_{\text{nonshared environ.}}$
Word Reading KG-2nd	United States	.18 [−.14, .81]	−.77 [−1.00, 1.00]	1.00* [.29, 1.00]
	Australia	−.01 [−1.00, 1.00]	−1.00 [−1.00, 1.00]	1.00* [.004, 1.00]
	Scandinavia	.76* [.15, 1.00]	−1.00* [−1.00, −.23]	.65* [.003, 1.00]
	United States	.91* [.41, 1.00]	−.75 [−1.00, 1.00]	1.00* [.42, 1.00]
	Australia	.69* [.12, 1.00]	−.76 [−1.00, 1.00]	1.00* [.50, 1.00]
	Scandinavia	.67 [−.06, 1.00]	−1.00* [−1.00, −.41]	1.00* [.18, 1.00]

Note: Word reading = TOWRE Sight Word Reading Efficiency subtest; Non-word Reading = TOWRE Phonemic Decoding Efficiency subtest; Correlations estimate extent of overlap of biometric influences between latent intercept and latent shape.

* $p < .05$ determined using 95% confidence intervals, shown in brackets.