Making Reading Easier: How Genetic Information can help.

Michelle Luciano
Psychology, University of Edinburgh, 7 George Square, EH8 9JZ, Edinburgh, UK

Acknowledgement
The author would like to thank Wendy Johnson for her critical comments on an earlier version.
Abstract

Reading is the cornerstone of all educational systems. Without adequate reading skills, learning in all other subject areas becomes very challenging. Children typically learn basic reading skills by age 6 or 7 years, but there is variation in timing of reading skill acquisition, and thereafter, in reading ability. A strong predictor of whether a child will have difficulties with reading is family history. Twin and family studies confirm that reading ability (including specific reading disorder) is substantively genetically influenced. Molecular genetic studies of reading ability have identified a number of candidate genes that are associated with reading disability and/or reading processes; many more are likely to be discovered. This review discusses implications that increased understanding of the genetic architecture of reading ability has for early identification and intervention for children at risk for reading difficulties.

Tweet:
To maximize children’s reading ability, consider the roles of genes and moderating environmental influences in its development [127 characters]

Key Points:

- Genes influence variation in reading ability (and conversely, disability). Even with adequate instruction and support, some children still experience difficulty with reading.
- Strong genetic correlations link specific reading disorder and other specific disorders (dyscalculia, language), as well as reading and cognitive abilities in the normal range.
- Quality of teaching moderates genes’ influences on reading achievement: Genetic influences are more prominent in children who receive higher-quality teaching.
- Candidate genes link with reading (dis)ability, and rapid advances in molecular genetics mean that discovering more is likely.
- Genetic research can inform initiatives to support reading acquisition and intervention
strategies for reading disability: it holds promise for early identification of children at risk of reading difficulty.

Keywords: reading impairment, dyslexia, gene by environment interaction, polygenic prediction, behavior genetics
Successful acquisition of reading skills in childhood positions individuals toward positive life courses. Difficulties in reading present barriers to engagement with and achievement in educational settings, which can then negatively impact socio-economic status (SES) and health (Ross & Wu, 1995). Despite adequate reading instruction, around 8% of children show reading impairment (i.e., reading achievement well below the average for their ages, and in some definitions, discrepant from their IQ; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990), diagnosed as a specific learning disorder within the DSM-5 (American Psychiatric, 2013). Specific deficits in word reading accuracy, reading fluency, or reading comprehension are typical. Reading scores lying, for instance, 1.5 standard deviations below the population mean might indicate impairment (Snowling, 2013).

Major risk factors for reading disorder include home learning environment and family history, which includes genetic inheritance. The specific genes influencing reading ability are still largely unknown, but twin and family studies reveal genetic and environmental structures underpinning reading ability, how it relates to development, and how environmental factors modify it. This knowledge can inform public policy so that educational programs are optimally developed to make learning to read easier for those who experience difficulty.

**The Genetics of Reading Ability**

Systematic twin studies compare a trait’s resemblance in identical (monozygotic, MZ) twins (who are effectively genetically identical) and fraternal (dizygotic, DZ) twins (who share, on average, half the genes on which humans tend to vary). Because twins reared together share the same family and other experiences growing up the higher resemblance of MZ than DZ twins is due to their increased genetic similarity. Trait differences between MZ twin pairs are due to environmental influences that are not shared or are uniquely experienced. The relative contributions of genes, shared environment and unique environment to variation in a trait are estimated by statistically modelling these expectations.
in the data (Rijsdijk & Sham, 2002). Such studies show that reading ability’s familial aggregation (similarity within families, differences between families) is primarily due to genetic differences (Defries, Fulker, & Labuda, 1987; Wadsworth, Corley, Hewitt, Plomin, & DeFries, 2002). Twin studies estimate heritability, or the proportion of observable variance in a trait due to genes, of specific reading impairment at around 70% (Fisher & Francks, 2006; Gayan & Olson, 2001).

Similarly, heritability of reading ability more generally is high, with estimates from twin and adoption studies ranging 51-76% for diverse reading measures (Bates et al., 2007; Davis, Knopik, Olson, Wadsworth, & DeFries, 2001; Reynolds et al., 1996; Wadsworth et al., 2002). Associations between varying components of reading ability (e.g., phonological awareness/decoding, comprehension, rapid automatized naming) are substantively due to overlapping genetic variance (Byrne et al., 2013; Davis et al., 2001; Keenan, Betjemann, Wadsworth, DeFries, & Olson, 2006). Whereas a slightly higher incidence of reading difficulties is observed in boys than girls (e.g., 1.7:1 at ~8 years, 2.3:1 at ~10 years; Wheldall & Limbrick, 2010), the genetic source and proportions of genetic and environmental variance in reading difficulties do not appear to vary by gender, at any level of severity (Hawke, Wadsworth, Olson, & DeFries, 2007). Similarly, the genetic sources of variance between low and normal variation in word reading do not seem to differ (Harlaar, Spinath, Dale, & Plomin, 2005).

Investigating the relative degrees of genetic and environmental influences on reading development requires longitudinal studies. Multivariate modelling enables the similarity in the sources of genetic and environmental influences over time to be quantified by respective genetic and environmental correlations. In a cross-cultural twin study sampling two time points (preschool print awareness, phonological awareness and rapid naming, and grade 2 word reading, reading comprehension and spelling) genetic influences correlated, although
new genetic influences emerged at grade 2 (Byrne et al., 2009). Another study additionally observed continuity in shared environmental influences on early literacy experience and pre-literacy knowledge with reading and writing skills 1-3 years later (Oliver, Dale, & Plomin, 2005; Petrill et al., 2007), particularly for measures (e.g., phonological awareness, letter/word knowledge) that likely benefit most from instruction (Petrill et al., 2007). This pattern of both stable and new genetic influences and stable shared environmental influences was supported for oral reading fluency (Hart et al., 2013), reading comprehension (Soden et al., 2015), and reading achievement (Harlaar, Dale, & Plomin, 2007). Into adolescence, an adoption study (Wadsworth, Corley, Hewitt, & DeFries, 2001) of single-word reading in genetically unrelated and related siblings showed virtually identical genetic influences across ages 7, 12, and 16 years (genetic correlations ranged .98 to 1). Whereas new genetic variance did not emerge at later ages, new unique environmental variance did, which may partly reflect the self-selection of established readers into diverse environments (e.g., good readers might read more for pleasure which in turn influences their reading skill). Nevertheless, at early ages (and even prior to formal reading instruction) genetic influences are present that continue to impact reading ability during the primary school years (and probably beyond).

Genetically influenced variation in reading ability can be detected even before reading is acquired: Individual differences in reading ability (or rather processes enabling its acquisition) emerge early and persist. The youngest age twins have been sampled is two years (and again at 3 and 4), deriving a language-ability measure based on parent-reported vocabulary and syntactic knowledge (Harlaar, Hayiou-Thomas, Dale, & Plomin, 2008). It correlated with teacher ratings of student’s reading across ages 7, 9, and 10 years, and showed both genetic and shared environmental correlations over time. Other early predictors of later reading difficulties include brain event-related potentials (measured by surface electrodes) tied to auditory/speech perception in newborns (Leppänen et al., 2012); these
would be useful to investigate within a longitudinal genetic context. Early predictors that show the greatest genetic continuity with reading ability might aid identification of children with more severe difficulties. The challenge, though, is to discriminate between early predictors that reflect low general cognitive ability and those specific to reading ability because different remedial pathways for each might be beneficial.

**Genetic Influence on Reading and Other Abilities**

Reading ability is associated to varying degrees with other cognitive abilities (Evans, Floyd, McGrew, & Leforgee, 2002). Reading achievement was correlated with memory, processing speed, verbal and spatial abilities in twins (6-12 years) from the general population (Thompson, Detterman, & Plomin, 1991). Genetic correlations indicated that genetic influences on reading ability largely overlapped those on processing speed, verbal and spatial abilities, and to a lesser extent, memory ability. In this same sample, the genetic correlations between reading and math achievement, and between reading and language achievement, respectively indicated complete overlap of such influences.

With regard to specific learning disorders, there is high comorbidity of reading and arithmetic disorders, although some comorbidity will include general achievement deficits (Dirks, Spyer, van Lieshout, & de Sonneville, 2008). Specific reading disorder and language impairment have even higher comorbidity (McArthur, Hogben, Edwards, Heath, & Mengler, 2000). Twin studies of low ability show that much of this comorbidity is genetically influenced, with genetic correlations between reading disability and math deficits (Knopik, Alarcón, & DeFries, 1997; Light & DeFries, 1995), and between reading and language impairments (Bishop, 2001; Bishop, Adams, & Norbury, 2004; Hayiou-Thomas, Harlaar, Dale, & Plomin, 2010; Logan et al., 2011). These studies typically excluded cases with low IQ, so the genetic correlation was not a function of general cognitive ability. Where population-based samples have been sampled without exclusions for low IQ, reading,
language and math skills (and general cognitive ability in one study) show substantial overlapping genetic influences in childhood (Haworth et al., 2009; Plomin & Kovas, 2005).

Genetic correlations can arise because the same genes have direct effects on the associated traits, for example, via shared underlying brain or cognitive processes (e.g., working memory). However, they are also compatible with causal mechanisms from one trait to the other. For example, genes influence development of reading ability, and reading ability influences learning math. Future research has the formidable challenge of teasing apart these competing (although not mutually exclusive) explanations.

Nevertheless, genetic overlap between reading and other abilities (however it arises) has implications for public policy. Most importantly genes predisposing to poor reading ability will likely contribute to difficulties with other cognitive abilities (and vice-versa, but to a lesser extent because most learning in other areas depends on the ability to read). Thus, any child identified at-risk for a specific learning disorder (especially if family history/genetic risk is present) would benefit not only from remediation for that learning disorder but for other specific learning disorders. Holistic teaching approaches have shown efficacy; for instance, training in reading comprehension should be prioritized within the mathematics teaching curriculum (Carter & Dean, 2006). Given the evidence of common genetic influences on all cognitive abilities across their distributions, arguably this combined teaching approach would be useful in both remedial and normal education.

Literacy and numeracy skills have shown higher heritability than general cognitive ability at ages 7 and 9, but not at 12 years in a British study (Kovas et al., 2013). General ability does not depend on any specific curriculum, whereas literacy and numeracy require specific instruction. Perhaps formal education systems (at least in the UK) provide standardized literacy and numeracy instruction that reduces environmental variation contributing to these traits, leaving higher variance proportions attributable to genetic
influence. With regard to academic skills, then, increased genetic variance in a trait might be evidence for more equitable socio-educational processes. One caveat, of course, is that the samples in which heritability is measured must be representative of the entire population.

Children from low SES are typically under-represented in twin and family studies, so results may simply indicate little variation in the quality of literacy programs in schools attended by middle-class children. In a US twin study that included families living close to or below the poverty line, 60% of the variance in general cognitive ability not shared with SES at age 7 was due to the shared environment in the most impoverished families but only 15% in the high SES group (Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). This effect has not been observed in samples from Western Europe and Australia, where social policies support fairly even access to high-quality education and medical care (Tucker-Drob & Bates, 2015). Factors, such as SES, that might moderate the genetic influence on reading ability are thus important to consider.

**Factors that Moderate the Genetic Influence on Reading Ability**

Moderation of the heritability in reading ability by SES has not been studied. However, parental years of education (highly correlated with SES) moderated genetic influences on reading disability in twins aged 8 to 20 years (Friend, DeFries, & Olson, 2008). In US twin pairs, where one member was classified as having reading difficulties, scores on a composite measure of word recognition, spelling, and reading comprehension showed greater proportional genetic influence at higher levels of parental education. The interpretation was that children with higher levels of environmental support (i.e., those with better educated parents) are enabled to reach their genetic potential. Or that environmental disadvantage suppresses children’s ability to reach their genetic potential. It becomes critical, then, to identify ways to remediate the blocks to learning created by poverty. Parental educational expectations and specific parenting behaviours (e.g., storybook reading, increased richness of
linguistic exchanges) support emergent literacy skills (Davis-Kean, 2005; Sénéchal & LeFevre, 2002). Also, parents with higher education are more likely to have access to higher quality schools and can provide enhanced support and more time (e.g., in homework; Guryan, Hurst, & Kearney, 2008; Patall, Cooper, & Robinson, 2008) and resources (e.g., books, e-readers) to facilitate their children’s achievement.

Of course, parental education can in part be affected by parents’ own genetic risk for reading difficulties, which can then lead to gene-environment correlation in the children (i.e., parents who are poorer readers transmit their poor-reading-risk genes and provide less supportive reading environments for their children). To aid remediation, children with reading difficulties whose parents are less educated (predominantly those in low SES school-catchment areas) should receive additional formal educational support to compensate for potential lack of parental support. Shared environmental influences on reading achievement in children with less-educated parents may be more prominent than in those with more-educated parents (Friend et al., 2008). If so, uniformly high-quality literacy instruction in both poorer and wealthier districts may help to suppress the low-SES disadvantage.

The moderating influence of child’s IQ may be similar to that of parental education on heritability of reading disability (Wadsworth, Olson, Pennington, & DeFries, 2000). Genetic influences on reading disability were much smaller at lower than at higher IQ levels. Given the significant positive correlations between child’s IQ and parental education, number of books, and parental book reading to children in this study, this could also arise because parents transmit both genetic and environmental influences on cognitive skills, such as reading to their children. Children with higher IQs are more likely to have home and school environments that support achievement (Sylva, 1994; Tong, Baghurst, Vimpani, & McMichael, 2007). Thus, their reading difficulties appear to be primarily genetically influenced.
Evidence of teacher (or classroom environment) quality moderating the heritability of reading achievement again highlights the importance of environmental influences on individual differences in reading ability. In a sample of 1st- and 2nd-grade twins representative of the ethnic and SES distributions in Florida (52% qualified for the US free/reduced-price lunch program), teacher quality—indexed by oral reading fluency test gains of the twins’ classmates—moderated the genetic variance in the twins’ end-of-year oral reading fluency test scores (Taylor, Roehrig, Hensler, Connor, & Schatschneider, 2010). Higher teacher quality was associated with greater genetic variance in reading performance, although amount of shared and unique environmental variance did not vary at different levels of teacher quality. At the poorest level of teacher quality, shared environmental influences were relatively more important than genetic influences, with the reverse being true at highest teacher quality level. Again, these results underscore the need for a standardized quality of teaching (or related to the classroom: classroom resources, blending of student populations, etc.) to enable children to reach their genetic potential. That shared environmental influences were present across all levels of teacher quality also reinforces the roles of factors beyond direct formal literacy instruction, which could be linked to wider school practices or family circumstances. In a utopian world where all children have equal environmental support to maximize reading achievement, the variation between children’s performance will be determined by their genetic differences. Children with biological disadvantage would thus need greater support to ensure that they reach levels of reading ability that will spur growth in other areas of their lives. What then do we know about the specific genes influencing reading ability?

**Specific Genes Influencing Reading**

Molecular genetic studies have identified no genes that have large associations with reading ability/difficulties. These studies have taken two basic forms: whole-genome linkage
and whole-genome association. *Genetic linkage studies* use samples of family groups to map regions of the genome containing genetic variants that co-segregate (co-occur) with a trait (e.g., reading disability). Such studies are generally well-powered to detect chromosomal regions where single genetic loci or clusters of genetic loci explain at least 10% of variance. Of eight independent genome-wide linkage studies of developmental dyslexia, none found evidence of gene effects this large (for a review of these findings see Scerri & Schulte-Körne, 2010).

*Genome-wide association studies* test for linear association between thousands of individual genetic variants and quantitative reading scores (or compare variant frequencies between reading-impaired cases and controls) in unrelated individuals. They have also failed to find any large associations (Carrion-Castillo et al., 2016; Luciano et al., 2013). Whereas strong associations might be important in rare cases of multigenerational families where reading disabilities are widespread (e.g., a linked region on the X chromosome appears to be specific to a Dutch family (de Kovel et al., 2004)), they are unlikely to be involved in reading difficulties more generally. Novel genetic variants uncovered by genome-wide association studies of reading ability/disability have not accounted for more than 0.4% of variance (Gialluisi et al., 2014; Luciano et al., 2013; Meaburn, Harlaar, Craig, Schalkwyk, & Plomin, 2008), so it appears that, like other complex traits, 1000s of genes can be associated to very small degrees with reading ability.

Candidate genetic variants associated with reading ability/dyslexia have mostly been identified by mapping genetic variants within broader chromosomal regions linked to reading disability and testing their associations with reading scores/disability. The few prospects show moderately robust replication and identification of relevant function. Evidence of their involvement has come from studies of both reading disability and full-range variation in reading ability where IQ is controlled (e.g., Bates et al., 2010; Luciano et al., 2007; Newbury
et al., 2011; Paracchini et al., 2008). None of the variants within these genes explains more
than 1% of variance in reading scores, and in some studies they have not been significant
(e.g., Brkanac et al., 2007; Harold et al., 2006). A bioinformatics investigation of the
purported/known function of 14 candidate genes for reading disability showed that 10 of
these had roles in a molecular network concerning neuronal migration and neurite outgrowth
(Poelmans, Buitelaar, Pauls, & Franke, 2011). Neuronal migration is a developmental process
occurring primarily during gestation whereby neurons migrate from their origin in the brain’s
ventricular zone to form layers of the cerebral cortex (Stiles & Jernigan, 2010). Each neuron
also undergoes neurite outgrowth, or emergence of neural axons and dendrites in response to
guidance signals and environmental input (Kiryushko, Berezin, & Bock, 2004). As new genes
associated with reading ability are discovered, we can gain more insight into their underlying
biological pathways. There is no understanding yet of whether prenatal or early postnatal
exposures might alter expression of the implicated genes, but it will be vital to investigate this
because therapeutic potential could be enormous.

Advances in understanding of genetic involvement in reading ability have been much
slower than for other complex traits because the samples typically focus on
children/adolescents with reading disability, thus are limited in size and population
representativeness. However, an international consortium on the genetics of language is
currently pooling genome-wide association results to overcome current power limitations.
Presently, variants in dyslexia candidate genes cannot be used for prediction purposes, but the
results of this collaboration may enable the development of polygenic scores to improve
prediction of reading difficulties. Polygenic scores are composites based on the effect sizes
(usually from a well-powered genome-wide association study) of individual variants
throughout the genome (Wray et al., 2014). Polygenic scores for educational attainment (a
correlate of reading ability) have explained 4% of observed variance in adult educational
attainment (Okbay et al., 2016). This same score explained up to 5% of observed variance in reading achievement at age 14, representing around 20% of the genetic variance in reading achievement due to common genetic variants (Selzam et al., 2017). In this study, the mean difference in reading achievement between children in the lowest 12.5% and highest 12.5% of the education polygenic score distribution equated to around 2 years of primary schooling. A reading-specific polygenic score will more precisely apply to reading achievement and could explain upwards of 20% of the genetic variance in reading. Rare genetic variants and structural (chromosomal) variants, for which reading-ability studies are lacking, could explain additional variance and improve prediction further.

**Policy Implications**

The future of genetic association studies in reading ability holds great promise for public policy. Polygenic scores could one day be used to predict which children are at greater risk of reading difficulties before they even acquire language. This method would be still more accurate when coupled with early pre-school correlates of reading-disability risk and with reading-skill indices following entry to school. Children with poor reading-related performance in the early years who also have high polygenic risk of reading difficulties might be especially targeted for intervention, such as more intense phonics instruction. The key to successful remediation is early intervention. Specific risk variants might respond best to specific interventions. In the same vein, children with high polygenic risk of reading disability, but lower polygenic risk for low general cognitive ability, might respond differently to remediation efforts that work best for children with high polygenic risk of both. Future research is sure to address this.

In the interim, educators need to be aware of the progress being made in the genetics of complex traits and what this means for students’ achievement. As we strive for equitable educational systems that provide high-quality instruction for all, we must recognize that such
systems will expose the genetic variance among people. But genetic influences on complex traits by no means mean determination. Height, for instance, is a highly heritable trait (~80%), but secular increases in height over the last century have been large, possibly due to better health care and nutrition (Stulp & Barrett, 2016). In the same way, improvements to the environments that foster reading acquisition and skill can result in similar population gains, and large strides in effective reading remediation have been made in recent years, based on observable patterns of difficulty.

At an individual level, a child who carries a larger burden of genetic variants that are associated with poorer reading (or rather with the cognitive processes on which learning to read depend) could be remediated to achieve normal/above normal reading levels. The exception would be in rare cases of pathology where language is severely affected. Where less egalitarian educational practices exist, environmental factors—such as pre-reading experience and instruction—explain substantial variability in reading achievement, and these are important indicators to identify and act on. Granted, it is difficult to control children’s pre-reading experiences (although public health literacy promotion—e.g., Reach Out and Read—that emphasises positive experiences of reading to children should continue to target low-income communities). Still, attaining uniformly high-quality systematic institutional phonics reading instruction is, in principle, achievable. Negative influences on reading ability related to SES need to be compensated by more intensive reading instruction in schools that service large proportions of low-income families. A national review of dyslexia and literacy difficulties in the UK recommended a model of intensive catch-up programs for children at risk of reading impairment, followed by individualized intervention for those most at risk (Rose, 2006). Such a model could be optimized in the future by use of genetic information; granted that ethical matters and cost-effectiveness of genotyping is evaluated. Those with higher genetic risk would likely benefit from individualized remediation and could be
allocated to suitable interventions at the outset. Additionally, the progress of such children in acquiring foundational reading skills could be monitored more closely within the formal education system. Both individuals and communities can benefit from targeted interventions informed by genetic analysis.
References


Heritable from Grades 1 to 6. *PLOS ONE*, 10(1), e0113807. doi: 10.1371/journal.pone.0113807


