Published in final edited form as:

Dev Psychopathol. 2012 November ; 24(4): 1345–1360. doi:10.1017/S0954579412000752.

Lexical decision as an endophenotype for reading comprehension: An exploration of an association

ADAM NAPLES^a, LEN KATZ^b, and ELENA L. GRIGORENKO^{a,c,d}

^aYale University

^bUniversity of Connecticut

^cMoscow State University

^dColumbia University

Abstract

Based on numerous suggestions in the literature, we evaluated lexical decision (LD) as a putative endophenotype for reading comprehension by investigating heritability estimates and segregation analyses parameter estimates for both of these phenotypes. Specifically, in a segregation analysis of a large sample of families, we established that there is little to no overlap between genes contributing to LD and reading comprehension and that the genetic mechanism behind LD derived from this analysis appears to be more complex than that for reading comprehension. We conclude that in our sample, LD is not a good candidate as an endophenotype for reading comprehension, despite previous suggestions from the literature. Based on this conclusion, we discuss the role and benefit of the endophenotype approach in studies of complex human cognitive functions.

“Everybody seems to agree that we should be studying comprehension, but how?” (Shankweiler, 1999, p. 113). This quote captures a sentiment shared by many, although the last decade in the psychology of reading, based on the number of publications, could truly be called the decade of the psychology of reading comprehension. Part of the problem invariably stems from difficulties in characterizing reading comprehension. Standardized measures of word decoding exhibit strong and robust intercorrelations, whereas standardized measures of reading comprehension demonstrate low and inconsistent associations (Cutting & Scarborough, 2006). This heterogeneity in measurement highlights the complexity of reading comprehension, a complexity that makes quantification at the cognitive level difficult and is thought to introduce heterogeneity at the genetic level of analysis.

Variability in word decoding is substantially heritable, ranging from 0.65 (Harlaar, Spinath, Dale, & Plomin, 2005) to 0.85 (Gayán & Olson, 2003), a widely replicated finding (for a review, see Grigorenko, 2004). Individual differences in reading comprehension also appear to be heritable, but there are far fewer genetically informed studies of comprehension, and the obtained heritability estimates tend to be lower than those of word decoding (e.g., Betjemann et al., 2008; Keenan, Betjemann, Wadsworth, DeFries, & Olson, 2006). Notwithstanding the inconsistency in methods and assessments mentioned above (see also Cain & Oakhill, 2006b; Keenan, Betjemann, & Olson, 2008), heritability for reading comprehension has been estimated at 0.50–0.76, with 27%–94% of this genetic variance shared with measures of word decoding (Betjemann et al., 2008; Byrne et al., 2007, 2009;

Keenan et al., 2006). A recent meta-analysis of behavior–genetic studies of reading comprehension has resulted in the estimate of narrow-sense heritability at 51%, with shared and nonshared environment accounting for ~24% of the variance in comprehension of each. It was stressed, however, that these estimates vary dramatically across countries, studies, and assessments (Pasisz & Schatschneider, 2010).

Despite the tight connection between words and text, decoding words is not equivalent to appreciating Goethe. It is recognized that reading comprehension (and the quality of lexical processing) is influenced by a variety of cognitive processes ranging from relatively low-level processes (i.e., working memory; Engle, Cantor, & Carullo, 1992) to relatively high-level processes (i.e., inferencing; Gernsbacher & Faust, 1991). Deficient comprehension might result from one of many weaknesses operating in isolation or in combination (Cain & Oakhill, 2006a). It is typical, however, that the fluency and accuracy of single-word lexical processing is a prerequisite of comprehension; success with text begins with the single word (Perfetti & Hogaboam, 1975). There is recognition, however, of the differentiation of typical and atypical developmental pathways. Whereas in typically developing individuals, indicators of word decoding generally highly correlate with indicators of reading comprehension, in atypically developing individuals, these indicators might demonstrate a substantial level of discrepancy, for example, individuals with “hyperlexia” are characterized by the presence of a pronounced gap between a higher level of word decoding and a lower level of reading comprehension (Grigorenko et al., 2002; Grigorenko, Klin, & Volkmar, 2003).

Yet, single-word decoding, although a prerequisite for comprehension, is itself a complex process emerging in the foundation of many other psychological processes (for a brief list, see Grigorenko et al., 2007). These other processes, typically described as sublexical, underlie the emergence of reading, which is a lexically based skill. This transition from sublexical to lexical processing appears to be faulty in disorders of reading acquisition (Zoccolotti et al., 2005). Modern models of reading acquisition (e.g., Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994) assume there are two major types of such sublexical processing, phonological and orthographic. At the level of the single word, phonological processes are typically captured by the task of word naming and orthographic processes, that is, lexical decision (LD). It has been observed that a LD heavily engages lexical knowledge, whereas a naming task does not (Martens & de Jong, 2006). Given that lexical knowledge is central to comprehension (Perfetti, 2007), it is reasonable to assume that performance of an LD task might be equally or more predictive of performance of a reading comprehension task than performance of a word naming task (Holmes, 2009; Michael, 2009). In summary, reading comprehension is a complex trait that assumes that multiple lower-level processes have been developed to support the emergence of reading comprehension. For typical developmental trajectories, indicators of all these processes tend to correlate with indicators of reading comprehension. Atypical developmental trajectories, however, could be characterized by peaks, valleys, and abysses among indicators of various comprehension-contributing processes, such that a lower level of comprehension may result from various patterns of deficiencies in processes, as per the principle of equifinality.

In response to the difficulty of the genetic analysis of complex traits, the endophenotype approach was developed and has gained considerable traction in the past few years (Gottesman & Gould, 2003). Endophenotypes are traits that are thought to be intermediate to complex neuropsychiatric disorders, simpler, and “closer” to the underlying biological deficits, thus making them more amenable to identification and localization in genetic studies (Almasy & Blangero, 2001; Iacono & Malone, 2011; Kendler & Neale, 2010; Peterson & Weissman, 2011). Studies of a number of neuropsychiatric disorders have utilized the endophenotype approach (e.g., on P50 inhibition in schizophrenia, see

Aulchenko & Struchalin, 2009; Freedman et al., 1997; on inhibitory control in ADHD, see Goos, Crosbie, Payne, & Schachar, 2009; on duration discrimination in ADHD, see Himpel et al., 2009; on motor inhibition in obsessive-compulsive disorder, see Morein-Zamir, Fineberg, Robbins, & Sahakian, 2010; on N400 in alcoholism, see Roopesh et al., 2010). The appeal of endophenotypes stems from the apparent simplicity of the concept. Genes for complex traits are difficult to localize because many of them may be contributing differentially across individuals; an endophenotype is a simpler trait, a “building block” of the deficit, presumably influenced by fewer genes that are easier to identify. Endophenotypes should not only be influenced by fewer genes than complex traits, but it is critical that these genes must also influence the complex trait under investigation, although other genes might also contribute to this trait (Kendler & Neale, 2010). Otherwise, an endophenotype that does not share genes with the complex trait, although it might be related to it, is simply a specific case of the phenomenon that virtually all traits are heritable (Johnson, Turkheimer, Gottesman, & Bouchard, 2009); thus, such a trait cannot be considered an endophenotype.

The problem then appears to be identifying endophenotypes. The field has reached something of a consensus in assigning general requirements for putative endophenotypes (Crosbie, Perusse, Barr, & Schachar, 2008; de Geus, Wright, Martin, & Boomsma, 2001; Glahn et al., 2012; Gottesman & Gould, 2003; Smit et al., 2010). Namely, endophenotypes (a) are lower-level processes (i.e., as basic a characteristic of a particular process as possible), (b) demonstrate reliable within-age individual differences, (c) show an empirical association with performance on the targeted task (e.g., reading comprehension) that is also theoretically meaningful, (d) are heritable, and (e) have a strong genetic correlation with the targeted task. However, despite the widespread adoption of the endophenotype approach at the conceptual level (Ritsner, 2009), the details of its implementation and specific recipes on how to pick an endophenotype among many phenotypes are still being discussed (Flint & Munafo, 2007; Glahn et al., 2012; Kendler & Neale, 2010; Meyer-Lindenberg, 2010).

Reading comprehension is a complex, heritable trait that exhibits significant variability and relies on multiple cognitive processes. Deficits in reading comprehension are present in many developmental disorders (Ricketts, 2011; van den Bos, Nakken, Nicolay, & van Houten, 2007; Verhoeven & Vermeer, 2006), although their etiology might be variable. In contrast, LD is a prototypically simple (or at least simpler, compared to comprehension) trait requiring only lexical access. Yet, LD has also been investigated as a source of group differences between individuals with and without developmental disorders (de Jong et al., 2009; Grossman, Bemis, Skwerer, & Tager-Flusberg, 2010; Rutherford, 2006; Stewart & Ota, 2008). In this article, we investigate whether LD is a suitable endophenotype for reading comprehension and test this hypothesis by determining to what extent LD and reading comprehension appear to have common genetic etiology. Although this investigation is carried out using an unselected sample of children and their families, the implications of this work are important for considering LD an endophenotype for comprehension difficulties in a variety of developmental disorders.

Study Objectives

Capitalizing on the literature linking reading comprehension and LD, we used a genetically informative design to evaluate LD as an endophenotype for comprehension. First, phenotypic correlations between comprehension and LD were obtained, with and without age and gender as covariates. Second, heritability coefficients were estimated for both comprehension and LD. Third, an investigation of the probable patterns of familial transmission of comprehension and LD indicators was carried out; that is, we estimated the

number of major genes contributing to reading comprehension and LD and the effect sizes of these genes.

As indicated above, the utilization of the concept of endophenotypes sets up specific requirements. Thus, we investigated whether LD met the requirements to serve as an endophenotype for comprehension. Specifically, we hypothesized that indicators of LD demonstrate a wide range of individual differences, which, in turn, predict individual differences in reading comprehension. Moreover, we predicted that both LD and reading comprehension are heritable and that the heritability of LD is greater than that of reading comprehension. Finally, we hypothesized that LD and reading comprehension “share” genes, and that LD is influenced by fewer genes overall than reading comprehension.

Methods

Participants

The sample consisted of children attending Grades 1–7 in public schools in a middle-sized city in Russia. The sample constituted a normative, community-based sample of reading skills in the general population of Russian speakers. It was ascertained through children whose families were invited to participate, and consenting families were assessed. Altogether, the sample included 502 children aged 7–13 years (mean = 9.6, $SD = 1.2$), consisting of 243 (48%) boys and 259 (52%) girls, clustered in 488 families (14 families had 2 children eligible to participate in the study, the remaining families had only 1 eligible child). In these families, at least 1 parent or adult care-giver was assessed. Altogether, we worked with 483 mothers (aged 24–52, mean = 33.9, $SD = 5.2$) and 345 fathers (aged 22–61, mean = 35.7, $SD = 5.6$). We also assessed 3 grandmothers (aged 53–65, mean = 62.2, $SD = 7.7$) and 6 grandfathers (aged 53–58, mean = 58.7, $SD = 7.7$). Overall, there were 350 complete families (with both parents and at least 1 child assessed) and 138 incomplete families (with at least 1 adult relative and 1 child assessed). Of the incomplete families, 5 families had fathers but not mothers, 132 families had mothers but not fathers, and 1 family had grandparents but not parents. Given that the statistical software utilized here are suitable for all kinds of family constellations and pedigree structures and can take into account all types of relatives, all 488 families were included in the analyses. Thus, the total number of consenting participants in this sample was 1,339.

Materials

Comprehension—Comprehension was assessed with overlapping sets of 11 connected text paragraphs with 43 comprehension questions (multiple-choice and open-ended, factual and inferential), varying in difficulty for children, adolescents, and adults (Grigorenko et al., 2007). A selected subset of paragraphs overlapped in adjacent age-appropriate sets, so that all responses could be scaled using the same ability scale (Dorans, Pommerich, & Holland, 2007; Loyd & Hoover, 1980). Participants read a series of age-appropriate paragraphs and were asked questions probing content. We counted the number of correct responses and then obtained ability scores using the item response theory methodology (Bond & Fox, 2001; Rasch, 1960; Smith & Smith, 2005). All materials were presented in Russian. The descriptive statistics of these scores are presented in Table 1. To control for possible age and gender effects, we used these two indicators as covariates. Table 2 presents correlations for all variables, both unadjusted and adjusted for age and gender.

LD—Visual LD is the dominant paradigm used for studying visual word recognition because it is both simple and flexible. A participant is presented with a string of letters and the participant must decide whether or not the string is a real word as quickly as possible. Response times in LD tasks reflect characteristics of both the stimuli and the test taker.

Characteristics of the stimuli, notably frequency or the wordlikeness of nonwords influence response times, as do the vocabulary and reading skill of the test taker.

In this study, the stimuli were single letter strings presented on a computer screen using E-Prime experiment presentation software from Psychology Software Tools. All stimuli were presented in Russian. Participants were seated facing a computer screen with a computer keyboard in front of them. Each participant completed 90 trials. Each session contained 30 nonwords and 60 words. The words varied in length (two and three syllables); there were 15 nonwords and 30 words of each length. The nonwords were generated so that all letter/sound combinations were permissible in Russian. The words were further divided by frequency (high, medium, and low) and regularity (regular and irregular), using conventional Russian frequency dictionaries. Thus, there were 10 words for each of the six frequency-by-regularity combinations. Each participant saw each word only once. The order of the stimuli was random for each participant.

Participants were asked to press the “z” key for a word response and the “/” key for a nonword response. The task ended after the participant reached the end of the word list. Response times were excluded from the experiment if they exceeded 4 s or were shorter than 400 ms (Ratcliff, 1993). The relatively long cutoff was chosen for two reasons: (a) because this was a normative sample, and our goal was to understand the etiology of LD response time, we only excluded the longest 5% of response times across all participants; and (b) there is some evidence that the long tail of response times may be psychologically informative (Schmiedek, Oberauer, Wilhelm, Suss, & Wittmann, 2007), particularly regarding individual differences (Gilden & Hancock, 2007; Gilden, Thornton, & Mallon, 1995).

Analyses of the data revealed no effect of regularity of the spelling to sound mapping on response time, which was not unexpected, because regularity effects have been shown to be transient and dependent on experimental manipulations (Gernsbacher, 1984). Note that, although errors were recorded, given the substantial range of ages in our sample (7–65) and corresponding age biases in the distribution of accuracy (following recommendations in the literature; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), accuracy indicators were not utilized in these analyses. As a result, only five indicators were generated for the LD task: average response time for words of different frequencies (high, medium, and low), average response time for nonwords, and an average response time across all stimuli. Similar to the comprehension indicator, Table 1 presents descriptive statistics for unadjusted variables and Table 2 contains correlations for both unadjusted (below the diagonal) and age- and gender-adjusted (above the diagonal) variables.

Resulting phenotypes—Our preliminary analyses of the data resulted in six indicators: comprehension and response times for LD for high (LDH), medium (LDM), and low-frequency words (LDL); nonwords (LDN); and all stimuli (LDT). These indicators were used in the analyses below.

Statistical analyses—The goal of this study was to evaluate the suitability of LD indicators as endophenotypes for reading comprehension. First, phenotypic correlations among the indicators of comprehension and LD were obtained, with and without covarying age but without gender. Second, heritability coefficients were estimated for indicators of comprehension and the LD task; genetic correlations among these variables were also estimated. Third, an investigation of the probable patterns of familial transmission of comprehension and LD indicators was conducted. These segregation analyses allowed us to quantify the number of genes involved in the transmission of the trait of interest across generations, and the magnitudes of the effects of these genes.

Heritability estimates—Heritability estimates were obtained for each phenotype using two software packages: SOLAR (Almasy & Blangero, 1998) and LOKI (Heath, 1997; Heath, Snow, Thompson, Tseng, & Wijsman, 1997).¹ SOLAR reports what is called “narrow-sense” heritability, which is the proportion of phenotypic (i.e., trait) variance that is attributable to additive genetic variance alone. LOKI reports what is called “broad-sense” heritability, which is the proportion of phenotypic variance that is attributable to both additive and dominant genetic variance.² Both LOKI and SOLAR allow the estimation of heritability based on family data of the kind we collected in this study (i.e., child–parent and other relative data from nuclear and small extended families). It is thought that these estimates represent the upper bounds for heritability estimates because it is possible that they are inflated by the contributions of family shared environment. Nevertheless, both pieces of software are widely used in genetic studies of complex disorders with family data, and the corresponding estimates of heritability are considered informative.

Segregation analysis—In addition to estimating the overall magnitude of genetic influences using heritability estimates, we were also interested in investigating modes of genetic transmission that might explain the familial aggregation of comprehension and LD indicators. The methodology for these analyses was adopted from published work (Hsu, Wijsman, Berninger, & Thomson, 2002; Wijsman et al., 2000). Specifically, we used the oligogenic segregation analysis Bayesian approach through the Markov chain Monte Carlo (MCMC) algorithm implemented in the LOKI software (Daw, Heath, & Wijsman, 1999).

In general, segregation analyses permit an investigation of the vertical familial transmission of the phenotype of interest; oligogenic segregation analyses allow for the modeling of such patterns of familial transmission, including both genetic and environmental factors. In this approach, it is assumed that the phenotype is related to an additive³ combination of the effects of the environmental covariates and genetic influences modeled through quantitative trait loci (QTL). The phenotypes and covariates are measured, whereas the parameters, including covariate effects, the number of QTL, and their specific effects, are estimated. Specifically, LOKI uses reversible jump MCMC techniques to incorporate moves that change the numbers of QTL and the parameters in the model. Each iteration contributes to the estimates of posterior probabilities. Thus, the estimation involves a stochastic element requiring many iterations, so that the probabilities of the competing models generated during the MCMC iterations can be estimated and compared. For our purposes, we were most interested in the estimates of how many QTLs contributed to each trait, and how much genetic variance each QTL contributed.

Despite the complexity of how LOKI goes about it, what LOKI does is rather simple. Given a data set of families and phenotypic data, LOKI estimates the number of genes and the effect sizes of these genes. The program starts out by proposing a putative QTL of a given effect size, estimates the probability of the data given the current QTL parameters and proposed model, and uses this as part of the decision as to whether or not to accept the proposed model parameters. The first proposal is usually not very good, so the next proposal might add another QTL, change the effect of the putative QTL, or both. This process of proposal, probability calculation, and revised proposal constitutes an iteration, and LOKI runs tens of thousands of iterations to estimate the posterior density of the underlying genetic model.

¹Prior to the analyses, we evaluated the data distribution to ensure that its characteristics match those required by the software (Blangero, 1995; Igo, Chapman, & Wijsman, 2006).

²These estimates are generated with default options and are well established; some technical issues exist for estimating narrow sense heritability in LOKI and broad-sense heritability in SOLAR.

³“Additive” in this context refers to the effects of multiple QTLs; in other words, the effects of multiple genes are summed to calculate the quantitative trait and are not “additive” in the sense of individual allele effects for each gene.

In these analyses, we used 25,000 iterations; each iteration contributed to the estimates of posterior probabilities. Test runs between 1,000 and 2 million iterations were investigated prior to these analyses and 25,000 iterations were found to be more than sufficient for a successful mixture and “burn-in” of a Markov Chain (Wijsman & Yu, 2004). Bayesian statistical approaches require the specification of a prior distribution for each of the estimated parameters. Although there is an option in which one specifies prior distributions for virtually all of the model parameters in LOKI, with a sufficiently informative data set these priors have little impact in practice on the posterior distribution. Specification of parameters in the prior distributions required for Bayesian analyses was done following previously developed procedures (Wijsman et al., 2000). All of the analyses presented here used the default prior specifications for number of genes, trait mean, and variance as specified in the LOKI documentation.

As Wijsman and Yu (2004) previously demonstrated, the results are insensitive to the choices of these prior parameters. However, there is one parameter that can have a considerable effect. This parameter, tau beta, determines the effect size of the genes influencing the trait of interest, or the amount of genetic variance that the sampler will posit for putative genes. In other words, when LOKI is proposing to add a new gene to the model (whether it is the first gene or any new additional gene), the effect size of the proposed gene is set by the tau beta parameter. Misspecification of tau beta may result in misspecification of the model. Thus, finding a range of tau beta(s) in which the results are robust and consistent is important. The preferred approach (Wijsman & Yu, 2004) is to run multiple tau beta(s) and compare the runs while gradually increasing the prior distribution of tau beta until the number of genes and estimated heritability plateau and begin to drop off. In addition, Wijsman and Yu advocate a cutoff for genes with an effect size less than 5%–10% of the total genetic variance. Their justification is that for genes with very small effect sizes (i.e., 1% of the total genetic variance), the related estimates are unlikely to be reliable. We used, therefore, an effect size cutoff of 5%.⁴ Thus, by running multiple tau beta(s), the robustness and stability of the obtained solution/model of the data can be evaluated. A set of posterior parameters that remains fairly stable across multiple iterations and multiple tau beta(s) provides further support for the stability of the solution in the face of multiple priors.

Power analysis—To investigate the power of other samples to identify and distinguish the presence of shared and nonshared genetic contributions to complex traits, a simulation study was conducted. As per the results, two traits of moderate broad-sense heritability (~50%), which is a common estimate in the field (Grigorenko, 2004), were simulated using the structure of the pedigrees from this sample. Two scenarios were investigated. In the first scenario, two traits were controlled, in the presence of a polygenic component, by four genes each (i.e., each gene contributed no more than 12.5% of the genetic variance, linearly and interactively), two of which overlapped and two of which did not. In the second scenario, there are no overlaps between the four genes controlling each of the two traits. We simulated six data sets ($n = 31, 162, 330, 497, 660$, and 995 families) across a range of possible contributing genes (zero to four) and a range of tau beta parameters. The analyses delineated above (i.e., the estimation of the number of genes involved in the trait) were completed 25,000 times for each combination of parameters and for each model. We established that for the univariate model, the smallest sample size ($n = 31$) underestimates the number of genes involved in the trait, but the modest sample size of 162 families was sufficient and as effective as the sample size of 1,000 families. For the covariate models, a range of

⁴Although genes with effect sizes of <5% can be included in such modeling exercises, such inclusions will call for, perhaps, greater power than what was available in this sample and require additional analyses across a range of tau beta(s). Thus, for the purposes of these analyses, we followed Wijsman and Yu's (2004) recommendations, leaving the possible explorations for inclusion of genes with effects <5% for the future.

observations were made. The most informative of these in the context of our work were the following: (a) with the number of families greater than 30 (and fewer than 162), the models with zero and three overlapping genes are easily distinguishable; (b) to distinguish the “intermediate” models (i.e., one and two overlapping models) from each other or from the no-overlap (zero genes in common) and substantially overlapping (three genes in common) models the sample size of 330+ families is adequate (the stability across a range of tau betas improves with greater sample size).

Results

The data presented in Table 1 adhere to well-established patterns in the literature (Cirrin, 1984; Morrison & Ellis, 1995). Specifically, children’s response times are slower than those of adults. However, the age effect is not linear. The correlations with age and the selected phenotypes for the entire sample are as follows: comprehension, $-.031$ ($p = .262$); LDH, $-.568$; LDM, $-.592$; LDL, $-.619$; LDN, $-.507$; and LDT, $-.586$ (all $ps = .000$). However, for children it is comprehension, $-.204$ ($p = .000$); LDH, $-.156$ ($p = .001$); LDM, $-.135$ ($p = .003$); LDL, $-.109$ ($p = .015$); LDN, $-.143$ ($p = .002$); and LDT, $-.134$ ($p = .003$); for adults it is comprehension, $-.086$ ($p = .066$); LDH, $-.026$ ($p = .479$); LDM, $-.034$ ($p = .364$); LDL, $-.055$ ($p = .135$); LDN, $-.013$ ($p = .733$); and LDT, $-.039$ ($p = .298$). For the LD task, adult females were quicker than males (paired t test $ps = .001, .000, .005, .117$, and $.003$, for LDH, LDM, LDL, LDN, and LDT, respectively), but the boys did as well as the girls (no $ps < .05$). The accuracy rates were approximately the same across both genders. Comprehension was also stronger in females ($ps = .000$ and $.021$ for females and girls, respectively). The table also shows pronounced word frequency effects. Specifically, it takes longer to respond to low frequency words; however, responses to nonwords were comparable to responses to lower-frequency words. Adult responses also increased, with the nonwords taking the longest time to process. It is also worth noting the pattern of accuracy indicators for the LD task. There were virtually no errors in decision accuracy for high-frequency words for both children and adults. Children’s accuracy correlated with word frequency, with the lowest accuracy being observed for low-frequency words (61%). The adult accuracy levels, however, were high and demonstrated little variability, with the lowest level at 89% for low-frequency words. Given such limited variability in adult accuracy on the LD measures, and in accordance with the literature (Coltheart et al., 2001; Seidenberg, Peersen, MacDonald, & Plaut, 1996), the subsequent analyses were carried out only on time, not accuracy, measures.

Correlational analyses

Table 2 presents correlations among the variables in this study, both unadjusted and adjusted for age and gender. A number of observations are of interest here. First, the pattern of correlations among the unadjusted and adjusted variables is very similar, indicating the stability of the relationships among variables. Second, for all three groups of participants (children, fathers, and mothers), there are robust significant negative correlations between indicators of LD reaction time and comprehension scores. LD latencies become shorter as comprehension increases, and the results are similar to those presented in the literature on LD. Third, familial correlations are of interest. For comprehension, the correlation between the performance of fathers and their children is not significant, but it is significant for mother–child pairs. None of the LD indicators correlate significantly between fathers and children, but there is a positive, although not substantial, pattern of correlations between LD indicators of mothers and children. Note, however, that indicators of both comprehension and LD correlated consistently and substantially for pairs of mothers and fathers. Using SOLAR, genetic correlations between the comprehension and LD indicators were estimated. Of the five estimated correlations (i.e., comprehension-LDH, comprehension-LDM,

comprehension-LDL, comprehension-LDN, and comprehension-LDT), none were significantly different from 0.

Heritability estimates

As indicated above, both narrow- and broad-sense heritability coefficients were estimated. (Broad-sense heritability estimates are depicted in Figure 1.) As anticipated, narrow-sense heritability estimates correspond to the magnitude of child–father and child–mother correlations shown in Table 2. These estimates are fairly low and range from 0.12 for LDH to 0.21 for comprehension, but they all are statistically different from 0. The broad-sense heritability estimates are substantially higher. Their values range across multiple tau(s), but they stabilize so that they range from 0.28 (for LDL) to 0.75 (for LDH). These coefficients are higher than what is implied by the correlations shown in Table 2; these heritability estimates presumably reflect dominant and overdominant effects of genes controlling these traits for each other. The greater values of broad-sense heritability estimates indicate the presence of substantial nonlinear effects (i.e., dominance and overdominance).

Segregation analyses

The segregation analysis methodology used here estimates the number and effect sizes of putative genes involved in the genetic control of a trait of interest. Two sets of analyses were conducted. In the first set of analyses, the number of genes⁵ putatively controlling each of the investigated traits was evaluated. The results of these analyses are shown in Figure 2 in black. Thus, the left panel of Figure 2 shows that the number of genes that appear to be controlling the trait of comprehension is estimated at one to two. These estimates are obtained across multiple tau beta(s). The right panel of Figure 2 shows that the number of major genes controlling different chronometric indicators of the LD task is variable (in black). It appears that two to three major genes are involved in the genetic control of LDH, one to three major genes are involved in the genetic control of LDM, one to three major genes control LDL, two to four major genes control LDN, and one to two genes generate genetic variability in LDT. Interpreting the LD analyses as a whole, it could be said that the estimates vary between one and four genes but, on average, they converge on two to three genes.

The second set of analyses was done with the same six phenotypes, but now the analyses for the phenotype of interest were carried out with the second phenotype used as a covariate. By including a covariate phenotype, we can “subtract” those genes that are common to both traits, and only those that are unique to the first trait will remain; by running sets of analyses with and without covariates, we can estimate how many genes are shared between a pair of traits. For example, the left panel of Figure 2 shows the number of genes that control for comprehension when indicators of LD are covaried out (shown in gray). These analyses intend to investigate whether there are shared genes that exert their influence on both phenotypes (comprehension and LD). Specifically, if the same genes influence the phenotypes of comprehension and LD, then when comprehension is analyzed in the presence of LD, indicators the gray line will drop to 0, showing that only shared genes control comprehension and LD, and that there are no specific genes that contribute to the genetic variance of only one trait. As shown in gray, however, these analyses indicated that the indicators of both comprehension and LD do not appear to have any genes in common. When comprehension is controlled for LD (left panel) and LD is controlled for comprehension (right panel), the estimated number of genes contributing to the genetic variance of the trait of interest (either comprehension or LD) does not appear to change

⁵Here the QTL, genetic variants, or sources of genetic variants can be used as well. We use the word “genes” for simplicity and to follow the tradition previously established in the literature for similar analyses.

noticeably, suggesting that there are no shared major gene effects between reading comprehension and LD. Again, this observation is true for a range of tau beta(s), attesting to the robust nature of these findings. Second, there are instances (e.g., LDM and LDN) where the model estimates more genes in the covariate condition. This is to be expected in situations where there is a phenotypic, but not genetic correlation between the traits; the presence of such environmental effects inflates the variability in the trait to the extent that more genes are now required to explain this variability.

There are also two general methodological points to be made from these graphs. The first is the increasing number of genes and the slowly decreasing plateau as the tau beta increases. The peak is the optimal tau beta, and the slow decline indicates that the tau beta(s) are getting progressively too big for the sampler to accept. This is in line with the recommendations for specifying parameters in LOKI, and indicates that the range of tau beta(s) examined was sufficient for the data set (Wijsman & Yu, 2004). The second observation is that although the overall relationship between the models with and without covariates is consistent across tau beta(s), any two runs with a particular tau will exhibit some randomness, with one being higher than the other. To ensure that the results were not the product of chance, it was necessary to compare the results across hundreds of tau beta(s) for a given model: any one pair of tau beta(s) might lead to inconsistent results. In the second set of analyses, we estimated the magnitudes of the effects of the genes involved in the control of each of the traits. The results of these analyses are shown in Figure 3. The y axis indicates the total genetic variance for the trait. Like the analyses presented earlier, these estimates were obtained across multiple tau beta(s) with each horizontal line representing a putative gene. These analyses capture the additive contribution of each of the genes detected by LOKI to the total genetic variance in the trait. Based on the details presented in Figure 3, it appears that there is a substantive-effect gene contribution to comprehension (with one or two other genes of minor effects), and there are small-effect (<0.20) and moderate-effect (0.25–0.50) gene contributions to LD response time.

Discussion

In this study, using a genetically informative design, we investigated the phenotypic and genetic relatedness among reading comprehension scores and LD reaction time in an unselected sample of children and their families. A few aspects of our investigation warrant discussion.

Note that in these data there is a robust (i.e., significant and reliable), although not very strong (4%–5% of the variance), correlation between comprehension and LD reaction time. This correlation is present for both children and adults, for males and females, for words of various frequencies, and for nonwords. This finding is not unexpected and is theoretically linked to other work in the field (Perfetti, 2007). To our knowledge, however, there is not a great deal of empirical data in the literature estimating shared variance between LD and reading comprehension with which these findings can be compared. Previously results indicated no substantial correlation between LD and comprehension in English, although there were associations between LD and pseudoword/word processing, which in turn predicated comprehension (e.g. Katz et al., 2012). It is also should be noted that there is a body of literature that has consistently failed to find relationships between tasks such as simple response time and reading ability (Stringer & Stanovich, 2000).

It is interesting that the profile of familial correlations indicates a rather tangential association between reading comprehension and LD in parents and children. Although weak, it is detectable for mother-child pairs; it is not statistically significant for father-child pairs. This finding needs to be substantiated further because it has been suggested in the literature

that the psychological texture of LD can vary depending on the stage of reading acquisition (Gijssels, Van Bon, & Bosman, 2004). It is interesting, however, that reading comprehension and LD correlate between parents rather strongly, suggesting either the presence of assortative mating or convergence for these markers of reading and lexicality. These findings are perhaps of no surprise, given the literature on spousal similarity on a variety of cognitive characteristics (Mascie-Taylor, 1989).

Nonetheless, although familial correlations might not be high, the estimated heritability coefficients are statistically significant for all phenotypes. The range of heritability estimates obtained for comprehension (0.20–0.60) is comparable to those in the literature (Pasisz & Schatschneider, 2010). To our knowledge, there are no published heritability coefficients for LD reaction time, but the range of the coefficients presented here is comparable to those presented for various chronometric indicators of cognitive tasks (Neubauer, Spinath, Riemann, Angleitner, & Borkenau, 2000). It is also important to mention that, for all of the investigated indicators, the broad-sense heritability estimates appear to be higher than the narrow-sense heritability estimates, indicating the potential importance of the nonlinear effects characteristic of the genetic forces controlling these traits.

Moreover, there are three observations generated by the segregation analyses. First, the findings suggest that there are more genes (or QTLs or sources of genetic variation) involved in the control of LD than in the control of comprehension. Second, there is evidence for major gene effect(s) in comprehension, but not in LD making (where the effects of the genes appear to be moderate at best). Third, the results of the segregation analyses suggest that there is no shared genetic etiology between LD response times and reading comprehension ability.

LD does not appear to be “simpler,” at least in terms of its genetic texture, than reading comprehension, nor do the traits appear to share a genetic etiology. As indicated above, the literature sets a number of requirements for the endophenotype–phenotype relationship. LD shows reliability within/cross-age individual differences and is associated with the target trait, an association that is theoretically supported by the connection between lexical representation and reading comprehension. LD is also heritable. However, the genetic correlations between LD indicators and reading comprehension are not statistically significant, and there is no evidence for shared genetic variance in the results of the segregation analyses. The most surprising finding from this work is that, if anything, the genetic machinery behind LD indicators appears to be at least as complex as the machinery behind reading comprehension. These findings suggest that LD is perhaps not a suitable endophenotype for reading comprehension if one uses the conceptual criteria for “being an endophenotype” discussed above. These findings concur with results suggesting that response-time measures might also be unsuitable endophenotypes for IQ (van Leeuwen, van den Berg, Hoekstra, & Boomsma, 2007).

Yet, the lack of genetic connection between LD and reading comprehension is surprising. Although there have been few studies investigating the shared genetic contribution between comprehension and single-word reading skills, those studies that do exist found significant and robust shared variance between the constructs. How can we reconcile our findings with those in the field? Furthermore, given that both tasks inevitably make demands on word decoding, it is surprising that there is not at least one shared gene of major effect between LD and reading comprehension. How can we reconcile the differences with prior literature, namely, that there is shared genetic variance between reading comprehension and word decoding?

With respect to reconciling our differences with the finding of shared genetic variance between reading comprehension and word decoding (Betjemann et al., 2008), it is possible that there are simply too many differences between the studies. Betjemann et al.'s (2008) study sampled both typically developing and reading disabled twins. Their measures of reading comprehension were different from those used in our study, and their study used single-word reading accuracy, not LD reaction time. In addition, their study was conducted using the English language, whereas our study was conducted in Russian. The increased orthographic complexity of printed English makes word decoding more difficult (Ziegler & Goswami, 2006) and may skew the relationship between decoding and comprehension. Decoding may act as a "gatekeeper" to comprehension, and may artificially inflate its contribution to comprehension in less transparent languages such as English. Finally, the different samples also led us to use very different analytic approaches with very different outcome variables. In their analysis, Betjemann et al. (2008) decomposed phenotypic variance into genetic and environmental effects using a twin design; the outcome demonstrated how much trait variance was attributable to genetic sources. Our study used a family design and, although we also estimated genetic variance, our approach was targeted to identify the number and effect sizes of the genes contributing to the traits. Furthermore, the twin design of Betjemann et al. assumes that, on average, dizygotic twins share 50% of their genes compared to 100% for monozygotic twins, but cannot account for dominance or epistasis, possibly inflating heritability estimates in comparison to our results (Eaves, Heath, Neale, Hewitt, & Martin, 1998).

A few points should also be made regarding LD as an endophenotype for reading comprehension in particular and the concept of endophenotypes in general. In our study, LD did not exhibit a shared genetic etiology with reading comprehension and did not have a simpler genetic texture than reading comprehension and therefore fails to qualify as an endophenotype. Given our considerable sample size, and the empirical support for a behavioral link between the two traits, this finding was surprising. What went wrong with the endophenotype approach?

Perhaps LD is too complex. Despite the significant overlap between LD and reading comprehension, LD has the added requirement of a speeded response and a decision stage. That the speed of the response is dependent on the processes involved in making the decision is critical, which may have more to do with strategic factors and decision criteria than word decoding. If a participant adopts a fast liberal response style or a slow, conservative one, the response time may be more reflective of these decision criteria than of lexical processing. These response styles or decision criteria might obscure the variance in LD as it relates to reading ability and instead reflect other traits that may be equally related to reading ability (e.g., working memory). There is definite evidence showing that characteristics of decision criteria can have a substantial effect on response times (Ratcliff & McKoon, 2008; Roach, Edwards, & Hogben, 2004; Spaniol, Madden, & Voss, 2006; Voss, Rothermund, & Voss, 2004), and it might be the case in our data that these other aspects of response time affect variability to such an extent that the impact of reading and lexical factors is lost. In this case, LD and reading comprehension are influenced by the same genes, but the LD task is much more sensitive to the effects of genes that influence attention and inhibition, in which case further investigation where these effects are estimated separately is warranted.

Another alternative explanation, mentioned above, is that the orthographic difficulty of English acts as a "gatekeeper" to comprehension, resulting in the high genetic correlations with comprehension and word decoding in twin studies, but not in our Russian speaking sample. However, we find it hard to believe that word decoding is so much easier in Russian that it shares no variance with reading comprehension.

Another possibility is that there is nothing wrong with the endophenotype approach, if we can successfully identify an endophenotype. The literature offers a number of approaches for identifying an endophenotype. One approach advocating the search for endophenotypes at the conceptual level casts a remarkably wide net over what can be considered an endophenotype and why (Cannon & Keller, 2005), nominating virtually any interindividual variability between gene and diagnosis. This liberal approach has two rather serious pitfalls. First, anything can potentially be considered an endophenotype, in which case perhaps we should just call them phenotypes again. Second, this pitfall is of greater concern: it is the tacit assumption that any complex neuropsychiatric disorder is decomposable into separate modules and functions, an assumption that may not reflect reality (Anderson, 2008, 2009). Another approach is data driven (Glahn et al., 2012) and is based on the statistical analyses of many phenotypes, resulting in an identification of few or one as an endophenotype. Both approaches (as well as other approaches) have their own proponents and the field has not accumulated enough data on either yet to impose a final judgment on the best strategy for identifying endophenotypes.

Throughout this article, we have assumed that reading comprehension is a hierarchical process, an outcome of subprocesses that combine additively or interactively. However, it might be that reading comprehension is simply not amenable to this sort of decomposition. Multiple factors can influence reading comprehension performance and perhaps reading comprehension is simply too complex to be decomposed in such a way. Vocabulary, attention, working memory, and phonological and orthographic skills all interact during a reading comprehension task. It is conceivable that in any reading comprehension task, in any individual, the limiting factor could differ. This interpretation still leaves the door open for an endophenotype approach, but coupled with the possible genetic complexity of the LD task, it is not immediately clear how to proceed.

Finally, given that LD and reading comprehension do share some neurobiological machinery, we see that there are three possible explanations, which are not mutually exclusive, for our finding of a lack of shared genetic variance. The first is that LD is a complex behavior that reflects more variance in motor and decision processes than in lexical access. This explanation is compatible with the endophenotype approach, provided that we can successfully fractionate and measure these component processes, perhaps using more sophisticated models of reaction times. The second possible explanation is that variability in reading comprehension is simply not amenable to the kind of decomposition that underlies the endophenotype approach. The failure of standardized measures of reading comprehension to converge suggests that there may be no reliably dissociable components of reading comprehension to begin with. We might be able to develop a reading assessment measure that shares significant variance with LD, but only after simplifying the measure so much that it no longer resembles what we would think of as reading comprehension. The third explanation is that it is possible that LD is an endophenotype for a specific (or a specific set of) developmental condition(s) (e.g., ADHD or reading disability) rather than another complex trait (e.g., reading comprehension) that itself might be an endophenotype for a specific (or a specific set of) developmental condition(s) (e.g., reading disability, autism, or intellectual disability).

These three explanations suggest three complementary courses of action. The first is the need to more carefully consider the mechanisms of our candidate endophenotypes, not simply their association with a complex trait. The second is the need to acknowledge the complexity of the complex trait, and decide whether it is decomposable into endophenotypes or better described as a collection of clinically significant “phenotypes.” The third is the need to utilize the endophenotype–phenotype relationship concept only in the context of the

quest for the genetic bases of a developmental disorder, not another complex trait. These courses of action are yet to be explored.

Acknowledgments

This work was supported by a grant from the International Dyslexia Association (E. Grigorenko and L. Katz, Principal Investigators [PIs]) and by Grants NICHD01994 (C. Fowler, PI), DC007665 (E. Grigorenko, PI), HD048830 (K. Pugh, PI), HD052120 (R. Wagner, PI), MH18268 (J. Leckman, PI), and MH081756 (A. Klin, PI) from the National Institutes of Health. Grantees undertaking such projects are encouraged to freely express their professional judgment. Therefore, this article does not necessarily represent the position or policies of the International Dyslexia Association or the National Institutes of Health, and no official endorsement should be inferred. The authors express sincere gratitude to Mr. Stephen Katz for programming the experimental procedures and managing the initial database, the team of research assistants who collected the data, and Ms. Robyn Rissman and Ms. Mei Tan for their editorial assistance.

References

- Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *American Journal of Human Genetics*. 1998; 62:1198–1211. [PubMed: 9545414]
- Almasy L, Blangero J. Endophenotypes as quantitative risk factors for psychiatric disease: Rationale and study design. *American Journal of Medical Genetics*. 2001; 105:42–44. [PubMed: 11424994]
- Anderson GM. The potential role for emergence in autism. *Autism Research*. 2008; 1:1–18. [PubMed: 19360644]
- Anderson GM. Conceptualizing autism: The role for emergence. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2009; 48:688–691. [PubMed: 19542823]
- Aulchenko, Y.; Struchalin, M. GenABEL: Genome-wide SNP association analysis. R package version 1.4-4 [Software]. 2009. Retrieved from <http://CRAN.R-project.org/package=GenABEL>
- Betjemann RS, Willcutt EG, Olson RK, Keenan JM, DeFries JC, Wadsworth SJ. Word reading and reading comprehension: Stability, overlap and independence. *Reading and Writing*. 2008; 21:539–558.
- Blangero J. Multivariate oligogenic linkage analysis of quantitative traits in general pedigrees. *American Journal of Human Genetics*. 1995; 57:50.
- Bond, T.; Fox, C. Applying the Rasch model. Hillsdale, NJ: Erlbaum; 2001.
- Byrne B, Coventry WL, Olson RK, Samuelsson S, Corley R, Willcutt EG, et al. 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, Samuelsson S, Wadsworth S, Hulslander J, Corley R, De-Fries JC, et al. Longitudinal twin study of early literacy development: Preschool through Grade 1. *Reading and Writing*. 2007; 20:77–102.
- Cain K, Oakhill J. Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*. 2006a; 76:683–696. [PubMed: 17094880]
- Cain K, Oakhill J. Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*. 2006b; 76:697–708. [PubMed: 17094881]
- Cannon T, Keller M. Endophenotypes in genetic analyses of mental disorders. *Annual Review of Clinical Psychology*. 2005; 2:267–290.
- Cirrin F. Lexical search speed in children and adults. *Journal of Experimental Child Psychology*. 1984; 37:158–175.
- Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*. 2001; 108:204–256. [PubMed: 11212628]
- Crosbie J, Perusse D, Barr CL, Schachar RJ. Validating psychiatric endophenotypes: Inhibitory control and attention deficit hyperactivity disorder. *Neuroscience & Biobehavioral Reviews*. 2008; 32:40–55. [PubMed: 17976721]

- Cutting L, Scarborough H. Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*. 2006; 10:277–299.
- Daw EW, Heath SC, Wijsman EM. Multipoint oligogenic analysis of age-at-onset data with applications to Alzheimer's disease pedigrees. *American Journal of Human Genetics*. 1999; 64:839–851. [PubMed: 10053019]
- de Geus EJC, Wright MJ, Martin N, Boomsma DI. Genetics of brain function and cognition. *Behavior Genetics*. 2001; 31:489–495. [PubMed: 11838528]
- de Jong CG, Van De Voorde S, Roeyers H, Raymaekers R, Oosterlaan J, Sergeant JA. How distinctive are ADHD and RD? Results of a double dissociation study. *Journal of Abnormal Child Psychology*. 2009; 37:1007–1017. [PubMed: 19488850]
- Dorans, NJ.; Pommerich, M.; Holland, PW. Linking and aligning scores and scales. New York: Springer Science; 2007.
- Eaves L, Heath A, Neale M, Hewitt J, Martin N. Sex differences and non-additivity in the effects of genes on personality. *Twin Research & Human Genetics*. 1998; 1:131–137.
- Engle R, Cantor J, Carullo J. Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1992; 18:975–992.
- Flint J, Munafo MR. The endophenotype concept in psychiatric genetics. *Psychological Medicine*. 2007; 37:163–180. [PubMed: 16978446]
- Freedman R, Coon H, Myles-Worsley M, Orr-Urtreger A, Olincy A, Davis A, et al. Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. *Proceeding of the National Academy of Sciences*. 1997; 94:587–592.
- Gayán J, Olson RK. Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology*. 2003; 84:97–123. [PubMed: 12609495]
- Gernsbacher MA. Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*. 1984; 113:256–281. [PubMed: 6242753]
- Gernsbacher MA, Faust ME. The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1991; 17:245–262.
- Gijssel MAR, Van Bon WHJ, Bosman AMT. Assessing reading skills by means of paper-and-pencil lexical decision: Issues of reliability, repetition, and word-pseudoword ratio. *Reading and Writing*. 2004; 17:517–536.
- Gilden DL, Hancock H. Response variability in attention-deficit disorders. *Psychological Science*. 2007; 18:796–802. [PubMed: 17760776]
- Gilden DL, Thornton T, Mallon MW. 1/f noise in human cognition. *Science*. 1995; 267:1837–1839. [PubMed: 7892611]
- Glahn DC, Curran JE, Winkler AM, Carless MA, Kent JW Jr, Charlesworth JC, et al. High dimensional endophenotype ranking in the search for major depression risk genes. *Biological Psychiatry*. 2012; 71:6–14. [PubMed: 21982424]
- Goos LM, Crosbie J, Payne S, Schachar R. Validation and extension of the endophenotype model in ADHD patterns of inheritance in a family study of inhibitory control. *American Journal of Psychiatry*. 2009; 166:711–717. [PubMed: 19448185]
- Gottesman II, Gould TD. The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*. 2003; 160:636–645. [PubMed: 12668349]
- Grigorenko EL. Genetic bases of developmental dyslexia: A capsule review of heritability estimates. *Enfance*. 2004; 3:273–287.
- Grigorenko EL, DeYoung CG, Getchell M, Haefel GJ, af Klinteberg B, Kaposov RA, et al. Exploring interactive effects of genes and environments in etiology of individual differences in reading comprehension. *Development and Psychopathology*. 2007; 19:1089–1103. [PubMed: 17931436]
- Grigorenko EL, Kiln A, Pauls DL, Sent R, Hooper C, Volkmar F. A descriptive study of hyperlexia in a clinically referred sample of children with developmental delays. *Journal of Autism and Developmental Disorders*. 2002; 32:3–11. [PubMed: 11916330]

- Grigorenko EL, Klin A, Volkmar F. Hyperlexia: Disability or superability? *Journal of Child Psychology and Psychiatry*. 2003; 44:1079–1091. [PubMed: 14626452]
- Grossman RB, Bemis RH, Skwerer DP, Tager-Flusberg H. Lexical and affective prosody in children with high-functioning autism. *Journal of Speech, Language, and Hearing Research*. 2010; 53:778–793.
- Harlaar N, Spinath FM, Dale PS, Plomin R. Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. *Journal of Child Psychology & Psychiatry*. 2005; 46:373–384. [PubMed: 15819646]
- Heath SC. Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genetics*. 1997; 61:748–760. [PubMed: 9326339]
- Heath SC, Snow G, Thompson E, Tseng C, Wijsman EM. MCMC segregation and linkage analysis. *Genetic Epidemiology*. 1997; 14:1011–1016. [PubMed: 9433616]
- Himpel S, Banaschewski T, Gruttner A, Becker A, Heise A, Uebel H, et al. Duration discrimination in the range of milliseconds and seconds in children with ADHD and their unaffected siblings. *Psychological Medicine*. 2009; 39:1745–1751. [PubMed: 19265568]
- Holmes VM. Bottom-up processing and reading comprehension in experienced adult readers. *Journal of Research in Reading*. 2009; 32:309–326.
- Hsu L, Wijsman E, Berninger V, Thomson J. Familial aggregation of dyslexia phenotypes. II: Paired correlated measures. *American Journal of Medical Genetics: Neuropsychiatric Genetics*. 2002; 114:471–478. [PubMed: 11992573]
- Iacono WG, Malone SM. Developmental endophenotypes: Indexing genetic risk for substance abuse with the P300 brain event-related potential. *Child Development Perspectives*. 2011; 5:239–247. [PubMed: 22247735]
- Igo RP, Chapman NH, Wijsman EM. Segregation analysis of a complex quantitative trait: Approaches for identifying influential data points. *Human Heredity*. 2006; 61:80–86. [PubMed: 16679774]
- Johnson W, Turkheimer E, Gottesman I, Bouchard T. Beyond heritability: Twin studies in behavioral research. *Current Directions in Psychological Science*. 2009; 18:217–220. [PubMed: 20625474]
- Katz L, Brancazio L, Irwin J, Katz S, Magnuson J, Whalen D. What lexical decision and naming tell us about reading. *Reading and Writing*. 2012; 25:1259–1282. [PubMed: 22754142]
- Keenan JM, Betjemann R, Wadsworth S, DeFries J, Olson R. Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading*. 2006; 29:75–91.
- Keenan JM, Betjemann RS, Olson RK. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*. 2008; 12:281–300.
- Kendler KS, Neale MC. Endophenotype: A conceptual analysis. *Molecular Psychiatry*. 2010; 15:789–797. [PubMed: 20142819]
- Loyd BH, Hoover HD. Vertical equating using the Rasch model. *Journal of Educational Measurement*. 1980; 17:179–193.
- Martens VE, De Jong PF. The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*. 2006; 98:140–149. [PubMed: 16690111]
- Mascie-Taylor C. Spouse similarity for IQ and personality and convergence. *Behavior Genetics*. 1989; 19:223–227. [PubMed: 2719625]
- Meyer-Lindenberg A. Intermediate or brainless phenotypes for psychiatric research? *Psychological Medicine*. 2010; 40:1057–1062. [PubMed: 20540175]
- Michael M. Responses on a lateralized lexical decision task relate to both reading times and comprehension. *Brain and Cognition*. 2009; 71:416–426. [PubMed: 19559514]
- Morein-Zamir S, Fineberg NA, Robbins TW, Sahakian BJ. Inhibition of thoughts and actions in obsessive-compulsive disorder: Extending the endophenotype? *Psychological Medicine*. 2010; 40:263–272. [PubMed: 19573261]
- Morrison C, Ellis A. Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21:116–133.

- Neubauer A, Spinath F, Riemann R, Angleitner A, Borkenau P. Genetic and environmental influences on two measures of speed of information processing and their relation to psychometric intelligence: Evidence from the German observational study of adult twins. *Intelligence*. 2000; 28:267–289.
- Pasisz, DJ.; Schatschneider, C. Unpublished manuscript. 2010. Genetic and environmental influences on reading comprehension ability: A research synthesis.
- Perfetti CA. Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*. 2007; 11:357–383.
- Perfetti CA, Hogaboam T. Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*. 1975; 67:461–469.
- Peterson BS, Weissman MM. A brain-based endophenotype for major depressive disorder. *Annual Review of Medicine*. 2011; 62:461–474.
- Rasch, G. Probabilistic models for some intelligence and attainment tests. Chicago: University of Chicago Press; 1960.
- Ratcliff R. Methods for dealing with reaction time outliers. *Psychological Bulletin*. 1993; 114:510–532. [PubMed: 8272468]
- Ratcliff R, McKoon G. The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*. 2008; 20:873–922. [PubMed: 18085991]
- Ricketts J. Research review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. 2011; 52:1111–1123.
- Ritsner, M., editor. The handbook of neuropsychiatric biomarkers, endophenotypes and genes. New York: Springer; 2009.
- Roach N, Edwards V, Hogben J. The tale is in the tail: An alternative hypothesis for psychophysical performance variability in dyslexia. *Perception*. 2004; 33:817–830. [PubMed: 15460509]
- Roopesh BN, Rangaswamy M, Kamarajan C, Chorlian DB, Pandey AK, Porjesz B. Reduced resource optimization in male alcoholics: N400 in a lexical decision paradigm. *Alcoholism: Clinical and Experimental Research*. 2010; 34:1905–1914.
- Rutherford BJ. Reading disability and hemispheric interaction on a lexical decision task. *Brain and Cognition*. 2006; 60:55–63. [PubMed: 16271819]
- Schmiedek F, Oberauer K, Wilhelm O, Suss HM, Wittmann WW. Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*. 2007; 136:414–429. [PubMed: 17696691]
- Seidenberg MS, Peersen A, MacDonald M, Plaut DC. Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1996; 22:48–62.
- Seidenberg MS, Plaut DC, Petersen AS, McClelland JL, McRae K. Nonword pronunciation and models of word recognition. *Journal of Experimental Psychology: Human Perception and Performance*. 1994; 20:1177–1196. [PubMed: 7844510]
- Shankweiler D. Words to meanings. *Scientific Studies of Reading*. 1999; 3:113–127.
- Smit D, Boersma M, van Beijsterveldt C, Posthuma D, Boomsma D, Stam C, et al. Endophenotypes in a dynamically connected brain. *Behavior Genetics*. 2010; 40:167–177. [PubMed: 20111993]
- Smith, EV.; Smith, RM., editors. Introduction to Rasch measurement. Maple Grove, MN: JAM Press; 2005.
- Spaniol J, Madden DJ, Voss A. A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology Learning, Memory, and Cognition*. 2006; 32:101–117.
- Stewart ME, Ota M. Lexical effects on speech perception in individuals with “autistic” traits. *Cognition*. 2008; 109:157–162. [PubMed: 18834977]
- Stringer R, Stanovich KE. The connection between reaction time and variation in reading ability: Unravelling covariance relationships with cognitive ability. *Scientific Studies of Reading*. 2000; 4:41–53.

- van den Bos KP, Nakken H, Nicolay PG, van Houten EJ. Adults with mild intellectual disabilities: Can their reading comprehension ability be improved? *Journal of Intellectual Disability Research*. 2007; 51:835–849. [PubMed: 17910536]
- van Leeuwen M, van den Berg SM, Hoekstra RA, Boomsma DI. Endophenotypes for intelligence in children and adolescents. *Intelligence*. 2007; 35:369–380.
- Verhoeven L, Vermeer A. Literacy achievement of children with intellectual disabilities and differing linguistic backgrounds. *Journal of Intellectual Disability Research*. 2006; 50:725–738. [PubMed: 16961702]
- Voss A, Rothermund K, Voss J. Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*. 2004; 32:1206–1220.
- Wijsman EM, Peterson D, Leutenegger AL, Thomson JB, Goddard KAB, Hsu L, et al. Segregation analysis of phenotypic components of learning disabilities: I. Nonword memory and digit span. *American Journal of Human Genetics*. 2000; 67:631–646. [PubMed: 10924405]
- Wijsman EM, Yu D. Joint oligogenic segregation and linkage analysis using Bayesian Markov Chain Monte Carlo Methods. *Molecular Biotechnology*. 2004; 28:205–226. [PubMed: 15542922]
- Ziegler JC, Goswami U. Becoming literate in different languages: Similar problems, different solutions. *Developmental Science*. 2006; 9:429–436. [PubMed: 16911438]
- Zoccolotti P, de Luca M, di Pace E, Gasperini F, Judica A, Spinelli D. Word length effect in early reading and in developmental dyslexia. *Brain and Language*. 2005; 93:369–373. [PubMed: 15862860]

Figure 1.

Broad-sense heritability estimates for comprehension and lexical decision. The x axis is not to scale and represents the increasing tau beta parameter. Each point is the mean heritability estimate for an analysis of 25,000 iterations. C, comprehension; LDL, lexical decision low frequency; LDH, lexical decision high frequency; LDN, lexical decision nonwords; LDM, lexical decision medium frequency; LDT, lexical decision total mean reaction time. [A color version of this figure can be viewed online at <http://journals.cambridge.org/dpp>]

Figure 2.

Number of genes estimated to control each of the six traits, when traits are investigated independently and when a different correlated trait is covaried out. The left column is lexical decision reaction times covarying for comprehension, and the right is comprehension covarying for lexical decision times. Black represents the original and gray the covariate models. C, comprehension; LDH, lexical decision high frequency; LDL, lexical decision low frequency; LDN, lexical decision nonwords; LDT, lexical decision total mean reaction time.

Figure 3.

The estimated magnitude of effects of the contributing genes (shown in different shades of grey) to the traits of interest. C, comprehension; LDH, lexical decision high frequency; LDM, lexical decision medium frequency; LDL, lexical decision low frequency; LDN, lexical decision nonwords; LDT, lexical decision total mean reaction time.

Table 1

Descriptive statistics for comprehension and lexical decision for children and adults

	<u>Reaction Time</u>		<u>Accuracy</u>	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Children (boys)				
Comp.			−0.08	1.06
LDH	1675.35	480.62	0.94	0.09
LDM	1870.55	501.92	0.86	0.12
LDL	2287.76	591.03	0.61	0.17
LDN	2266.12	535.68	0.80	0.19
LDT	2008.22	490.67	0.80	0.09
Children (girls)				
Comp.			0.13	0.99
LDH	1598.50	432.32	0.95	0.12
LDM	1808.42	467.21	0.85	0.13
LDL	2227.24	547.90	0.61	0.16
LDN	2208.95	513.56	0.81	0.19
LDT	1949.22	457.60	0.80	0.11
Fathers				
Comp.			−0.28	0.92
LDH	1118.88	331.55	0.99	0.06
LDM	1245.91	366.66	0.98	0.07
LDL	1478.82	437.50	0.89	0.11
LDN	1634.95	499.64	0.93	0.11
LDT	1386.59	381.00	0.95	0.07
Mothers				
Comp.			0.20	0.97
LDH	1039.61	277.94	0.99	0.09
LDM	1156.05	294.95	0.98	0.10
LDL	1390.84	374.90	0.89	0.12
LDN	1575.06	456.00	0.92	0.15
LDT	1306.59	317.24	0.94	0.10

Note: Comp., comprehension; LDH, lexical decision high frequency; LDM, lexical decision medium frequency; LDL, lexical decision low frequency; LDN, lexical decision nonwords; LDT, lexical decision total mean reaction time reflects correct responses for word and nonword stimuli.

Table 2
Family correlations (Pearson r) for lexical decision and reading comprehension measures

		Child					Father					Mother				
		Comp.	LDH	LDL	LDN	LDT	Comp.	LDH	LDL	LDN	LDT	Comp.	LDH	LDL	LDN	LDT
Child																
Comp.	<i>r</i>	.985	-.187**	-.179**	-.177**	-.189**	.05	-.02	0	-.03	-.02	.191**	-.05	-.06	-.09	-.07
	<i>p</i>		0	.01	.01	0	.38	.74	.97	.56	.8	0	.26	.17	.05	.15
	<i>n</i>		491	494	488	494	353	291	291	291	291	493	472	472	472	472
LDH	<i>r</i>	-.185**	.996	.893**	.825**	.951**	.02	.07	.06	.05	.06	-.03	.08	.09	.098*	.108*
	<i>p</i>	0	0	0	0	0	.67	.24	.3	.43	.28	.49	.08	.06	.04	.02
	<i>n</i>	491	493	488	493	493	347	286	286	286	286	484	464	464	464	464
LDL	<i>r</i>	-.158**	.893**	.997	.871**	.956**	.03	.06	.06	.04	.06	-.02	.109*	.110*	.107*	.128**
	<i>p</i>	0	0	0	0	0	.59	.31	.29	.47	.32	.74	.02	.02	.02	.01
	<i>n</i>	495	493	490	490	496	348	286	286	286	286	487	466	466	466	466
LDN	<i>r</i>	-.178**	.825**	.870**	.998	.937**	.05	.06	.05	.05	.06	-.02	.092*	.095*	.120**	.124**
	<i>p</i>	0	0	0	0	0	.38	.36	.41	.38	.33	.64	.05	.04	.01	.01
	<i>n</i>	495	488	490	490	490	346	284	284	284	284	481	460	460	460	460
LDT	<i>r</i>	-.178**	.951**	.955**	.937**	.997	.03	.06	.05	.04	.06	-.02	.104*	.111*	.116*	.132**
	<i>p</i>	0	0	0	0	0	.54	.33	.37	.5	.35	.68	.02	.02	.01	0
	<i>n</i>	488	493	496	490	286	348	286	286	286	286	487	466	466	466	466
Father																
Comp.	<i>r</i>	.04	.02	.03	.05	.03	.999	-.152**	-.195**	-.172**	-.190**	.370**	-.128*	-.156**	-.01	-.139*
	<i>p</i>	.44	.66	.57	.38	.53		.01	.09	0	0	0	.02	.01	.07	.01
	<i>n</i>	331	347	348	346	348		287	287	287	287	348	331	331	331	331
LDH	<i>r</i>	-.03	.07	.07	.06	.06	-.125*	.999	.889**	.739**	.934**	-.1	.162**	.166**	.11	.164
	<i>p</i>	.66	.22	.27	.31	.29	.04		0	0	0	.1	.01	.01	.07	.01
	<i>n</i>	280	280	280	278	280	280		292	292	292	285	268	268	268	268
LDL	<i>r</i>	-.01	.07	.07	.06	.06	-.166**	.889**	.998	.753**	.939**	-.11	.140*	.159**	0.1	.149*
	<i>p</i>	.92	.27	.25	.36	.32	.01	0	0	0	0	.07	.02	.01	.12	.02
	<i>n</i>	280	280	280	278	280	280		292	292	292	285	268	268	268	268

\$watermark-text

\$watermark-text

\$watermark-text

		Child			Father			Mother		
	Comp.	LDH	LDL	LDN	LDT	Comp.	LDH	LDL	LDN	LDT
LDN	<i>r</i>	-.03	.05	.07	.05	-.158**	.741**	.759**	.999	.904**
	<i>p</i>	.66	.38	.28	.4	.01	0	0	.19	.04
	<i>n</i>	280	280	278	280	280	286	286	268	268
LDT	<i>r</i>	-.05	.07	.07	.06	-.139*	.934**	.940**	.906**	.998
	<i>p</i>	.39	.24	.27	.29	.02	0	0	.09	.01
	<i>n</i>	280	280	278	280	280	286	286	268	268
Mother										
Comp.	<i>r</i>	.187**	-.04	-.02	-.03	.370**	-.1	-.11	-.11	-.09
	<i>p</i>	0	.37	.64	.53	0	.1	.07	.08	.14
	<i>n</i>	481	477	481	474	326	272	272	272	272
LDH	<i>r</i>	-.05	.07	.103*	.09	.383**	.172**	.149*	.137*	.160*
	<i>p</i>	.28	.12	.03	.07	0	.01	.02	.03	.01
	<i>n</i>	461	453	455	449	325	258	258	258	258
LDL	<i>r</i>	-.07	.08	.111*	.111*	-.118*	.165**	.162**	.170**	.179**
	<i>p</i>	.13	.08	.02	.04	.04	.01	.01	.01	0
	<i>n</i>	461	453	455	449	308	258	258	258	258
LDN	<i>r</i>	-.07	.09	.098*	.113*	-.148**	.12	.11	.169**	.146*
	<i>p</i>	.17	.07	.04	.02	.01	.06	.07	.01	.02
	<i>n</i>	461	453	455	449	308	258	258	258	258
LDT	<i>r</i>	-.08	.099*	.121**	.119*	-.139*	.172**	.161**	.179**	.183**
	<i>p</i>	.08	.04	.01	.01	.02	.01	.01	0	0
	<i>n</i>	461	464	455	449	308	258	258	258	258

Note: *n*, number of data points for a particular analysis; Comp., comprehension; LDH, lexical decision high frequency; LDM, lexical decision medium frequency; LDL, lexical decision low frequency; LDN, lexical decision nonwords; LDT, lexical decision total. Correlations are shown for both unadjusted (below the diagonal) and age- and gender-adjusted (above the diagonal) variables.

* $p < .05$.

** $p < .01$, indicates significant correlation.