Published in final edited form as:
Sci Stud Read. 2011 January 1; 15(4): 363–382. doi:10.1080/10888438.2010.493965.

Choice of Reading Comprehension Test Influences the Outcomes of Genetic Analyses

Rebecca S. Betjemann[1], **Janice M. Keenan**[2], **Richard K. Olson**[3], and **John C. DeFries**[3]

[1] Regis University

[2] University of Denver

[3] University of Colorado at Boulder

Abstract

Does the choice of test for assessing reading comprehension influence the outcome of genetic analyses? A twin design compared two types of reading comprehension tests classified as primarily associated with word decoding (RC-D) or listening comprehension (RC-LC). For both types of tests, the overall genetic influence is high and nearly identical. However, the tests differed significantly in how they covary with the genes associated with decoding and listening comprehension. Although Cholesky decomposition showed that both types of comprehension tests shared significant genetic influence with both decoding and listening comprehension, RC-D tests shared most genetic variance with decoding, and RC-LC tests shared most with listening comprehension. Thus, different tests used to measure the same construct may manifest very different patterns of genetic covariation. These results suggest that the apparent discrepancies among the findings of previous twin studies of reading comprehension could be due at least in part to test differences.

Background

Recent years have seen significant growth in understanding genetic influences on individual differences in reading and reading disability, both at the behavioral and molecular level (for reviews see Pennington & Olson, 2005; Fisher & DeFries, 2002). One thing that often varies across these different studies is how reading is being assessed. The question we address in this paper is: does a researcher's choice of test instrument for assessing the reading phenotype influence the outcome of their analyses?

Oftentimes the choice of how to assess the reading phenotype can be dictated by administrative concerns, e.g., how long a test takes to administer may be a deciding factor when administering a large battery of cognitive tests. As a result, tests of single word reading are frequently used because they take much less time than assessing passage fluency and passage comprehension. However, while often related, word reading and comprehension are also separable components of reading skill (e.g., Gough, Hoover, & Peterson, 1996; Gough & Tunmer, 1986; Hoover & Gough, 1990), having unique genetic influences (Keenan, Betjemann, Wadsworth, DeFries & Olson, 2006). Thus, it matters considerably whether reading is assessed by tests of single word reading or tests of comprehension.

In this paper we examine the question of whether it also matters which particular reading comprehension test is used in assessing genetic influences on comprehension processes.

Address correspondence to: Rebecca S. Betjemann, Department of Psychology, D-12, Regis University, 3333 Regis Blvd., Denver, CO 80221, rbetjema@regis.edu.

Non-genetic comparisons of reading comprehension tests have recently shown differences across tests (Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997). If genetic results also vary depending on the test, these differences have the potential to impact how we understand the genetics of reading comprehension. The goal of this paper is to assess that impact.

Assessing Genetic Influences on Word Reading & Comprehension

Studies with genetically sensitive designs using twins allow for an assessment of the genetic, shared environmental, and non-shared environmental contributions to etiology. The classic twin design compares the similarity within identical, or monozygotic (MZ) twin pairs who share 100% of their genes with that of fraternal, or dizygotic (DZ) twin pairs, who share on average 50% of their segregating genes (Plomin, DeFries, McClearn, & McGuffin, 2008). To the extent that individual differences in a trait are caused by genetic influences, MZ twin pairs should be more similar than DZ twin pairs. Shared environmental influences are those factors which make twins pairs more similar to each other, such as having a common home or school environment. In contrast, nonshared environment independently influences members of twin pairs, such as one twin having an illness or accident that the other did not have, and also typically includes measurement error. While there have been a number of studies demonstrating genetic influences on word reading, with estimates ranging from .30 to .85 (e.g., Byrne, Wadsworth, et al., 2005; DeFries et al., 1997; Gayán & Olson, 2003; Harlaar, Spinath, Dale, & Plomin, 2005; Petrill, Deater-Deckard, Thompson, DeThorne, & Schatschneider, 2006; Stevenson, Graham, Fredman & McLoughlin, 1987; Tiu, Wadsworth, Olson, & DeFries, 2004; Wadsworth, DeFries, Olson, & Willcutt, 2007), only a handful of studies have also investigated genetic influences on comprehension, and they find similar estimates ranging from .41 to .76 (e.g., Betjemann, Willcutt, Olson, Keenan, DeFries, & Wadsworth, 2008; Byrne, Olson, et al., 2005; Harlaar, Dale, & Plomin, 2007; Keenan et al., 2006; Petrill, Deater-Deckard, Thompson, Schatschneider, & DeThorne, 2007).

In addition to investigating overall genetic influences on a behavior by using univariate analyses within a measure, we can also use multivariate analyses to investigate the genetic and environmental influences on component skills underlying the behavior. This allows us to assess how component skills are related by determining which genetic and environmental factors are shared across measures, and which are unique. As we discuss below, only a couple of studies have examined the multivariate relations between decoding and comprehension (e.g., Betjemann et al., 2008; Byrne et al., 2007; Keenan et al., 2006). Because there are some discrepancies across these studies in whether there are separate genetic influences on word decoding and comprehension, the present paper attempts to help reconcile the results.

Separate Genetic Influences on Word Reading & Comprehension

Until recently, twin studies of reading tended to focus on single word reading and its component skills (e.g., Gayán & Olson 2001, 2003) or a composite score that included reading comprehension, word reading, and spelling (e.g., DeFries, Fulker, & Labuda,1987). It was common to assume that the basic difficulty was in word identification and that any difficulties in reading comprehension were byproducts of word decoding problems (e.g., Perfetti, 1985). Although comprehension and word reading skills are highly correlated (e.g., Catts, Fey, Zhang, & Tomblin, 1999; Gough et al., 1996; Leach, Scarborough, & Rescorla, 2003), their theoretical separation was proposed long ago in the Simple View of Reading (e.g., Gough et al., 1996; Gough & Tunmer, 1986; Hoover & Gough, 1990), and there is now considerable empirical support for it. For example, studies have found dissociations of the clinical profiles of children with reading problems, such that some children have comprehension deficits despite normal word reading skills (Cain, Oakhill, & Bryant, 2000;

Catts, Hogan, & Fey, 2003; Nation, 2005; Oakhill, Cain, & Bryant, 2003) and others have normal comprehension but selective word reading deficits (e.g., Leach et al., 2003).

This independence of word reading and comprehension skills was shown to have a genetic etiology by Keenan et al. (2006). They used multivariate genetic analyses to investigate the genetic and environmental contributions to reading comprehension that are shared with listening comprehension and word decoding. Keenan and colleagues found one significant genetic factor for word decoding that was shared with reading comprehension, and a second genetic factor for reading comprehension and listening comprehension, independent of word reading. This finding has now been replicated both in a longitudinal sample of Colorado twins (Betjemann et al., 2008) and in a study using both the Twins Early Development Study (TEDS) and the Western Reserve Reading Project (WRRP) samples of twins (Harlaar & Petrill, 2009).

In contrast, Byrne et al. (2007) reported that in first-graders tested with the Woodcock-Johnson Passage Comprehension subtest, the genetic influences involved in reading comprehension were almost completely overlapping with those of word decoding; there was not a significant separable factor for comprehension. One reason Byrne et al.'s results may differ from Keenan et al. (2006), Betjemann, et al. (2008), and Harlaar & Petrill (2009) is that his participants were so much younger – the average age was 7 years, as opposed to 10 to 13 years in the other studies. Many first-graders' word decoding skills are so limited that comprehension assessment at this age is not much more than assessing word identification.

Another critical difference between the studies was the specific test used to assess reading comprehension. A number of studies have now shown that reading comprehension tests vary considerably in how much individual differences in performance reflect differences in decoding versus oral comprehension skills (Cutting & Scarborough, 2006; Keenan et al., 2008; Nation & Snowling, 1997). Keenan et al. showed also that some tests vary in the relative contributions of decoding and oral language skills depending on the age and reading ability of the child, whereas other tests are more consistent across age.

All these differences raise the question of whether the particular tests used to assess reading comprehension could result in different outcomes for behavior genetic analyses. Specifically, if a reading comprehension test phenotypically is very sensitive to decoding ability, such that one would incorrectly answer comprehension questions if they make small errors on decoding, then the genetic covariation between reading comprehension and decoding could be very high, and a unique genetic or environmental influence for comprehension might not be observed. On the other hand, if the test is not as sensitive to decoding, so that more global comprehension is assessed, then separate influences might be observed. The goal of the current study is thus to examine the impact of the specific reading comprehension test on behavioral genetic analyses so that choices of assessment instruments may be guided by a theoretical understanding and not just the convenience of test administration and scoring. Furthermore, if we show that different tests claiming to measure the same construct actually have different patterns of genetic covariation, this could be important to understanding the failures to replicate that often occur in molecular genetic research (e.g., Abbott, 2008).

Current Study

The current study uses multivariate genetic analyses to investigate how the phenotypic differences between five tests of reading comprehension first reported by Keenan et al. (2008) impact our understanding of the genetic and environmental etiology of individual differences. Keenan et al. found that two tests (Woodcock-Johnson Passage Comprehension and Peabody Individual Achievement Test; Woodcock, McGrew, & Mather, 2001 and Dunn

& Markwardt, 1970, respectively) were most strongly associated with decoding, and three other tests (Gray Oral Reading Test, Qualitative Reading Inventory Questions, and Qualitative Reading Inventory Retell; Wiederholt & Bryant, 1992 and Leslie & Caldwell, 2001, respectively) were more associated with listening comprehension. We evaluate these same five reading comprehension tests, and confirm with a factor analysis on our sample that the tests indeed define two classes. To reduce the effects of test error on our analyses, we use latent traits for the two types of tests. The first latent trait for reading comprehension consists of the two measures that loaded most highly on decoding (RC-D), and the second latent trait for reading comprehension includes the three measures which load more strongly on listening comprehension (RC-LC).

One question we will examine is whether the amount of variance accounted for by genetic, shared environment, and nonshared environment factors differs for the RC-D and RC-LC reading comprehension tests. We also assess whether there is a difference in the patterns of the multivariate genetic and environmental paths across the two groups of tests. Will the patterns of the path loadings be different for RC-D tests than for RC-LC tests? More specifically, will we see unique genetic factors for both decoding and comprehension in the RC-D measures, or will there only be one genetic factor for these RC-D measures, reflecting that what is being assessed as comprehension in these tests is actually mostly decoding? Similarly, will the RC-LC measures load onto two separate genetic factors, or will there only be one common genetic factor for these measures? If both RC-D and RC-LC factors show significant independent genetic variance from word decoding, is there significant variation in the amount of independent genetic variance? And will the patterns of environmental influence differ across the tests so that we find separable environmental covariation on decoding and comprehension in the RC-D tests, or will the same environmental factors covary with both types of tests? Differences between tests would show the importance of qualifying genetic results in terms of the test used, and of using identical measures when attempting to replicate genetic results.

Method

Participants

All participants were tested as part of the Colorado Learning Disabilities Research Center (CLDRC; see DeFries et al. 1997; Olson, 2004). Twins were identified from school records in 27 different Colorado school districts. One group of twin pairs (affected pairs) was invited to participate if one or both twins were identified by school records or parent report to have a school history of reading problems. A comparison group of twin pairs with no history of reading problems was also invited to participate (control pairs). Twins included in the current analyses spoke English as their first language and had no uncorrected sensory deficits. Participating children were ages 8 – 18, with a mean age of 11.0. There were 322 twin pairs included: 117 MZ pairs, and 205 DZ pairs (101 same-sex DZ pairs, 105 opposite-sex DZ pairs)1. Zygosity of same-sex pairs was determined by a checklist of items based on the questionnaire by Nichols and Bilbro (1966). In ambiguous cases, zygosity is confirmed by analysis of blood or buccal samples. Of the 644 individual participants, 230 (36%) had a reported history of reading difficulty. The sample was 46% male and 54% female. Similar to the local population, approximately 77% of the overall CLDRC sample is Caucasian, 10% Hispanic, 2% African American, and other groups made up 1% or less.

[1]This current sample overlaps with the samples used in some previous studies referenced in this paper. A total of 50 participants in this sample (7.8%) were in the Betjemann et al. (2008) sample, 354 (55%) were in the Keenan et al. (2006) sample, and 422 (65%) were in the Keenan at al. (2008) sample.

Procedures and Measures—Each participant completed one full day of testing at the University of Colorado and a second full day of testing at the University of Denver. The test battery included the measures of Word Reading, Listening Comprehension, and Reading Comprehension listed below. Scores from most measures were normally distributed, but raw scores from three of the measures were transformed by squaring to correct for non-normality, as indicated below. For analysis, each measure was z-scored across all participants, adjusted for age, age-squared, and sex. Outlier datapoints were trimmed to +/− 4 standard deviations; a total of four data points were adjusted in the affected group (less than 0.01% of the data), and no data points were trimmed in the control group.

To confirm that these reading tests define two different types of tests, we repeated the factor analysis done by Keenan et al. (2008) on the data for the current sample. We performed an exploratory principal components factor analysis using oblique rotation, to allow for correlations between the factors. The variables included in the analysis were the five reading comprehension measures, a listening comprehension composite and word and nonword decoding measures. Table 1 presents the pattern matrix of factor loadings. Like Keenan et al., we too found separable factors within reading comprehension tests for decoding and listening comprehension. We also replicated their finding that the Woodcock-Johnson Passage Comprehension and Peabody Individual Achievement Test loaded more highly on decoding than comprehension, while the other tests loaded more highly on comprehension. It is this finding that justified the division of our measures into two separate latent traits for reading comprehension: RC-D defined by the Woodcock-Johnson Passage Comprehension and Peabody Individual Achievement Test, and RC-LC defined by the Gray Oral Reading Test, Qualitative Reading Inventory Questions, and Qualitative Reading Inventory Passage Retell. The correlations between all measures included in the latent variables are presented in Table 2.

RC-D Reading Comprehension

Woodcock-Johnson Passage Comprehension subtest (WJ-PC) from the Woodcock-Johnson Tests of Achievement-III (Woodcock, McGrew, & Mather, 2001). This test uses a cloze procedure; participants read short passages silently, and are asked to provide the missing word that completes the sentence. The mean of age-standardized scores in this sample was 101.2 (*SD* = 10.9). One-year retest reliability is .92.

Peabody Individual Achievement Test, Reading Comprehension subtest (PIAT; Dunn & Markwardt, 1970). Participants read sentences silently, and for each pick the one of four pictures that best represents the meaning of the sentence. The mean of standard scores in this sample was 106.4 (*SD* = 13.2). Test- retest reliability is .64.

RC-LC Reading Comprehension

Gray Oral Reading Test - 3 (GORT; Wiederholt & Bryant, 1992). Participants read passages of increasing difficulty aloud and answer multiple-choice questions which are read to them by the tester after each passage. The mean of standard scores in this sample was 10.8 (*SD* = 3.1). Test-retest reliability with alternate forms is .75.

Qualitative Reading Inventory - 3 (QRI; Leslie & Caldwell, 2001). Participants read longer passages aloud (approx. 350–750 words), and then completed the following two assessments (Test-retest reliability for instructional-level placement based on assessments is 78%, and standard error of measurement averages 15%):

Passage retellings—Participants were asked to retell as much of the passage as they could. Retellings were scored by the proportion of concepts on a checklist that were recalled.

Open-ended comprehension questions—Three literal and three inferential questions were asked per passage, scored as correct or incorrect. Raw scores were transformed to adjust for non-normality.

Listening Comprehension

Woodcock-Johnson Oral Comprehension subtest (WJ-OC) from the Woodcock-Johnson Tests of Achievement (Woodcock, McGrew, & Mather, 2001). Children listened to short passages of generally two sentences each, and are asked to provide one word that completes each passage. The mean of age-standardized scores in this sample was 106.2 (*SD* = 11.1). One-year retest reliability is .88.

Qualitative Reading Inventory – 3 (QRI; Leslie & Caldwell, 2001). Comparable to the reading task version, participants listened to longer passages, and then were asked to a) retell the passage and b) answer open-ended comprehension questions, which were transformed to adjust for non-normality.

The KNOW-IT Test (Barnes & Dennis, 1996; Barnes, Dennis, & Haefele-Kalvaitis, 1996). Participants learned a novel knowledge base and then listened to a story (approx. 830 words) which incorporated that knowledge. After the story they answered 18 literal and inferential comprehension questions that require integration of the new knowledge base with the story that they heard. Raw scores were transformed to adjust for non-normality.

Word Reading

Timed Word Recognition Test (TWRT; Olson, Forsberg, Wise, & Rack, 1994). Participants read aloud words that were presented on a computer screen in order of increasing difficulty. Only responses initiated within two seconds of stimulus onset were considered correct. Test-retest reliability is .93.

Peabody Individual Achievement Test, Word Recognition subtest (PIAT; Dunn & Markwardt, 1970). Children read single words of increasing difficulty aloud, and are scored for accuracy. The mean of standard scores in this sample was 104.6 (*SD* = 12.8). Test- retest reliability is .89.

Analyses

We estimated the proportion of the observed variance due to genetic effects (A or a^2), shared environmental effects (C or c^2), and nonshared environmental effects (E or e^2) for each individual trait. The multivariate Cholesky decomposition performed on the data partitions the phenotypic variance of traits into these genetic, shared environmental, and nonshared evironmental factors, as shown in the top half of Figure 1. It also allows us to investigate the genetic and environmental influence on the *covariance* between measures (e.g., Neale & Cardon, 1992;Plomin & DeFries, 1979). For example, the Cholesky decomposition can show the amount of genetic variance that is shared between word reading and reading comprehension measures, to indicate what proportion of the genetic effects related to word reading were also related to comprehension. Latent traits were used for word reading, listening comprehension, and reading comprehension in the models. The variables contributing to each latent trait can be seen in the lower boxes in Figure 1.

Univariate estimates of genetic and environmental components of variance

From the standardized path coefficients of the multivariate model, the univariate proportion of variance due to genetic and environmental influences in each of the three latent traits can be computed by summing the squared paths from the factors to each trait. For example, the heritability (a^2; i.e., the proportion of phenotypic variance attributable to genetic influences) of listening comprehension is the square of the path from A1 to listening comprehension, plus the square of the path from A2 to listening comprehension. Estimates of shared (c^2) and nonshared (e^2) environmental influences are obtained in an analogous fashion.

Multivariate path coefficients and correlations

The standardized path coefficients also indicate which factors are unique or common to different latent traits. Confidence intervals are computed for each path coefficient, to determine which paths contribute significantly to the model. We can also determine the degree to which the same genetic and environmental influences affect the different latent traits. For example, the genetic correlation is the extent to which common genes are influencing the two traits. Estimates of these genetic, shared environmental, and nonshared environmental correlations are also computed from the standardized paths.

Results

Univariate Genetic Results

Univariate heritability estimates are presented in Table 3. We observed significant genetic, shared environmental, and nonshared environmental influences on each of the four latent traits, as indicated by confidence intervals above zero. For reading comprehension in particular, we see in Table 3 that the a^2 estimates are nearly identical for both the RC-D and RC-LC latent traits (.60 and .66, respectively), with their confidence intervals almost completely overlapping. We further see that the c^2 and e^2 estimates are equally similar. Thus, the overall proportions of influence from genes and environment are almost identical for these two different types of reading comprehension measures.

Multivariate Genetic Results

Although the univariate additive genetic influence on the two types of reading comprehension measures appear to be about the same, this does not necessarily mean that the patterns of multivariate genetic influences contributing to those heritabilities must be the same. In Table 4, we present the standardized path coefficients from the multivariate Cholesky decomposition. We see that there is a first genetic factor that is significant for all four latent traits, as determined by the confidence intervals remaining above zero. This factor includes all of the genetic variance for word reading, which likely also includes variance for some amount of general intelligence and semantic knowledge, as suggested by the significant path from this factor to listening comprehension, as well. A second genetic path also obtained, independent of the variance accounted for by word reading, which is significant for listening comprehension and also significant for both measures of reading comprehension.

It is interesting that both types of tests require both word reading and listening comprehension to explain their variance. However, there are differences between the two types of reading comprehension tests in the strengths of those paths. The standardized path coefficient of .31 between listening comprehension and RC-D reading comprehension is much smaller than the path of .64 from listening comprehension to RC-LC reading comprehension; this is a significant difference as indicated by the fact that the 95% confidence intervals for these coefficients are not overlapping and that the two values cannot be equated without significant loss of fit to the model ($\Delta \chi^2 = 6.17$, $\Delta df = 1$, $p < .05$).

Similarly, the path from the word reading factor is much larger (.71) to RC-D reading comprehension than to RC-LC comprehension (.50). While their confidence intervals do overlap slightly, the two parameters cannot be equated without significant loss of fit to the model, indicating they are also significantly different ($\Delta \chi^2 = 23.35$, $\Delta df = 1$, $p < .001$). Overall, these results show that word reading shares a stronger genetic influence with RC-D reading comprehension tests than with RC-LC reading comprehension tests. Conversely, the listening comprehension factor shares a greater genetic influence with RC-LC reading comprehension than with RC-D reading comprehension.

This discrepancy between the two types of reading comprehension tests is further supported by the genetic correlations presented in Table 5. As indicated by non-overlapping confidence intervals, the genetic correlation between word reading and reading comprehension is significantly higher for RC-D tests (.92) than for RC-LC tests (.61). In addition, the genetic correlation of reading comprehension with listening comprehension is significantly higher for RC-LC tests (.96) than for RC-D tests (.70). This indicates that word reading shares a higher proportion of genetic influence with the RC-D reading comprehension, while listening comprehension shares a higher proportion of genetic influence with the RC-LC reading comprehension. However, it should be noted that the genetic correlation between RC-LC and RC-D is also very high (.88), suggesting that individual differences in these two measures are due substantially to the same genetic influences.

As for environmental influences on individual differences in reading comprehension, the analysis revealed only a single significant shared environmental factor and a single nonshared environmental factor, each common to all four latent traits. The standardized path coefficients for these environmental factors were extremely similar for both RC-D and RC-LC reading comprehension. So, not only are the same environmental influences affecting all components of reading, but they are also influencing reading comprehension to an equal extent regardless of how it is assessed.

Assessing Group Differences in Multivariate Genetic Analyses

Because Keenan et al. (2008) found that reading ability tended to be associated with the degree to which word decoding or listening comprehension accounted for performance especially on the RC-D measures of reading comprehension, we hypothesized that the genetic findings presented above could vary according to reading ability. To examine this possibility, the raw scores for each test were re-standardized within the affected and control groups (defined in the method section) for analysis, and a new multivariate Cholesky decomposition was performed. The variables and latent traits in the Cholesky decomposition were the same as presented above and shown in Figure 1, but separate solutions were computed for the affected and control groups. The solutions for the two groups were then equated to determine if this would result in a significant loss of fit compared to the un-equated model. The solutions for the affected and control groups could indeed be equated without loss of fit ($\Delta \chi^2 = 37.5$, $\Delta df = 30$, $p = .16$), indicating that there is not a significant difference in the pattern of genetic and environmental influences on these measures in children with reading problems compared to controls, in our sample.

Because both members of each MZ and DZ twin pair are included in the same group, the difference between the group means is assumed to be due primarily to shared environmental influences when the data are combined without regard to group membership. In contrast, when analyses of twin data are conducted within group, the etiology of the difference between the group means is not analyzed. Consequently, when the data were standardized within group, the proportions of the variance due to genetic influences (a^2) increased by .18 on average compared to the estimates in Table 3, and those due to shared environmental

influences (c^2) decreased by .20 and were no longer significantly greater than zero. Nevertheless, the genetic and environmental correlations between the measures were almost identical to those presented in Table 5 (results are available from the first author).

While some previous phenotypic studies have found differences in tests by the reading level of the participants (e.g., Keenan et al., 2008; Nation & Snowling, 1997), we did not find differences by reading ability here in the genetic results. Though this could be due to insufficient power to detect group differences with the current sample size, the pattern of results also did not suggest a trend of differences by reading ability.

Discussion

The current study used a twin study design to investigate how phenotypic differences between five tests of reading comprehension first reported by Keenan et al. (2008) impact our understanding of the etiology of individual differences in reading comprehension. The behavioral genetic analyses that can be performed on twin data reveal two main types of information. One is the relative influence of genes, shared environment, and nonshared environment on the behavior, in this case reading comprehension. The other is the extent to which the covariation between specific component factors is due to shared genes or shared environments.

When we examined the relative influence of genes, shared environment, and nonshared environment on individual differences in reading comprehension, we found that there was very little difference between the two types of reading comprehension tests that Keenan et al. (2008) found to be phenotypically so different. Regardless of whether reading comprehension is assessed by tests that load most highly on decoding skill or by tests that load most highly on listening comprehension, individual differences in these measures are due substantially to genetic influences, and the estimated percentages of genetic variance are nearly identical for the different types of tests.

In contrast, when we examine what factors underlie this very similar amount of genetic influence, we find striking differences. RC-D and RC-LC reading comprehension tests differ significantly in how they are impacted by the genetic factors associated with word decoding and listening comprehension. The coefficients for decoding and for listening comprehension paths were both significantly greater than zero for both RC-D and RC-LC tests; however, the strengths of the paths were quite different across the two types of reading comprehension measures. The path from the genetic factor for decoding was significantly greater to the RC-D tests than to the RC-LC tests; in contrast, the genetic path from the listening comprehension factor was significantly smaller to RC-D reading comprehension than it was to RC-LC reading comprehension. Thus, there are clear differences between these tests in how they covary with genetic factors associated with comprehension and decoding.

These results inform previous phenotypic results, providing evidence for the etiology of differences found among tests. Phenotypic test differences could have been due to either environmental influences, such as one test being more culturally dependent than another, or to genetic influences. The current results provide evidence for highly similar environmental influences associated with the two types of tests. The only significant differences we found across tests were in the genetic contributions. Our findings thus show that previously reported phenotypic differences between comprehension tests reflect cognitive differences that vary not as a function of different environmental influences but in their genetic basis. This finding has important implications for understanding late emergent reading disability - children who readily learn to decode words but who have difficulties understanding what they read (e.g., Catts, Hogan, & Adlof, 2005; Compton, Fuchs, Fuchs, Elleman, & Gilbert,

2008; Leach et al., 2003). As reading comprehension assessments shift from dependence on word decoding to dependence on listening comprehension, we see different genetic influences coming on line to influence performance. Moreover, detecting these children will be more likely when using RC-LC comprehension tests rather than RC-D tests.

The current results also may account for the apparent discrepancies in the literature as to whether reading comprehension has unique genetic components for comprehension and decoding, as found by Keenan et al. (2006), or a single genetic factor that is associated with both decoding and comprehension, as was found by Byrne et al. (2007). While differences across the ages of the participants may be responsible for some of this discrepancy, the current study indicates that at least part of the reason for the differences between studies may be differences in the tests used to assess reading comprehension. The Woodcock-Johnson Passage Comprehension test, used by Byrne et al., shares a much stronger genetic influence with decoding and weaker influence with listening comprehension, compared to RC-LC reading comprehension tests. Thus, when using this test with younger participants, whose reading comprehension depends even more on decoding, it is not surprising that Byrne et al. did not find separate genetic influence for comprehension unique to that of decoding, in contrast to the results of Keenan et al.

In conclusion, this study offers further evidence that not all reading comprehension tests are the same (e.g., Cutting & Scarborough, 2006; Keenan et al., 2008; Nation & Snowling, 1997) by finding differential genetic covariances with comprehension depending on the particular test used to assess it. Despite the fact that the overall proportion of genetic influence was the same across the different types of reading comprehension tests, we found further evidence that these tests are definitely not interchangeable. We have shown that the differences between tests in the relative importance of word decoding and comprehension skills are reflecting unique genetic influence. Thus we now have a genetic basis for encouraging reading comprehension assessment that goes beyond tests which mainly reflect word decoding ability. Different genes impact different neurological systems, and by demonstrating how the genetic covariation between word decoding and comprehension depends on the specific test used to assess reading comprehension, we have further reason for clinicians and researchers to attend to their choice of test. In doing so, they will be able to identify not only children who suffer from word decoding difficulties, but also those who suffer from comprehension deficits despite adequate word reading. Furthermore, the present findings may provide molecular genetic researchers with another avenue for understanding some of the failures to replicate which genes are involved in reading and reading disability (e.g., Abbott, 2008). Because the present study showed differential patterns of genetic covariance depending on the reading test used, it may be that when different genes are identified in molecular studies, they are reflecting differences in the cognitive processes tapped by the specific test used to assess the reading phenotype.

Acknowledgments

Initial work on this manuscript was done when R. Betjemann was at the Institute for Behavioral Genetics at the University of Colorado as a Postdoctoral trainee on NIMH training grant T32 MH016880-25. This research was also supported by a grant from NIH HD27802 to the Colorado Learning Disabilities Research Center. Preliminary analyses of these data were presented by J. Keenan at the Reading Comprehension: Theory to Practice workshop, April 2007, Oxford, U.K. and at The 1st Stavanger Reading and Writing Conference, September 2007, Stavanger, Norway, and the final analyses were presented at the Pacific Coast Research Conference, February 2009. We are grateful to all the participants and their families, and to all the testers and scorers.

References

Abbott A. The brains of the family. Nature. 2008; 454:154–157. [PubMed: 18615056]

Barnes, MA.; Dennis, M. Reading comprehension deficits arise from diverse sources: Evidence from readers with and without developmental brain pathology. In: Cornoldi, C.; Oakhill, J., editors. Reading Comprehension Difficulties: Proceses and Intgervention. Mahwah, NJ: Erlbaum; 1996. p. 251-278.

Barnes MA, Dennis M, Haefele-Kalvaitis J. The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. Journal of Experimental Child Psychology. 1996; 61:216–241. [PubMed: 8636665]

Betjemann RS, Willcutt EG, Olson RK, Keenan JM, DeFries JC, Wadsworth SJ. Word reading and reading comprehension: Stability, overlap and independence. Reading and Writing: An Interdisciplinary Journal. 2008; 29:539–558.

Byrne, B.; Olson, RK.; Wadsworth, S.; Corley, R.; Samuelsson, S.; Quain, P. Longitudinal twin study of literacy and language: The first three years. Paper presented at the meeting of the Society for the Scientific Study of Reading; Toronto, Canada. June 26, 2005; 2005.

Byrne B, Samuelsson S, Wadsworth S, Hulslander J, Corley R, DeFries JC, Quain P, Willcutt E, Olson RK. Longitudinal twin study of early literacy development: Preschool through Grade 1. Reading and Writing: An Interdisciplinary Journal. 2007; 20:77–102.

Byrne B, Wadsworth S, Corley R, Samuelsson S, Quain P, DeFries JC, Willcutt E, Olson RK. Longitudinal twin study of early literacy development: Preschool and kindergarten phases. Scientific Studies of Reading. 2005; 9:219–235.

Cain K, Oakhill J, Bryant. Investigating the causes of reading comprehension failure: A test of the phonological processing deficit hypothesis. Reading and Writing. 2000; 13:31–56.

Catts H, Fey M, Zhang X, Tomblin JB. Language basis of reading and reading disabilities: Evidence from a longitudinal study. Scientific Studies of Reading. 1999; 3:331–361.

Catts, HW.; Hogan, TP.; Adlof, SM. Developmental changes in reading and reading disabilities. In: Catts, HW.; Kamhi, AG., editors. The connections between language and reading disabilities. Mahwah, NJ: Erlbaum; 2005. p. 25-40.

Catts HW, Hogan TP, Fey ME. Subgrouping poor readers on the basis of individual differences in reading-related abilities. Journal of Learning Disabilities. 2003; 36(2):151–164. [PubMed: 15493430]

Compton DL, Fuchs D, Fuchs LS, Elleman AM, Gilbert JK. Tracking children who fly below the radar: Latent transition modeling of students with late-emerging reading disability. Learning and Individual Differences. 2008; 18:329–337.

Cutting LE, Scarborough HS. 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.

DeFries JC, Filipek PA, Fulker DW, Olson RK, Pennington BF, Smith SD. Colorado Learning Disabilities Research Center. Learning Disabilities. 1997; 8:7–19.

DeFries JC, Fulker DW, LaBuda MC. Evidence for a genetic aetiology in reading disability of twins. Nature. 1987; 329:537–539. [PubMed: 3657975]

Dunn, LM.; Markwardt, FC. Examiner's manual: Peabody individual achievement test. Circle Pines, Minnesota: American Guidance Service; 1970.

Fisher SE, DeFries JC. Developmental dyslexia: genetic dissection of a complex cognitive trait. Nature Reviews Neuroscience. 2002; 3:767–780.

Gayán J, Olson RK. Genetic and environmental influences on orthographic and phonological skills in children with reading disabilities. Developmental Neuropsychology. 2001; 20:487–511.

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]

Gough, PB.; Hoover, WA.; Peterson, CL. Some observations on a simple view of reading. In: Cornoldi, C.; Oakhill, J., editors. Reading Comprehension Difficulties: Processes and Intervention. Mahwah, N.J: Lawrence Erlbaum Associates; 1996. p. 1-13.

Gough PB, Tunmer WE. Decoding, reading, and reading disability. Remedial and Special Education. 1986; 7:6–10.

Harlaar N, Dale PS, Plomin R. From learning to read to reading to learn: Substantial and stable genetic influence. Child Development. 2007; 78:116–131. [PubMed: 17328696]

Harlaar, N.; Petrill, SA. A genetic basis for the Simple View of Reading?. Paper presented at the Pacific Coast Research Conference; San Diego, CA. 2009 February.

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 and Psychiatry. 2005; 46:373–384. [PubMed: 15819646]

Hoover WA, Gough PB. The simple view of reading. Reading & Writing: An Interdisciplinary Journal. 1990; 2:127–160.

Keenan JM, Betjemann RS, Wadsworth SJ, DeFries JC, Olson RK. Genetic and environmental influences on reading and listening comprehension. Journal of Research in Reading. 2006; 29:79–91.

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.

Leach JM, Scarborough HS, Rescorla L. Late-emerging reading disabilities. Journal of Educational Psychology. 2003; 95:211–224.

Leslie, L.; Caldwell, J. Qualitative Reading Inventory. Vol. 3. New York: Addison Wesley Longman, Inc; 2001.

Nation, K. Children's reading comprehension difficulties. In: Snowling, M.; Hulme, C., editors. The science of reading: A handbook. Oxford, England: Blackwell; 2005. p. 248-266.

Nation K, Snowling M. Assessing reading difficulties: the validity and utitlity of current measures of reading skill. British Journal of Educational Psychology. 1997; 67:359–370. [PubMed: 9376312]

Neale, MC.; Cardon, LR. Methodology for genetic studies of twins and families. Dordrecht, The Netherlands: Kluwer Academic Publishers; 1992.

Nichols RC, Bilbro WC. The diagnosis of twin zygosity. Acta geneticae medicae et gemellologiae. 1966; 16:265–275.

Oakhill J, Cain KE, Bryant PE. The dissociation of word reading and text comprehension: Evidence from component skills. Language and Cognitive Processes. 2003; 18:443–468.

Olson RK. SSSR, environment, and genes. Scientific Studies of Reading. 2004; 8(2):111–124.

Olson, R.; Forsberg, H.; Wise, B.; Rack, J. Measurement of word recognition, orthographic, and phonological skills. In: Lyon, GR., editor. Frames of reference for the assessment of learning disabilities: New views on measurement issues. Baltimore: Paul H. Brookes Publishing Co; 1994. p. 243-277.

Pennington, BF.; Olson, RK. Genetics of dyslexia. In: Snowling, M.; Hulme, C., editors. The Science of Reading: A Handbook. Oxford, England: Blackwells; 2005. p. 453-472.

Perfetti, CA. Reading ability. New York: Oxford University Press; 1985.

Petrill SA, Deater-Deckard K, Thompson LA, DeThorne L, Schatschneider C. Reading skills in early readers: Genetic and shared environmental influences. Journal of Learning Disabilities. 2006; 39(1):48–55. [PubMed: 16512082]

Petrill SA, Deater-Deckard K, Thompson LA, Schatschneider C, DeThorne LS. Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. Reading & Writing: An Interdisciplinary Journal. 2007; 20:127–146.

Plomin R, DeFries JC. Multivariate behavioural genetic analysis of twin data on scholastic abilities. Behavior Genetics. 1979; 9:505–517. [PubMed: 263636]

Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, P. Behavioral Genetics. 5. New York: Worth Publishers; 2008.

Stevenson J, Graham P, Fredman G, McLoughlin V. A twin study of genetic influences on reading and spelling ability and disability. Journal of Child Psychology and Psychiatry. 1987; 28:229–247. [PubMed: 3584294]

Tiu RD Jr, Wadsworth SJ, Olson RK, DeFries JC. Causal models of reading disability: A twin study. Twin Research. 2004; 7:275–283. [PubMed: 15193172]

Wadsworth SJ, DeFries JC, Olson RA, Willcutt EG. Colorado Longitudinal Twin Study of Reading Disability. Annals of Dyslexia. 2007; 57:139–160. [PubMed: 18060583]

Wiederholt, L.; Bryant, B. Examiner's manual: Gray Oral Reading Test. Vol. 3. Austin, TX: Pro-Ed; 1992.

Woodcock, RW.; McGrew, KS.; Mather, N. Woodcock-Johnson III tests of achievement. Itasca, IL: Riverside Publishers; 2001.

Figure 1.
Multivariate Cholesky Model using Latent Traits.

Table 1

Pattern Matrix Showing the Factor Loadings of the Reading Comprehension Tests (Bold), the Word and Nonword Decoding Composites, and the Listening Comprehension Composite

	Comprehension Factor	Decoding Factor
Listening Comprehension Composite	.86	−.02
GORT-3 Comprehension	.64	.11
QRI Reading – Questions	.79	−.04
QRI Reading – Passage Retell	.73	.00
PIAT Comprehension	.25	.70
W-J Passage Comprehension	.37	.62
Word Decoding Composite	.02	.94
Nonword Decoding	−.16	.99

Table 2

Correlation Matrix of the Individual Measures Included in the Latent Variables

Latent Variable	Measure	PIAT Word Rec.	TWRT	Know- It	W-J Oral Comp.	QRI - L Qs	QRI - L Retell	PIAT Comp.	W-J Passage Comp.	GORT Comp.	QRI - R Qs
Word Reading	PIAT Word Recognition	--									
	Timed Word Recognition Test	.85	--								
Listening Comp.	Barnes Know-It	.29	.29	--							
	W-J Oral Comprehension	.46	.45	.45	--						
	QRI Listen - Questions	.28	.27	.34	.36	--					
	QRI Listen – Passage Retell	.28	.27	.34	.29	.53	--				
RC-D	PIAT Comprehension	.75	.70	.36	.52	.31	.32	--			
	W-J Passage Comprehension	.74	.74	.41	.58	.34	.36	.73	--		
RC-LC	GORT Comprehension	.45	.42	.40	.46	.31	.28	.50	.56	--	
	QRI Reading – Questions	.40	.37	.36	.45	.46	.34	.44	.48	.38	--
	QRI Reading – Passage Retell	.35	.38	.36	.34	.28	.57	.43	.45	.31	.42

Note: All correlations are significant at the $p < .01$ level. RC-D = Reading Comprehension – Decoding, RC-LC = Reading Comprehension – Listening Comprehension

Table 3

Univariate Heritability Estimates Calculated from Multivariate Cholesky Decomposition with 95% Confidence Intervals in Parentheses.

	Heritability a^2	Shared Environment c^2	Nonshared Environment e^2
Word Reading	.68 (.50, .85)	.25 (.09, .41)	.07 (.03, .13)
Listening Comprehension	.55 (.32, .81)	.40 (.16, .61)	.04 (.01, .12)
Reading Comprehension (RC-D)	.60 (.40, .83)	.32 (.10, .50)	.08 (.03, .15)
Reading Comprehension (RC-LC)	.66 (.38, .87)	.23 (.06, .48)	.10 (.03, .22)

Table 4

Standardized Path Coefficients from Cholesky Decomposition

Genetic Path Coefficients	**A1**	**A2**	**A3**	**A4**
Word Reading	.82 (.71, .92)			
Listening Comprehension	.27 (.09, .45)	.69 (.53, .82)		
Reading Comp. (RC-D)	**.71 (.57, .84)**	**.31 (.19, .41)**	0 (−.25, .25)	
Reading Comp. (RC-LC)	**.50 (.31, .65)**	**.64 (.47, .74)**	0 (−.30, .30)	0 (−.27, .27)
Shared Environmental Paths	C1	C2	C3	C4
Word Reading	.50 (.29, .64)			
Listening Comprehension	.59 (.32, .76)	.24 (−.50, .50)		
Reading Comp. (RC-D)	.55 (.30, .70)	−.13 (−.25, .25)	0 (−.25, .25)	
Reading Comp. (RC-LC)	.45 (.19, .67)	.18 (−.41, .41)	0 (−.29, .29)	0 (−.28, .28)
Nonshared Environmental Paths	E1	E2	E3	E4
Word Reading	.27 (.19, .35)			
Listening Comprehension	.17 (.03, .31)	.11 (−.28, .28)		
Reading Comp. (RC-D)	.22 (.11, .34)	.18 (−.29, .29)	0 (−.23, .23)	
Reading Comp. (RC-LC)	.21 (.05, .39)	.24 (.−.41, .41)	0 (−.30, .30)	0 (−.19, .19)

Table 5

Genetic and Environmental Correlations

	Word Reading	Listening Comp.	R. Comp. (RC-D)	R. Comp. (RC-LC)
Genetic Correlations				
Word Reading	--			
Listening Comprehension	.36 (.14, .55)	--		
Reading Comp. (RC-D)	**.92 (.86, .97)**	**.70 (.52, .84)**	--	
Reading Comp. (RC-LC)	**.61 (.45, .74)**	**.96 (.87, .99)**	.88 (.76, .95)	--
Shared Environmental Correlations				
Word Reading	--			
Listening Comprehension	.93 (.62, 1.0)	--		
Reading Comp. (RC-D)	.98 (.86, 1.0)	.82 (.51, .99)	--	
Reading Comp. (RC-LC)	.93 (.59, 1.0)	1.0 (.78, 1.0)	.82 (.43, .99)	--
Nonshared Environmental Correlations				
Word Reading	--			
Listening Comprehension	.83 (.18, 1.0)	--		
Reading Comp. (RC-D)	.76 (.41, 1.0)	.99 (.55, 1.0)	--	
Reading Comp. (RC-LC)	.66 (.17, 1.0)	.97 (.47, 1.0)	.99 (.78, 1.0)	--