Etiology of the Stability of Reading Difficulties: The Longitudinal Twin Study of Reading Disabilities

Raven L. Astrom,[1,2] Sally J. Wadsworth,[1] and John C. DeFries[1]

[1] *Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado, United States of America*
[2] *Department of Psychology, University of Colorado, Boulder, Colorado, United States of America*

Results obtained from previous longitudinal studies of reading difficulties indicate that reading deficits are generally stable. However, little is known about the etiology of this stability. Thus, the primary objective of this first longitudinal twin study of reading difficulties is to provide an initial assessment of genetic and environmental influences on the stability of reading deficits. Data were analyzed from a sample of 56 twin pairs, 18 identical (monozygotic, MZ) and 38 fraternal (dizygotic, DZ), in which at least one member of each pair was classified as reading-disabled in the Colorado Learning Disabilities Research Center, and on whom follow-up data were available. The twins were tested at two time points (average age of 10.3 years at initial assessment and 16.1 years at follow-up). A composite measure of reading performance (PIAT Reading Recognition, Reading Comprehension and Spelling) was highly stable, with a stability correlation of .84. Data from the initial time point were first subjected to univariate DeFries-Fulker multiple regression analysis and the resulting estimate of the heritability of the group deficit (h^2_g) was .84 (± .26). When the initial and follow-up data were then fitted to a bivariate extension of the basic DF model, bivariate heritability was estimated at .65, indicating that common genetic influences account for approximately 75% of the stability between reading measures at the two time points.

The heritable nature of reading disabilities (RD) has been well established. Heritability estimates for deficits in reading performance range from .37 to .72 for subjects aged 7 to 20 years (e.g., DeFries & Alarcón, 1996; DeFries & Gillis, 1991; Harlaar et al., 2005; Stevenson et al., 1987). Moreover, results obtained from longitudinal studies indicate that reading deficits are generally stable (e.g., Satz et al., 1998), with stability correlations over intervals of 1 to 8 years ranging from .23 to .96 (e.g., Bast & Reitsma, 1998; DeFries, 1988; DeFries & Baker, 1983; Shaywitz et al., 1992; Spira et al., 2005; Wagner et al., 1997; Williams & McGee, 1996). However, because few previous longitudinal studies of RD have utilized genetically informative designs, little is known about the etiology of this stability.

The evidence that deficits in reading are both stable and heritable suggests that genetic influences may be largely continuous throughout development, that is, the genetic factors which are important in early reading development may also be important for later reading performance. To our best knowledge, no previous studies have assessed the etiology of the stability of reading *deficits;* however, a few studies have examined the etiology of stability of *individual differences* in reading performance. As an early first step in assessing the etiology of the stability of reading performance, DeFries and Baker (1983) tested 102 RD and control probands (i.e., 51 pairs matched for sex and age) in the Colorado Family Reading Study at average ages of 9.5 and 14.9 years. Reading performance was measured using Reading Recognition, Reading Comprehension, and Spelling subtests of the Peabody Individual Achievement Test, (PIAT; Dunn & Markwardt, 1970). Results of structural equation modeling indicated that for families of reading-disabled children, over 60% of the longitudinal stability was attributable to parental influences. However, as this was a family study, rather than a twin or adoption study, genetic effects were not distinguishable from shared family environmental effects.

A few subsequent studies have used twin and adoption data to assess the etiology of the stability of genetic and environmental influences on reading performance within the normal range. Recently, Harlaar et al. (2007) assessed the stability of genetic influences on reading achievement in participants of the Twins Early Development Study (TEDS), a longitudinal

Received 26 August, 2006; accepted 7 December, 2006.

Address for correspondence: Sally J. Wadsworth, Institute for Behavioral Genetics,University of Colorado, 447 UCB, Boulder, CO 80309, USA. E-mail: Sally.Wadsworth@Colorado.edu

study of twins ascertained from population records of twin births in England and Wales. The reading achievement of 4291 twin pairs was evaluated by teacher assessment at ages 7, 9, and 10 years, using a rating scale of general reading achievement based on UK National Curriculum (NC) achievement goals for literacy. In addition, at age 10, participants completed a web-based test at home, which included an adaptation of the reading comprehension subtest of the Peabody Individual Achievement Test — Revised (PIAT-R; Markwardt, 1997). Heritability estimates of .67, .65, and .57 were obtained for reading performance at 7, 9, and 10 years respectively. Results from this study confirm that individual differences in reading performance are stable (r = .59 – .63) and suggest that 68% to 77% of the phenotypic stability correlation is genetically mediated.

Similar results were recently obtained by Wadsworth et al. (2006) using data from participants in the Colorado Adoption Project (CAP), an ongoing longitudinal study examining genetic and environmental influences on behavioral development. Reading performance data (PIAT Reading Recognition subtest, Dunn & Markwardt, 1970) from adoptive and nonadoptive sibling pairs who participated in the CAP at ages 7, 12, and 16 years, were subjected to Cholesky decomposition analysis. Similar to the findings of Harlaar et al. (2007), stability correlations were substantial and ranged from .58 between ages 7 and 16 years to .71 between ages 12 and 16 years. Moreover, between 53% and 86% of these stability correlations were due to genetic influences, suggesting that those genetic factors influencing reading performance at age 7 are also operating at ages 12 and 16.

Although a few previous twin and adoption studies have assessed the etiology of stability of reading performance within the normal range, we know of no other studies that have examined the etiology of stability of reading *deficits* using genetically informative designs. Thus, the primary objective of this first longitudinal twin study of reading difficulties was to assess the genetic and environmental etiologies of the stability of reading deficits using data from twin pairs tested initially in the Colorado Learning Disabilities Research Center (CLDRC; DeFries et al., 1997), and retested 5 to 6 years later in the Longitudinal Twin Study of Reading Disabilities (LTSRD; Wadsworth et al., 2006). Based on previous evidence of genetic influence on the stability of reading performance within the normal range of scores, we hypothesized that genetic influences on reading difficulties are stable with largely the same genes influencing reading deficits at both time points.

Materials and Methods

Participants and Measures

The subjects were tested in the CLDRC between September 1996 and August 2000 and also participated in follow-up testing 5 to 6 years later in the LTSRD. Initially, CLDRC twin pairs were identified by personnel of 27 school districts in Colorado. If either member of the pair had a positive school history of reading problems, both twins were invited to participate in the study at the University of Colorado, Boulder, where they completed an extensive battery of psychometric tests, including measures of reading, language and perceptual processes, as well as the PIAT (Dunn & Markwardt, 1970), and the Wechsler Intelligence Scale for Children — Revised (WISC-R; Wechsler, 1974), or the Wechsler Adult Intelligence Scale — Revised (WAIS-R; Wechsler, 1981). In addition, a comparison sample of control twin pairs was also tested in which neither member of the pair had a school history of reading problems. Control twins were matched to the probands by age, gender, and school district. For the current analyses, a composite measure of reading performance was used based on age-standard scores from the Reading Recognition, Reading Comprehension, and Spelling subtests of the PIAT at initial testing, and the PIAT-R (Markwardt, 1989) at follow-up testing. A discriminant function score was computed for each subject, employing discriminant weights estimated from an analysis of PIAT data from an independent sample of 140 nontwin children with reading disabilities and 140 nontwin controls (DeFries, 1985).

Currently in the LTSRD, we are recontacting all subjects who participated in the CLDRC during the previous 5 to 6 years. As of May 31, 2006, 156 twin pairs and 46 siblings have participated in the follow-up study. The average age of the subjects was 10.3 years at initial assessment and 16.1 years at follow-up. Data from only those twin pairs meeting criteria for inclusion in the proband sample are included in the current analyses. Twin pairs were included in the proband sample if at least one member of the pair met the following criteria at initial assessment: (1) a positive school history of reading problems, (2) classification as affected by the discriminant score, (3) no serious behavioral or emotional problems, and (4) no uncorrected visual or auditory acuity deficits. Zygosity of the twin pairs was determined using selected items from the Nichols and Bilbro (1966) zygosity questionnaire, which has a reported accuracy of 95% for same-sex twin pairs. In cases which were questionable, blood or buccal samples were obtained and twin pairs were genotyped using polymorphic DNA markers. The analysis sample of 18 MZ and 38 DZ pairs (18 same-sex, 20 opposite-sex) meeting criteria for inclusion in the initial proband sample included 35 male probands and 37 female probands, a ratio of .95:1. The ethnic composition of the sample is approximately representative of that in the state of Colorado from which subjects were ascertained (i.e., 84% non-Hispanic Caucasian, 8% Hispanic, 5% African American, 2% Native American, and 1% Asian).

Analyses

Multiple Regression Analysis

While a comparison of MZ and DZ concordance rates facilitates a test for the genetic etiology of a dichotomous variable (e.g., the presence or absence of a psychiatric disorder), RD is diagnosed using quantitative measures with arbitrary cut-off points. Thus, this transformation of a continuous measure into a categorical variable such as RD versus normal, results in a loss of information regarding the continuum of variation in reading performance. Consequently, DeFries and Fulker (1985, 1988) proposed fitting a multiple regression model to continuous data from twin pairs in which at least one member of each pair has a deviant score on the variable of interest. This method fully utilizes the available information and also provides direct estimates of group heritability *(h^2_g)*, a standardized measure of the extent to which performance deficits are due to genetic influences. When analyzing data from probands who are selected because of deviant scores on a continuous variable such as reading performance, the differential regression of MZ and DZ co-twin scores toward the mean of the unselected population provides an appropriate test of genetic etiology (DeFries & Fulker, 1985). Because MZ twins are genetically identical and DZ twins share only half of their segregating genes on average, the scores of DZ co-twins should regress more toward the mean of the unselected population if the condition is heritable. Therefore, when the MZ and DZ proband means are approximately equal, a simple *t* test of the difference between the MZ and DZ co-twin means provides a test of genetic etiology. However, DeFries and Fulker (1985, 1988) proposed that a multiple regression analysis of such data, in which a co-twin's score is regressed on both the proband's score and the coefficient of relationship (DF analysis), facilitates a more flexible and statistically powerful test. Further, they demonstrated that a simple transformation of the data prior to multiple regression analysis yields a direct estimate of h^2_g, an index of the extent to which the observed proband deficit is heritable.

In the current study, reading composite data at initial assessment were fitted to the following basic multiple regression model:

$$C = B_1P + B_2R + A \qquad [1]$$

where C is the co-twin's score, P is the proband's score, R is the coefficient of relationship (R = 1.0 for MZ twins and .5 for DZ twins), and A represents the regression constant. The B_1 coefficient is the partial regression of the co-twin's score on the proband's score, a measure of the average MZ and DZ twin resemblance (DeFries & Fulker, 1985, 1988). The B_2 coefficient is the partial regression of the co-twin's score on the coefficient of relationship and equals twice the difference between the MZ and DZ co-twin means after covariance adjustment for any difference between MZ and DZ proband means. Consequently, the B_2 coefficient provides a direct test for genetic etiology. When the data are transformed prior to multiple-regression analysis (i.e., each score is expressed as a deviation from the mean of the unselected population and then divided by the difference between the proband and population means), B_2 directly estimates h^2_g. In the current study, the CLDRC control sample of 1264 subjects represents the unselected population. Because subjects were selected based on their initial reading composite scores, and not reselected at follow-up, only data from the initial assessment were fitted to the univariate DF model.

Etiology of Stability

To assess the heritable nature of the stability of reading deficits, a bivariate extension of the basic DF model was employed in which data from the initial and follow-up sessions were fitted to the following regression model:

$$C_y = B_1P_x + B_2R + A \qquad [2]$$

where C_y is the co-twin's follow-up composite reading score, P_x is the proband's initial composite reading score, R is the coefficient of relationship, and A is the regression constant. B_1 is the partial regression of the co-twin's follow-up reading score on the proband's initial reading score and is a measure of average MZ and DZ cross-variable twin resemblance. Thus, B_1 estimates the extent to which co-twin scores on the follow-up measure are related to proband scores on the initial measure across zygosity. B_2 is the partial regression of the co-twin's follow-up reading score on the coefficient of relationship. Because the data were transformed prior to multiple-regression analysis, the bivariate B_2 coefficient is a function of the square roots of the group heritabilities for reading performance at the two time points and the genetic correlation (r_G) between them (i.e., $h_{\text{initial}} \times h_{\text{follow-up}} \times r_G$; Light & DeFries, 1995). Therefore, B_2 provides an estimate of 'bivariate heritability', an index of the extent to which the proband reading deficit at the initial participation is due to genetic factors which also influence the reading deficit at follow-up. Further, the proportion of the phenotypic stability correlation (r_P) attributable to genetic influences can be obtained by dividing the B_2 estimate by r_P.

For the present study, data were analyzed from twin pairs in which at least one member of the pair met the criteria for reading disability at their initial participation, and in which both members of the pair underwent follow-up testing. Because truncate selection was employed (DeFries & Gillis, 1991), pairs in which both members were diagnosed as reading-disabled were double-entered for all regression analyses. This is analogous to the computation of probandwise concordance rates, in which both affected members of concordant pairs are included as probands. Standard error estimates and tests of significance were adjusted accordingly.

Table 1

Proband and Co-Twin Standardized Means (*M*) and Standard Deviations (*SD*) of Reading Composite Scores at Initial and Follow-up Testing[1]

	MZ			DZ		
	M		*SD*	*M*		*SD*
Initial						
Proband	–2.15	±	.84	–2.11	±	.82
Co-twin	–2.12	±	1.03	–1.03	±	1.41
Follow-up						
Proband	–2.30	±	.85	–2.24	±	.88
Co-twin	–1.94	±	1.36	–1.16	±	1.53

Note: [1]Initial scores have been standardized against the mean of 1264 control twins participating in the CLDRC; follow-up scores have been standardized against the mean of 93 control twins participating in the LTSRD.

Results

Table 1 presents the standardized MZ and DZ proband and co-twin mean reading performance score, at each time point. Scores at initial assessment were standardized against the mean of all 1264 control twins participating in the CLDRC, whereas follow-up scores were standardized against the mean of the 93 control twins tested at follow-up. The MZ and DZ proband scores are highly similar at both time points and are more than two standard deviations below the control twin mean, suggesting that the deficit of the probands is highly stable. In addition, at both time points there is a differential regression of the MZ and DZ co-twin scores towards the control mean. In the initial sample, MZ co-twins regress only 0.03 standard deviation units and DZ co-twins regress 1.08 standard deviation units toward the control mean. Similarly, for the follow-up sample, the MZ co-twins regress 0.36 standard deviation units, whereas DZ co-twins regress 1.08 standard deviation units toward the control mean. Corresponding transformed proband and co-twin means, wherein each score is expressed as a deviation from the mean of the control population and then divided by the difference between the proband and control means, are presented in Figure 1.

When the transformed proband and co-twin initial scores were fitted to the basic regression model (Equation 1), $B_2 = h_g^2 = .84 \pm .26$ ($p \leq .002$), indicating that the proband reading deficit in this subsample is due substantially to genetic influences. Further, when the transformed proband initial scores and co-twin follow-up scores were fitted to the bivariate model (Equation 2), $B_2 = .65 \pm .32$ ($p < .05$), suggesting that about two-thirds of the proband deficit in reading at initial assessment is due to genetic factors that also influence reading difficulties at follow-up. Moreover, the ratio of B_2 to the observed correlation (.84) between initial and follow-up scores suggests that common genetic influences account for approximately 75% (.65/.84 = .77) of the stability between reading difficulties at the initial and follow-up sessions.

Discussion

Although a few previous studies have shown that reading deficits are stable and heritable, the genetic and environmental etiologies of this stability have not been previously investigated. The goal of this first longitudinal twin study of RD was to assess the etiology of the stability of reading deficits at two time points using behavioral genetic methods. Accordingly, data from twin pairs first tested in the CLDRC, and again 5 to 6 years later in the LTSRD, were subjected to bivariate DF analysis (Light & DeFries, 1995).

In the current study, the reading composite scores were highly stable over the 5 to 6 year interval ($r_P = .84$), somewhat higher than the stability correlations reported by Harlaar et al. (2007) and Wadsworth et al. (2006). In addition, the reading deficit of the probands was remarkably stable, with proband means more than two standard deviations below those of the controls at each assessment. This is especially noteworthy given that two different test versions were administered at initial and follow-up sessions (i.e., the PIAT at initial assessment and the PIAT-R at the follow-up).

When composite reading performance data collected from twin pairs at their initial assessment were

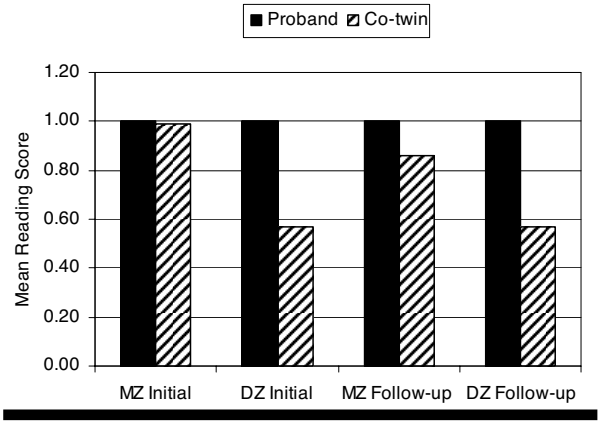

Figure1

Proband and co-twin transformed means of reading composite scores at initial and follow-up testing.

subjected to DF multiple regression analysis, an h^2_g estimate of .84 (± .26) was obtained. Although this h^2_g estimate is somewhat higher than that obtained from the full CLDRC proband sample of 283 MZ and 402 DZ pairs (h^2_g = .61 ± .06) for same-sex and opposite-sex pairs combined, the two estimates are not significantly different ($p > .30$).

When the bivariate extension of the multiple regression model (Equation 2) was fitted to proband scores at initial assessment and co-twin scores at follow-up, the resulting estimate of bivariate heritability was .65 ± .32 ($p < .05$), indicating that about two thirds of the proband deficit at initial assessment was due to genetic influences which also influence reading deficits at follow-up. Further, these results suggest that approximately 75% of the observed stability correlation is due to shared genetic influences. These findings are highly consistent with those of Harlaar et al. (2007), and Wadsworth et al. (2006), who found that 53% to 86% of the stability of reading performance in the normal range between ages 7 and 16 was due to genetic influences.

The preliminary results of this first longitudinal twin study of reading difficulties suggest that reading deficits are not only stable, but that this stability is due largely to heritable influences. Although this and other studies of reading disability suggest significant stability, about 30% of the variation in reading performance at follow-up is independent of initial reading performance in this sample, and 25% of the stability of reading performance is due to environmental influences. Thus, although our finding of significant genetic influence suggests that persistent reading deficits may be somewhat difficult to remediate, novel teaching/remediation methods may be highly effective.

It should be noted that the current sample of twin pairs meeting criteria for inclusion in the proband sample and on whom we have follow-up data is still very small, and results should be interpreted with caution. Nevertheless, these findings are highly consistent with those of previous longitudinal studies of reading performance within the normal range and both the univariate and bivariate h^2_g estimates are statistically significant. Moreover, follow-up testing of RD and control twin pairs continues in the LTSRD, thereby eventually facilitating more rigorous assessments of the etiology of difficulties in various reading-related cognitive processes and their stabilities.

Acknowledgments

The Colorado Learning Disabilities Research Center is supported by grant HD-27802 from the National Institute of Child Health and Human Development (NICHD). The Longitudinal Twin Study of Reading Disability is supported by grant DC-05190 from the National Institute on Deafness and other Communication Disorders (NIDCD). The continued cooperation of the many families participating in the CLCRC and LTSRD, as well as the work of the staff members of these projects is gratefully acknowledged.

References

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology, 34,* 1373–1399.

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

DeFries, J. C. (1988). Colorado Reading Project: Longitudinal analysis. *Annals of Dyslexia, 38,* 120–130.

DeFries, J. C., & Alarcón, M. (1996). Genetics of specific reading disability. *Mental Retardation and Developmental Disabilities Research Reviews, 2,* 39–47.

DeFries, J. C., & Baker, L. A. (1983). Parental contributions to longitudinal stability of cognitive measures in the Colorado Family Reading Study. *Child Development, 54*, 388–395.

DeFries, J. C., Filipek, P. A., Fulker, D. W., Olson, R. K., Pennington, B. F., Smith, S. D., & Wise, B. W. (1997). Colorado Learning Disabilities Research Center. *Learning Disabilities, 8,* 7–19.

DeFries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics, 15,* 467–473.

DeFries, J. C., & Fulker, D. W. (1988). Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologiae, 37,* 205–216.

DeFries, J. C., & Gillis, J. J. (1991). Etiology of reading deficits in learning disabilities: Quantitative genetic analysis. In J. E. Obrzut & G. W. Hynd (Eds.), *Neuropsychological foundations of learning disabilities: A handbook of issues, methods and practice* (pp. 29–47). Orlando, FL: Academic Press.

Dunn, L. M., & Markwardt, F. C. (1970). *Examiner's manual: Peabody Individual Achievement Test,* Circle Pines, MN: American Guidance Service.

Light, J. G., & DeFries, J. C. (1995). Comorbidity of reading and mathematics disabilities: Genetic and environmental etiologies. *Journal of Learning Disabilities, 28,* 96–106.

Harlaar, N., Dale, P. S., & Plomin, R. (2007). From learning to read to reading to learn: Substantial and stable genetic influence. *Child Development, 78,* 116–131.

Harlaar, N., Spinath, F. M., Dale, P. S., & Plomin, R. (2005). Genetic influences on early word recognition abilities and disabilities: A study of 7-year-old twins. *Journal of Child Psychology and Psychiatry, 46,* 373–384.

Markwardt, Jr., F. C. (1989). *Examiner's manual: Peabody Individual Achievement Test — Revised,* Circle Pines, MN: American Guidance Service.

Markwardt, Jr., F. C. (1997). *Examiner's manual: Peabody Individual Achievement Test — Revised (Normative Update)*. Circle Pines, MN: American Guidance Service.

Nichols, R. C., & Bilbro Jr., W. C. (1966). The diagnosis of twin zygosity. *Acta Geneticae Medicae et Gemellologiae, 16,* 265–275.

Satz, P., Buka, S. L., Lipsitt, L. P., & Seidman, L. (1998). The long-term prognosis of learning disabled children: A review of studies (1954–1993). In B.K. Shapiro, P. J. Accardo & A. J. Capute (Eds.), *Reading disability: A view of the spectrum*. Timonium, MD: York Press.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326,* 144–150.

Spira, E. G., Bracken, S. S., & Fischel, J. E., (2005). Predicting improvement after first-grade reading difficulties: The effects of oral language, emergent literacy, and behavioral skills. *Developmental Psychology, 41*, 225–234.

Stevenson, J., Graham, P., Fredman, G., & McLoughlin, V. (1987). A twin study of genetic influences on reading and spelling ability and disability. *Journal of Child Psychology and Psychiatry, 28,* 229–247.

Wadsworth, S. J., Corley, R. P., Plomin, R., Hewitt, J. K., & DeFries, J. C. (2006). Genetic and environment influences on continuity and change in reading achievement in the Colorado Adoption Project. In A. C. Huston & M. N. Ripke (Eds.), *Developmental contexts in middle childhood*. New York: Cambridge University Press.

Wadsworth, S. J., DeFries, J. C., Olson, R. K., & Willcutt, E. G. (2006). *Colorado Longitudinal Twin Study of Reading Disability*. Manuscript submitted for publication.

Wagner, R. K., Torgeson, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*, 468–479.

Wechsler, D. (1974). *Examiner's manual: Wechsler Intelligence Scale for Children — Revised.* New York: Psychological Corp.

Wechsler, D. (1981). *Examiner's manual: Wechsler Adult Intelligence Scale — Revised*. New York: The Psychological Corp.

Williams, S., & McGee, R. (1996). Reading in childhood and mental health in early adulthood. In J. H. Beitchman, N. J. Cohen, M. M. Konstantareas, & R. Tannock (Eds.), *Language, learning, and behavior disorders: Developmental, biological, and clinical perspectives*. Cambridge University Press: Cambridge.