Published in final edited form as:
J Learn Disabil. 2015 July ; 48(4): 433–445. doi:10.1177/0022219413508323.

Gender Differences in Reading Impairment and in the Identification of Impaired Readers: Results From a Large-Scale Study of At-Risk Readers

Jamie M. Quinn, MS[1] and **Richard K. Wagner, PhD**[1]
[1]Florida State University, Tallahassee, FL, USA

Abstract

Reading impairment is more common in males, but the magnitude and origin of this gender difference are debated. In a large-scale study of reading impairment among 491,103 beginning second-graders, gender differences increased with greater severity of reading impairment, peaking at a ratio of 2.4:1 for a broad measure of fluency and a ratio of 1.6:1 for a narrow measure of decoding. Results from three tests indicate that gender differences in reading impairment are attributable primarily to male vulnerability rather than ascertainment bias. Correspondence between identification as an impaired reader by our study criteria and school identification as learning disabled was poor overall and worse for girls: Only 1 out of 4 boys and 1 out of 7 girls identified as reading impaired in our study was school identified as learning disabled.

Keywords

gender differences; reading; dyslexia

Reading impairment is more common in males than in females. However, the magnitude and origin of this gender difference are in doubt. Ratios reported as the number of males to females with reading impairment range from a low of 1.2:1 to a high of 6.78:1 (e.g., Finucci & Childs, 1981; Miles, Haslum, & Wheeler, 1998; Rutter et al., 2004). Previous studies of gender differences in reading are presented in Table 1.

Differences in previously reported ratios may have arisen from differences in study characteristics (Hawke, Wadsworth, Olson, & DeFries, 2007; Limbrick, Wheldall, & Madelaine, 2008; Siegel & Smythe, 2005). For example, studies using a "discrepancy" definition between a student's intelligence quotient (IQ) and scores on a reading task have varied in the type of aptitude test used and the magnitude of the required discrepancy (Limbrick et al., 2008; Siegel, 1992). Different levels of severity of the reading problem required to count as having reading impairment across studies could account for variability

---

Corresponding Author: Jamie M. Quinn, Florida State University, 1107 West Call Street, Tallahassee, FL 32306-4301, USA. quinn@psy.fsu.edu.

Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

in reported gender differences if gender differences vary with severity. Sampling error associated with small studies could also produce considerable variability in apparent gender differences. With the exception of one large study conducted in Australia with data from approximately 1 million students (Wheldall & Limbrick, 2010), only a few studies have had sample sizes in the thousands (e.g., Flannery, Liederman, Daly, & Schultz, 2000; Jimenez et al., 2011; Miles et al., 1998; Rutter et al., 2004), and most studies have had small samples with fewer than 50 individuals with reading disability (e.g., Berger, Yule, & Rutter, 1975; Finucci & Childs, 1981; Jorm, Share, Matthews, & MacLean, 1986; Lewis, Hitch, & Walker, 1994; S. Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). Finally, studies of gender differences in reading disability have been published over a 60-year period. Conceptualizations of reading impairment and methods of identification have changed considerably over this time frame, which also could contribute to differences in the magnitude of reported gender differences.

Regarding origins of gender differences in reading impairment, two competing views exist. One view considers gender differences to reflect a methodological artifact arising from ascertainment bias. Ascertainment bias refers to males being more likely to be referred for evaluation than females with equivalent reading problems. Ascertainment bias could occur if males with reading problems are more likely than females to express frustration associated with their reading problem by exhibiting disruptive behavior, and if disruptive students are more likely to be noticed and referred for evaluation. Support for this view comes from reports that gender differences are larger for referred samples than for samples identified by applying objective criteria to all students (Mirkin, 1982; Prior, Sanson, Smart, & Oberklaid, 1995; S. Shaywitz et al., 1990) and from reports that females who are referred for evaluation tend to be older and to have more serious reading impairments than males (Vogel, 1990). The alternative view is that gender differences are real and represent a genuine male vulnerability that is not a product of the ascertainment bias (Liederman, Kantrowitz, & Flannery, 2005; Rutter et al., 2004). Both views could be partially correct if reading impairment is more common in males than in females and ascertainment bias exacerbates observed ratios of males to females with reading disability.

An issue of importance to both public policy and practice is the accuracy with which existing identification procedures used in public school and clinic settings identify individuals who meet objective criteria for having reading disability. Accuracy of existing identification procedures is largely unknown. Most studies of gender differences in reading impairment use objective criteria to identify samples that meet researcher-specified criteria for reading impairment, but data are not obtained on whether researcher-identified individuals are also identified by their schools or clinics. Having data on both researcher-based and school- or clinic-based identification for the same sample would permit analysis of the correspondence between researcher and school identification procedures. Two small-scale studies of gender differences in reading disability that did obtain these data used them to compare gender differences in school identification and research-based identification to study ascertainment bias, but did not use them to assess the correspondence between school- or clinic-based identification procedures (Mirkin, 1992; S. Shaywitz et al., 1990).

The Present Study

The goals of the present study were to (a) examine the magnitude of gender differences in reading impairment for a large sample of second grade students, (b) determine whether the magnitude of gender differences varies as a function of severity of the reading impairment or how reading impairment is operationally defined, (c) determine whether gender differences are attributable to genuine male vulnerability, ascertainment bias, or both, and (d) examine the correspondence between researcher- and school-based identification procedures for students with reading impairment.

Method

Participants

The sample consisted of 491,103 beginning second-grade students in elementary schools throughout the state of Florida. The data were obtained in the United States from the state of Florida's statewide Progress Monitoring and Reporting Network. This database was created to monitor the performance of students in the state's Reading First schools in concordance with the 2004 reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA). Reading First was a large, federally funded initiative that was designed to improve the reading performance of students in kindergarten through third grade who were at risk for reading problems based largely on poverty. Although primarily a database of Reading First schools, a small number of non–Reading First schools are also included in the database. The sample was 48.5% female and 51.5% male. The race and ethnicity composition was 41% White, 26.8% Black, 26.2% Hispanic, 1.7% Asian, 0.3% Native American/Pacific Islander, and 4.0% Mixed. Of students, 64% qualified for free or reduced lunch, and 21% had a history of receiving services for limited English proficiency.

Those students in the sample classified as learning disabled in the database were school identified in their respective elementary schools. In concordance with the reauthorization of IDEIA in 2004, the state of Florida adopted guidelines for identifying students with specific learning disabilities (SLD). Adopted in September 2004, rule 6A-6.0331 of the Florida Administrative Codes states for students in kindergarten through Grade 12 that the local school board is responsible for attempting to address and resolve a student's behavioral issue prior to referral for evaluation for a SLD. The following activities must be completed prior to referral: (a) parent-teacher conferences addressing areas of concern, (b) behavioral observations of the student by at least two people, (c) evaluation of medical and educational records, (d) review of attendance records, (e) screening for vision, speech, language, and hearing to rule out sensory deficits, and (f) attempt of a minimum of two interventions (e.g., increased instruction, change in schedule).

If, after completing the previous activities, the student requires formal evaluation, the referral process may begin. This process includes writing a formal request for evaluation of a student's eligibility for education services. Student evaluation begins after parental consent is given. Some of the requirements of evaluation include the following: Standardized tests used are validated for the intended usage of diagnosis and are given by trained personnel, and no single assessment will be used as the only criterion for eligibility. The primary criteria used to determine eligibility during the period of our study were a significant discrepancy between verbal IQ and achievement and a level of achievement that was impaired to the extent that participation in normal educational settings was not possible without special education services.

Measures

Two measures of reading were available, and a measure of vocabulary served as a proxy for verbal IQ for discrepancy-based operational definitions of reading impairment. Trained examiners administered each of the measures individually in concordance with administration standards. We selected measures obtained at the beginning of second grade. By second grade, students would have had a good opportunity to acquire the reading skills we were measuring.

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency (ORF; Good & Kaminski, 2002)—Students were given three passages to read aloud, each for 1 min. Scores were calculated as number of words read correctly in 1 min, averaged across the three passages. Alternate-form reliability exceeds .9, and validity coefficients for predicting reading comprehension exceed .7 (Good, Simmons, & Kame’enui, 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003a, 2003b).

DIBELS Nonsense Word Fluency (NWF; Good & Kaminski, 2002)—Participants were presented with 60 single-syllable pseudo-words with short vowel sounds and asked to read them aloud. The score was the number of correct pronunciations in a 1-min interval. Alternate-form reliability exceeds .8, and criterion-related validity coefficients with reading range from .4 to .9 (Good et al., 2001; Speece, Mills, Ritchey, & Hillman, 2003).

Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 1997)—This measure of receptive vocabulary requires pointing to a named picture. Alternate-form reliability exceeds .88, and criterion-related validity coefficients with reading range from .69 to .91 (Williams & Wang, 1997). PPVT scores were available for 206,485 of the 401,103 participants, approximately half of the sample.

Procedure

Magnitude and Variability of Gender Differences in Reading Impairment—To examine the magnitude of gender differences in reading impairment, we analyzed reading performance and gender data from a large sample of beginning second-grade students. Using the measures above, four operational definitions to identify students at risk for reading impairment were examined, each at four levels of severity of the reading problem (30th, 15th, 5th, and 3rd percentile ranks) to check for variability in the gender differences. Absolute low achievement in NWF was operationally defined as scoring at four percentile ranks. Absolute low achievement in ORF was defined identically. Using these scores to identify students who potentially have RD represents a “final benchmark” version of a response to intervention model of identification (Good et al., 2001). Under a final benchmark approach, all students are tested at a specific time point (e.g., the beginning of the school year), and those who score below the benchmark are considered to have a reading problem. To represent the traditional aptitude–achievement discrepancy definition of reading impairment, the two reading measures were regressed on the students' PPVT scores. The residuals from these regression analyses represent reading that is unexpected or discrepant from vocabulary performance, which served as a proxy for verbal IQ. The distribution of residuals was examined to identify scores corresponding to the same percentile ranks used above. These data were available for only approximately half of the sample (206,485 of the 401,103). Children who were school identified as having a learning disability were also analyzed for gender differences. Since these data come from a large, preexisting database, data on how these children were identified as learning disabled by their schools are not available.

Male Vulnerability or Ascertainment Bias?—Four sets of analyses were carried out to examine genuine male vulnerability and ascertainment bias as origins of observed gender differences in reading impairment.

The first set of analyses provided a direct estimate of the degree of male vulnerability by applying identification criteria based on operational definitions to the entire sample. Because ascertainment bias requires selective referral of individuals for identification, it is eliminated when identification criteria are applied to everyone in the sample as opposed to only individuals who had been referred for evaluation. Consequently, the magnitude of gender differences observed in an unreferred sample provides a direct estimate of the degree of male vulnerability.

Three additional sets of analyses examined ascertainment bias as an origin of observed gender differences in reading impairment. The first set of analyses involved comparing the gender ratio of members of our sample who met criteria for reading impairment to that of members of our sample who were school identified as learning disabled. Because we avoided ascertainment bias by applying our criteria to all members of the sample, and ascertainment bias was possible for school identification, comparing the gender ratio of members of our sample who met criteria for reading impairment to that of the school-identified members of our sample provided a direct estimate of the amount of ascertainment bias.

The second set of analyses of ascertainment bias involved determining whether school-identified females had more severe reading problems than school-identified males. A previous report of more severe reading problems for school-identified females than for males has been cited as evidence of ascertainment bias (Vogel, 1990). The results would follow this hypothesis if males are more likely to be noticed and referred for evaluation because of disruptive behavior associated with a mild or moderate reading problem, and females are only noticed when they exhibit a severe reading problem. This was analyzed using a *t* test to test for differences in mean scores.

The final set of analyses involved determining whether there was a closer correspondence between school identification and our identification for females than for males. Closer correspondence between school and our identification for females than for males could result from ascertainment bias if disruptive behavior played more of a role in school identification for males than females. The closer correspondence for females between both methods of identification would occur because our identification was based on reading performance alone.

Furthermore, we created quantile–quantile plots to compare the representation of males versus females over the entire distribution of scores. This allowed us to determine whether the greater representation of males in the lower tail of the distribution was better characterized as a gender difference localized in the lower tail or as a shift in the entire distribution of the female distribution to the right of the male distribution. A shift in the entire distribution would mean that females would be underrepresented in the lower tail but overrepresented in the upper tail.

Correspondence With School-Based Identification—To examine the accuracy of currently used school-identification procedures, we compared school-based identification status with our identification status based on applying our criteria to the entire sample. We used four statistics to quantify accuracy of school-based identification procedures. The first was the affected-status agreement statistic (ASAS; Waesche, Schatschneider, Maner, Ahmed, & Wagner, 2011; Wagner, Waesche, Schatschneider, Maner, & Ahmed; 2011). This statistic quantifies correspondence, or agreement between two methods as the ratio of the number of individuals identified by both methods over the number identified by either method. For reference purposes, we also calculated overall proportion of agreement, the relative observed agreement between the two identification methods (used to calculate Cohen’s kappa); sensitivity, the proportion of students identified as meeting an operational definition of reading impairment who also were school identified as learning disabled; and Cohen’s kappa, which measures the degree to which the two identification methods agree while taking chance agreement into consideration (Cohen, 1960).

Results

Descriptive Statistics

Descriptive statistics for both ORF and NWF included means, standard deviations, skewness, and kurtosis values are included in Table 2. Overall, the sample had a relatively normal distribution of scores for both NWF and for ORF. Males had slightly lower ORF scores and smaller standard deviations ($\mu = 46.03$, $SD = 28.82$) than females ($\mu = 53.06$, $SD = 30.08$). The means for NWF were similar for both genders (males: $\mu = 50.4$, $SD = 29.86$; females: $\mu = 50.33$, $SD = 28.09$), though the distribution for females was slightly more leptokurtic (1.73) than males (1.109).

Magnitude of Gender Differences in Reading Impairment Attributable to Male Vulnerability

Observed gender ratios, gender ratios corrected for differences in the total numbers of males and females in the sample, odds ratios and their 95% confidence intervals, and corrected odds ratios are presented in Table 3. For all 16 male-female ratios, males were affected significantly more than females based on the fact that none of the 95% confidence intervals for the corresponding odds ratios contained 1. Because these data come from applying research-based criteria to the entire sample, the corrected gender ratios provide unbiased estimates of male vulnerability for reading impairment without contamination from ascertainment bias. For all four operational definitions, gender ratios increased with worsening severity of the reading problem. For the relatively narrow measure of decoding pseudo-words, the corrected gender ratio peaked at approximately 1.6 to 1 for both low achievement and aptitude–achievement discrepancy operational definitions at the most severe level of reading impairment. Higher gender ratios were obtained for the broader measure of ORF, peaking at approximately 2.1:1 for the low-achievement and 2.4:1 for the aptitude–achievement discrepancy operational definitions at the most severe level of reading impairment. This relation between gender ratio and severity of the reading impairment can be seen clearly in Figure 1, which shows corrected gender ratios for the four operational definitions of reading impairment at the four levels of severity.

The quantile–quantile plots, presented in Figures 2 through 5 for absolute low ORF, absolute low NWF, discrepant ORF, and discrepant NWF, respectively, are largely linear with males falling below the line in the lower tail, especially for absolute low ORF. This corresponds to the results presented in Figure 1: Gender differences were greatest for this operational definition. The absence of an upward curve in the upper tail of the distribution indicates that the observed gender differences are largely localized to the lower tail of the distribution.

To examine potential variability in gender differences across ethnic or racial groups, the results were broken down by group. Males outnumbered females for all four definitions of reading impairment 23 out of 24 times (see Figure 6) in the fifth percentile. Males outnumbered females by 3:1 in Asian children, by 2.35:1 in Black children, by 2.8:1 in White children, by 2.5:1 in Native American children, by 1.67:1 in Hispanic children, and by 2.48:1 in mixed race children. The only definition and racial group for which results were inconclusive were for Hispanic children, where females outnumbered males 1.07:1 for the NWF achievement discrepancy definition at the fifth percentile.

Evidence of Ascertainment Bias

Comparison of Gender Ratios—Of the total sample of 491,103 students, 5% or 25,257 were school identified as learning disabled. The observed gender ratio for school-identified members of the sample was 2.25:1. After adjustment for slightly different numbers of males and females in the sample, the corrected gender ratio for school identification was 2.11:1. We compared this ratio to those obtained for the members of the sample we identified using the four operational definitions of reading impairment. We used the 5th percentile level of severity to match the 5% incidence rate for school identification. The corrected gender ratio of 2.11:1 was (a) marginally greater than that obtained for the low-achievement and discrepant-achievement in NWF (1.44:1 and 1.47:1, respectively) operational definitions, (b) comparable to that obtained for low-achievement in ORF (1.97:1) operational definition, and (c) marginally less than the discrepant achievement in ORF (2.24:1) operational definition. The fact that the corrected gender ratio for the school-identified members of the sample fell within the range of corrected gender ratios for the members of the sample we identified does not support the view that gender differences result primarily from ascertainment bias.

Comparison of Mean Performance—The second test of ascertainment bias was to determine whether school-identified females had more severe reading problems than school identified males. These results are presented in Table 4. For NWF, there were no significant differences between male and female mean scores ($t = -0.926$, $p > .05$). For the other three definitions (NWF residuals, ORF, and ORF residuals), male mean scores were significantly *lower* that female mean scores (p values < .001), which is opposite of what would be predicted on the basis of ascertainment bias.

Comparison of Correspondence Rates for Both Genders—The third test of ascertainment bias was to test whether there was a closer correspondence between school identification and research-based identification for females than for males. For these analyses, we carried out a classification analysis with children who were school identified as having a learning disability used to predict research-based identification as meeting one of the operational definitions of being at-risk for reading impairment. For present purposes, the critical comparison was between sensitivity, kappa, and affected status agreement values for females and males. These results are presented in Table 5. Female sensitivity scores range from .083 to .197, whereas male sensitivity scores ranged from .183 to .291. Kappa scores were lower, with female scores ranging from .063 to .178, and male scores ranging from .119 to .232. ASAS values were the lowest, ranging from .048 to .113 for females and .094 to .168 for males. In contrast to predictions based on ascertainment bias, the correspondence between school identification and research-based identification appears to be lower for females than for males.

Correspondence Between Study- and School-Based Identification

Overall Correspondence—The classification results presented in Table 5 address the accuracy of existing school-based identification procedures. The overall sensitivity values ranged from .14 to .26, with a median of .19. The values of kappa and the ASAS were much lower. The overall values of kappa ranged from .063 to .232, with a median of .142. The values of the ASAS ranged from .048 to .168, with a median of .102. These results suggest little correspondence between identification using the four operational definitions examined and school identification. When broken out by ethnic or racial groups, average sensitivity values ranged from .07 to .19 for Asian children, .13 to .23 for Black children, .15 to .22 for Hispanic children, .10 to .30 for Native American children, .14 to .35 for mixed race children, .16 to .36 for White children (see Table 6).

Gender Differences in Correspondence—Although the correspondence between identification using the study operational definitions and school identification was remarkably low overall, the correspondence rates were worse for girls than for boys. For example, in Table 5, sensitivity, kappa, and ASAS values were all lower for females than for males for the NWF IQ–achievement discrepancy definition.

Correspondence Between Alternative Study-Based Identification

Statistics representing the correspondence between alternative research-based identification procedures used are presented in Table 7. Overall, the kappa values ranged from low to moderate depending on the procedures examined. The correspondence, or overlap, between identification procedures was worst when calculated using the ASAS, ranging from .11 to .38, for an average correspondence of .24. The best correspondence rate found was that between the low-achievement definitions for both NWF and ORF using both the kappa statistic (.530) and the ASAS (.380) values. In general, there was relatively low correspondence between alternative identification procedures used.

General Discussion

In summary, the magnitude of gender differences based on a large sample of students increased with increasing severity of the reading problem, which is consistent with the results of several other studies (Hawke et al., 2007; Olson, 2009; Rutter et al., 2004; Wheldall & Limbrick, 2010). In the Rutter et al. (2004) study, gender ratios ranged from 1.35:1 to 2.76:1. In the Wheldall and Limbrick (2010) study of a sample of 1,133,988 students in either third or fifth grade, it was found that the gender ratios of children at risk for a reading problem ranged from 1.51:1 to 1.94:1 for third grade students in Band 1 (approximately the lowest 12th to 16th percentiles), and ranged from 1.17:1 to 1.32:1 for third grade students in Band 2 (approximately the lowest 16th to 25th percentiles). The results were largely comparable across ethnic and racial groups represented in the sample.

Three tests were carried out to determine whether referral bias explained the observed gender differences. None of these tests provided clear evidence for substantial referral bias. First, there were minimal differences in gender ratios for the school-identified sample versus research-identified samples. The ratio for the school identified subset, 2.26:1, was only marginally higher than the ratio seen for the research-identified samples using ORF and only moderately higher than the ratio using NWF. Second, if referral bias were operating, female mean performance should have been lower than males for each of the tests of reading. In three *t* tests to compare average scores, female mean performance was higher than male mean performance (*p*s < .001), and in the fourth there were no significant differences between genders. Third, there was no evidence of a lower correspondence between school-identified and research-identified samples for males than for females. In fact, the correspondence appeared to be lower for females.

What might account for gender differences in reading impairment beyond ascertainment bias cannot be determined from our study. Some evidence exists of early developmental effects of fetal hormone levels that might increase risk in males for reading impairment and other developmental disorders (Beech & Beauvois, 2005; Geshwind & Galaburda, 1985a, 1985b; Stevenson et al., 2007). Finding a male vulnerability for reading impairment is consistent with a male preponderance for other disorders such as attention-deficit/hyperactivity disorder and autism spectrum disorder (e.g., American Psychiatric Association, 2000; Fombonne, 2009; Gomez, Harvey, Quick, Scharer, & Harris, 1999; Graetz, Sawyer, & Baghurst, 2005; Ramtekkar, Reiersen, Todorov, & Todd, 2010; Szatmarti, Offord, & Boyle, 1989). Auyeung et al. (2009) found that fetal testosterone levels were related to quantitative variability in autistic traits in the general population. This is marked by individuals with autism spectrum disorder having reduced 2D:4D (second digit to fourth digit) ratios, a proxy marker for increased fetal testosterone levels (Manning, Baron-Cohen, Wheelwright, & Sanders, 2001).

Perhaps the most surprising and potentially troubling result from the study was how few students who met research-based criteria for reading impairment were identified as learning disabled by their schools. Fewer than 20% of students we identified as meeting an operational definition of reading impairment were also identified as learning disabled by their schools. Because the analyses are based on the 5th percentile level of severity so as to match the 5% incidence of school-identified learning disabilities, the low correspondence between study-based and school-based identification was not the result of different numbers of students identified as having a reading problem. Although some of the lack of correspondence can be attributed to the fact that school-identified learning disabilities include learning disabilities in areas other than reading such as mathematics, this does not explain the surprisingly low correspondence observed because learning disabilities in reading account for 80% of all learning disabilities (S. Shaywitz, 1998; S. Shaywitz, Morris, & Shaywitz, 2008). Another difference that might have reduced correspondence between researcher- and school-based identification is that we would have picked up students who were high in vocabulary and only average in reading whereas school-identification requires low performance in reading. This would have affected the two discrepancy-based criteria only, and indeed the correspondence rates appear to be somewhat lower for the discrepancy-based compared to the low-achievement-based operational definitions. However, the correspondence still remained low for the low-achievement-based operational definitions. Finally, the primary measures of reading used in the present study were brief, speeded assessments. Tests typically used for school identification tend to be more comprehension and nonspeeded assessments. This might also have contributed to lessened correspondence between research- and school-based identification.

Although each of the factors just described might have contributed, the most likely explanation for the lack of correspondence between research-based and school-based identification is that it simply reflects the general pattern of a lack of agreement among any operational definitions of reading disability that are based on a single criterion (e.g., IQ–achievement discrepancy, poor response to intervention). The reason for this general pattern of lack of agreement among operational definitions of reading disability that are based on a single criterion is well understood: When a cut point is placed on any continuous distribution, measurement error will result in unreliable identification of individuals near the cut point (Francis et al., 2005; Waesche et al., 2011; Wagner et al., 2011). The problem is that for low base rate conditions including reading disability, nearly all affected individuals are close to the cut point. A promising solution to this problem, which also would address the lack of correspondence between research-based and school-based identification, is to explore moving from operational definitions based on a single criterion to multivariate definitions based on multiple criteria. An example of this approach would be a hybrid model that combines criteria such as poor achievement, poor response to intervention, familial risk, and even gender in assessing the probability that an individual suffers from reading disability.

Previous studies of gender differences in reading impairment have varied in whether they excluded individuals with IQ scores below a cutoff value. No individuals were excluded from reported analyses. However, when analyses were run again after excluding individuals who obtained standard scores below 80, the overall pattern of results was unchanged.

Gender differences were slightly smaller for two operational definitions, unchanged for a third, and somewhat larger for the remaining operational definition.

Although we did not find evidence for ascertainment bias in the present study, the observed gender ratios were lower than previously reported ratios of 3 or 4 to 1, or even higher. It is likely that ascertainment bias did exist and explained part of the previously observed gender differences. Perhaps that increased attention paid to data-based decision making has reduced the role of ascertainment bias.

Limitations

One potential limitation of our study is that we had brief measures of reading and vocabulary compared to the typical assessments available in other studies and those used to identify students with reading impairment in schools and clinics. This is, of course, what allowed us to obtain such a large sample size. Furthermore, our usage of a receptive vocabulary measure (PPVT) as a proxy for verbal IQ in IQ–achievement discrepancy-based operational definitions is an additional limitation of our study. It would be important to replicate the present study using a measure of verbal IQ, which has been shown to be a better predictor of reading achievement in referred and nonreferred samples than Performance IQ (Greenblatt, Mattis, & Trad, 1990; Swanson, Carson, & Sasche-Lee, 1996; Vellutino, Scanlon, & Tanzman, 1991). It is important to note, however, that our results are consistent with other studies using more typical assessments (e.g., Rutter et al., 2004; Wheldall & Limbrick, 2010).

A final limitation of our study is the sample was obtained from a database that was largely, but not exclusively, comprising students served in Reading First schools in the state of Florida. Consequently, our sample had more children living in poverty and from minority backgrounds compared to the populations of the state or nation. Despite these potentially important differences in both tasks and populations, we are reassured by our results being remarkably consistent with those of other relatively recent studies that used different tasks and populations (Hawke et al., 2007; Olson, 2009; Rutter et al., 2004; Wheldall & Limbrick, 2010).

In summary, gender differences in reading impairment exist and are attributable largely to male vulnerability as opposed to ascertainment bias. The lack of correspondence between study-based and school-based identification is an important topic for future research.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Support for this research was provided by Grant P50 HD052120 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and a Predoctoral Interdisciplinary Fellowship, funded by the Institute of Education Sciences, U.S. Department of Education (R305B04074 & R305B090021).

References

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: Author; 2000. text revision

Auyeung B, Baron-Cohen S, Ashwin E, Knickmeyer R, Taylor K, Hackett G. Fetal testosterone and autistic traits. British Journal of Psychology. 2009; 100:1–22. [PubMed: 18547459]

Beech J, Beauvois M. Early experience of sex hormones as a predictor of reading, phonology, and auditory perception. Brain and Language. 2005; 96:49–58. [PubMed: 16203032]

Berger M, Yule W, Rutter M. Attainment and adjustment in two geographical areas: II. The prevalence of specific reading disability. British Journal of Psychiatry. 1975; 126:510–519. [PubMed: 1174768]

Chan DW, Ho CS, Tsang S, Lee S, Chung KKH. Prevalence, gender ratio and gender differences in reading-related cognitive abilities among Chinese children with dyslexia in Hong Kong. Educational Studies. 2007; 33:249–265.

Cohen J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement. 1960; 20:37–46.

Dunn, D.; Dunn, DM. Peabody Picture Vocabulary Test. Circle Pines, MN: American Guidance Service; 1997. Third Edition

Finucci, JM.; Childs, B. Are there really more dyslexic boys than girls?. In: Ansora, A.; Geschwind, A.; Galaburda, A.; Albert, M.; Gartrell, N., editors. Sex differences in dyslexia. Towson, MD: Orton Dyslexia Society; 1981. p. 1-9.

Flannery K, Liederman J, Daly L, Schultz J. Male prevalence for reading disability is found in large sample of Black and White children free from ascertainment bias. Journal of the International Neuropsychological Society. 2000; 6:433–442. [PubMed: 10902412]

Flynn JM, Rahbar MH. Prevalence of reading failure in boys compared with girls. Psychology in the Schools. 1994; 31:66–71.

Fombonne E. Epidemiology of pervasive developmental disorders. Pediatric Research. 2009; 65:591–598. [PubMed: 19218885]

Francis DJ, Fletcher JM, Stuebing KK, Lyon GR, Shaywitz BA, Shaywitz SE. Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. Journal of Learning Disabilities. 2005; 38:98–108. [PubMed: 15813593]

Geshwind N, Galaburda A. Biological mechanisms, associations, and pathology: II. A hypothesis and program for research. Archives of Neurology. 1985a; 42:521–552. [PubMed: 3890812]

Geshwind N, Galaburda A. Cerebral lateralization: Biological mechanisms, associations, and pathology: I. A hypothesis and program for research. Archives of Neurology. 1985b; 42:428–459. [PubMed: 3994562]

Gomez R, Harvey J, Quick C, Scharer I, Harris G. DSM-IV AD/HD: Confirmatory factor models, prevalence, and gender and age differences based on parent and teacher ratings of Australian primary school children. Journal of Child Psychology and Psychiatry and Allied Disciplines. 1999; 40:265–274.

Good, RH.; Kaminski, RA., editors. Dynamic Indicators of Basic Early Literacy Skills. 6. Eugene, OR: Institute for the Development of Educational Achievement; 2002. Retrieved from http://dibels.uoregon.edu/

Good R, Simmons D, Kame'enui E. The importance and decision making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. Scientific Studies of Reading. 2001; 5:257–288.

Graetz BW, Sawyer MG, Baghurst P. Gender differences among children with DSM-IV ADHD in Australia. Journal of the American Academy of Child and Adolescent Psychiatry. 2005; 44:159–168. [PubMed: 15689729]

Greenblatt E, Mattis S, Trad P. Nature and prevalence of learning disabilities in a child psychiatric population. Developmental Neuropsychology. 1990; 6:71–83.

Hawke J, Wadsworth S, Olson R, DeFries J. Etiology of reading difficulties as a function of gender and severity. Reading and Writing. 2007; 20:13–25.

Individuals with Disabilities Education Improvement Act, 20 U.S.C. §1415(k) (2004)

Jenkins J, Fuchs L, van den Broek P, Espin C, Deno S. Accuracy and fluency in list and context reading of skilled and RD groups: Absolute and relative performance levels. Learning Disabilities Research and Practice. 2003a; 18:237–245.

Jenkins J, Fuchs L, van den Broek P, Espin C, Deno S. Sources of individual differences in reading comprehension and reading fluency. Journal of Educational Psychology. 2003b; 95:719–729.

Jimenez JE, Garcia de LC, Siegel LS, O'Shanahan I, Garcia E, Rodriguez C. Gender ratio and cognitive profiles in dyslexia: A cross-national study. Reading and Writing. 2011; 24:729–747.

Jorm A, Share D, Matthews R, MacLean R. Behaviour problems in specific reading retarded and general reading backwards children: A longitudinal study. Journal of Child Psychology and Psychiatry. 1986; 27:33–43. [PubMed: 3949905]

Lewis C, Hitch G, Walker P. The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. Journal of Child Psychology. 1994; 35:283–292.

Liederman J, Kantrowitz L, Flannery K. Male vulnerability to reading disability is not likely to be a myth: A call for new data. Journal of Learning Disabilities. 2005; 38:109–129. [PubMed: 15813594]

Limbrick L, Wheldall K, Madelaine A. Gender ratios for reading disability: Are there really more boys than girls who are low-progress readers? Australian Journal of Learning Difficulties. 2008; 13:161–179.

Manning JT, Baron-Cohen S, Wheelwright S, Sanders G. The 2nd to 4th digit ratio and autism. Developmental Medicine and Child Neurology. 2001; 43:160–164. [PubMed: 11263685]

Miles TR, Haslum MN, Wheeler TJ. Gender ratio in dyslexia. Annals of Dyslexia. 1998; 48:27–55.

Mirkin, P. Direct and repeated measurement of academic skills: An alternative to traditional screening, referral, and identification of learning disabled students. Washington, DC: Office of Special Education and Rehabilitative Services; 1982. (Rep. No. IRLD-RR-75). Retrieved from ERIC database (ED224191)

Olson R. Dyslexia: Nature and nurture. Dyslexia. 2009; 8:143–159. [PubMed: 12222731]

Pennington B, Gilger J, Olson R, DeFries J. The external validity of age-versus IQ-discrepancy definitions of reading disability: Lessons from a twin study. Journal of Learning Disabilities. 1992; 25:562–573. [PubMed: 1431540]

Prior M, Sanson A, Smart D, Oberklaid F. Reading disability in an Australian community sample. Australian Journal of Psychology. 1995; 47:32–37.

Ramtekkar UP, Reiersen AM, Todorov AA, Todd RD. Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. Journal of the American Academy of Child and Adolescent Psychiatry. 2010; 49:217–228. [PubMed: 20410711]

Rutter M, Caspi A, Fergusson D, Horwood JL, Goodman R, Maughan B, Carroll J, et al. Sex differences in developmental reading disability: New findings from four epidemiological studies. Journal of the American Medical Association. 2004; 291:2007–2012. [PubMed: 15113820]

Shaywitz B, Shaywitz S, Pugh K, Constable R, Skudlarski P, Fulbright R, Gore JC, et al. Sex differences in functional organization of the brain for language. Nature. 1995; 373:607–609. [PubMed: 7854416]

Shaywitz S. Dyslexia. New England Journal of Medicine. 1998; 338:307–312. [PubMed: 9445412]

Shaywitz SE, Morris R, Shaywitz BA. The education of dyslexic children from childhood to young adulthood. Annual Review of Psychology. 2008; 59:451–475.

Shaywitz SE, Shaywitz B, Fletcher J, Escobar MD. Prevalence of reading disability in boys and girls. Journal of the American Medical Association. 1990; 264:998–1002. [PubMed: 2376893]

Siegel LS. An evaluation of the discrepancy definition of dyslexia. Journal of Learning Disabilities. 1992; 25:618–629. [PubMed: 1460383]

Siegel L, Smythe I. Reflections on research on reading disability with special attention to gender issues. Journal of Learning Disabilities. 2005; 38:473–477. [PubMed: 16329447]

Speece D, Mills C, Ritchey K, Hillman E. Initial evidence that letter fluency tasks are valid indicators of early reading skill. Journal of Special Education. 2003; 36:223–233.

Stevenson J, Everson P, Williams D, Hipskind G, Grimes M, Mahoney E. Attention deficit/hyperactivity disorder (ADHD) symptoms and digit ratios in a college sample. American Journal of Human Biology. 2007; 19:41–50. [PubMed: 17160985]

Swanson HL, Carson C, Sasche-Lee C. A selective synthesis of intervention research for students with learning disabilities. School Psychology Review. 1996; 25:370–391.

Szatmarti P, Offord DR, Boyle MH. Ontario child health study: Prevalence of attention deficit disorder with hyperactivity. Journal of Child Psychology and Psychiatry. 1989; 30:219–230. [PubMed: 2708462]

Vellutino F, Scanlon D, Tanzman M. Bridging the gap between cognitive and neuropsychological conceptualizations of reading disabilities. Learning and Individual Differences. 1991; 3:181–203.

Vogel SA. Gender differences in intelligence, language, visual-motor abilities, and academic achievement in students with learning disabilities: A review of the literature. Journal of Learning Disabilities. 1990; 23:44–52. [PubMed: 2404079]

Waesche JB, Schatschneider C, Maner J, Ahmed Y, Wagner R. Examining agreement and longitudinal stability among traditional and RTI-based definitions of reading disability using the affected-status agreement statistic. Journal of Learning Disabilities. 2011; 44:296–307. [PubMed: 21252372]

Wagner, RK.; Waesche, JB.; Schatschneider, C.; Maner, JK.; Ahmed, Y. Using response to intervention for identification and classification. In: McCardle, P.; Lee, JR.; Miller, B.; Tzeng, O., editors. Dyslexia across languages: Orthography and the brain-gene-behavior link. Baltimore, MD: Brookes; 2011. p. 202-213.

Wheldall K, Limbrick L. Do more boys than girls have reading problems? Journal of Learning Disabilities. 2010; 43:418–429. [PubMed: 20375293]

Williams, KT.; Wang, JJ. Technical references to the Peabody Picture Vocabulary Test–Third Edition (PPVT-III). Circle Pines, MN: American Guidance Services; 1997.

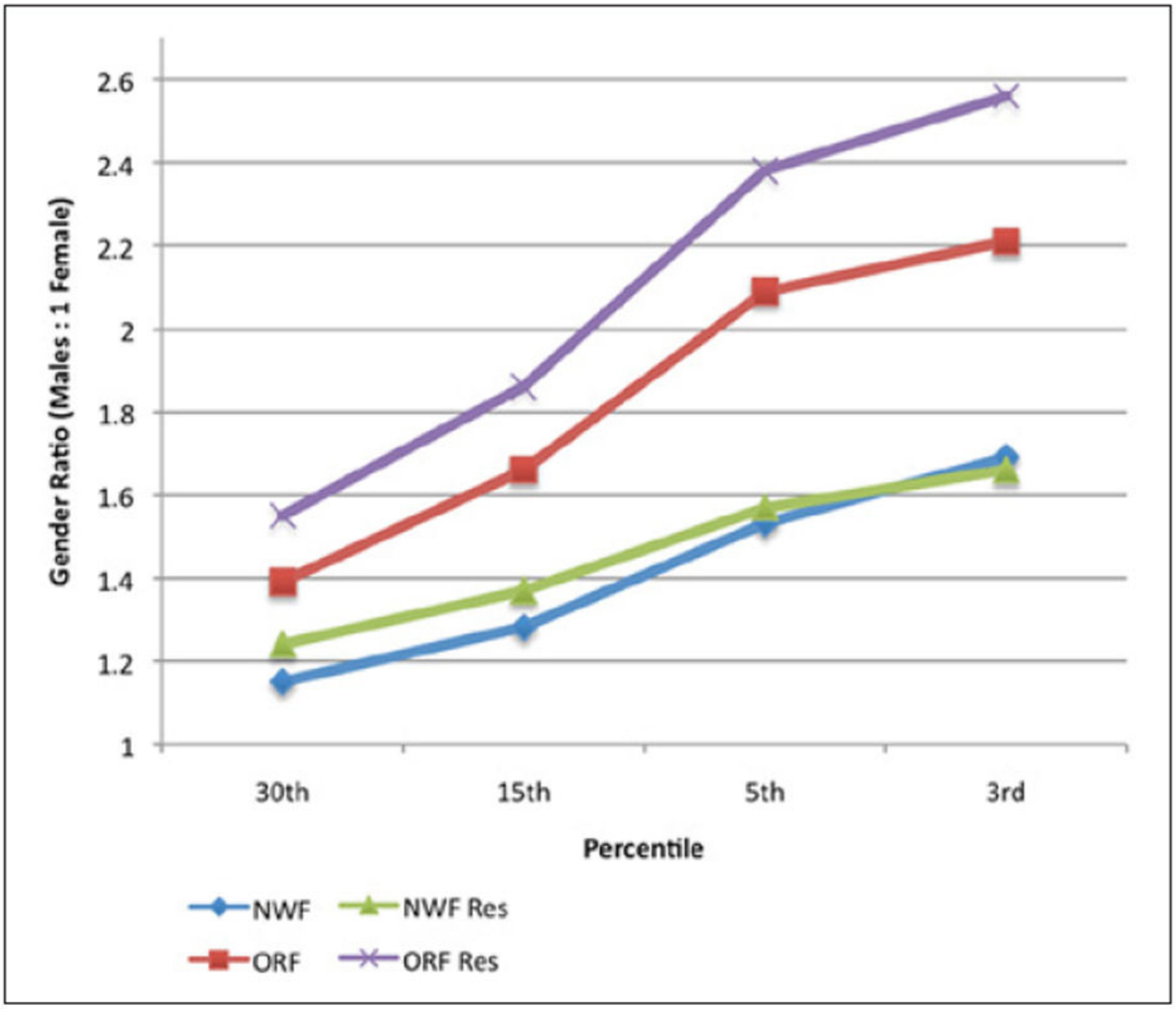

Figure 1.

Gender ratios by definition and severity.

Note: This figure illustrates the corrected gender ratios as a function of the operational definition of reading impairment used and the severity of the reading problem.

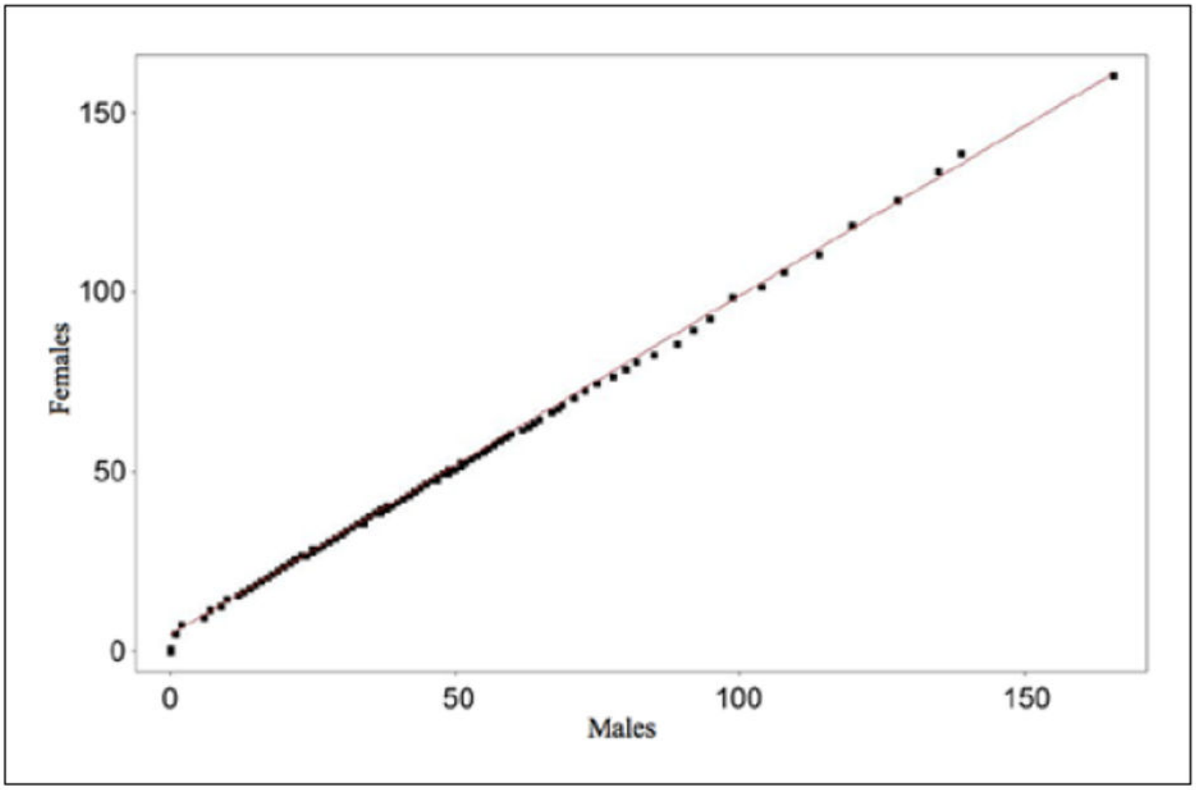

Figure 2.

Quantile–quantile plot for the ORF low achievement definition.

Note: This figure illustrates a quantile–quantile plot for the scores on the *Dynamic Indicators of Basic Early Literacy Skills* Oral Reading Fluency (ORF) task using the absolute low definition.

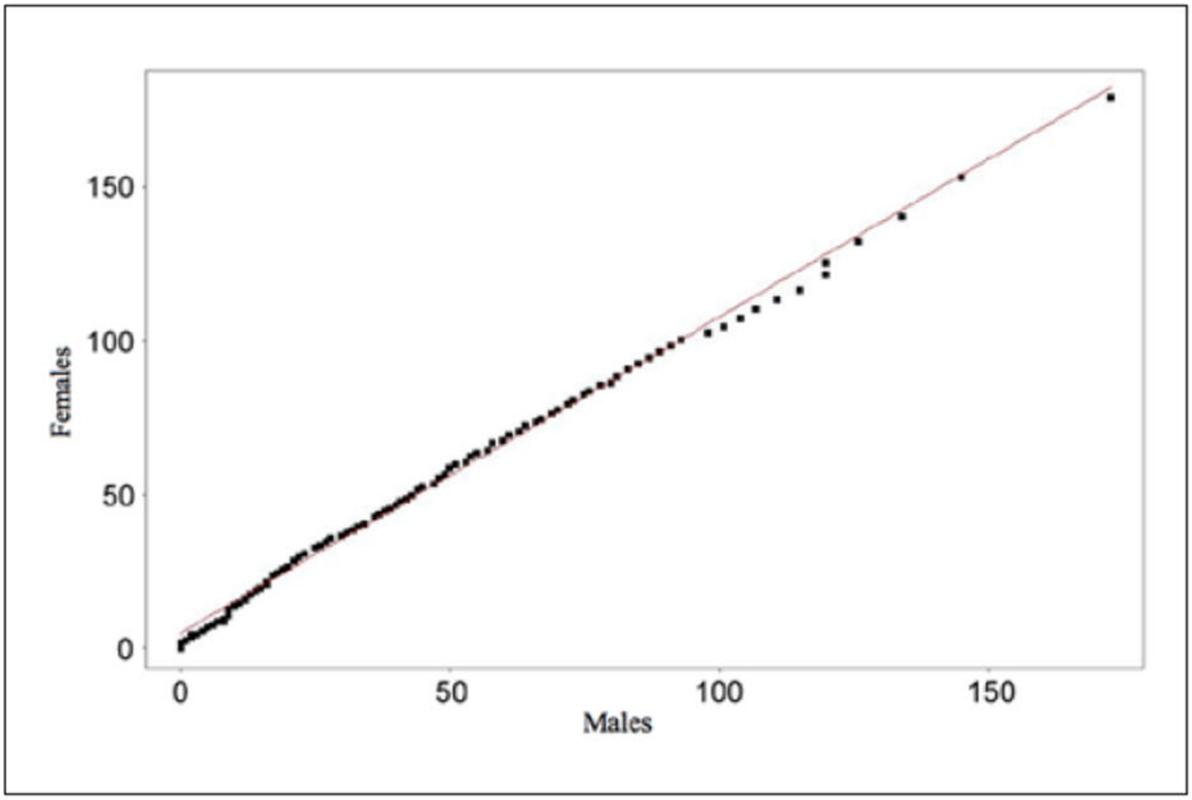

Figure 3.

Quantile–quantile plot for the NWF low achievement definition.

Note: This figure illustrates a quantile–quantile plot for the scores on the *Dynamic Indicators of Basic Early Literacy Skills* Nonsense Word Fluency (NWF) task using the absolute low definition.

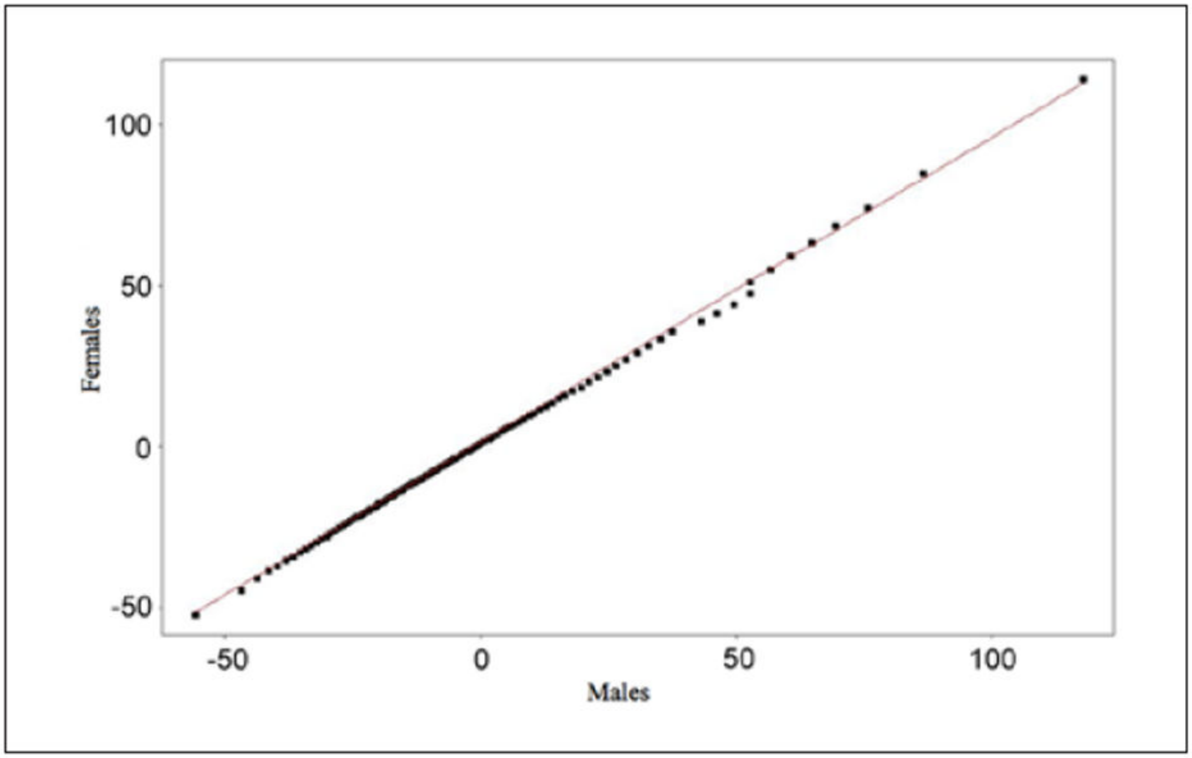

Figure 4.

Quantile–quantile plot for the ORF discrepancy definition.

Note: This figure illustrates a quantile–quantile plot of the residual scores on the *Dynamic Indicators of Basic Early Literacy Skills* Oral Reading Fluency (ORF) task used to calculate the IQ–achievement discrepancy definition of reading impairment.

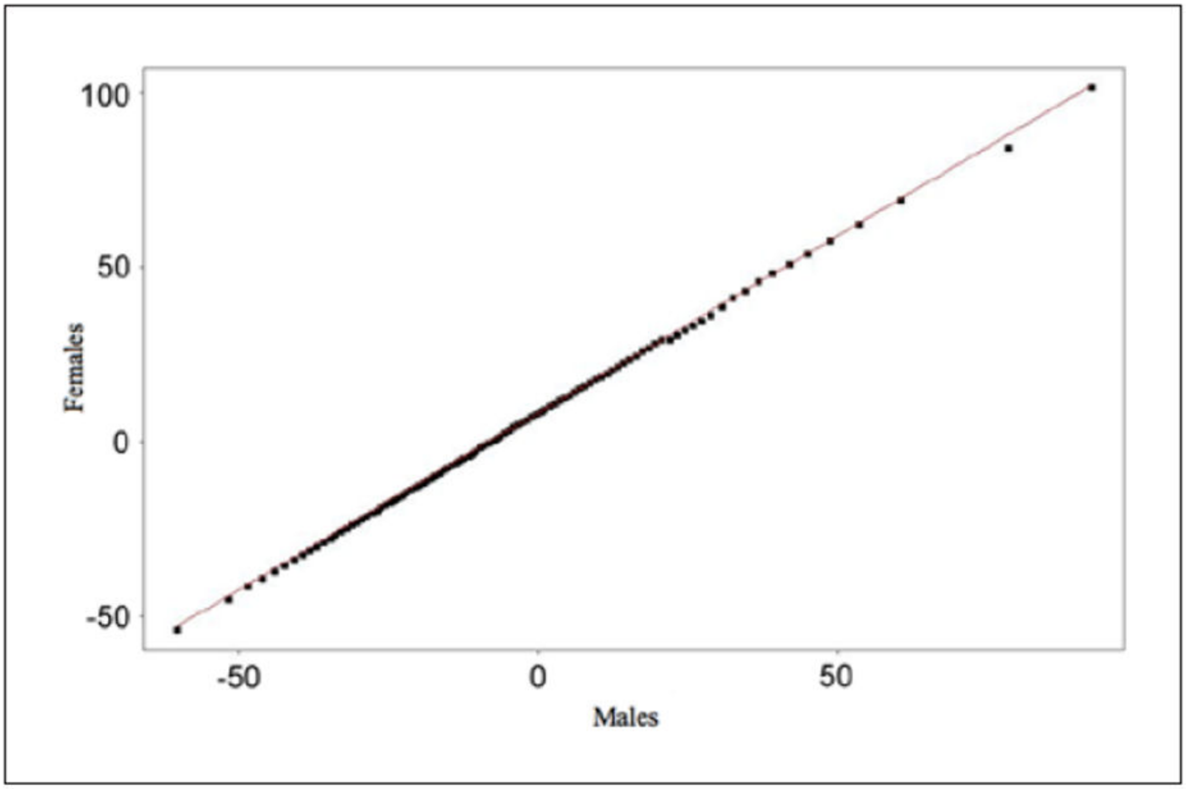

Figure 5.

Quantile–quantile plot for the NWF discrepancy definition.

Note: This figure illustrates a quantile–quantile plot of the residual scores on the *Dynamic Indicators of Basic Early Literacy Skills* Nonsense Word Fluency (NWF) task used to calculate the IQ–achievement discrepancy definition of reading impairment.

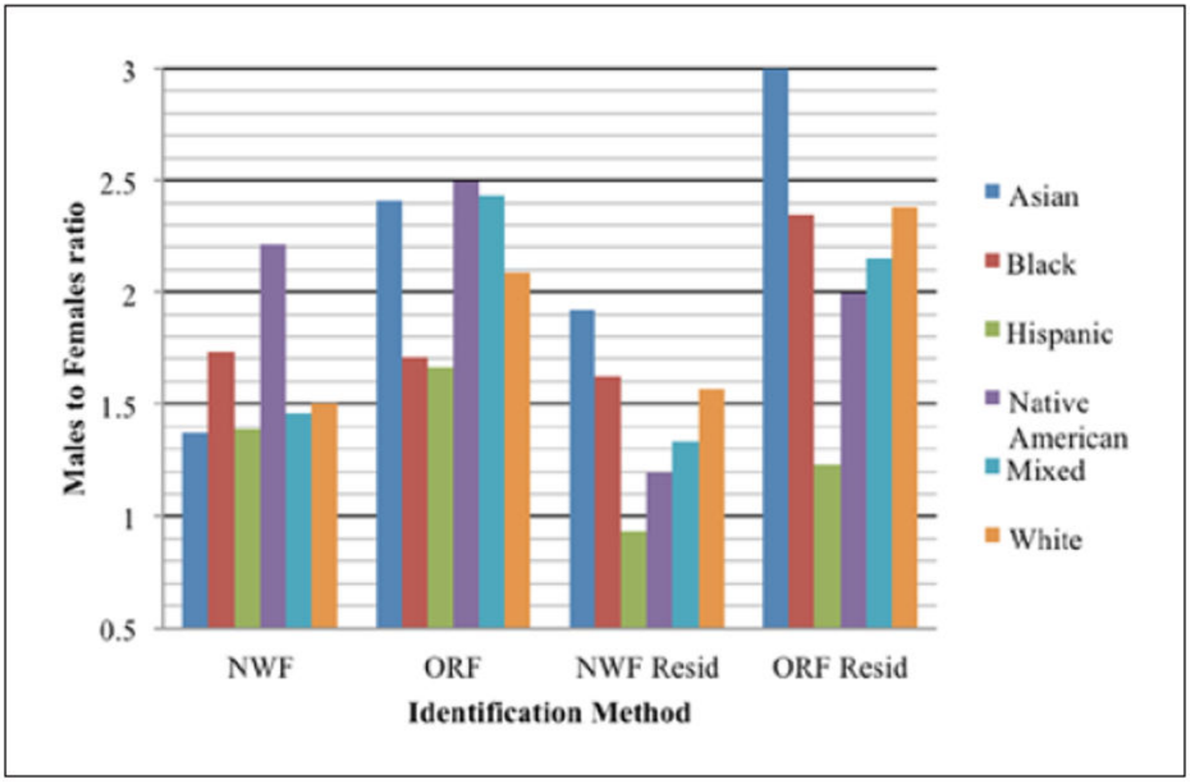

Figure 6.

Gender ratios by ethnicity and identification method.

Note: This figure illustrates that even when broken down in to ethnic groups, males outnumber females at the fifth percentile of the four researcher-based methods of identification for risk of reading impairment 23 out of 24 times.

Table 1

Results From Previous Studies Looking at Gender Differences in Reading Impairment.

		Criteria		Gender Ratio	
Citation	**Sample**	**Inclusion/Exclusion**	**RD Diagnosis**	**NRD**	**RD**
Berger et al. (1975) Sample 1	Isle of Wight *N* = 1,142	Age: 10 years IQ cutoff: average IQ on WISC Criteria: Children from the 1964 Isle of Wight survey	Whole sample given SRA; those with >5 or <10 randomly selected for individual testing. Individual test scores for accuracy or comprehension scales of Neale ≥2 standard errors of prediction below IQ or 30 months or more below age-expected score	0.94	2.67
Berger et al. (1975) Sample 2	Inner London Boroughs *N* = 1,660	Age: 10 years IQ cutoff: Average IQ WISC Criteria: Children from the 1970 Inner London Boroughs Survey; at least one parent British or Irish	Whole sample given SRA; those with >10 or <15 randomly selected for individual testing. Individual test scores for accuracy or comprehension scales of Neale ≥2 standard errors of prediction below IQ or 30 months or more below age-expected score	N/A	3.00
Chan et al. (2007)	Hong Kong, *N* = 702	Age: 6–10.5 years IQ cutoff: 85 on the Hong Kong WISC Criteria: Excluded those children who recently moved from Mainland China to Hong Kong, slow learners, and the mentally handicapped	HKT-SpLD composite score of 7 or below on the literacy test domain and on one of the cognitive test domains	N/A	1.60
Finucci & Childs (1981)	Two Baltimore parochial and public schools *N* = 241	Age: 6–13 years IQ cutoff: 95 on school IQ test Criteria: Parental consent	Combine reading/spelling quotient of ≤ .8 on GORT and WRAT scores relative to IQ, grade, and age	N/A	1.2–1.9^{t}
Flannery et al. (2000)	National Collaborative Perinatal Project (NCPP) *N* = 32,223	Age: 6–7 years IQ cutoff: 80 WISC Criteria: English primary language, no child blind, deaf, or with severe behavioral problems per NCPP behavior checklist	WRAT reading score 1.5 standard errors of prediction below expectation based on IQ and grade	0.97	1.91–2.89
Flynn & Rahbar (1994)	Gunderson Reading Study: 13 Midwestern public and parochial school districts *N* = 708	Grade 3 IQ cutoff: none Criteria: children present at 1987–1988 kindergarten screening	Severe reading disabilities: scores ≤ 10% on ITBS, CAT, and SAT	1.01	1.3^{t}
Jimenez et al. (2011)	Cross-national comparison of Spanish and Guatemalan children, *N* = 1048; *N* = 557	Age: 7 years, 9 months to 12 years, 8 months IQ cutoff: IQ > 75 Criteria: selected from state and private schools in the country of origin; all	Percentile score below 25 on accuracy and pseudo-word reading from naming task of SICOLE-R, or a percentile above 75 on reading time on pseudo-word or word reading from naming task for SICOLE-R	1.5 (Spain) 1.3 (Guatemala)	1.48t (Spain) 1.44^{t} (Guatemala)

		Criteria		Gender Ratio	
Citation	**Sample**	**Inclusion/Exclusion**	**RD Diagnosis**	**NRD**	**RD**
		spoke Spanish as native language; no sensory/neurological problems spoke Spanish as native language; no sensory/neurological problems			
Jorm et al. (1986)	Nine schools in and around Geelong, Australia, *N* = 453	Age: 7.11 years mean IQ cutoff: average IQ for age on CMMS Criteria: children available for 3-year longitudinal study	Neale scores 1.5 standard errors of prediction below expectation based on age and IQ	N/A	5.25
Lewis, Hitch, & Walker (1993)	51 primary schools in Lancashire, England, *N* = 1,056	Age: 9–10 years IQ cutoff: 90 on CPM (nonverbal intelligence scale) Criteria: no children with sensory or perceptual handicaps or psychiatric disturbances; native English speakers	Arithmetic scores above 90 (GMT) and reading (SPAR) scores below 85	1.11	3.20
Lovell et al. (1964) Sample 1	22 junior schools in England, *N* = 1,205	Age: 9–10 years IQ Cutoff: 90 NFER for nonverbal IQ Criteria: third-year students who have not changed schools	Reading quotient of <.80 on NFER Sentence Reading Test	1.03	6.78
Lovell et al. (1964) Sample 2	16 junior schools in England, *N* = 1,464	Age: 7–8 years IQ cutoff: 90 NFER for nonverbal IQ Criteria: pupils in the second half of the first year of junior school	Reading quotient of <.80 on NFER Sentence Reading Test	0.97	3.07
Miles et al. (1998)	British Birth Survey, *N* = 11,804	Age: 10 years IQ cutoff: 90 combined score on similarities and matrices from British ability scales Criteria: 10-year follow-up British Births Survey	Single word and Edinburgh reading test performance less than 1.5 or more standard errors of prediction based on IQ	1.01	1.74
Pennington et al. (1992)	Colorado Twin Study of Reading Disability, *N* = 1,076	Age: 12.42 years (mean) IQ cutoff: 90 on WISC/WAIS Criteria: Twin pairs with one member with reading difficulties; no neurological, emotional, behavioral, visual, or auditory difficulties; no neurological, emotional, behavioral, visual, or auditory difficulties	MRD: PIAT score 1.64–1.95 standard errors of prediction below IQ expectative and age–reading discrepancy = 1 SD. SRD: PIAT score < 1.96 standard errors of prediction below IQ expectation and age–reading discrepancy > 1 SD	N/A	1.34/1.65
Rutter et al. (2004) Sample 1	Dunedin Multidisciplinary Health & Development Study, *N* = 989	Age: 7, 9, 11 (years of testing) IQ cutoff: none stated	Those scoring in lowest 15% of BWRT as low-achievement definition, and those 1 SD below predicted reading score based on WISC PIQ as the IQ–achievement discrepancy definition	N/A	2.73/2.76

Citation	Sample	Criteria		Gender Ratio	
		Inclusion/Exclusion	**RD Diagnosis**	**NRD**	**RD**
		Criteria: Cohort formed at participants' age of 3, investigators enrolled 91% of births in New Zealand from April 1972 to March 1973			
Rutter et al. (2004) Sample 2	Christchurch Health and Development Study, *N* = 895	Age: 8–10 years IQ cutoff: none stated Criteria: part of an unselected birth cohort in a longitudinal study	Those scoring in lowest 15% of BWRT as low-achievement definition, and those 1 SD below predicted reading score based on WISC PIQ as the IQ–achievement discrepancy definition	N/A	2.10/2.40
Rutter et al. (2004) Sample 3	Office for National Statistics Study, *N* = 5,752	Age: 9–15 years IQ cutoff: none stated Criteria: children involved in a national survey of child mental health	Lowest 15% on the BAS-II for low achievement definition; and those in lowest 15% for discrepancy between BAS-II and BPVS-II. Also had a more stringent cutoff of 5% for both definitions because of sufficient power.	N/A	1.35 and 1.91/1.59 and 2/61
Rutter et al. (2004) Sample 4	Environmental Risk Longitudinal Twin Study, *N* = 2,163	Age: 7 years IQ cutoff: none stated Criteria: Part of a twin study on environmental risk factors of development	Lowest 15% on TOWRE for low achievement definition; and those 1 SD below predicted TOWRE score based on WPPSI-R IQ scores	N/A	1.38/1.73
S. Shaywitz et al. (1990)	Research-identified sample from Connecticut Longitudinal Study, *N* = 412	Grade 3 IQ cutoff: 80 on WISC-R Criteria: no emotional, behavioral, visual, auditory, or neurological impairments; native English speakers	WJ scores 1.5 standard errors of prediction below expectation based on IQ	0.9	1.34^{t}
B. Shaywitz et al. (1995)	Schools, parent groups, media announcement, *N* = 186	Age: 7.5–9.5 years IQ cutoff: none Criteria: Native English speakers; no emotional, visual, auditory, or neurological impairments; children with ADHD inattentive type excluded	Regression discrepancy of 1.5 standard errors between reading achievement and IQ; scored below 25th percentile on reading, decoding, and reading cluster	N/A	2.31–5.56
Wheldall & Limbrick (2010)	Australian representative sample, *N* = 1,133,988	Age: 8 years (3rd grade) or 10 years (5th grade) IQ cutoff: none Criteria: secondary data set from New South Wales Educational Measurement and School Accountability Directorate, data collected from 1997 to 2006	Scoring in the lowest two bands of the BST for third graders or the lowest three bands of the BST for fifth graders.	0.89/0.94	1.22–2.26

Note: Special thanks to Liederman, Kantrowitz, and Flannery (2005) for producing much of this table. ADHD = attention-deficit/hyperactivity disorder; BAS-II = *British Ability Scales–II*; BST = *Basic Skills Test*; BPVS-II = *British Picture Vocabulary Scales–II*; BWRT = *Burt Word Reading Test*; CAT = *California Achievement Test*; CMMS = *Columbia Mental Maturity Scale*; CPM = *Raven's Colored Progressive Matrices*; GMT = *Group Mathematics Test*; GORT = *Gray Oral Reading Test*; HKT-SpLD = *Hong Kong Test of Specific Learning Difficulties in Reading and Writing*; ITBS = *Iowa Test of*

Basic Skills; MRD = *moderate reading disability*; NA = *not available*; Neale = *Neale Analysis of Reading Disability*; NFER = *National Foundation for Educational Research*; NRD = *no reading disabilities*; PIAT = *Peabody Individual Achievement Test*; PIQ = *performance intelligence quotient*; RD = *reading disabilities*; SAT = *Stanford Achievement Test*; SICOLE-R = *Spanish Standardized Multimedia Battery*; SPAR = *Spelling and Reading Test*; SRA = *Group Reading Test*; SRD = *severe reading disability*; TOWRE = *Test of Word Reading Efficiency*; WAIS = *Wechsler Adult Intelligence Scale*; WISC = Wechsler *Intelligence Scale for Children*; WJ = *Woodcock–Johnson Tests of Achievement*; WPPSI-R = *Wechsler Preschool and Primary Scales of Intelligence–Revised*; WRAT = *Wide Range Achievement Test*.

[t] Nonsignificant difference.

Table 2

Descriptive Statistics for Oral Reading Fluency (ORF) and Nonsense Word Fluency (NWF) by Gender.

Measure	***M***	***SD***	**Skewness**	**Kurtosis**
NWF				
Males	50.4	29.86	1.02	1.109
Females	50.33	28.09	1.181	1.73
ORF				
Males	46.03	28.82	0.833	0.917
Females	53.06	30.08	0.813	0.834

Table 3

Gender Ratios and Odds Ratios for Absolute and Discrepancy Definitions of Reading Impairment.

Percentile	Gender Ratio	Corrected Ratio	Odds Ratio	95% Confidence Interval	Corrected Odds Ratio
Nonsense Word Fluency (absolute)					
30th	1.15:1	1.08:1	1.12	1.10–1.13	1.10
15th	1.28:1	1.21:1	1.25	1.23–1.27	1.25
5th	1.53:1	1.44:1	1.47	1.43–1.50	1.47
3rd	1.69:1	1.59:1	1.62	1.56–1.67	1.62
Oral Reading Fluency (absolute)					
30th	1.39:1	1.30:1	1.47	1.45–1.48	1.47
15th	1.66:1	1.56:1	1.69	1.67–1.72	1.69
5th	2.09:1	1.97:1	2.03	1.98–2.09	2.04
3rd	2.21:1	2.09:1	2.13	2.06–2.20	2.13
Nonsense Word Fluency residuals (discrepancy)					
30th	1.24:1	1.16:1	1.24	1.21–1.26	1.24
15th	1.37:1	1.28:1	1.34	1.31–1.37	1.34
5th	1.57:1	1.47:1	1.50	1.44–1.56	1.50
3rd	1.66:1	1.56:1	1.58	1.50–1.67	1.58
Oral Reading Fluency residuals (discrepancy)					
30th	1.55:1	1.46:1	1.71	1.68–1.74	1.71
15th	1.86:1	1.75:1	1.92	1.88–1.97	1.91
5th	2.38:1	2.24:1	2.32	2.23–2.43	2.32
3rd	2.56:1	2.40:1	2.46	2.33–2.60	2.46

Table 4

Comparing Performance of Female and Male School-Identified Leading-Disabled Students.

Gender	*N*	*M* (*SD*)	*t* Test
Nonsense Word Fluency			
Females	989	31.63 (18.015)	*t*(2920) = −0.926
Males	1,933	30.92 (20.141)	
Oral Reading Fluency			
Females	988	24.96 (19.344)	*t*(2919) = −5.263**
Males	1,933	21.14 (18.158)	
Nonsense Word Fluency residuals			
Females	665	−12.623 (18.582)	*t*(1911) = −3.138**
Males	1,248	−15.564 (20.006)	
Oral Reading Fluency residuals			
Females	665	−17.472 (20.126)	*t*(1911) = −6.741**
Males	1,248	−23.830 (19.386)	

** *p* <.001.

Table 5

Sensitivity, Overall Proportion of Agreement, Kappa, and Affected-Status Agreement Statistic Values by Gender.

					Number Identified			
Gender	**Sensitivity**	**OPA**	**Kappa**	**ASAS**	**SI & RI**	**SI & Not RI**	**Not SI & RI**	**Not SI & Not RI**
Nonsense Word Fluency								
Females	.123	.934	.109	.076	1,288	6,480	9,142	221,091
Males	.222	.896	.154	.118	3,532	13,940	12,409	223,055
Oral Reading Fluency								
Females	.197	.946	.178	.113	1,622	6,151	6,618	223,643
Males	.291	.902	.232	.168	4,997	12,484	12,196	223,318
Nonsense Word Fluency residuals								
Females	.083	.935	.063	.048	340	2,929	3,781	95,993
Males	.183	.896	.119	.094	1,182	6,143	5,279	97,203
Oral Reading Fluency residuals								
Females	.138	.946	.115	.073	434	2,836	2,710	97,068
Males	.220	.895	.167	.125	1,644	5,682	5,843	96,655

Note: ASAS = affected-status agreement statistic; OPA = overall proportion of agreement; RI = identified as having a reading impairment in the research; SI = school-identified LD status.

Table 6

Sensitivity Values by Racial/Ethnic Group.

Race	A	B	H	I	M	W
NWF	.07	.16	.17	.24	.20	.24
ORF	.12	.23	.22	.30	.35	.35
NWF residuals	.07	.13	.16	.10	.14	.16
ORF residuals	.19	.17	.19	.14	.19	.22

Note: A = Asian/Pacific Islander; B = Black; H = Hispanic; I = American Indian/Alaska Native; M = Mixed Race; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency; W = White.

Table 7

Overall Proportion of Agreement, Kappa, and Affected-Status Agreement Statistics for Research-Based Criteria.

				Number Identified			
Method	**OPA**	**Kappa**	**ASAS**	**1 & 2**	**1 & Not 2**	**2 & Not 1**	**Neither 1 Nor 2**
Poor NWF (1) and Poor ORF (2)	.953	.530	.552	14,267	12,091	11,152	453,383
Poor NWF (1) and ORF residuals (2)	.916	.154	.109	2,189	9,491	8,439	192,721
Poor NWF (1) and NWF residuals (2)	.947	.465	.330	5,520	6,164	5,063	196,106
Poor ORF (1) and ORF residuals (2)	.93	.317	.217	4,106	8,321	6,525	193,923
Poor ORF (1) and NWF residuals (2)	.923	.247	.171	3,358	9,065	7,223	193,194
ORF residuals (1) and NWF residuals (2)	.938	.345	.231	3,973	6,655	6,608	195,604

ASAS = affected-status agreement statistic; NWF = Nonsense Word Fluency; OPA = overall proportion of agreement; ORF = Oral Reading Fluency; RI = identified as having a reading impairment in the research; SI = school-identified LD status.