

Differential identification of females and males with reading difficulties: A meta-analysis

Jamie M. Quinn¹

Published online: 2 February 2018

© Springer Science+Business Media B.V., part of Springer Nature 2018

Abstract Males are more likely than females to be identified as having reading difficulties, but it is unclear if this is a result of sample ascertainment or identification bias. The purpose of this meta-analysis was to determine the magnitude of gender differences in reading difficulties using available studies in which researchers investigated this difference and an additional dataset with a representative U.S. sample. After conducting a literature search, sixteen studies and a restricted use dataset were included in the present analysis ($N = 552,729$). A random-effects odds ratio model indicated that males are 1.83 times more likely than females to have reading difficulties. Moderator analyses revealed that the gender ratio is greater when the identified reading difficulties were more severe. Further, this difference in identification rates across males and females was found without evidence of publication bias. Implications for the identification of students with reading difficulties are discussed.

Keywords Gender differences · Reading · Meta-analysis · Identification

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11145-018-9827-8>) contains supplementary material, which is available to authorized users.

✉ Jamie M. Quinn
jqquinn@fcrr.org

¹ The Meadows Center for Preventing Educational Risk, The University of Texas at Austin, Austin, TX, USA

Introduction

Best practices for identifying students with reading difficulties remains a topic of considerable concern (see Badian, 1999; Spencer et al., 2014). One reason for the prevailing concerns is there remains no consensus on how best to define reading difficulties. A variety of terms have been used that encompass students demonstrating low-level reading skills: dyslexia and reading disability have been used for more severe cases; terms like garden-variety poor readers, reading difficulties, reading problems, and reading struggles imply less severe levels of low reading achievement. For the purposes of this review, I use “reading difficulties” to encompass all of these terms. Another reason for the lack of agreement among researchers is due to differences in identification criteria. Criteria for identifying students have mostly fallen into three distinct categories: researcher identification criteria based on low achievement, IQ-discrepancy definitions, and response-to-intervention (RTI) or other school-based approaches.

A recent advancement is a constellation model approach (e.g., Fletcher, Stuebing, Morris, & Lyon, 2013; Spencer et al., 2014) which uses markers of low achievement, achievement-discrepancy, and RTI within one distinct model. Constellation models use multiple sources of information to predict which students are more likely to struggle with reading, but ignore whether the reader is male or female. Males are more likely to be identified as having difficulties with reading, with the ratio of males to females with reading difficulties ranging from a low of 1.2:1 to a high of 6.78:1 (e.g., Finucci & Childs, 1981; Miles, Haslum, & Wheeler, 1998; Quinn & Wagner, 2015; Rutter, Caspi, Fergusson, Horwood, Goodman, Maughan, Moffitt, Meltzer, & Carroll, 2004). The purpose of this study was to conduct a thorough search of the existing literature on differences in identification rates of males and females with reading difficulties, to determine the magnitude of this gender difference using an odds ratio (OR) meta-analysis, and to discuss potential implications this difference may have for the prediction of children who struggle with reading.

Why are more males than females identified as having reading difficulties?

In previous studies, researchers investigating gender differences with reading difficulties have discovered a wide range of gender ratios, but the origin of this gender difference in prevalence rates is unknown. Several characteristics of the individual studies may have impacted previous findings, including differences in identification criteria, the types of reading measures used in identification procedures, and differences in sample sizes and year of publication (e.g., Hawke, Wadsworth, Olson, & DeFries, 2007; Liederman, Katrowitz, & Flannery, 2005; Limbrick, Wheldall, & Madelaine, 2012; Siegel & Smythe, 2005).

Researcher-based and school-based definitions of reading difficulties

The most prominent methods for identifying struggling readers have included low achievement definitions, IQ-discrepancy definitions, and school-based identification procedures.

Low achievement definitions

Low achievement (LA) definitions require the researcher to choose a particular cutoff score (e.g., below the 30th percentile) on a relevant task or measure in determining if a student has a reading difficulty. Several large studies using this approach have chosen cutoffs from a large range of the lower tail of the distribution, from as low as the 3rd percentile to as high as the 30th percentile on their respective reading measures (e.g., Flynn & Rahbar, 1994; Jiménez et al., 2011; Quinn & Wagner, 2015). Other similar studies have used composite or standard score cutoffs (Chan, Ho, Tsang, Lee, & Chung, 2007, 2008; Lingren, De Renzi, & Richman, 1985; Wong, McBride-Chang, Lam, Chan, Lam, & Doo, 2012) or standard deviation cutoffs (Chan et al., 2008; Donfrancesco et al., 2010). The resulting gender ratios using this definition have ranged from 1.2:1 to 6.78:1 (see Online Supplemental Materials A).

Using low achievement on a reading measure may be a matter of convenience if no other methods are available; however, students near the cutoff value may be incorrectly categorized if there is a large amount of measurement error associated with the reading task (Cotton, Crewther, & Crewther, 2005). Researchers may inadvertently exclude a student whose true ability is below the cutoff but whose score on the measure is above the cutoff. Unless a highly reliable measure or a latent variable modeling method is used to estimate a latent ability score (theta score) that has significantly reduced measurement error, using a single achievement score method of identification has been discouraged (e.g., Spencer et al., 2014). Further, there is evidence that no distinct cutoff point exists that will correctly distinguish between children with and without reading difficulties, making the decision of which cutoff value to choose both difficult and arbitrary (Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992).

IQ-discrepancy definitions

IQ-discrepancy (IQ-D) definitions require a discrepancy between the student's intelligence (as measured by a full-scale IQ test or a proxy variable for IQ, such as vocabulary knowledge) and their expected score on a reading measure. Researchers have used different aptitude tests (Limbrick et al., 2012; Siegel, 1992), and in cases when full-scale IQ scores were not available, vocabulary knowledge was chosen as a proxy variable (e.g., Quinn & Wagner, 2015). Additionally, these definitions have imposed a relaxed discrepancy of one standard deviation between IQ and reading scores (e.g., Lindgren, de Renzi, & Richman, 1985) or a more stringent discrepancy of two standard deviations or standard errors of prediction between IQ and reading

scores (e.g., Berger, Yule, & Rutter 1975). The magnitude of the gender differences reported using this method have ranged from 1.34:1 to 5.25:1 (see Online Supplemental Materials A), consistently resulting in larger numbers of males identified.

IQ-D definitions over-identify males relative to females with respect to reading difficulties (Share & Silva, 2003). Lindgren et al. (1985) reported results from a cross-national comparison of rates of dyslexia in similar cities in Italy and the United States. The authors used three separate definitions for identifying readers with dyslexia, including a low-achievement definition (standard score [SS] on a reading comprehension test of less than 85 with average intelligence), an IQ-D definition (reading SS less than 85 and the WISC Full Scale IQ SS 1 SD below the mean), and a regression equation (using WISC IQ as an independent variable). Although the gender ratios in both Italy and the US were approximately 1:1 for students classified using the LA definition (Italy: 50% male; US: 56% male), once IQ-D was used, the ratio increased to almost 2:1 (Italy: 65% male; US: 72% male). Similar results were found in a large-scale study of readers from an at-risk population in the US (Quinn & Wagner, 2015), where an IQ-D definition over-identified males relative to using a LA definition. However, using either the LA definition or the IQ-D definition resulted in a significant ratio in favor of males (range 1.28:1–1.86:1).

IQ-D definitions are also subject to measurement unreliability (Cotton et al., 2005). Psychometricians have criticized using difference score techniques as the effects of measurement unreliability are doubled when two sources of variance are considered (Cattell, 1982). Instead of considering one source of measurement unreliability, unreliability using a discrepancy definition is sourced through both the IQ measure and the reading measure (Lord, 1958). Further, as the correlation between the IQ test and the reading test increases, the reliability of the difference score decreases (Caruso & Witkiewitz, 2002).

School- or clinician-based definitions

School-based methods of identification vary widely based on differences in state-level criteria. For example, Quinn and Wagner (2015) described the criteria used by the state of Florida to refer a student for screening of learning disabilities is a multi-staged process, including parent–teacher meetings, medical evaluations, and a history of non-response to classroom intervention. Students are further referred if difficulties persist after targeted classroom interventions were attempted. These types of procedures have been reported as biased in ways that increase the number of males referred (Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). These methods rely on teacher recommendations for evaluation, which may be affected by the typically more challenging behaviors exhibited by males rather than solely by their reading difficulty. Males are more likely to externalize their frustrations, potentially calling more attention to their difficulties and increasing the likelihood that teachers would recognize them. Additionally, reading difficulties often co-occur with attention deficit/hyperactivity disorder (ADHD; Willcutt & Pennington, 2000; Willcutt, Pennington, & DeFries, 2000), leading to more male referrals and

increasing the clinical ratio of males to females, which may reflect a difference in behavior instead of a difference in reading.

Differences in study characteristics

In addition to definitions, previous studies have also differed in the types of reading measures used, the sample ascertainment methods, and the date of publication.

Differences in types of reading measures

The type of reading measures used to qualify students with reading difficulties could have important implications for proper identification. For example, multiple studies used word-level decoding measures (e.g., Jiménez et al., 2011; Quinn & Wagner, 2015). Difficulties in decoding, particularly phonological or non-word decoding, has long been used to identify students with broader reading difficulties (e.g., Gough & Tunmer, 1986; Shaywitz, Morris, & Shaywitz, 2008). However, previous studies have also used measures of reading comprehension (e.g., Lingren et al., 1985; Wheldall & Limbrick, 2010) or reading fluency (Quinn & Wagner, 2015) for identifying individuals with reading difficulties. Quinn and Wagner used both decoding and fluency measures and both LA and IQ-D definitions. Both IQ-D definitions resulted in larger ratios than the LA definitions, and using a measure of decoding resulted in larger ratios (1.31:1 for LA and 1.86:1 for IQ-D) than using a measure of fluency (1.28:1 for LA and 1.66:1 for IQ-D, Quinn & Wagner, 2015). Chan et al. (2007) used the Hong Kong Test of Specific Learning Disabilities in Reading and Writing (HKT-SpLD), a multi-component literacy skills test and a cognitive abilities test. Students with reading difficulties were identified when they scored one standard deviation below average in both the literacy and the cognitive domains. This multi-component LA definition resulted in a gender ratio of two males to one female (see Online Supplemental Materials A).

Differences in sample sizes

The size of the ascertained sample could produce considerable variability in apparent gender differences. With the exception of a few large studies (e.g., Arnett et al., 2017; Quinn & Wagner, 2015; Wheldall & Limbrick, 2010), only a few studies have had sample sizes in the thousands (e.g., Chan et al., 2008; Flannery, Liederman, Daly, & Schultz, 2000; Jiménez et al., 2011; Miles et al., 1998; Rutter et al., 2004; Undheim & Sund, 2008). Studies published before 1995 have typically had small samples of fewer than 50 with a reading difficulties (e.g., Berger et al., 1975; Finucci & Childs, 1981; Jorm, Share, Matthews, & MacLean, 1986; Lewis, Hitch, & Walker, 1994; Shaywitz et al., 1990).

Differences in year of publication

Finally, studies of gender differences in reading disability have been published over a 60-year period. Conceptualizations of reading difficulties measures of reading, and methods of identification have changed considerably, which also could contribute to differences in the magnitude of reported gender differences. Older studies tended to show larger gender differences (e.g., Berger et al., 1975; Lovell, Shapton, & Warren, 1964).

A reflection of underlying distributions or ascertainment bias?

The gender ratio may be artificially inflated due to differences in the underlying distributions of skills for males and females (e.g., Arnett et al., 2017; Hawke, Olson, Willcut, Wadsworth, & DeFries, 2009) or that more males are being referred than females as a result of ascertainment bias (Shaywitz et al., 1990).

Underlying distributions

The gender ratio increases as the severity of the difficulty increases. Pennington, Gilger, Olson, and DeFries (1992) reported a smaller ratio of males to females when using a moderate definition of reading disability (1.34:1) versus a severe definition (1.65:1) of reading disability. Quinn and Wagner (2015) also reported that as the severity of the reading difficulty increased (from below the 30th percentile to below the 3rd percentile), the gender ratio increased significantly (e.g., 1.30:1 in the 30th percentile to 2.09:1 in the 3rd percentile for a LA definition). Olson (2002) grouped readers identified with dyslexia by their full-scale IQ and the severity of their deficit in reading single words to estimate potential gender differences related to both IQ and severity of the reading problem. When there was no IQ selection criteria, as the severity of the word recognition (WR) deficit increased, the ratio of males to females increased from 1.11 (at 1 SD below mean on WR) to 2.01 (at 3.5 SD below the mean on WR). However, when children were selected for above-average intelligence (> 100 on full-scale IQ), the ratio of males to females increased from 1.36 (at one SD below the mean on WR) to a ratio of 9.6 (at 3.5 SD below the mean on WR). Alternatively, Chan et al. (2007) reported no gender differences according to three categories of increasing severity (mild, moderate, and severe).

Researchers have argued that the distributions of scores for males and females are inherently different (see Arnett et al., 2017; Hawke et al., 2007). Standard deviations for males tend to be larger than for females in measures of reading, and therefore choosing an arbitrary cutoff point will automatically include more males. This ratio is increased when choosing a lower cutoff, as fewer females are found in the tails of the distribution. It might follow, then, that males would maintain greater representation in the upper tail of the distribution. However, Quinn and Wagner (2015) found no evidence of gender differences in the upper tail using a large sample of students with reading difficulties according to quantile–quantile (QQ) plots.

Ascertainment bias

Referred samples may have been artificially selected because of ascertainment bias. Shaywitz et al. (1990) proposed that when children are referred for impaired reading, males are more highly represented because they are also referred for externalizing behavioral problems. In their epidemiological study, males were 2–4 times more likely to be identified as having reading difficulties when based on school or clinical referrals. However, when tested on a battery of standardized reading measures, no gender differences emerged (Shaywitz et al., 1990).

Quinn and Wagner (2015) found a large gender difference in favor of males (2.11:1) when they considered school-determined learning disability (LD) status in relation to reading skills. To assess the presence of ascertainment bias, the authors postulated that if males were only being referred due to behavior problems, but females were being referred due to genuine struggles with their reading, females with a LD should have lower scores on tests of reading than males with a LD. Yet, there were no significant differences between males and females with a LD on a measure of non-word reading fluency, and males scored significantly *worse* than females on a measure of oral reading fluency. The authors concluded they did not find support for school-level ascertainment bias in selecting students with a LD (Quinn & Wagner, 2015).

Why a deeper understanding of differential identification matters

Differential identification of females and males with a reading difficulty may have important implications for how we help children who develop these difficulties. Between 5 and 10% of children, and as much as 17% of children, are estimated to have developmental dyslexia (Shaywitz, 1998), with an even larger number of students affected by general reading difficulties, otherwise known as “garden-variety poor readers” (Stanovich, 1988). Understanding even one minor source of individual differences in reading may enhance predictive models of reading difficulties. Additionally, although studies have estimated the gender ratio with their own samples of students, these studies may be biased to detect differential rates of reading difficulties in males and females through either sample ascertainment or method bias. A decisive meta-analysis that considers these possible sources of bias and examines potential confounding factors is warranted.

The present study

The purpose of the present meta-analysis was to estimate the magnitude of gender differences in reading difficulties across a large range of abilities, samples, measures, and years. The odds ratio of males versus females having reading difficulties was estimated using a random effects meta-analysis package in R. Additionally, moderator analyses were conducted to determine if publication year, average age of the sample, type of reading measure used (word reading out of

context or within context), severity of the reading difficulty, or identification method (LA vs. IQ-D vs. school-based definitions) affected the findings. Finally, steps were taken to identify and control for potential publication bias effects.

Methods

Literature base

Initial identification

A comprehensive search of the literature was conducted on May 19, 2017 to create a literature base for this meta-analysis. The following terms were entered into EBSCOhost using the PsycINFO, ERIC, Psychology and Behavioral Sciences Collection, and PsycArticles databases:

```
TI('reading disab*' OR 'reading impairment*' OR dyslexia OR
'reading problem*' OR 'reading difficult*' or 'struggling
reader*')
```

```
AND (AB(((gender OR sex) differences) OR ((gender OR sex)
ratio)))
```

```
OR (SU(((gender OR sex) differences) OR ((gender OR sex) ratio)))
```

The title must have contained a search term related to reading difficulties, and either the subject or the abstract must have pertained to gender differences. The terms “*sex differences*” and “*sex ratio*” were also included to capture studies that may have used this term in place of “*gender differences*” or “*gender ratio*.” After removing duplicates, 186 studies were identified.

Inclusion criteria

The abstracts for the 186 studies were screened for inclusion eligibility. Articles were assessed for relevance to the study, such that they mention measuring text reading and have a term for gender/sex differences or gender/sex ratios. Accordingly, twenty-six articles were irrelevant and were discarded. The full texts of the remaining articles were assessed for further eligibility. In order to be included in the meta-analysis, the following criteria must have been met. Studies must have been published in English ($k = 4$ excluded). Studies that were not directly related to text reading ($k = 42$; i.e., did not measure text reading in an attempt to identify students with reading difficulties), studies that investigated brain imaging or morphometry ($k = 27$), studies that investigated behavioral or molecular genetics ($k = 14$), and studies that focused on physical or physiological data ($k = 23$) were excluded. Studies that did not include children in grade 6 or younger ($k = 7$) or studies that used gender/sex-matched controls in analyses ($k = 11$) were excluded. Studies with participants from special populations were excluded (e.g., those with an intellectual disability or hearing or vision impairment, $k = 16$); however, students with ADHD were included due to the high comorbidity with dyslexia.

Studies that only included males, only included females, or did not involve the identification of students with reading difficulties were excluded ($k = 11$).

In sum, 26 studies were identified as eligible for this study. Additionally, a restricted use dataset was included: The Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999 (ECLS-K, Tourangeau, Nord, Lê, Sorongon, & Najarian, 2009). In this longitudinal study, researchers measured cognitive, social, emotional, and physical development and gathered information regarding the home and schooling environments for a nationally representative sample of kindergarteners. A 2×2 contingency table of males and females school-identified with reading difficulties was created from the first grade (1999) and third grade (2001) measurement occasions (total unweighted $n = 21,348$) and included within the coding scheme described below. Ten studies included data that could not be disaggregated across males and females; however, five of these studies were included in the descriptive table included in the Online Supplemental Materials A for reporting purposes; the other five were discarded due to missing information. Twenty-two samples [total $N = 552,729$, average age = 9.18 (range 6.50–13.70)] from 17 were included in the meta-analysis. A table provided in the Online Supplemental Materials A presents the authors, year of publication, sample information, identification criteria, and reported gender ratios for students with and without reading difficulties.

Coding procedures

For analyses in R , a coding scheme was created with the following: Author(s), year of publication, average age of the sample, the identification criteria (i.e., school-based criteria, IQ-discrepancy criteria, or low-achievement criteria), type of reading measure used (i.e., word reading in context or word reading out-of-context), severity of the reading problem, and columns for total n of males and females with and without an identified reading difficulty. A column was coded to specify if samples came from the same article. Odds ratios, log odds ratios, and the variance and standard error (SE) of the log odds ratio were calculated in Microsoft Excel prior to being imported into R to ensure accuracy of the estimates.

Coding reliability

Inter-rater reliability of coding was calculated as the rate of absolute disagreement versus agreement per cell in the coding scheme. A random selection of 10 studies were selected for double-coding to ensure accuracy of the coding scheme; an inter-rater reliability of .97 was achieved across the 10 double-coded studies for all cells within the coding scheme as specified above (e.g., disaggregated sample size, identification criteria).

Calculating the effect size: the odds ratio

The effect size of interest for this study was the odds ratio (OR; i.e., the odds that a male had reading difficulties relative to the odds that a female had reading difficulties). The OR was calculated as:

$$\frac{a * d}{b * c}$$

where a , b , c , and d refer to the top left, top right, bottom left, and bottom right cells of a 2×2 matrix. In this matrix, the columns represent the number of students without or with reading difficulties (respectively) and the rows represent number of males and females (respectively). Therefore, cell a refers to the number of males without a reading problem, cell b refers to the number of males with a reading problem, etc. If males and females were equally likely to be identified as having a reading problem, the OR would not be significantly different from one. Since males are the reference group, if females were more likely to be identified, the OR would be less than 1, and if males were more likely to be identified, the OR would be greater than 1. Only studies for which all data were available were included in the calculation of the OR, as it was necessary to know the total number of each gender who were and were not identified as having reading difficulties. Five studies were excluded but are still reported in the table in Supplemental Materials A (Chan et al., 2008; Finucci & Childs, 1981; Jorm et al., 1986; Lindgren et al., 1985; Wheldall & Limbrick, 2010).

Handling variability in effect sizes across studies

A random-effects model using the R package `metafor` was conducted (Viechtbauer, 2010). This package provided an estimate of heterogeneity across studies using the Q test for heterogeneity, and provided an estimate of the percentage of heterogeneity using I^2 . The `metafor` package used a maximum likelihood estimator of *tau-squared* (τ^2) within a random-effects model framework (Viechtbauer, 2010). The parameter τ^2 provides an estimate of the variance *within* studies, which can be used to account for or to help explain the amount of heterogeneity in effect sizes *across* studies.

Moderator analyses

In the event that the OR was calculated from a heterogeneous selection of studies, moderator analyses were conducted. These moderators included the average age of the sample, year of publication, type of diagnostic criteria used to identify students struggling with reading (dummy coded as three categories), severity of the reading impairment (dummy coded as two categories, with empty cells for studies that used school-identification procedures), and type of reading measure used within the researcher-identified studies (dummy coded as two categories, with empty cells for studies that used school-identification procedures).

Publication bias

Recent focus on publication bias has shown that psychological science is prone to only publishing studies with significant results (Fanelli, 2010; Franco, Malhotra, & Simonovits, 2014). Publication bias can have a negative impact on the results of this meta-analysis, such that the true population odds-ratio may be smaller (or larger) than estimated. Multiple tests for publication bias were consulted using `metafor`.

Funnel plot

The first procedure was to examine the funnel plot of the effect sizes versus their standard errors. An investigation of the funnel plot that yields slight asymmetry around the estimate, such that there are more studies above or below the estimated effect size, would support publication bias. In order to parametrically test for this bias, the Egger test (Egger, Smith, Schneider, & Minder, 1997) and the rank correlation test were conducted on the resulting funnel plot. These tests give an indication for asymmetry using a z -estimate with an associated p value.

Fail-safe N

The fail-safe N procedure was chosen as one sensitivity analysis to determine the degree of publication bias. Fail-safe N calculates the number of studies with null results (i.e., an equal OR for males and females) that would be necessary for the results of the meta-analysis to be null. The `metafor` package was used for this method.

Trim and fill

The ‘trim and fill’ is a second sensitivity analysis method that aims to identify and correct for potential funnel plot asymmetry arising from publication bias. The method ‘trims’ (i.e., removes) the smaller studies causing funnel plot asymmetry, uses the trimmed funnel plot to estimate the true center of the funnel, then ‘fills’ (i.e., replaces) the omitted studies and their missing ‘counterparts’ around the center (Duval & Tweedie, 2000a, b). As well as providing an estimate of the number of missing studies, an adjusted gender effect is derived by performing a meta-analysis including the filled studies. This method was also performed using the `metafor` package.

Hierarchical model

In order to account for stochastically dependent effect sizes, whereby effect sizes within the same study are related, `robumeta` (robust variance estimation package) was used to estimate a hierarchical model (Fisher, Tipton, & Hou, 2016). This package is useful for determining if varying the values of ρ (ρ ; the correlation between effect sizes from a single sample within a study) results in different estimates of the population OR and its variance (Hedges, Tipton, & Johnson, 2010).

Results

Pooled odds ratio

A random-effects OR meta-analysis was conducted on the available samples. Sample 1 from Lovell et al. (1964) was determined to be an extreme outlier (OR 6.55) and was removed. A forest plot of the ORs of the remaining samples is presented in Fig. 1, with the overall random effects model estimate presented at the bottom of the figure (OR 1.83 [1.62–2.06], $z = 9.947$, $p < .001$). Given that the confidence interval of the pooled OR did not contain one, it can be inferred that males are 1.83 times more likely than females to be identified as struggling with reading within these studies.

Also presented in Fig. 1 are the pooled estimates for each subgroup of identification criteria. The OR was indistinguishable between the studies that used school-identification criteria (OR 1.69) and studies that used low-achievement definitions (OR 1.68). Studies using IQ-D definitions tended to estimate the highest male to female ratios (OR 2.01). Studies using school-identification criteria had lower OR estimates than studies using IQ-D methods, but estimated similar ORs compared to studies using LA definitions.

The test for heterogeneity was significant ($Q [44] = 2145.62$, $p < .001$), with the total heterogeneity of the OR estimated to be $I^2 = 99.10\%$. Further, the variance within studies was significant ($\tau^2 = 0.13 [0.07–0.23]$). The I^2 statistics for each subgroup analysis were all larger than 89%, indicating these subgroup models did not account for significant additional heterogeneity in the effect sizes. Therefore, multiple moderator analyses were conducted to determine if year of publication, category of identification, severity of the reading difficulty, or age of participants significantly predicted the pooled OR.

Moderator analyses

Year of publication

A mixed-effects meta-analysis indicated that year of publication did not significantly predict the OR ($\beta = -0.0076 [-0.0186 \text{ to } 0.0034]$, $p = .1773$). This moderator accounted for no additional variance in the model ($R^2 = .0179$; $Q_M(1) = 1.8201$, $p = .1773$). Although year was not a significant moderator, a cumulative meta-analysis was conducted in the Supplemental Materials B (Figure A). This figure shows that beginning with the oldest study, and as studies are added to the estimation of the OR, the OR estimate decreased until around year 1998, where it stabilized to the 1.83 estimate found with the random-effects meta-analysis.

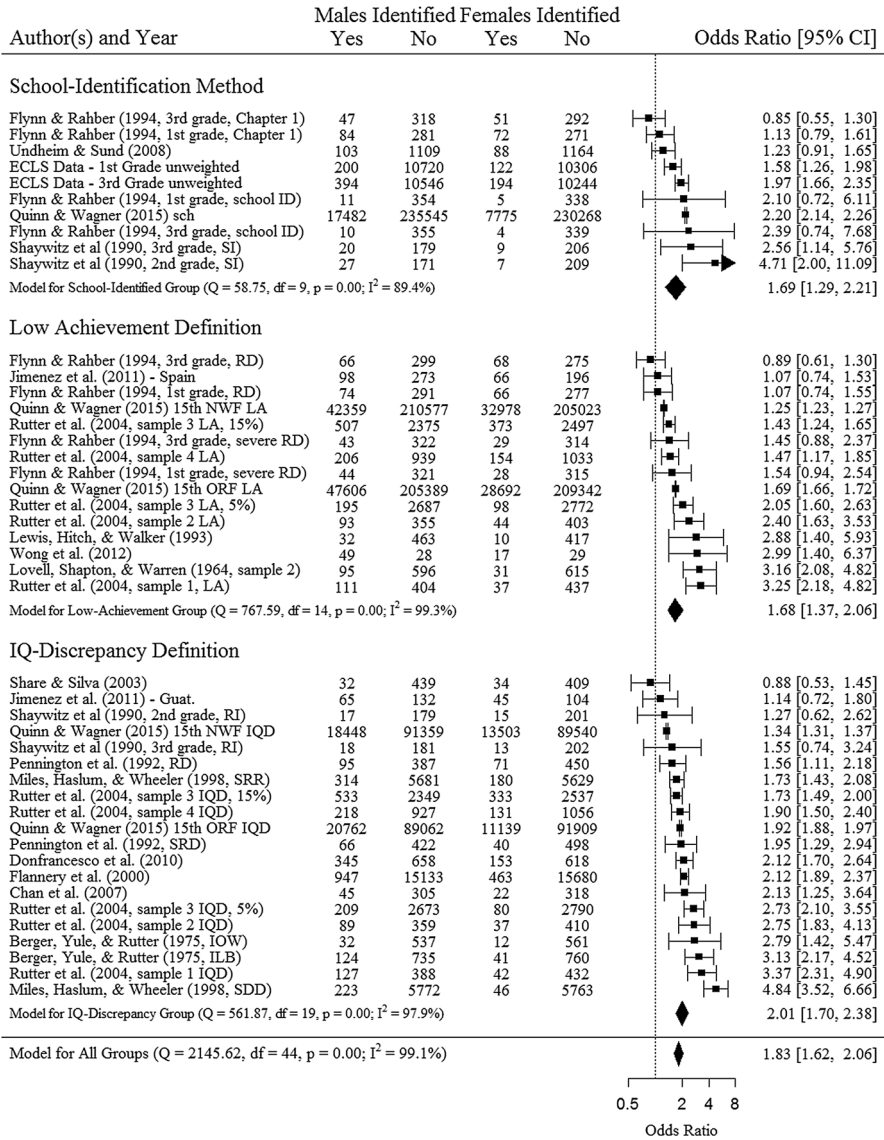


Fig. 1 Forest plot for the odds ratios using a random effects model and separated by identification criteria category. The dotted line is the reference value of one, where the odds do not favor males nor females. The random effects model for each category of identification is presented beneath each subgroup, and the final model for all studies is presented in the bottom of the figure

Category of identification

A visual interpretation of Fig. 1 suggested that the category of identification resulted in different pooled ORs. To test this, I conducted a mixed-effects meta-analysis with category of identification as the moderator. The intercept of this

model, or the studies that used an IQ-D definition as the reference group, was significant ($\beta^0 = 0.6965$ [0.5249–0.8681], $p < .0001$). As compared to the reference level, neither studies that used LA definitions ($\beta^1 = -0.1779$ [–0.4411 to 0.082], $p = .1851$) nor the studies that used school-based criteria ($\beta^2 = -0.1754$ [–0.4911 to 0.1404], $p = .2763$) had significantly different log ORs. These results also held when the reference group was changed. Differences in identification criteria did not affect the estimation of the OR ($R^2 = .038$, $Q_M(2) = 2.1945$, $p = 0.3338$).

Type of reading measure

Only studies using researcher-based criteria were included in this moderator analysis. As compared to studies that used measures of word decoding out of context (i.e., reading lists of words), studies that used measures of word reading in context (i.e., fluency of sentences or passages) did not have significantly different ORs ($\beta = -0.0759$ [–0.3728 to 0.2209], $p = .6161$). This result suggested that the type of reading measure used did not significantly predict the OR ($R^2 = 0$; $Q_M(1) = 0.2513$, $p = .6161$).

Age of participants

The average age of the participants was not a significant moderator ($\beta = 0.022$ [–0.04 to 0.09], $p = .4932$), and this moderator accounted for no additional variance in the model ($R^2 = 0$; $Q_M(1) = 0.4696$, $p = .4932$). There were no significant differences in ORs due to heterogeneous ages across the included samples.

Severity of the reading problem

As compared to studies using more relaxed criteria for reading difficulties, studies that used more strict criteria (i.e., IQ-discrepancies of more than 1.5 SD, or low achievement below the 10th percentile) produced significantly larger ORs ($\beta^1 = 0.2489$ [0.0162–0.4817], $p = .0361$). Severity of the reading difficulty accounted for 12.61% of the heterogeneity in the OR ($Q_M(1) = 4.3935$, $p = .0361$). Studies that used less severe criteria produced smaller ORs (OR 1.71) than studies that used more severe criteria (OR 2.19).

Publication bias

Several parametric and nonparametric tests were conducted to determine if there was publication bias affecting the estimation of the pooled OR.

Funnel plot

The funnel plot of the ORs calculated from each study was consulted (see Online Supplemental Materials B, *Figure B*). There were multiple studies outside of the shaded area, suggesting slight publication bias; however, the Egger Test for funnel plot asymmetry was not significant ($z = 1.1915$, $p = .2335$), suggesting that the funnel plot was statistically symmetrical over the pooled OR estimate. Additionally, the rank correlation test of funnel plot asymmetry was not significant (Kendall's $\tau = -0.1111$, $p = .2881$). An additional funnel plot was created that was conditioned on the severity of diagnosis (see Online Supplemental Materials B, *Figure C*). This funnel plot was also symmetrical according to the Egger test ($z = 1.0669$, $p = .2860$) and the rank correlation test (Kendall's $\tau = -0.1606$, $p = .1544$).

Sensitivity analyses

Trim-and-fill

The trim-and-fill procedure estimated that eight studies were missing from below the estimated OR (see Online Supplemental Materials B, *Figure D*). The empty circles represent the filled-in studies, whereas the black circles represent the remaining studies that were not trimmed from the analysis. A meta-analysis on these included hypothesized studies resulted in a lower adjusted estimate (OR 1.64 [1.44–1.87], $p < .0001$), but the confidence interval still did not contain one.

Fail-safe N

According to the results of the fail-safe N test using the Rosenthal approach, in order to achieve null population results (i.e., population OR 1), an additional 50,735 studies with null results or results favoring females over males are needed to achieve the target null p value of $> .05$. To achieve a p value greater than $.01$, and additional 25,342 studies with null results (OR 1) or results favoring females over males (OR < 1) are needed.

Hierarchical models to handle stochastic dependence

Since multiple point estimates for the OR were pulled from certain studies, an analysis that considered this stochastic dependence was conducted. When accounting for dependent effect sizes, the OR was estimated to be 1.96 (95% CI 1.66–2.33), which was somewhat larger and less accurate than the pooled OR without handling the dependency (OR 1.83 [1.62–2.06]). The estimated between-study variance was $\tau^2 = 0.26$, larger than the estimate from the pooled OR analysis without handling stochastic dependence ($\tau^2 = 0.12$). After accounting for heterogeneity due to stochastic dependence, there was still a large portion of unexplained variance ($I^2 = 95.68$), and adjusting the ρ value (that is, adjusting the correlation between

effect sizes from a single sample within a study) did not significantly affect the estimated pooled OR.

Discussion

This meta-analysis provided support that there is a disproportionate number of males, compared with females, who exhibit reading difficulties. Males were 1.83 times more likely than females to be identified as having a reading problem, regardless of method of identification, reading measure, publication year, and age of the participant. Only the severity of the reading difficulty emerged as a significant predictor of the OR: the more severe the definition, the more likely males were to be identified relative to females at that level of difficulty. Subsequently, it is critical to review the problems with previous methods of identification, discuss the implications for using this information in predictive models of reading, and consider potential sources of variation beyond methodological effects that may be of future interest to researchers interested in understanding and predicting difficulties in reading.

The identification of children with reading difficulties

Researcher-based methods for identifying children with reading difficulties have previously been criticized, particularly as it relates to measurement error and variance. Such criticisms and limitations have relevance for the current analysis.

The problem with single criterion methods

Previous studies that have estimated the ratio of males to females who struggle with reading have used various researcher-based criteria for identification or have used a referred sample to estimate the gender ratio. However, these methods, especially those based on a single criterion, yield unreliable results as a result of measurement error (Cattell, 1982; Cotton et al., 2005; Lord, 1958). Hybrid or constellation models that use multiple criteria through many sources of information to identify children with reading difficulties are being implemented with moderate success (e.g., Erbeli, Hart, Wagner, & Taylor, 2017; Fletcher et al., 2013; Spencer et al., 2014), but these methods also have reliability challenges and often have poor agreement (see Quinn & Wagner, 2015).

Differences in identification methods

School-based identification methods did not have significantly different ORs in this meta-analysis as compared to two researcher-based methods. However, previous research has suggested that the method of identification matters with respect to observed gender ratios, whereby IQ-discrepancy methods over-identify male students (Share & Silva, 2003) and school- or clinician-based identification methods over-identify male students because of ascertainment bias (Shaywitz et al.,

1990). The IQ-D definition produced an OR of 2.01, suggesting that males are 2 times as likely as females to be identified as having a difficulty under this method. However, the confidence intervals around the estimates were large; enough that the OR estimates were not significantly different across methods. At present, these analyses cannot support a hypothesis that IQ-D or school-based definitions are more biased than other methods at identifying males versus females.

What the results of this study mean for the identification of students with reading difficulties

It is known that the use of single criterion methods can lead to biased outcomes. The more that researchers use hybrid models of reading to predict reading difficulties in children, the more insight we gain into the reliability and accuracy of these methods. The present results pointed to gender as a factor to include in these constellation models as an additional source of information to improve model accuracy. With more accurate predictive models, researchers will be more precise with forecasting which students are at the highest risk for difficulties in reading.

Potential sources of variability unrelated to identification methods and distributions

As this meta-analysis has shown, males are more likely to be identified with a reading difficulty outside of methodological and statistical influences. However, there exists evidence for genetic and environmental influences that may help to explain the existence of male vulnerability to reading difficulties. Three such examples are genetic heritability, prenatal testosterone exposure, and environmental influences such as stereotype threat.

Genetic heritability and prenatal influences

Reading difficulties, particularly developmental dyslexia, tend to run in families, suggesting there is a genetic involvement in the development of these difficulties. As much as 50% of the variance in developmental dyslexia is explained by genetic factors (Shaywitz et al., 2008), and a recent application of behavioral genetics using a hybrid model of reading disability showed a large heritable influence (55% of the variance) in addition to a significant shared environmental influence (19% of the variance) in the etiological explanation of reading disability (Erbeli et al., 2017). From a molecular view, a recent longitudinal study of Dutch-speaking children showed that two single nucleotide polymorphisms (SNPs) on one target gene (KIAA0319) were nominally but consistently associated with rapid naming abilities in children (Carrion-Castillo et al., 2017).

In addition to genetic influences, there is support for the early developmental effects of hormones on the development of dyslexia, particularly through the differential effects prenatal testosterone levels have on the development of brain areas responsible for auditory temporal processing (Beech & Beauvois, 2005; Galaburda, 1999; Geschwind & Galaburda, 1985a, b). The brain areas for auditory

temporal processing are responsible for language and phonological processing, both of which are critical components of skilled reading.

Stereotype threat

Girls are more motivated to read and have a better attitude towards reading (Logan & Johnston, 2009; McGeown, Goodwin, Henderson, & Wright, 2012). As a result, teachers may inadvertently show more interest and support for girls' reading over boys, negatively affecting boys' reading performance (Retelsdorf, Shchwartz, & Asbrock, 2015). This reflects a conundrum termed *stereotype threat* (S. Spencer, Steele, & Quinn, 1999): since girls are more likely to be regarded as more interested in reading and potentially better readers than boys are, boys then perceive that reading is a less desirable activity and may even perceive that the consequences for poor reading are greater for them. Spencer and colleagues reported that boys might be threatened by the possibility of being negatively treated should they fail a reading test (Spencer, Logel, & Davies, 2016). Pansu and colleagues examined the reading scores of eighty children assigned to a stereotype threat condition (a diagnostic reading test) versus a threat-reduced condition (a game). Males underperformed relative to females in the threat condition, but the results reversed in the reduced-threat condition (Pansu, Régner, Max, Colé, Nezlek, & Huhuet, 2016). Retelsdorf and colleagues corroborated these findings by showing that teachers' negative stereotypes affect boys' self-concept of their own reading skills above their reading performance (Retelsdorf et al., 2015). Stereotype threat and gender beliefs held by teachers may be an additional source of the differences in males' and females' reading skills.

Study limitations

Though the present meta-analysis showed evidence of an over-identification of males with reading difficulties, several limitations prevail. First, previous research has claimed that males are more likely to be identified with reading problems because of differences in the underlying distributions of reading skills across genders. Arnett et al. (2017) concluded that due to larger standard deviations for males, more males are represented in the lower tail of the distribution, and therefore the gender ratio was larger for lower performing students. There was no way to control for study-level variance between male and female performance, as the data for the present meta-analysis was taken directly from information available within the manuscripts, and student-level data was not available.

Secondly, this study does not address comorbid disorders and problems that are typically seen with students who have reading problems, such as ADHD (Willcutt & Pennington, 2000; Willcutt et al., 2000) and writing disability (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008). These comorbid disorders tend to occur more often in males than in females (Gomez, Harvey, Quick, Scharer, & Harris, 1999; Graetz, Sawyer, & Baghurst, 2005), which could affect the gender difference ratio, should behaviors associated with the disorders affect the ascertainment of samples.

However, there was no evidence for ascertainment bias in many of the larger studies included in this meta-analysis.

Thirdly, multiple terms have been used to define students with low levels of reading skills (e.g., dyslexia, reading disability, students with reading problems or difficulties, garden-variety poor readers). This meta-analysis was inclusive of as many terms as possible, but the primary search may have missed studies for which the identification of poor readers was not the primary goal. However, given the substantial fail-safe N required to result in a non-significant or reversed OR ($OR \leq 1$), and given the lack of support for publication bias, these results show a preponderance of males identified with a problem, even if the terms used to define and identify this problem are heterogeneous in nature.

Implications and future directions

This meta-analysis only included primary studies that were specifically investigating gender ratios in reading difficulties. Although the aggregated sample size was very large ($N = 552,729$), the analyses were powered by the number of primary studies (45 separate estimates from the 22 samples from $k = 17$ studies). The non-significant moderator analyses could be a result of low power to detect these effects. If this meta-analysis were expanded to any study that included gender differences (without that being the primary goal of the study), the power to detect moderator effects may increase. It should be noted that had sample 1 from Lovell et al. (1964) been retained, the results of the moderator analysis for year of publication would have been significant ($\beta = -0.015, p < .01$), but this is confounded with the fact it was published over 50 years ago and also included the highest ratio of any included study.

Males are more likely to be identified as having a reading difficulty, and males are more likely to have this difficulty because of genetic, prenatal, and environmental influences. Although it was not possible to directly assess the effects of these influences, this study provides a strong rationale for the inclusion of gender as a predictor variable in identification methods. Further work is needed regarding the inclusion of this covariate in multi-dimensional hybrid models or constellation models used to predict and identify children with reading difficulties.

Conclusions

The purpose of this meta-analysis was to determine if males are more likely than females to have difficulties with reading. Results indicated that there is a statistically significant difference in identification rates, suggesting that more males are located in the bottom of the distribution of reading ability. School administrators might consider incorporating gender as a predictor for difficulties to assure that they are adequately identifying all students with reading difficulties, whether those difficulties be the result of intrinsic or extrinsic factors. Researchers may incorporate gender into predictive models of reading failure to provide an additional source of information for identifying reading difficulties in children.

Acknowledgements This research was supported by grant P50 HD052117-07 from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development or the National Institutes of Health. I would like to express my immense gratitude to Dr. Greg Roberts, Dr. Sharon Vaughn, and Dr. Richard Wagner for their incredible mentorship.

References

- *Denotes that a reference was used in the meta-analyses.
- Arnett, A. B., Pennington, B. F., Peterson, R. L., Willcutt, E. G., DeFries, J. C., & Olson, R. K. (2017). Explaining the sex difference in dyslexia. *Journal of Child Psychology and Psychiatry*, *58*, 719–727. <https://doi.org/10.1111/jcpp.12691>.
- Badian, N. A. (1999). Reading disability defined as a discrepancy between listening and reading comprehension: A longitudinal study of stability, gender differences, and prevalence. *Journal of Learning Disabilities*, *32*, 138–148. <https://doi.org/10.1177/002221949903200204>.
- Beech, J., & Beauvois, M. (2005). Early experience of sex hormones as a predictor of reading, phonology, and auditory perception. *Brain and Language*, *96*, 49–58. <https://doi.org/10.1016/j.bandl.2005.08.006>.
- *Berger, M., Yule, W., & Rutter, M. (1975). Attainment and adjustment in two geographical areas: II. The prevalence of specific reading disability. *British Journal of Psychiatry*, *126*, 510–519. <https://doi.org/10.1192/bjp.126.6.510>.
- Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Gender differences in writing and reading disabilities. *Journal of School Psychology*, *46*, 151–172. <https://doi.org/10.1016/j.jsp.2007.02.007>.
- Carrion-Castillo, A., Maassen, B., Franke, B., Heister, A., Naber, M., Van Der Leij, A., et al. (2017). Association analysis of dyslexia candidate genes in a Dutch longitudinal sample. *European Journal of Human Genetics*, *25*, 452–460. <https://doi.org/10.1038/ejhg.2016.194>.
- Caruso, J. C., & Witkiewitz, K. (2002). Increasing the reliability of ability-achievement difference scores: An example using the Kaufman Assessment Battery for Children. *Journal of Educational Measurement*, *39*, 39–58. <https://doi.org/10.1111/j.1745-3984.2002.tb01134.x>.
- Cattell, R. B. (1982). The clinical use of difference scores: Some psychometric problems. *Multivariate Experimental Clinical Research*, *6*, 87–98. <http://psycnet.apa.org/psycinfo/1983-31658-001>.
- *Chan, D. W., Ho, C. S., Tsang, S., Lee, S., & Chung, K. K. H. (2007). Prevalence, gender ratio and gender differences in reading-related cognitive abilities among Chinese children with dyslexia in Hong Kong. *Educational Studies*, *33*, 249–265. <https://doi.org/10.1080/03055690601068535>.
- Chan, D. W., Ho, C. S., Tsang, S., Lee, S., & Chung, K. K. H. (2008). Estimating incidence of developmental dyslexia in Hong Kong: What differences do different criteria make? *Australian Journal of Learning Difficulties*, *13*, 1–16. <https://doi.org/10.1080/19404150802093695>.
- Cotton, S. M., Crewther, D. P., & Crewther, S. G. (2005). Measurement error: implications for diagnosis and discrepancy models of developmental dyslexia. *Dyslexia*, *11*, 186–202. <https://doi.org/10.1002/dys.298>.
- *Donfrancesco, R., Iozzino, R., Caruso, B., Ferrante, L., Mugnaini, D., Talamo, A., et al. (2010). Is season of birth related to developmental dyslexia? *Annals of Dyslexia*, *60*, 175–182. <https://doi.org/10.1007/s11881-010-0037-6>.
- Duval, S. J., & Tweedie, R. L. (2000a). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- Duval, S. J., & Tweedie, R. L. (2000b). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98. <https://doi.org/10.1080/01621459.2000.10473905>.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. <https://doi.org/10.1136/bmj.315.7109.629>.

- Erbeli, F., Hart, S. A., Wagner, R. K., & Taylor, J. (2017). Examining the etiology of reading disability as conceptualized by the hybrid model. *Scientific Studies of Reading, 22*, 167–180. <https://doi.org/10.1080/10888438.2017.1407321>.
- Fanelli, D. (2010). “Positive” results increase down the hierarchy of the sciences. *PLoS ONE, 5*(4), e10068. <https://doi.org/10.1371/journal.pone.0010068>.
- Finucci, J. M., & Childs, B. (1981). Are there really more dyslexic boys than girls? In A. Ansora, A. Geschwind, A. Galaburda, M. Albert, & N. Gartrell (Eds.), *Sex differences in dyslexia* (pp. 1–9). Townson, MD: Orton Dyslexia Society.
- Fisher, Z., Tipton, E., & Hou, Z. (2016). *Robumeta: Robust variance meta-regression*. R package version 1.8. <https://CRAN.R-project.org/package=robumeta>.
- *Flannery, K. A., Liederman, J., Daly, L., & Schultz, J. (2000). Male prevalence for reading disability is found in a large sample of black and white children free from ascertainment bias. *Journal of the International Neuropsychological Society, 6*, 433–442.
- Fletcher, J. M., Stuebing, K. K., Morris, R. D., & Lyon, G. R. (2013). Classification and definition of learning disabilities: A hybrid model. In H. L. Swanson & K. Harris (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 33–50). New York, NY: Guilford Press.
- *Flynn, J. M., & Rahbar, M. H. (1994). Prevalence of reading failure in boys compared with girls. *Psychology in the Schools, 31*, 66–71. [https://doi.org/10.1002/1520-6807\(199401\)31:1%3C66::AID-PITS2310310109%3E3.0.CO;2-J](https://doi.org/10.1002/1520-6807(199401)31:1%3C66::AID-PITS2310310109%3E3.0.CO;2-J).
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science, 345*, 1502–1505. <https://doi.org/10.1126/science.1255484>.
- Galaburda, A. M. (1999). Developmental dyslexia: A multilevel syndrome. *Dyslexia, 5*, 183–191. [https://doi.org/10.1002/\(SICI\)1099-0909\(199912\)5:4%3C183::AID-DYS147%3E3.0.CO;2-C](https://doi.org/10.1002/(SICI)1099-0909(199912)5:4%3C183::AID-DYS147%3E3.0.CO;2-C).
- Geschwind, N., & Galaburda, A. M. (1985a). Cerebral lateralization: Biological mechanisms, associations, and pathology: I. A hypothesis and program for research. *Archives of Neurology, 42*, 428–459. <https://doi.org/10.1001/archneur.1985.04060050026008>.
- Geschwind, N., & Galaburda, A. M. (1985b). Cerebral lateralization: Biological mechanisms, associations, and pathology: II. A hypothesis and a program for research. *Archives of Neurology, 42*, 521–552. <https://doi.org/10.1001/archneur.1985.04060060019009>.
- Gomez, R., Harvey, J., Quick, C., Scharer, I., & Harris, G. (1999). 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, 40*, 265–274. <https://doi.org/10.1111/1469-7610.00440>.
- Graetz, B. W., Sawyer, M. G., & Baghurst, P. (2005). Gender differences among children with DSM-IV ADHD in Australia. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*, 159–168. <https://doi.org/10.1097/00004583-200502000-00008>.
- Hawke, J. L., Olson, R. K., Willcutt, E. G., Wadsworth, S. J., & DeFries, J. C. (2009). Gender ratios for reading difficulties. *Dyslexia, 15*, 239–242. <https://doi.org/10.1002/dys.389>.
- Hawke, J., Wadsworth, S., Olson, R., & DeFries, J. (2007). Etiology of reading difficulties as a function of gender and severity. *Reading and Writing, 20*, 13–25. <https://doi.org/10.1007/s11145-006-9016-z>.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39–65. <https://doi.org/10.1002/jrsm.5>.
- *Jiménez, J. E., de la Cadena, C. G., Siegel, L. S., O’Shanahan, I., García, E., & Rodríguez, C. (2011). Gender ratio and cognitive profiles in dyslexia: A cross-national study. *Reading and Writing, 24*, 729–747. <https://doi.org/10.1007/s11145-009-9222-6>.
- Jorm, A., Share, D., Matthews, R., & MacLean, R. (1986). Behaviour problems in specific reading retarded and general reading backwards children: A longitudinal study. *Journal of Child Psychology and Psychiatry, 27*, 33–43. <https://doi.org/10.1111/j.1469-7610.1986.tb00619.x>.
- *Lewis, C., Hitch, G., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. *Journal of Child Psychology, 35*, 283–292. <https://doi.org/10.1111/j.1469-7610.1994.tb01162.x>.
- Liederman, J., Kantowitz, L., & Flannery, K. (2005). Male vulnerability to reading disability is not likely to be a myth: A call for new data. *Journal of Learning Disabilities, 38*, 109–129. <https://doi.org/10.1177/00222194050380020201>.
- Limbrick, L., Wheldall, K., & Madelaine, A. (2012). Reading and related skills in the early school years: Are boys really more likely to struggle? *International Journal of Disability, Development and Education, 59*, 341–358. <https://doi.org/10.1080/1034912X.2012.723939>.

- Lindgren, S. D., de Renzi, E., & Richman, L. C. (1985). Cross-national comparisons of developmental Dyslexia in Italy and the United States. *Child Development, 56*, 1404. <https://doi.org/10.2307/1130460>.
- Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading, 32*, 199–214. <https://doi.org/10.1111/j.1467-9817.2008.01389.x>.
- Lord, F. M. (1958). The utilization of unreliable difference scores. *Journal of Educational Psychology, 49*, 150–152. <http://psycnet.apa.org/doi/10.1037/h0045456>
- *Lovell, K., Shapton, D., & Warren, N. S. (1964). A Study of some cognitive and other disabilities in backward readers of average intelligence as assessed by a non-verbal test. *British Journal of Educational Psychology, 34*, 58–64. <https://doi.org/10.1111/j.2044-8279.1964.tb00605.x>.
- McGeown, S., Goodwin, H., Henderson, N., & Wright, P. (2012). Gender differences in reading motivation: Does sex or gender identity provide a better account? *Journal of Research in Reading, 35*, 328–336. <https://doi.org/10.1111/j.1467-9817.2010.01481.x>.
- *Miles, T. R., Haslum, M. N., & Wheeler, T. J. (1998). Gender ratio in dyslexia. *Annals of Dyslexia, 48*, 27–55. <https://doi.org/10.1007/s11881-998-0003-8>.
- Olson, R. (2002). Dyslexia: Nature and nurture. *Dyslexia, 8*, 143–159. <https://doi.org/10.1002/dys.228>.
- Pansu, P., Régner, I., Max, S., Colé, P., Nežlek, J. B., & Huguët, P. (2016). A burden for the boys: Evidence of stereotype threat in boys' reading performance. *Journal of Experimental Social Psychology, 65*, 26–30. <https://doi.org/10.1016/j.jesp.2016.02.008>.
- *Pennington, B. F., Gilger, J. W., Olson, R. K., & DeFries, J. C. (1992). The external validity of age-versus IQ-discrepancy definitions of reading disability: Lessons from a twin study. *Journal of Learning Disabilities, 25*, 562–573. <https://doi.org/10.1177/002221949202500904>.
- *Quinn, J. M., & Wagner, R. K. (2015). Gender differences in reading impairment and in the identification of impaired readers: Results from a large-scale study of at-risk readers. *Journal of Learning Disabilities, 48*, 433–445. <https://doi.org/10.1177/0022219413508323>.
- Retelsdorf, J., Schwartz, K., & Asbrock, F. (2015). “Michael can't read!” Teachers' gender stereotypes and boys' reading self-concept. *Journal of Educational Psychology, 107*, 186–194. <https://doi.org/10.1037/a0037107>.
- *Rutter, M., Caspi, A., Fergusson, D., Horwood, J. L., Goodman, R., Maughan, B., et al. (2004). Sex differences in developmental reading disability: New findings from four epidemiological studies. *Journal of the American Medical Association, 291*, 2007–2012. <https://doi.org/10.1001/jama.291.16.2007>.
- *Share, D. L., & Silva, P. A. (2003). Gender bias in IQ-discrepancy and post-discrepancy definitions of reading disability. *Journal of Learning Disabilities, 36*, 4–14. <https://doi.org/10.1177/00222194030360010201>.
- Shaywitz, S. (1998). Dyslexia. *New England Journal of Medicine*. <https://doi.org/10.1056/NEJM199801293380507>.
- 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*, 145–150. <https://doi.org/10.1056/NEJM199201163260301>.
- Shaywitz, S. E., Morris, R., & Shaywitz, B. A. (2008). The education of dyslexic children from childhood to young adulthood. *Annual Review of Psychology, 59*, 451–475. <https://doi.org/10.1146/annurev.psych.59.103006.093633>.
- *Shaywitz, S. E., Shaywitz, B. A., Fletcher, J., & Escobar, M. D. (1990). Prevalence of reading disability in boys and girls. *JAMA*. <https://doi.org/10.1001/jama.1990.03450080084036>.
- Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25*, 618–629. <https://doi.org/10.1177/002221949202501001>.
- Siegel, L., & Smythe, I. (2005). Reflections on research on reading disability with special attention to gender issues. *Journal of Learning Disabilities, 38*, 473–477. <https://doi.org/10.1177/00222194050380050901>.
- Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual Review of Psychology, 67*, 415–437. <https://doi.org/10.1146/annurev-psych-073115-103235>.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*(1), 4–28. <https://doi.org/10.1006/jesp.1998.1373>.
- Spencer, M., Wagner, R. K., Schatschneider, C., Quinn, J. M., Lopez, D., & Petscher, Y. (2014). Incorporating RTI in a hybrid model of reading disability. *Learning Disability Quarterly, 37*(3), 161–171. <https://doi.org/10.1177/0731948714530967>.

- Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*, 590–604. <https://doi.org/10.1177/002221948802101003>.
- *Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (ECLS-K), Combined User's Manual for the ECLS-K Eighth-Grade and K–8 Full Sample Data Files and Electronic Codebooks* (NCES 2009–004). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from <http://eric.ed.gov/?id=ED511826>.
- *Undheim, A. M., & Sund, A. M. (2008). Psychosocial factors and reading difficulties: Students with reading difficulties drawn from a representative population sample. *Scandinavian Journal of Psychology, 49*, 377–384. <https://doi.org/10.1111/j.1467-9450.2008.00661.x>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1–48. <http://www.metafor-project.org/doku.php/metafor>.
- Wheldall, K., & Limbrick, L. (2010). Do more boys than girls have reading problems? *Journal of Learning Disabilities, 43*, 418–429. <https://doi.org/10.1177/0022219409355477>.
- Willcutt, E. G., & Pennington, B. F. (2000). Comorbidity of reading disability and attention-deficit/hyperactivity disorder: Differences by gender and subtype. *Journal of Learning Disabilities, 33*, 179–191. <https://doi.org/10.1177/002221940003300206>.
- Willcutt, E. G., Pennington, B. F., & DeFries, J. C. (2000). Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *American Journal of Medical Genetics, 96*, 293–301. [https://doi.org/10.1002/1096-8628\(20000612\)96:3%3C293::AID-AJMG12%3E3.0.CO;2-C](https://doi.org/10.1002/1096-8628(20000612)96:3%3C293::AID-AJMG12%3E3.0.CO;2-C).
- *Wong, S. W. L., McBride-Chang, C., Lam, C., Chan, B., Lam, F. W. F., & Doo, S. (2012). The joint effects of risk status, gender, early literacy and cognitive skills on the presence of dyslexia among a group of high-risk Chinese children: Predicting the presence of dyslexia. *Dyslexia, 18*, 40–57. <https://doi.org/10.1002/dys.1434>.