Research Article

Associations Between the 2D:4D Proxy Biomarker for Prenatal Hormone Exposures and Symptoms of Developmental Language Disorder

Sean M. Redmond^a and Andrea C. Ash^a

Purpose: Relative lengths of the index (2D) and ring (4D) fingers in humans represent a retrospective biomarker of prenatal hormonal exposures. For this reason, the 2D:4D digit ratio can be used to investigate potential hormonal contributions to the etiology of neurodevelopmental disorders. This study tested potential group differences in 2D:4D digit ratios in a sample of boys with and without developmental language disorder (DLD) and examined the strength of associations between 2D:4D digit ratio and a battery of verbal and nonverbal measures.

Method: A group of 29 boys affected by DLD and a group of 76 boys with typical language abilities participated (age range = 5;6–11;0 years). Scanned images were used to

measure finger lengths. Language measures included the core language subtests from the Clinical Evaluation of Language Fundamentals–Fourth Edition (Semel, Wiig, & Secord, 2003), a nonword repetition task, a sentence recall task, and the Test of Early Grammatical Impairment (Rice & Wexler, 2001).

Results: Significant group differences indicated lower 2D:4D digit ratios in the group with DLD. Modest associations were found between 2D:4D digit ratios and some Clinical Evaluation of Language Fundamentals–Fourth Edition subtests.

Conclusions: Prenatal hormone exposures may play a role in the etiology of some language symptoms.

Biological sex contributes to both physical and psychiatric risks. For example, as a general rule, neurodevelopmental disorders occur more often in boys than girls, although the level of asymmetric risk varies across conditions. On one end, a considerably elevated male–female ratio of 4.5:1 exists among individuals with autism spectrum disorder (Centers for Disease Control and Prevention, 2016), and ratios ranging from 2:1 to 3:1 have been provided for the prevalence of attention-deficit/hyperactivity disorder (ADHD; Ramtekhar, Reiersen, Todotov, & Todd, 2010). Likewise, an estimated ratio of 2:1 exists for individuals with intellectual disability (Ropers, 2008). On the other end, smaller but still elevated probabilities associated with male sex have been reported for developmental reading and language disorders. A systematic review by Liederman, Kantrowitz, and Flannery (2005)

estimated the sex ratio for dyslexia to be between 1.74:1 and 2:1, and a meta-analysis by Whitehouse (2010) reported a pooled sex ratio of 1.73:1 for siblings and a 1.54:1 for parents across familial aggregation studies of specific language impairment (SLI), the most common form of developmental language disorder (DLD).

Potential biological mechanisms behind sex differences in neurodevelopmental risk include, but are not limited to, direct sex chromosome effects, sex-specific activation patterns over the course of development of gonadal and nongonadal hormone secretion, and prenatal gonadal hormone exposures (see Becker et al., 2008; Knickmeyer & Baron-Cohen, 2006; and Neave, 2008, for reviews). Research suggests that fetal testosterone exposures in particular may have direct organizational effects on human corpus callosum size and asymmetry that in turn may help shape structural brain development, cognition, and behavior (e.g., Chura et al., 2010). In this report, we utilized relative finger lengths, a retrospective biomarker of prenatal hormone exposures, to examine possible associations between prenatal hormone exposures and linguistic abilities in a sample of boys with and without DLD.

^aUniversity of Utah, Salt Lake City

Correspondence to Sean M. Redmond: sean.redmond@health.utah.edu

Editor-in-Chief: Krista Wilkinson

Editor: Carol Miller

Received April 17, 2017

Revision received May 26, 2017

Accepted June 22, 2017

https://doi.org/10.1044/2017_JSLHR-L-17-0143

Disclosure: The authors have declared that no competing interests existed at the time of publication.

Male and female humans are exposed to very different perinatal hormone environments (see Aueyung, Lombardo, & Baron-Cohen, 2013, for a review). For women, in most cases, the circulating levels of testosterone are relatively constant and set to very low levels throughout prenatal and neonatal development. In contrast, for men, in most cases, the course of circulating testosterone levels during the prenatal and neonatal periods are characterized by two pronounced surges. The first and larger of the two surges takes place between 8 and 24 weeks of gestation, peaking around 16 weeks. During this period, circulating testosterone in the male fetal environment spikes to levels 2.5 times higher than those in the female counterpart. The second neonatal testosterone surge takes place shortly after birth, peaking around 10–20 weeks. Sometimes referred to as “mini-puberty” (Neave, 2008), the second surge is characterized by a smaller peak, roughly a third of the magnitude of the first surge. By week 30, levels of circulating testosterone have returned to baseline levels for men and remain at levels slightly higher than women until puberty.

Experiments on animals document the organizational effects of prenatal testosterone on the hypothalamus, limbic system, and neocortex as well as its effects on sexually dimorphic behaviors such as aggression, activity level, and spatial navigation (Neave, 2008). Direct manipulation in humans is impossible, and direct measurement of prenatal gonadal hormone exposures would involve risky invasive procedures. Fortunately, proxy biomarkers have been suggested that are readily available for human research that, although retrospective, have links to prenatal gonadal hormone exposures within mammalian and avian species. In the course of fetal development, gonads, brains, fingers, and toes are built around the same time, and variation in these structures reflects intrauterine gonadal hormone exposures (Galis, Ten Broek, Van Dongen, & Wijnaendts, 2010). Mammalian variation in digit length in particular and its potential association with other sexual dimorphisms have generated considerable interest (Manning, 2002).

Digit length proportions are similar in mice and humans. In both species, males generally have shorter second digits (2D) relative to their fourth digits (4D), whereas females generally present with relatively longer 2D relative to 4D. Epigenetic mechanisms responsible for these sex differences have been identified in a study of mice through direct measurement of fetal digit cartilage condensations and manipulation of prenatal hormone exposures at different periods of gestation (Zheng & Cohn, 2011). Several key findings from this study are relevant here. First, Zheng and Cohn showed that individual digits differed in their distribution of testosterone and estrogen receptors. In particular, the developing 4D had higher concentrations of testosterone receptor cells than 2D in fetuses from both sexes. Second, experimentally induced increases and decreases in hormone exposures produced digit-specific accelerations and decelerations of cartilage growth. Third, the effects of hormone exposures on differential digit cartilage growth were limited to the digit-forming stages of embryonic development.

Postnatal exposures had no effect. Finally, the observed digit-forming stages corresponded to the stages when gonadal hormones are known to masculinize and feminize the mouse brain. Thus, the ratio of 2D:4D in mice (and presumably in humans) appears to be an embryonic artifact that provides a retrospective window into the relative balance of estrogen-to-testosterone exposures encountered within the uterine environment at a critical stage of neural development.

Human sexual dimorphism in finger lengths has been known for over a century (Ecker, 1875), and although fingers obviously grow throughout the course of physical maturation, an individual's 2D:4D digit ratio is established by the 14th week of gestation and, for the most part, does not change significantly over the life span (see Galis et al., 2010). The 2D:4D digit ratio captures variation within men and women as well, such that individuals can have more masculinized (lower) or more feminized (higher) values for their digit ratios relative to other members of the same sex (Manning, 2002). One demonstration of this variability comes from multiple births where hormonal exposures are influenced by the shared intrauterine environment. Women from opposite-sex fraternal twin pairs have been shown to have significantly lower 2D:4D digit ratios than women from same-sex fraternal twin pairs due to testosterone exposures brought in by their co-twin (Voracek & Dressler, 2007).

As a retrospective marker of relative prenatal hormonal exposures, the 2D:4D digit ratio has been successfully linked to an extensive set of human sex-typed behaviors, traits, attributes, competencies, and clinical phenotypes (see Manning, 2002, for a review of this research program). However, the absence of investigations of the 2D:4D digit ratio and phenotypes associated with SLI and other DLDs represents a noticeable gap.

In addition to the presence of asymmetric sex ratios in familial risk rates, there are other indirect lines of research that encourage suspicion that a link might exist between low 2D:4D digit ratios and linguistic deficits. For example, studies of typical adults have found associations between lower 2D:4D digit ratio values and lower verbal IQ scores within the normal range of ability (Burton, Henninger, & Hafetz, 2005: $r = .41$, Cohen's $d = 0.90$; Luxen & Buunk, 2005: $r = .52$, Cohen's $d = 0.52$). In a study sample of seventy-five 6- to 7-year-old schoolchildren enrolled in regular education, Brosnan (2008) found that, although 2D:4D digit ratio did not correlate with individual academic assessments of numeracy and literacy, it did correlate with the magnitude of the discrepancy between children's numeracy and literacy assessments. Lower digit ratios were associated with larger relative advantages in numeracy to literacy performance, especially in boys ($r = -.37$, Cohen's $d = 0.88$).

Another reason to suspect a possible 2D:4D link to DLD is the presence of more masculinized ratios relative to normative expectations in other neurodevelopmental disorders. Although there have been some negative results (e.g., Guyatt, Heron, Le Cornu Knight, Golding, & Rai,

2015; Lemiére, Boets, & Danckaerts, 2010), multiple studies report significant group differences between typically developing (TD) controls and individuals with autism (Al-Zaid, Alhader, & Al-Ayadhi, 2015; de Bruin, Verheij, Wigeman, & Ferdinand, 2006; Manning, Baron-Cohen, Wheelwright, & Sanders, 2001; Milne et al., 2006; Teatero & Netley, 2013) as well as individuals with ADHD (de Bruin et al., 2006; Stevenson et al., 2006). In contrast, there is currently little support for links between lower 2D:4D digit ratios and dyslexia and intellectual disability. For example, both van Gelder, Tijms, and Hoeks (2005) and Boets, De Smedt, Wouters, Lemay, and Ghesquiere (2007) failed to find significant group differences between their participants with and without dyslexia (but see Manning, 2005, for a critique), and Ypsilanti, Ganou, Koidou, and Grouios (2008) found significant group differences between their participants with and without intellectual disabilities but in the opposite direction. In other words, Ypsilanti et al. reported that 2D:4D values were significantly higher in their participants with intellectual disability.

When reported, links between individuals' 2D:4D ratios and their clinical symptoms have generally been more robust in men than women and stronger in measurements collected on the right hand (Martel, Gobrogge, Breedlove, & Nigg, 2008; McFadden, Westhafer, Pasanen, Carlson, & Tucker, 2005). It is not clear why there have been sex and hand differences in these observed effects (see Hönekopp & Watson, 2010, for a discussion). The clinical values of observed group differences have also been open to interpretation. For example, Al-Zaid et al. (2015) suggested that 2D:4D digit ratios should be collected as part of routine screening for autism in young boys based on the magnitude of observed differences between their group of thirty-one 3- to 8-year-old participants with autism and their control group. Other investigators have cautioned against overinterpreting 2D:4D group differences when trying to estimate individual risk (Guyatt et al., 2015).

Perhaps, the most direct evidence available linking variations in early hormonal exposures to linguistic deficits comes from Whitehouse et al. (2012). In this study, umbilical cord blood samples collected on 767 births were used to measure testosterone concentrations. These values were used to predict variation in parental reports of communication development at 1, 2, and 3 years old. Results indicated that boys in the highest quartile of testosterone levels were at an increased risk for language delay (odds ratio = 2.47, 95% CI [1.12, 5.47]), whereas this relationship was not observed in girls. The results of this study are provocative in implicating a potential role for variations in perinatal testosterone levels in boys in the etiology of SLI and other DLDs, but they are not conclusive. Variation in gonadal hormone levels at birth does not necessarily correspond to variation in exposure levels during prenatal development, and hormone levels can be influenced by stress brought in by the birth process itself. The behavioral evidence is also currently limited to the earliest stages of development. Stronger evidence for the contribution of prenatal testosterone exposures to persistent idiopathic language disorders would

come from associations between children's 2D:4D digit ratios and their proficiencies with established clinical markers (e.g., nonword repetition [NWR], tense marking, and sentence recall; see Pawlowska, 2014).

In summary, although somewhat contentious, evidence gathered from a variety of clinical and nonclinical study samples suggests that potential links between the 2D:4D digit ratio and DLDs in school-age children warrant investigation.

Research Questions

In this study, we collected right-hand index finger (2D) and ring finger (4D) length measurements on a study sample of 105 boys who had participated in community-based screenings for DLDs. We restricted our investigation to boys and to their right hands to reduce confounds and because these conditions have provided the largest recorded 2D:4D digit ratio effects associated with neurodevelopmental disorders and other sex-linked conditions (see Hönekopp & Watson, 2010, for a meta-analysis). We then compared our participants' 2D:4D digit ratios with their performances on a battery of linguistic and nonlinguistic measures to address the following questions:

1. Are there 2D:4D digit ratio differences between boys with DLDs and boys with typical language (TL)?
2. What are the associations between 2D:4D digit ratios and psycholinguistic, nonverbal, and behavioral measures?

Method

Approval for this study was received by the University of Utah Institutional Review Board. Written parental consent and child assent were obtained from each participant. All testing was conducted by examiners naïve to the participants' developmental status.

Participants

Boys recruited to participate in this study had taken part in a larger community-based study examining verbal screening measures for students from kindergarten to Grade 3 ($N = 1,060$). All potential participants passed a hearing screening (at 25 dB across 500, 1000, 2000, and 4000 Hz) and were monolingual English speakers. Potential dialectal variations were not systematically considered. Parents completed a questionnaire about their son's health history and their interactions with developmental specialists. Children who were receiving services for learning difficulties or emotional disorders or had positive histories of developmental or health difficulties were included in the sample.

Group Assignment

In this study, DLD status (DLD, TL) was determined using the Clinical Evaluation of Language Fundamentals—Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003)

core language standard score and a cutoff at or below 85. The CELF-4 represents a widely used omnibus language measure (Betz, Eickhoff, & Sullivan, 2013) that has demonstrated adequate levels of sensitivity and specificity (sensitivity = 0.87, specificity = 0.96; Spaulding, Plante, & Farinella, 2006). The CELF-4 manual reports test–retest reliability ranges from .67 to .94 on the subtests administered to the age ranges in this study. Because epidemiological reports consistently indicate that children who would meet research criteria for SLI and other DLDs are frequently overlooked in school settings (Beitchman, Wilson, Brownlie, Walters, & Lancee, 1996; Norbury et al., 2016; Tomblin & Nippold, 2014; Zhang & Tomblin, 2000), receipt of services from a speech-language pathologist was not used as an inclusionary criterion. Parents reported that 75.9% (22/29) of the children in the group with DLD and 3.9% (3/76) of children in the TL group had received services from a speech-language pathologist.

Estimates of participants' nonverbal abilities were based on their performances on the Naglieri Nonverbal Abilities Test (NNAT; Naglieri, 2003), and parents completed the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), providing estimates of their son's socioemotional behavioral difficulties. The NNAT and CBCL were included in this report for descriptive purposes and did not influence group placement. This allowed for both groups to include participants with weaknesses in these areas. The NNAT has a similar format to Raven's Standard Progressive Matrices with a moderately high correlation between the two instruments reported in the manual ($r = .78$; p. 51). Sensitivity and specificity rates associated with the capacity of the NNAT to differentiate intellectual disabilities from cases of typical development are 0.69 and 0.98, respectively (p. 51). The CBCL has been shown to have acceptable levels of reliability with excellent divergent and convergent validity across a variety of socioemotional and behavioral symptoms when compared with other behavioral measures (Nakamura, Ebesutani, Bernstein, & Chorpita, 2009).

Combining parental reports of services and diagnoses with our testing data (viz. CELF-4 and NNAT standard scores ≤ 85) indicated that 58.6% of our cases of DLD would be characterized as having a profile consistent with SLI (SLI = 17, low nonverbal IQ + LI = 3, ADHD + LI = 3, autism + LI = 1, behavioral disorder + LI = 1, low nonverbal IQ + ADHD + LI = 2, ADHD + behavior disorder + LI = 2). Similarly, most of our cases of TL (86.8%) would be characterized as TD, as the term has been commonly used in previous research (TD = 66, ADHD = 4, behavior disorder = 2, autism = 3, autism + behavior disorder = 1).

Characteristics of participants from the group with DLD and the TL group are summarized in Table 1. Age distributions within the group with DLD and the TL group were nearly identical, $t(103) = -0.064$, $p = .95$.

Proportionally, there were more participants parentally identified as non-White in the group with DLD than the TL group, but this difference did not reach statistical

significance—race: $\chi^2(5, N = 105) = 7.23$, $p = .204$; ethnicity: $\chi^2(1, N = 105) = 0.406$, $p = .526$. Specific minority community distributions for each group were as follows: Hispanic: DLD = 2/29 (6.9%), TL = 3/76 (3.9%); American Indian or Alaskan Native: DLD = 0/29 (0%), TL = 1/76 (1.3%); Asian: DLD = 4/29 (13.8%), TL = 4/76 (5.3%); Black: DLD = 2/29 (6.9%), TL = 2/76 (2.6%); Native Hawaiian or Pacific Islander: DLD = 1/29 (3.4%), TL = 1/76 (1.3%); and mixed race: DLD = 1/29 (3.4%), TL = 0/76 (0%).

Consistent with an extensive body of research (see Leonard, 2014, for a review), significant differences were observed between the affected and unaffected groups in several areas. The TL group displayed advantages relative to the group with DLD in reported levels of maternal education, $t(103) = -3.40$, $p = .001$, Cohen's $d = 0.79$, and in their performance levels on standardized measures of verbal and nonverbal abilities—CELF-4: $t(103) = -12.18$, $p < .001$, Cohen's $d = 2.94$; NNAT: $t(103) = -5.04$, $p < .001$, Cohen's $d = 1.17$. In addition, extracted census tract data based on participants' addresses indicated that participants in the TL group tended to reside in census tracts with lower rates of family poverty than the group with DLD, $t(103) = 3.31$, $p = .001$, Cohen's $d = 0.74$. In contrast, the affected and unaffected groups were similar in parentally reported levels of socioemotional behavioral difficulties, $t(103) = 1.52$, $p = .131$, Cohen's $d = 0.34$.

Although the range of NNAT scores associated with both groups included both low and high scores, overall, the TL group's performance would be characterized as “moderately above average,” whereas the nonverbal abilities of the group with DLD were more consistent with the instrument's mean. The observed nonverbal advantage in our TL group relative both to published test norms and to our participants with DLD is a commonly reported phenomenon (see Gallinat & Spaulding, 2014, for a meta-analysis of study samples of SLI).

As indicated by obtained CELF-4 core language standard scores, general severity levels of language disorder associated with our DLD group would be characterized as “moderate–severe,” but across individual participants, the range went from “mild–moderate” to “severe.” The TL group's verbal standard scores ranged from “borderline/low normal” to “gifted” with a mean value very close to the instrument's mean.

Measures

Language

Three additional measures of psycholinguistic performance were collected that correspond to emerging phenotypic markers of SLI in English-speaking children: nonword repetition (Estes, Evans, & Else-Quest, 2007), tense marking (Rice & Wexler, 1996), and sentence recall (Archibald & Joanisse, 2009; Conti-Ramsden, Botting, & Faragher, 2001).

NWR. We administered Dollaghan and Campbell's (1998) NWR task to capture children's phonological processing abilities. NWR represents a well-established area of

Table 1. Group means (standard deviations) and ranges for participant characteristics.

Group	Age ^a	Race ^b	Ethnicity ^c	Maternal ed. ^{d,*}	Census tract ^{e,**}	Verbal ^{f,**}	Nonverbal ^{g,**}	Behavioral ^h
DLD (<i>n</i> = 29)	7;8 (1;10) 5;6–11;0	72.4%	93.1%	3.24 (0.79) 1–5	25.05 (5.55) 11.0–35.4	66.10 (15.50) 40–85	96.59 (15.62) 73–122	52.88 (11.32) 34–80
TL (<i>n</i> = 76)	7;8 (1;7) 5;6–10;10	89.5%	96.1%	3.86 (0.89) 2–5	20.98 (5.61) 9.30–37.90	104.39 (10.23) 87–132	114.49 (13.92) 88–161	50.11 (10.22) 26–76

Note. DLD = developmental language disorder; TL = typical language.

^aAge in years;months. ^bPercent White. ^cPercent non-Hispanic. ^dMaternal educational level scale: 1 = *no high school diploma*, 2 = *high school graduate*, 3 = *some college*, 4 = *college graduate*, and 5 = *postgraduate*. ^eThe percentage of families within participant's census tract living in poverty. ^fClinical Evaluation of Language Fundamentals–Fourth Edition Core Language Score (*M* = 100, *SD* = 15). ^gNaglieri Nonverbal Abilities Test standard score (*M* = 100, *SD* = 15). ^hChild Behavior Checklist Total Problem T score (*M* = 50, *SD* = 10); higher scores indicate elevated levels of difficulty.

p* < .01. *p* < .001 (differences of the group with DLD and the TL group).

weakness for children with SLI and other DLDs (see Estes et al., 2007, for a meta-analysis). The nonword stimuli were presented via digital audio recording of a female speaker presenting the items. Items on the NWR include 16 nonwords that gradually increase from one to four syllables over the duration of the task, yielding a percentage of correct phonemes.

Test of Early Grammatical Impairment. The Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001) was administered to measure children's tense marking abilities. The Elicited Grammar Composite provides a percent correct use of the targeted finiteness structures (third person singular present tense *-s*, past tense, BE and DO copulas and auxiliaries). Adequate levels of sensitivity (0.81) and specificity (0.95) are associated with the TEGI Elicited Grammar Composite when the recommended cut-offs provided by the manual are used (Spaulding et al., 2006).

Redmond Sentence Recall. We used the Redmond (2005) sentence recall (RSR) task that consists of eight simple passive and eight simple active declarative sentences matched for number of words (9–12). Children listened to a digital audio presentation of the sentences by an adult female speaker. Sentences received a score of 0 (four or more errors), 1 (three or fewer errors), or 2 (correct), with a maximum of 32 possible points. Previous work documents adequate levels of differentiation (sensitivity = 0.85, specificity = 0.90) when the RSR is used to segregate cases of language impairment from TL in the age range associated with this study sample (Archibald & Joanisse, 2009).

Finger Measurement

Participants were asked to fan out their fingers and place their right hand onto a digital scanner. A soft cloth bag filled with rice was placed over each participant's hand to provide a standard pressure weight. Finger lengths were derived from the scanned images using digital calipers calibrated before each measurement on a reference image. Measurements were taken down the center of each finger, from the highest point on the tip to the basal crease.

Reliability

Finger lengths were measured independently by two examiners naïve to participants' status from the scanned image to calculate interrater consistency. The association between raters was highly positive and statistically significant ($r = .92, p < .001$). Estimates for reliability associated with the NWR, RSR, and TEGI measures were calculated using 20% of the sample (21/105). A second examiner watched video recordings of original testing sessions and rescored each measure. Pearson correlation coefficients were calculated between the original testing session and the second scoring for the NWR ($r = .92, p < .001$), RSR ($r = .99, p < .001$), and TEGI ($r = .99, p < .001$), indicating high level of interrater scoring consistency.

Results

Complete data were available for all 105 participants.

Analysis Strategy

A planned independent *t* test (two-tailed) was conducted to examine potential 2D:4D digit ratio differences between the group with DLD and the TL group. The group comparison was followed by a series of planned zero-order Pearson product correlations (two-tailed) to examine potential associations between 2D:4D digit ratios and each psycholinguistic, nonverbal, and behavioral measure using the entire study sample. We did not apply Bonferroni adjustments because these are widely recognized as overly conservative, would not provide a fair test, and apply only in situations of excessive unplanned post hoc comparisons (Kusuoka & Hoffman, 2002; Perneger, 1988; Rice, 1989; Rothman, 1990; Savitz & Olshan, 1995).

Raw scores on the NWR, sentence recall, and tense marking indices were transformed into standard scores, (z score $\times 15$) + 100, before analyses using mean and standard deviation values obtained from the larger community-based sample associated with the verbal screening project. Because the distribution of TEGI scores was heavily skewed,

these data were transformed using square root transformation before analyses. However, for ease of interpretation, standard scores are provided in Table 2. It is important to note that TEGI standard scores in our study sample also needed to be truncated to accommodate for the presence of extremely low scores in some cases ($n = 7$) relative to age expectations in the affected group. For example, for children aged 7 years, a composite tense score of 84 on the TEGI would be more than 6 *SDs* below the expected mean for this age (98) providing a negative standard score value. In these cases, derived standard scores less than 1 were replaced with 1.

Group Differences

Within-group variations are presented for comparison as box plots in Figure 1. Following conventional definitions of outlier designations (i.e., ± 3.0 *SDs*; see Osborne & Overbay, 2008), no outliers were present in the study sample.¹ Thus, all cases were included in the analyses. As expected, significant advantages for the group with TL were observed on each of the language indices—NWR: $t(103) = -8.19, p < .001$, Cohen's $d = 1.98$; RSR: $t(103) = -10.60, p < .001$, Cohen's $d = 2.57$; TEGI: $t(103) = -10.09, p < .001$, Cohen's $d = 2.63$ (Table 2). There was also a significant difference between the TL group and the group with DLD in their 2D:4D digit ratios, $t(103) = -2.20, p = .03$, Cohen's $d = 0.48$. On average, boys in the group with DLD presented with a lower 2D:4D digit ratio than boys in the TL group (0.9345 vs. 0.9516). However, the range of observed 2D:4D values associated with each group was very similar.

Associations Among 2D:4D Digit Ratios and Psycholinguistic, Nonverbal, and Behavioral Measures

Data from all 105 participants were pooled to examine potential associations between participants' 2D:4D digit ratios and their linguistic proficiencies treated as continuous variables as well as their nonverbal and behavioral scores (Table 3). A small but significant positive correlation between our participants' digit ratios and their CELF-4 Core Language standard score was observed ($r = .22, p = .03$). In contrast, correlations between 2D:4D digit ratios and the other language measures were closer to zero and did not reach statistical significance (range of r values = .07–.18). Similarly, associations between 2D:4D digit ratios and the nonverbal and behavioral measures were not significant.

¹The observed 2D:4D digit ratio value of 1.05 in one of our cases in the TL group has been regularly reported in other study samples (e.g., de Bruin et al., 2006; Guyatt et al., 2015; Noipayak, 2009). Furthermore, according to Hönemann and Watson's (2010) meta-analysis, a 2D:4D digit ratio value of 1.05 falls within 3.0 *SDs* based on pooled population estimates. Nonetheless, to consider the potential impact of this particular score on our outcomes, we ran all analyses excluding this individual case. Results were essentially the same.

Associations Among 2D:4D Digit Ratios and CELF-4 Subtests

The presence of a significant correlation between our participants' 2D:4D digit ratios and their CELF-4 core language standard score motivated a follow-up analysis examining the contributions of individual CELF-4 subtests to this association (Table 4). For children aged 5–8 years, this involved the Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences scaled scores. For children aged 9–11 years, this involved the Concepts and Following Directions, Recalling Sentences, Formulated Sentences, and Word Classes scaled scores. Results revealed a significant positive correlation between the participants' 2D:4D digit ratio and their scores on the Concepts and Following Directions ($r = .20, p = .04$) and Formulated Sentences ($r = .23, p = .02$). Performance on Word Classes had a higher correlation with 2D:4D digit ratio ($r = .37, p = .16$) than either Concepts and Following Directions or Formulated Sentences but did not reach statistical significance, probably the result of the smaller number of our participants in the older age range ($n = 18$).

Discussion

Male sex is widely recognized as a general risk factor for neurodevelopmental disorder—most notably for autism and ADHD. To a lesser extent, this has been true for idiopathic language disorders as well. In this study, we examined the extent to which evidence could be found for a role for prenatal hormone exposures in school-age children's linguistic symptoms. We used the 2D:4D digit ratio, a retrospective biomarker of estrogen exposure levels relative to testosterone levels during the second trimester, and found a significant group difference in our study sample between boys with DLD, a designation that included cases of SLI as well as cases that did not meet the criteria for SLI, and boys with TL skills. Consistent with previous study samples of boys with autism and ADHD, participants with DLD presented with lower (i.e., more masculinized) 2D:4D digit ratios relative to the control group, suggesting that, as a group, boys with DLD had experienced relatively elevated levels of prenatal testosterone exposure. The observation of differences between groups based on their linguistic abilities was supplemented by correlational analyses. Consistent with previous studies of other neurodevelopmental disorders, a significant association was observed on some but not all aspects of the developmental domain considered. In our case, a modest but nonetheless significant correlation was found between 2D:4D digit ratios and our participants' CELF-4 core language standard scores. At first glance, this suggests that more masculinized digit ratios were associated with generally lower levels of language abilities. However, follow-up analyses revealed that this association was, to a large extent, the result of participants' scores on the Concepts and Following Directions, Formulated Sentences, and, perhaps, Word Class subtests. In contrast, there was little evidence that scores on the CELF-4 Recalling

Table 2. Group means (standard deviations) and ranges for 2D:4D digit ratio and psycholinguistic indices.

Group	2D:4D ^a	NWR ^{a,**}	TEGI ^{b,**}	RSR ^{c,**}
DLD (<i>n</i> = 29)	0.9345 (0.04) 0.87–1.01	65.08 (21.10) 40–105.52	28.48 (35.67) 1–107.26	65.90 (17.64) 40–106
TL (<i>n</i> = 76)	0.9516 (0.03) 0.88–1.05	101.01 (14.90) 60.88–123.24	100.68 (14.53) 57.98–116.04	103.58 (12.64) 58–129

Note. DLD = developmental language disorder; TL = typical language.

^aDollaghan and Campbell's nonword repetition task (NWR): standard score (*M* = 100, *SD* = 15). ^bTest of Early Grammatical Impairment (TEGI): Elicited Grammar Composite standard score (*M* = 100, *SD* = 15).

^cRedmond Sentence Recall (RSR): standard score (*M* = 100, *SD* = 15).

p* < .05. *p* < .001 (differences of the group with DLD and the TL group).

Sentences or Word Structure subtests were contributing to this association. Similarly, there were no significant associations between 2D:4D digit ratios and our participants' standard scores on the three most frequently suggested clinical markers for English-speaking children (see Pawlowska, 2014). This discrepancy might indicate an etiological difference between what could be characterized as primary symptoms of language disorder and secondary or peripheral symptoms associated with language disorder. Put differently, variation across measures might have reflected the presence of different neurodevelopmental processes underlying grammatical processing and verbal memory resources on one hand and conceptual/semantic domains on the other.

Alternatively, our null findings with these particular clinical markers of language disorder might have been due to various measurement limitations. Specifically, the NWR, tense marking, and sentence recall measures we used might not have sufficiently captured the range of linguistic performance needed to find robust correlations because, as essentially criterion measures, they have been

primarily designed to segregate affected from unaffected groups. However, this explanation is complicated by a few observations associated with our data and appears to only be plausible with the tense marking measure we used. First, we utilized community-based means and standard deviations to transform our participants' performance on the clinical measures into standard scores before analyses. Only the TEGI required additional transformations to adjust for skewed distribution. Second, the range of standard scores as well as group standard deviations associated with the nonword and sentence recall measures were both consistent with the range of scores and standard deviations associated with the CELF-4 core language score. Thus, the CELF-4, nonword, and sentence recall measures appeared to capture a similar range of linguistic performance.

Another potential measurement limitation complicating the case for etiological differences across language domains was the presence of variability in test–retest reliability associated with CELF-4 subtests. Unfortunately, we are not in a position to readily resolve this issue with our data because CELF-4 subtest reliability also varies as a function of age.

The results of this study were limited further by a variety of factors, and they should therefore be interpreted with caution. The size of effect associated with the observed 2D:4D digit ratio difference between the group with DLD and the TL group means (Cohen's *d* = 0.48) would be characterized as small or perhaps approaching the threshold

Figure 1. Box plots of 2D:4D digit ratios for groups with developmental language disorder (DLD) and typical language (TL). **Table 3.** Associations among 2D:4D digit ratios and psycholinguistic, nonverbal, and behavioral measures (*N* = 105)

Digit ratio	CELF-4	NWR	TEGI	RSR	NNAT	CBCL
2D:4D	.220*	.039	.192	.132	.134	.032

Note. CBCL = Child Behavior Checklist total *t* score; CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition: Core Language standard score; NNAT = Naglieri Nonverbal Abilities Test: standard score; NWR = Dollaghan and Campbell's nonword repetition task: standard score; RSR = Redmond Sentence Recall: standard score; TEGI = Test of Early Grammatical Impairment: Elicited Grammar Composite transformed scores.

**p* < .05.

Table 4. Associations among 2D:4D digit ratio and CELF-4 subtests.

Digit ratio	CELF-4				
	CFD (<i>N</i> = 105)	WS (<i>n</i> = 87)	RS (<i>N</i> = 105)	FS (<i>N</i> = 105)	WC (<i>n</i> = 18)
2D:4D	.204*	.191	.178	.240*	.363

Note. CELF-4 = Clinical Evaluation of Language Fundamentals–Fourth Edition; CFD = Concepts and Following Directions subtest; FS = Formulated Sentences subtest; RS = Recalling Sentences subtest; WC = Word Classes subtest; WS = Word Structure subtest.

**p* < .05.

of a moderate designation. Similarly, the associations between our participants' 2D:4D digit ratios and their language scores were consistently small. Replications are needed to provide stronger evidence of a role for prenatal hormone exposures in the etiology of SLI and other DLDs in male humans and to further test the characterization that these modest effects are limited to either peripheral or conceptual/semantic aspects of language. Similarly, power limitations could have compromised our ability to detect some of the smaller effects of prenatal hormone exposures that would have been detected within a larger sample. This seems to likely be the case with our observation of nonsignificant associations between the 2D:4D digit ratio and the Word Classes subtest on the CELF-4 administered to our older participants. Additional research focusing on older affected children and young adults is needed to determine if there are links between 2D:4D digit ratio and advanced semantic abilities. Previous research on unaffected adults using verbal IQ measures suggests a possible link (Burton et al., 2005).

Another consequence of the modest effect sizes and their related power limitations is that we were not able to explore potential contributors to observed group differences and associations. For example, in addition to large and statistically significant group differences in participants' linguistic abilities, the TL group presented with significant advantages relative to the group with DLD in nonverbal abilities and maternal educational levels, and either or both of these variables may have contributed to our results. The 2D:4D digit ratio is known to vary across racial and ethnic groups (Hönekopp & Watson, 2010). Because both the group with DLD and the TL group were primarily White and non-Hispanic and group differences were not statistically significant, this threat to internal validity was minimized. However, the presence of a 17.1% difference in proportional representations of non-White participants between groups suggests that it may not have been entirely eliminated. Furthermore, unaccounted for dialectal differences associated with nonsignificant racial and ethnic differences could still have influenced the results of this study. However, this concern is mitigated by key aspects of the language measures used in this study. First, the linguistic measures most likely to be affected by dialectal variation would be the Word Structure subtest of the CELF-4 (Semel et al., 2003, pp. 305–313) and the TEGI,

and neither of these measures were significantly associated with participants' 2D:4D digit ratio. Moreover, there was no significant association between participants' performance on the NWR task and their 2D:4D digit ratios, and NWR tasks have been developed to provide a dialect neutral index of language processing (e.g., Dollaghan & Campbell, 1998). Thus, if dialectal differences were introducing error into our results, they would be more likely in the direction of underestimating rather than overestimating links between linguistic abilities and the 2D:4D digit ratio.

Our results were limited to the specific measures of linguistic ability we selected. We based our selection on emerging phenotypes of SLI and other DLDs (Pawlowska, 2014), but this did not exhaust the possible aspects of language disorder that could have been explored. Likewise, longitudinal data would provide a stronger test of our research questions and might also reveal additional effects on the course of development. In a longitudinal study of receptive vocabulary growth in children with and without SLI covering development from 2 to 21 years old, Rice and Hoffman (2015) documented an early female advantage in vocabulary that is replaced later by a male advantage around 12 years old. This outcome undermined the commonly held generalization that female advantage in language skills was constant across development. Furthermore, the observed age-sensitive sex advantages operated the same in both affected and unaffected groups indicating that they did not interact with children's linguistic deficits. The general timing of the shift in semantic growth patterns between men and women suggests that it could reflect the influence of postnatal hormonal exposures associated with pubescence rather than prenatal exposures.

When it arrives in the empirical record, the full story of testosterone's effects across the life span on different aspects of language development will undoubtedly be complex. Additional research is needed to uncover details behind the causal chain from genes and embryonic environment interactions to the biochemical underpinnings of neural organization and function to the development of cognitive and behavioral differences between and within the sexes. The results of this first study of the associations between linguistic abilities in school-age boys and the 2D:4D digit ratio proxy measure when combined with the results of previous studies of variation in early lexical development that directly measured young children's

circulating testosterone levels either through saliva (Kung et al., 2016) or umbilical cord blood (Whitehouse et al., 2012) or from amniotic fluid in high-risk pregnancies (Lutchmaya, Baron-Cohen, & Raggatt, 2002) suggest that perinatal testosterone exposures represent a viable biological contributor to variation in language skills. Although concerns have been raised regarding the extent to which the 2D:4D digit ratio underestimates human prenatal hormone exposures (e.g., Berenbaum, Bryk, Nowak, Qugley, & Moffat, 2009; Wallen, 2009), we agree with Hönekopp and Watson's (2010) observation that the apparent limitations of the 2D:4D digit ratio in this regard are offset by its considerable advantages as a widely accessible biomarker for initial explorations into different aspects of development.

Acknowledgments

Funding for this research was provided by the National Institute on Deafness and Other Communication Disorders Grant R01DC011023. Portions of this study were presented at the June 2016 Symposium for Research in Child Language Disorders, Madison, WI. We are greatly indebted to the children and their families for their participation. Appreciation is also extended to Lisa Holmstead (Salt Lake City School District) and Deb Luker (Jordan School District) for their assistance with recruitment. Several graduate and undergraduate students from the University of Utah's Department of Communication Sciences and Disorders assisted in various aspects of the project and deserve recognition for their contributions: David Aamodt, Peter Behnke, Hannah Caron, Kimber Campbell, Jessica Carrizo, Tyler Christopoulos, Faith Denzer, Olivia Erickson, Micah Foster, Elizabeth Hafen, Lyssandra Harker, Kristin Hatch, Nathan Lily, Amy Ludlow, Kristi Moon, Elie Muyankindi, Theresa Pfaff, McKenzie Rohde, Michelle Stettler, and Amy Wilder.

References

- Achenbach, T. M., & Rescorla, L. A., (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families.
- Al-Zaid, F. S., Alhader, A., & Al-Ayadi, L. Y. (2015). The second to fourth digit ratio (2D:4D) in Saudi boys with autism: A potential screening tool. *Early Human Development*, 91, 413–415.
- Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, 52, 899–914.
- Aueyung, B., Lombardo, M. V., & Baron-Cohen, S. (2013). Prenatal and postnatal hormone effects on the human brain and cognition. *Pfunders Archive: European Journal of Physiology*, 455, 557–571.
- Becker, J. R., Berkley, K. J., Geary, N., Hampson, E., Herman, J. P., & Young, E. A. (2008). *Sex differences in the brain: From genes to behavior*. New York, NY: Oxford University Press.
- Beitchman, J. H., Wilson, B., Brownlie, E. B., Walters, H., & Lancee, W., (1996). Long-term consistency in speech/language profiles: I. Developmental and academic outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry*, 35(6), 804–814.
- Berenbaum, S. A., Bryk, K. K., Nowak, N., Qugley, C. A., & Moffat, S. (2009). Fingers as a marker of prenatal androgen exposure. *Endocrinology*, 150, 5119–5124.
- Betz, S. K., Eickhoff, J. R., & Sullivan, S. F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language, Speech, and Hearing Services in Schools*, 44, 133–146.
- Boets, B., De Smedt, B., Wouters, J., Lemay, K., & Ghesquiere, P. (2007). No relation between 2D:4D fetal testosterone marker and dyslexia. *NeuroReport*, 18(14), 1487–1491.
- Brosnan, M. J. (2008). Digit ratio as an indicator of numeracy relative to literacy in 7-year-old British schoolchildren. *British Journal of Psychology*, 99, 75–85.
- Burton, L. A., Henninger, D., & Hafetz, J. (2005). Gender differences in relations of mental rotation, verbal fluency, and SAT scores to finger length ratios as hormonal indexes. *Developmental Neuropsychology* 28(1), 493–505.
- Centers for Disease Control and Prevention. (2016). *Key findings from the ADDM network: A snapshot of autism spectrum disorder*. Community Report on Autism. Retrieved from <http://www.cdc.gov/features/new-autism-data/community-report-autism-key-findings.pdf>
- Chura, L. R., Lombardo, M. V., Ashwin, E., Auyeung, B., Chakrabarti, B., Bullmore, E. T., & Baron-Cohen, S. (2010). Organizational effects of fetal testosterone on human corpus callosum size and asymmetry. *Psychoneuroendocrinology*, 35, 122–132.
- Conti-Ramsden, G., Botting, N., & Fragar, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, 42, 741–748.
- de Bruin, E. I., Verheij, F., Wigman, T., & Ferdinand, R. F. (2006). Differences in finger length ratio between males with autism, pervasive developmental disorder- not otherwise specified, ADHD, and anxiety disorders. *Developmental Medicine and Child Neurology*, 48, 962–965.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136–1146.
- Ecker, A. (1875). Some remarks about a varying character in the hands of humans. *Archives fur Anthropologie* 8, 67–75.
- Estes, K. G., Evans, J., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50, 177–195.
- Galis, F., Ten Broek, C. M., Van Dongen, S., & Wijnaendts, L. C. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D). *Archives of Sexual Behavior*, 39(1), 57–62.
- Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 57(4), 1363–1382.
- Guyatt, A. L., Heron, J., Le Cornu Knight, B., Golding, J., & Rai, D. (2015). Digit ratio and autism spectrum disorders in the Avon longitudinal study of parents and children: A birth cohort study. *BMJ Open*, 5, 1–8.
- Hönekopp, J., & Watson, S. (2010). Meta-analysis of digit ratio 2D:4D shows greater sex differences in the right hand. *American Journal of Human Biology*, 22(5), 619–630.
- Knickmeyer, R. C., & Baron-Cohen, S. (2006). Fetal testosterone and sex differences. *Early Human Development*, 82, 755–760.
- Kung, K. T., Browne, W. V., Constantinescu, Noorderhoven, R. M., & Hines, M. (2016). Early postnatal testosterone predicts sex-related difference in early expressive vocabulary. *Psychoneuroendocrinology*, 68, 111–115.

- Kusuoka, H., & Hoffman, J. I.** (2002). Advice on statistical analysis for circulation research. *Circulation Research*, 91, 662–671.
- Lemiere, J., Boets, B., & Danckaerts, M.** (2010). No association between the 2D:4D fetal testosterone marker and multi-dimensional attentional abilities in children with ADHD. *Developmental Medicine and Child Neurology*, 52, e202–e208. <https://doi.org/10.1111/j.1469-8749.2010.03684.x>
- Leonard, L. B.** (2014). *Children with specific language impairment*. Boston, MA: MIT Press.
- 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(2), 109–129.
- Lutchmaya, S., Baron-Cohen, S., & Raggatt, P.** (2002). Foetal testosterone and vocabulary size in 18- and 24-month-old infants. *Infant Behavior and Development*, 24, 418–424.
- Luxen, M. F., & Buunk, B. P.** (2005). Second-to-fourth digit ratio related to verbal and numerical intelligence and the big five. *Personality and Individual Differences*, 39, 959–966.
- Manning, J. T.** (2002). *Digit ratio: A pointer to fertility, behavior, and health*. New Brunswick, NJ: Rutgers University Press.
- Manning, J. T.** (2005). Reply to van Gelder, Tijms, & Hoeks (2005). *Developmental Medicine and Child Neurology*, 47, 719.
- Manning, J. T., Baron-Cohen, S., Wheelwright, S., & Sanders, G.** (2001). The 2nd to 4th digit ratio and autism. *Developmental Medicine and Child Neurology*, 43(3), 160–164.
- Martel, M. M., Gobrogge, K. L., Breedlove, S. M., & Nigg, J. T.** (2008). Masculinized finger-length ratios of boys, but not girls, are associated with attention-deficit/hyperactivity disorder. *Behavioral Neuroscience*, 122(2), 273–281.
- McFadden, D., Westhafer, G., Pasanen, E. G., Carlson, C., & Tucker, D. M.** (2005). Physiological evidence of hypermasculinization in boys with inattentive type of attention-deficit/hyperactivity disorder (ADHD). *Clinical Neuroscience Research*, 5, 233–245.
- Milne, E., White, S., Campbell, R., Swettenham, J., Hansen, P., & Ramus, F.** (2006). Motion and form coherence detection in autism spectrum disorder: Relationship to motor control and 2:4 digit ratio. *Journal of Autism and Developmental Disorders*, 36(2), 225–237.
- Naglieri, J. A.** (2003). *Naglieri Nonverbal Ability Test: Individual administration*. San Antonio, TX: Harcourt Assessment.
- Nakamura, B. J., Ebesutani, C., Bernstein, A., & Chorpita, B. F.** (2009). A psychometric analysis of the Child Behavior Checklist DMS-Oriented Scales. *Journal of Psychology and Behavioral Assessment*, 31, 178–189.
- Neave, N.** (2008). *Hormones and behavior: A psychological approach*. New York, NY: Cambridge Press.
- Noipayak, P.** (2009). The ratio of 2nd and 4th digit length in autistic children. *Journal of the Medical Association of Thailand*, 92, 1040–1045.
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... Pickles, A.** (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *The Journal of Child Psychology and Psychiatry*, 57(11), 1247–1257.
- Osborne, J. W., & Overbay, A.** (2008). Best practices in data cleaning: How outliers and “fringelers” can increase error rates and decrease the quality and precision of your data. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 205–214). Thousand Oaks, CA: Sage Publishing.
- Pawlowska, M.** (2014). Evaluation of three proposed markers for language impairment in English: A meta-analysis of diagnostic accuracy studies. *Journal of Speech, Language, and Hearing Research*, 57, 2261–2273.
- Perneger, T. V.** (1998). What’s wrong with Bonferroni adjustments. *British Medical Journal*, 18, 1236–1238.
- Ramtekkar, U., Reiersen, A., Todotov, A. A., & Todd, R.** (2010). 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*, 49(3), 217–228.
- Redmond, S. M.** (2005). Differentiating SLI from ADHD using children’s sentence recall and production of past tense morphology. *Clinical Linguistics and Phonetics*, 19(2), 109–127.
- Rice, M. L., & Hoffman, L.** (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research*, 58(2), 345–359.
- Rice, M. L., & Wexler, K.** (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research*, 39, 1239–1257.
- Rice, M. L., & Wexler, K.** (2001). *Test of Early Grammatical Impairment*. San Antonio, TX: The Psychological Corporation.
- Rice, W. R.** (1989). Analyzing tables of statistical tests. *Evolution*, 43, 223–225.
- Ropers, H. H.** (2008). Genetics of intellectual disability. *Current Opinion in Genetics and Development*, 18(3), 241–250.
- Rothman, K. J.** (1990). No adjustments are needed for multiple comparisons. *Epidemiology*, 1, 43–46.
- Savitz, D. A., & Olshan, A. F.** (1995). Multiple comparisons and related issues in the interpretation of epidemiologic data. *American Journal of Epidemiology*, 142, 904–908.
- Semel, E., Wiig, E. H., & Secord, W. A.** (2003). *The Clinical Evaluation of Language Fundamentals—Fourth Edition*. Toronto, Canada: The Psychological Corporation.
- Spaulding, T. J., Plante, E., & Farinella, K. A.** (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, 37, 61–72.
- Stevenson, J. C., Everson, P. M., Williams, D. C., Hipkind, G., Grimes, M., & Mahoney, E. H.** (2006). Attention deficit/hyperactivity disorder (ADHD) symptoms and digit ratios in a college sample. *American Journal of Human Biology*, 19, 41–50.
- Teatero, M. L., & Netley, C.** (2013). A critical review of the research on the extreme male brain theory and digit ratio (2D:4D). *Journal of Autism and Developmental Disorders*, 43, 2664–2676.
- Tomblin, J. B., & Nippold, M. A.** (2014). *Understanding individual differences in language development across the school years*. New York, NY: Taylor & Francis.
- van Gelder, M., Tijms, J., & Hoeks, J.** (2005). Second to fourth digit ratio and dyslexia: No evidence for an association between reading disabilities and the 2D:4D ratio. *Developmental Medicine and Child Neurology*, 47, 718–719.
- Voracek, M., & Dressler, S. G.** (2007). Digit ratio (2D:4D) in twins: Heritability estimates and evidence for a masculinized trait expression in women from opposite sex pairs. *Psychological Reports*, 100, 115–126.
- Wallen, K.** (2009). Does finger fat produce sex differences in second to fourth digit ratio? *Endocrinology*, 150, 4819–4822.
- Whitehouse, A. J. O.** (2010). Is there a sex ratio difference in the familial aggregation of specific language impairment? A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53, 1015–1025.
- Whitehouse, A. J. O., Matters, E., Maybery, M. T., Sawyer, M. G., Jacoby, P., Keelan, J. A., & Hickey, M.** (2012). Sex-specific associations between umbilical cord blood testosterone levels and

-
- language delay in early childhood. *Journal of Child Psychology and Psychiatry*, 53(7), 726–734.
- Ypsilanti, A., Ganou, M., Koidou, I., & Grouios, G.** (2008). Digit ratio (2D:4D) in individuals with intellectual disability: Investigating the role of testosterone in the establishment of cerebral lateralisation. *Laterality: Asymmetries of Body, Brain and Cognition*, 13(6), 527–544.
- Zhang, X., & Tomblin, J. B.** (2000). The association of intervention receipt with speech-language profiles and social-demographic variables. *American Journal of Speech-Language Pathology*, 9, 345–357.
- Zheng, Z., & Cohn, M.** (2011). Developmental basis of sexually dimorphic digit ratios. *Proceedings of the National Academy of Sciences*, 108(39), 16289–16294.