Published in final edited form as:

J Speech Lang Hear Res. 2008 August ; 51(4): 967–982. doi:10.1044/1092-4388(2008/071).

Vocabulary Abilities of Children with Williams Syndrome: Strengths, Weaknesses, and Relation to Visuospatial Construction Ability

Carolyn B. Mervis and Angela E. John

University of Louisville, Louisville, KY

Abstract

Purpose—This project was designed to identify relative strengths and weaknesses in vocabulary ability for children with Williams syndrome (WS) and to demonstrate the importance of stringent matching criteria for cross-group comparisons.

Methods—Children with WS and typically developing (TD) children completed standardized assessments of intellectual and language ability. Children with WS also completed a visuospatial construction ability assessment.

Results—Study 1: Concrete and relational vocabulary standard scores were significantly lower for 5 – 7-year-olds with WS than for TD children. Children with WS earned significantly higher standard scores for concrete than for relational vocabulary. When groups were stringently matched for relational vocabulary size, children with WS did not evidence a specific weakness in spatial vocabulary. Study 2: Standard scores for relational vocabulary were similar to those for visuospatial construction ability for 5 – 7-year-olds with WS. Study 3: 9 – 11-year-olds with WS demonstrated very limited relational vocabulary ability; relational vocabulary ability at 5 – 7 years was highly correlated with later relational language ability.

Conclusions—Concrete vocabulary is a relative strength for children with WS; relational vocabulary ability is very limited and is at about the level of visuospatial construction ability. Accurate determination of group comparison results depends on stringent group matching.

Keywords

Williams syndrome; intellectual disability; language acquisition; vocabulary; visuospatial construction

Williams syndrome (WS) is a neurodevelopmental disorder caused by the deletion of ~25 genes on chromosome 7q11.23 (Hillier et al., 2003; Osborne, 2006). The WS deletion, which has a prevalence of 1/7500 (Strømme, Bjørnstad, & Ramstad, 2003), results in a characteristic pattern of physical characteristics, including a distinctive facial appearance, cardiovascular disease (especially supralvalvar aortic stenosis), connective tissue abnormalities, and growth deficiency (see Mervis & Morris, 2007 for a review). Individuals with WS typically have mild to moderate intellectual disability, but some individuals have low average to average intellectual ability and a smaller minority have severe intellectual disability. WS is associated with a specific cognitive profile, including relative strengths in verbal short term memory and in language and

severe weakness in visuospatial construction (Mervis et al., 2000). The personality profile was one of the first diagnostic clues provided for WS (Bennett et al., 1978; Cassidy & Morris, 2002) and includes overfriendliness, gregariousness, and high levels of empathy, with an undercurrent of anxiety (Klein-Tasman & Mervis, 2003). The psychiatric profile includes ADHD, specific phobia, and generalized anxiety disorder beginning in early adolescence (Leyfer et al., 2006).

WS came to the attention of researchers in the United States because of the pioneering work of Ursula Bellugi (e.g., Bates, 1990; Bellugi, Marks, Bihle, & Sabo, 1988) arguing that WS provided a paradigmatic example of the independence of language from cognition. In particular, the language abilities of individuals with WS were described as largely intact, including excellent grammar and unusually strong vocabulary in the face of severe mental retardation. More recent research has tempered this view. In particular, although researchers agree that language is a relative strength for individuals with WS, grammatical abilities are generally considered to be at the level expected for mental age (MA); in languages that have complex morphology, morphological development may even be at a level slightly below that expected for MA (see reviews in Mervis, 2006 and Mervis & Becerra, 2007). In addition, despite their gregariousness, the pragmatic abilities of children with WS are often quite limited (Klein-Tasman, Mervis, Lord, & Phillips, 2007; Laws & Bishop, 2004; Philofsky, Fidler, & Hepburn, 2007; Stojanovic, 2006). Furthermore, language abilities are strongly correlated with non-verbal cognitive ability (e.g., Lukács, Pléh, & Racsmany, 2004; Mervis, 1999) and overall level of intellectual ability is almost always well above the level of severe mental retardation.

Vocabulary ability is generally considered to be the greatest language strength for children with WS. Bellugi and her colleagues (e.g., 1988, 2000) showed that age equivalent (AE) scores on the Peabody Picture Vocabulary Test-Revised (PPVT-R; Dunn & Dunn, 1981) were above the level expected for overall IQ. Karmiloff-Smith et al. (1997) found that AE scores on the British Picture Vocabulary Scale (BPVS; Dunn, Dunn, Whetton, & Pintilie, 1982) were higher than those for the Test for Reception of Grammar (TROG; Bishop, 1989). Mervis and Morris (2007) reported that mean standard score on the PPVT-III (Dunn & Dunn, 1997) was 9 points higher than on the TROG-2 (Bishop, 2003). This difference is especially striking given that whereas the lowest possible standard score on the PPVT-III is 20, the lowest possible standard score on the TROG-2 is 55, and 39% of a large sample of children with WS earned a 55 on this test (Mervis, 2007). Thus, if the lowest possible standard score on the TROG-2 was the same as on the PPVT-III, there likely would have been a considerably larger discrepancy between vocabulary and grammatical ability.

Vocabulary is not monolithic, however. The PPVT, which has been the focus of most studies of the vocabulary abilities of individuals with WS, emphasizes concrete vocabulary, primarily names for objects, actions, and descriptors. In the present project, we expand our examination of vocabulary ability to include not only concrete vocabulary but also conceptual/relational vocabulary. This research addresses potential differences in concrete and relational vocabulary ability, including the question of whether, given their extreme difficulty with many aspects of spatial cognition, especially visuospatial construction (Landau et al., 2006), children with WS evidence a specific difficulty with spatial vocabulary. We also demonstrate that the answer to the latter question differs depending on the strictness of the criterion for considering the target group (in this case, children with WS) and the control/contrast group to be matched on the control variable.

The spatial language abilities of individuals with WS have been addressed in several studies. Bellugi et al. (2000) argued that because spatial language provides important cues to nonverbal spatial representation (e.g., Bowerman, 1996), and individuals with WS have extreme difficulty with visuospatial construction, they would be expected to have much more difficulty with

spatial language than with other types of language. In a comparison of the spatial language of adolescents and young adults with WS to that of 11-year-old typically developing (TD) children, Bellugi et al. found that the WS group made significantly more errors on spatial prepositions (11%) than the TD group (0%) and was significantly more likely to confuse the figure and ground when describing the spatial position of a target object. Bellugi et al. concluded that individuals with WS have specific difficulty with spatial language.

Landau and Zukowski (2003) compared the spatial language of children with WS (mean CA 9.6 years) to that of a TD group (mean CA 5.0 years) matched for raw score on the Matrices subtest of the Kaufman Brief Intelligence Test (KBIT; Kaufman & Kaufman, 1990). Participants watched a series of video clips of events containing spatial relations and described what they saw. In contrast to Bellugi et al.'s (2000) result, the WS group confused figure and ground only 1% of the time. The groups tended to use the same verbs, and the most common path descriptions were also the same. However, the WS group was significantly more likely to omit the path term; such omissions were especially likely for bounded-from and via paths, which require memory for two locations. Omissions were less common for bounded-to paths, which require memory for only one location. Landau and Zukowski argued that children with WS have good control over much of the language needed to describe spatial events, and that the difficulty with path description was due to problems with spatial short-term memory, which is significantly weaker than verbal short-term memory (Rowe & Mervis, 2006).

Lukács (2005) compared the spatial language abilities of a group of children and young adults with WS (mean CA 15 years) to those of a group of TD children (mean CA 7 years) individually matched for raw score on the Hungarian version of the PPVT (Csányi, 1974). In Hungarian, the simplest spatial relations (in, on) are encoded as suffixes; more complex spatial relations (e.g., under, behind, between) are encoded as postpositions. The TD group performed significantly better on production of both suffixes and postpositions and also on comprehension of postpositions. The pattern of errors was the same for both groups.

Phillips, Jarrold, Baddeley, Grant, and Karmiloff-Smith (2004) compared the performance of a group of children and adults with WS (mean CA 21 years) to that of a TD group (mean CA 7 years) and a group with moderate learning difficulties (MLD; mean CA 13.8 years). Each participant with WS was individually matched to one person in each of the other groups for raw score on the last 11 blocks of the TROG. The WS group performed significantly worse on the three blocks that the authors argued contain a spatial component. In a follow-up study, individuals in a WS group (mean CA 19 years) were individually matched for BPVS AE to participants in a TD group (mean CA 8 years) and a MLD group (mean CA 13 years) and performance on a measure developed by Phillips et al. that included both spatial and non-spatial terms was compared. Both the TD and MLD groups performed significantly better on the spatial terms than the non-spatial terms. In contrast, the WS group performed significantly better on the non-spatial items than on the spatial items. Furthermore, the WS group performed significantly worse than the control groups on the spatial items and on one non-spatial item: lighter/darker. The WS and MLD groups made the same types of errors. The authors concluded that individuals with WS have specific problems comprehending spatial descriptions and that these problems are conceptual rather than linguistic. They argued that it is likely that individuals with WS have difficulty constructing mental models for verbal descriptions of space. They add that to be sure that the difficulty is with spatial mental model construction, it would be necessary to show that people with WS do not have general difficulty with relational reasoning.

In the present project, we address the question of whether the difficulty children with WS have with spatial concepts is specific to spatial concepts or instead characterizes relational concept ability more generally. One clue that the difficulty may be more general is contained in Phillips et al.'s (2004) finding that the WS group performed significantly worse than both the TD and

the MLD groups on the non-spatial comparative adjectives lighter-darker. In Study 1, we address the question of whether any difficulties 5 – 7-year-old children with WS have with relational vocabulary are specific to spatial concepts or apply to relational language more generally and provide a methodological demonstration of the impact of how closely groups are matched on the control variable on the answer to the previous question. In Study 2, we consider the relation between visuospatial construction ability and relational vocabulary ability for children with WS. In Study 3, we consider the performance of 9 – 11-year-old children with WS on a measure that assesses more advanced relational vocabulary. Because a subset of children in Study 3 had previously participated in Study 1, we also were able to assess the continuity of relative relational language ability over a 4-year period for children with WS.

Study 1: Relation between Receptive Concrete and Conceptual/Relational Vocabulary Ability

As reviewed above, the results of the few studies comparing the spatial language abilities of individuals with WS to those of matched controls have indicated that spatial language abilities are more limited than expected for level of concrete vocabulary and level of grammatical ability. The interpretation of results is somewhat complicated by the fact that the control groups were much younger than the WS groups (see methodological discussion in Mervis & Klein, 2005; Mervis & Robinson, 2003, 2005; Rowe & Mervis, 2006). In the present study, we consider the performance of children with WS and TD children on two measures of vocabulary, one assessing primarily concrete vocabulary and the other assessing conceptual/relational vocabulary. Mean CA difference between groups was <1 year. Because the measure of relational vocabulary included both spatial and non-spatial concepts, we also were able to address the possibility raised by prior researchers (e.g., Bellugi et al., 2000) that the extreme difficulty individuals with WS evidence with visuospatial construction tasks suggests that individuals with this syndrome should have particular difficulty with spatial vocabulary relative to other types of vocabulary.

This study also addresses an important methodological issue that frequently arises in research on the language and cognitive abilities of children with disabilities: How closely do groups have to score on the control variable for it to be reasonable to treat the groups as matched? The logic underlying group matching designs requires that the groups do not differ on the control variable(s). This point is often interpreted as meaning that if there is no significant difference between groups on the control variable, then the groups are matched. Based on this interpretation, groups have been considered to be matched even when the p level for the analysis of the control variable was only slightly above .05 (e.g., Paterson, 2001 adult vocabulary study; Sigman & Ruskin, 1999). However, finding that the null hypothesis that the groups do not differ on the control variables cannot be rejected (e.g., because $p > .05$) is very different from concluding that the null hypothesis can be accepted (the foundation for the group-matching design). As Cohen (1990) has pointed out, the null hypothesis is almost never literally true. Therefore, researchers using the group-matching design must consider the question, “How close is close enough?” We consider several different levels of match and demonstrate that these differences lead to differences in conclusions regarding whether spatial vocabulary is inordinately difficult for children with WS relative to other types of relational vocabulary.

Method

Participants—Two groups of children participated in this study. The WS group included 92 children (56 girls, 36 boys) aged 5.00 – 7.95 years whose diagnosis had been genetically confirmed. Six additional children were excluded because they had a comorbid diagnosis on the autism spectrum based on clinical judgment following administration of the ADOS-G (Lord, Rutter, & DiLavore, 1998) and ADI-R (Le Couteur, Lord, & Rutter, 2003). Some of the

children were enrolled in a longitudinal project in which the measures used in the present study were administered annually. For these children, the data from their first assessment in this age range were included. The TD group included 72 children (29 girls, 43 boys) aged 4.01 – 6.97 years.

Measures—Three standardized assessments were administered.

The *Kaufman Brief Intelligence Test* (KBIT; Kaufman & Kaufman, 1990) includes two subscales: Vocabulary and Matrices. Standard scores (mean = 100, SD = 15) are available for each subscale and for composite IQ. The KBIT is normed for ages 4 – 90+ years.

The *Peabody Picture Vocabulary Test, 3rd edition* (PPVT-III; Dunn & Dunn, 1997) measures receptive vocabulary. Most items assess concrete vocabulary, including object names, action words, and descriptors. An overall standard score (mean = 100, SD = 15) is provided. The PPVT-III is normed for ages 2.5 – 90+ years.

The *Test of Relational Concepts* (TRC; Edmonston & Litchfield Thane, 1988) measures conceptual/relational language. Five types of relational concepts are included: temporal (e.g., before/after), quantitative (e.g., many/few), dimensional (e.g., tall/short), spatial (e.g., under/over), and other (e.g., (same/different). An overall scaled score (T score; mean = 50, SD = 10) is provided. The TRC is normed for ages 3.0 – 7.99 years.

Procedure—The three assessments were administered according to the test authors' instructions and were almost always completed on the same day or within a day of each other.

Results

Between-group comparisons of concrete and relational vocabulary ability—To compare the children's concrete vocabulary and relational vocabulary abilities, we began by converting the TRC scaled scores to the standard score format (mean = 100, SD = 15) used for the PPVT-III. Based on the norms for the assessments, the lowest possible standard score was 40 for the KBIT, 20 for the PPVT-III, and 25 for the TRC. As expected, no child in the TD group earned the lowest possible standard score on any of the assessments. No child in the WS group earned the lowest possible standard score on the KBIT or PPVT-III, but 1 child earned the lowest possible standard score on the TRC and 24 (26%) earned TRC standard scores below 40 (the lowest possible standard score on the KBIT).

Descriptive statistics for both the WS and the TD groups, including CA, KBIT composite IQ, and raw and standard scores for the PPVT-III and the TRC are provided in Table 1. As indicated in the table, children in the WS group were on average ~8 months older than the children in the TD group. This difference in CA was significant. For the TD group, mean standard score for both the KBIT and the PPVT-III was very close to 100 (the mean for the general population), indicating that this sample of TD children is likely representative of TD children in general. Mean standard scores for the TD group were significantly higher than for the WS group for all three tests. In addition, mean PPVT-III and TRC raw scores were significantly higher for the TD group than for the WS group.

To confirm that the significant between-group difference in relational language ability was not simply a reflection of the TD group's stronger concrete vocabulary, an ANCOVA controlling for PPVT-III standard score was conducted, with TRC standard score as the dependent variable. Results indicated that the significant between-group difference in relational vocabulary ability remained even after concrete vocabulary ability was taken into account [$F(1,164) = 71.62, p < .001$, partial $\Omega^2 = .30$].

Within-group relations between concrete and relational vocabulary ability—To determine if concrete vocabulary ability was related to relational vocabulary ability, correlations between PPVT-III and TRC raw scores were computed. Correlations were strong for both groups (for WS, $r(90) = .71, p < .001$; for TD, $r(70) = .80, p < .001$) and remained strong even after controlling for CA (for WS, $r(89) = .69, p < .001$; for TD, $r(69) = .69, p < .001$).

Between-group comparisons of spatial relational language ability—As is clear from the analyses above, children with WS have more difficulty with relational vocabulary than expected given their concrete vocabulary abilities. Prior research has established that children with WS have particular difficulty with some types of spatial abilities, especially pattern construction, drawing, and spatial memory. The possibility that this difficulty extends to spatial language has also been raised (e.g., Bellugi et al. 2000; Landau & Zukowski, 2003; Phillips et al., 2004; Lukács, 2005). To determine if children with WS have more difficulty comprehending spatial concepts than do TD children, we began by comparing the percentage of correct responses on TRC items testing spatial concepts for the WS and TD groups, for all children whose overall raw scores on the TRC were within the range that was common to the two groups [from the lowest raw score obtained by a TD child (4 of 56) through the highest raw score obtained by a child with WS (45 of 56)]. We did not compare the full samples because 14 TD children earned higher raw scores than any child with WS, which made the outcome of any analyses of the full samples a foregone conclusion. Descriptive statistics for TRC raw scores and percent correct for spatial concepts and p levels for between-group t -tests, are reported in Table 2. As indicated in Table 2a, the TD group performed significantly better on the TRC spatial concepts than did the WS group. At the same time, however, the TD group responded correctly to significantly more items on the TRC as a whole than did the WS group, indicating that the two groups differed significantly on number of relational concepts known, independent of type of relational concept. Given that the TD group comprehended significantly more relational concepts than the WS group did, it is not surprising that the TD group outperformed the WS group on the spatial language concepts.

A better test of the question of whether children with WS have specific difficulty with spatial concepts requires that the groups be matched on level of receptive relational vocabulary. To address the question of the impact of the stringency of this match on the outcome of the test of between-group differences in percentage of spatial relational concepts known, we compared performance at three different levels of match. Descriptive statistics and p levels for between-group comparisons are included in Table 2b-d. In all cases, the p value for the match indicated that the two groups did not differ significantly. As indicated in Mervis and Robinson (2003) and Mervis and Klein-Tasman (2004), p values only slightly greater than .05 are commonly taken to indicate that groups are matched on the control variable(s). In contrast, Frick (1995) has argued that p values $< .20$ indicate that groups are *not* matched. The first level of match we included fell in this range: $p < .123$. To achieve this level of match, we followed the usual procedure for matching groups that were initially not matched: We restricted the range of scores on the control variable, removing from the sample everyone whose scores did not fall within the narrower range, such that the p value for the control variable comparison was $> .05$. Thus, we included only children whose TRC raw scores were between 15 and 40, eliminating the children who scored lowest (primarily children with WS) and the children who scored highest (primarily TD children). Once again, the results of a between-group t test comparing the spatial concepts percent correct indicated a significant difference favoring the TD group.

Frick has argued that p values between .20 and .50 indicate that groups *may* be matched on the control variable. To obtain a match within this p range, we restricted the lower end of the raw score range to 20 (once again eliminating primarily children with WS) while maintaining 40 as the top of the raw score range. The p value for the between-group comparison on the control

variable was .276. As indicated in Table 2c, even when the groups were matched at this level, there was still a significant between-group difference on spatial concepts percent correct, favoring the TD group, although the p value was now very close to .05.

According to Frick, p values greater than .50 indicate that the groups *are* matched on the control variable. To ensure as close a match as possible while maintaining the full range of scores, we followed a different procedure from the usual one. Rather than restricting the range of scores on the control variable, we preserved as much of the range as possible but restricted the number of children who obtained a particular set of scores to the same number in each group. We refer to this method as equating the samples. To equate the WS and TD groups on TRC raw scores, we included the entire range of TRC raw scores for which there was overlap between the WS and TD groups and divided the scores into bins of 3 points. Thus, the range from 4 – 45 items correct was covered by 14 3-point bins. Within each bin, the number of children included corresponded to the smaller of the number of children included in that bin for each group. For the group with the larger number of children in a particular bin, a random sample of the children in that bin was selected to correspond to the number of children for that bin in the other group. For example, if a bin contained 5 children with WS and 3 TD children, 3 of the 5 children with WS were randomly selected for inclusion in the equated sample. This procedure yielded samples of 31 children in each group, with a p value of .964 for the between-group t test for TRC raw score. As indicated in Table 2d, once the groups were well matched, the between-group t test comparing percent correct for spatial concepts did not indicate a significant between-group difference ($p = .239$). Thus, children with WS do not evidence more difficulty with spatial relational concepts than would be expected given their level of relational vocabulary ability.

An alternative explanation for why there is not a significant between-group difference in spatial relational vocabulary for the equated sample comparison even though there were significant differences when the groups were not as closely matched on the control variable is that the number of participants in the comparison using the equated sample procedure is smaller than in the previous comparisons, resulting in less power. To determine if that explanation is correct, samples of 31 children per group (the number in the equated sample) were randomly selected from the larger samples that had been matched for raw score range. As indicated in Table 2, the smaller samples were slightly better matched for TRC raw score. Nevertheless, for the two comparisons for which the p value for the control variable comparison did not meet Frick's possibly-matched criterion (4 – 45 raw score and 15 – 40 raw score), the groups still differed significantly on spatial concepts percent correct. For the 20 – 40 raw score comparison, the p value for the control variable comparison was higher (although still within the possibly-matched range) than the p value for the original (larger) sample within this raw score range and the difference between groups on spatial concepts percent correct was no longer significant. This lack of significance may be due to the reduced sample size (resulting in less power); the somewhat closer between-groups match on the control variable, or a combination of the two.

Discussion

The group of children with WS in this study includes all children we have tested who completed both the PPVT-III and the TRC as part of the same assessment. As a group, this sample was higher functioning than is usually reported for WS. For example, relative to the considerably larger sample of children and adolescents (which includes the children in the present study) reported in Mervis and Morris (2007), mean KBIT IQ was 9 points higher and mean PPVT-III standard score was 7 points higher. For the KBIT, some of the difference may be due to the wider age range in the Mervis and Morris sample; Mervis, Kistler, Peregrine, Rowe, and John (2006), in a longitudinal study, found that KBIT IQ was negatively related to CA for children with WS. However, this is not the entire explanation, because PPVT-III standard score was

not negatively related to CA. Even though the present group's standard scores are relatively high for WS, their performance indicates that even receptive concrete vocabulary ability, the area of language previously identified as the strongest for individuals with WS, is not "intact." Mean performance was ~1 standard deviation below the means for both the TD group and the norming sample, and the WS group included both a clear excess of children whose scores were $\leq 2^{\text{nd}}$ percentile (13% earned standard scores ≤ 70) relative to either the TD group (1.3%) or the norming sample (2%) and a considerably smaller percentage of children whose scores were $\geq 50^{\text{th}}$ percentile (19%) than either the TD group (57%) or the norming sample (50%).

At the same time, receptive concrete vocabulary is considerably stronger than receptive conceptual/relational vocabulary for the WS group. Mean PPVT-III standard score was 30.5 points (2 standard deviations) higher than mean TRC standard score. TRC standard score was $< 2^{\text{nd}}$ percentile for 74% of the WS group and $\geq 50^{\text{th}}$ percentile for only 2%. The great difficulty that children with WS have with conceptual/relational concepts, however, is not due primarily to difficulty with spatial concepts. Although the results of analyses for which the WS and TD groups were not well matched on overall relational language ability (even though $p > .05$ on the control variable comparison) suggested that children with WS did have inordinate difficulty with spatial concepts, once the groups were well matched on number of items correct on the TRC, the WS group did not differ from the TD group on percent correct for spatial concepts.

Study 2: Relation between Conceptual/Relational Language and Visuospatial Construction

WS is generally characterized as involving a relative strength in language and extreme weakness in visuospatial construction. Within language, vocabulary is considered to be the area of greatest strength. However, the results of Study 1 demonstrated a clear discrepancy between receptive concrete vocabulary and receptive conceptual/relational vocabulary. On the receptive concrete vocabulary test, no child scored in the moderate impairment range and only 13% scored in the mild impairment range. In contrast, on the receptive relational vocabulary test, 31% scored in the mild impairment range, 17% in the moderate impairment range, and 26% in the severe impairment range. This pattern of performance evokes the pattern that has previously been reported for visuospatial construction. In Study 2 we considered the conceptual/relational language abilities of children with WS relative to their visuospatial construction abilities.

Method

Participants—Participants were the 92 children with WS who participated in Study 1.

Measures—In addition to the measures included in Study 1, the Differential Ability Scales (DAS; Elliott, 1990) Pattern Construction subtest was administered. Successful performance on this measure requires the child to mentally parse the target pattern into its component parts and then locate these parts on the squares or cubes provided and orient the correct parts in the same manner as in the target pattern. A scaled score (T score) with a mean of 50 and standard deviation of 10 is provided. Norms are available for ages 3.5 – 17.99 years.

Procedure—All assessments were administered according to the test authors' instructions. Children completed the assessments either on the same day or over a 2-day span.

Results

To compare performance on the PPVT-III, TRC, and DAS Pattern Construction, the TRC and Pattern Construction T scores were converted to the same scale as the PPVT-III (mean = 100, SD = 15). The lowest possible score on Pattern Construction (T = 20, standard score = 55) is

considerably higher than the lowest possible standard score for the PPVT-III (20) or the TRC (25). To equate the floor for the three measures, all standard scores < 55 were recoded as 55. As the lowest PPVT-III standard score obtained by the children in the study was 55, none needed to be recoded. However, the TRC standard scores for 43% of the children had to be recoded.

Overall, performance was weak on both assessments: 63% of children earned a standard score of 55 on one or both assessments (37% on both assessments, 10% on the TRC only, and 16% on Pattern Construction only). In strong contrast, only one child earned a standard score of 55 on the PPVT-III. Means and standard deviations for these measures, with floor set at 55, were 86.73 (SD = 13.67) for the PPVT-III, 64.09 (SD = 12.99) for the TRC, and 60.43 (SD = 7.95) for DAS Pattern Construction. Histograms of standard score distributions for the three assessments for the full sample of 92 children are shown in the left column of Figure 1. Pearson correlations between standard scores were .63 for TRC and PPVT-III, .56 for TRC and DAS Pattern Construction, and .39 for PPVT-III and DAS Pattern Construction (all $ps < .001$).

To address the possibility that relational vocabulary ability is more similar to visuospatial construction ability than to concrete vocabulary ability for children with WS, a repeated measures ANOVA with two planned contrasts was conducted. Results indicated a significant main effect of test [$F(2, 182) = 282.26, p < .001$]. Results of the first planned contrast indicated that performance on the TRC was significantly worse than performance on the PPVT-III [$F(1, 91) = 366.17, p < .001, \eta_p^2 = .80$]. Although results of the second planned contrast indicated that performance on the TRC was significantly better than performance on DAS Pattern Construction [$F(1, 91) = 10.55, p = .002, \eta_p^2 = .10$], the effect size for this contrast was much smaller than for the TRC – PPVT-III contrast.

We also completed the same analyses for the group of children with WS who were in the equated sample in Study 1 ($n = 31$). Histograms of standard score distributions for the equated sample on the three assessments are shown in the right column of Figure 1. Results of the ANOVA and planned contrasts were similar: There was a significant main effect of test [$F(2, 60) = 91.75, p < .001$]. The planned contrast comparing performance on the TRC and PPVT-III also was significant, with a large effect size [$F(1, 30) = 131.73, p < .001, \eta_p^2 = .82$]. Although the planned contrast comparing performance on the TRC and DAS Pattern Construction was significant, the effect size was much smaller [$F(1, 30) = 5.96, p = .02, \eta_p^2 = .17$].

Discussion

Consistent with prior reports, the participants in the present study evidenced considerably stronger receptive concrete vocabulary ability than visuospatial construction ability. Their receptive conceptual/relational vocabulary also was significantly stronger than their visuospatial construction ability. However, examination of both the effect sizes and the difference in means (3½ points for the TRC – Pattern Construction comparison vs. 22 points for the PPVT-III – TRC comparison) indicated that the relational vocabulary ability of children with WS is more aligned with their visuospatial construction ability than with their concrete vocabulary ability. This provides further evidence that not only is WS characterized by clear relative strengths and weaknesses within the vocabulary component of language, but the weakness in relational vocabulary is considerable, with the result that relational vocabulary ability is only slightly higher than that of with the hallmark weakness associated with WS, visuospatial construction.

Study 3: Knowledge of More Advanced Relational Concepts

The types of relational concepts measured by the TRC are relatively simple and are expected to be mastered by TD children before age 8 years, as reflected in the fact that the TRC is normed

only through age 7.99 years. The acquisition of more advanced relational concepts, however, continues through early adolescence. Ideally, the study of the performance of 5 – 7-year-olds with WS on a test of receptive knowledge of the basic relational concepts measured by the TRC would be followed up by a study of the receptive knowledge of relational language in somewhat older children with WS on more complex relational concepts. Unfortunately, there are no assessments that measure receptive knowledge of more advanced relational concepts. However, the Formulated Sentences subtest of the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4; Semel, Wiig, & Secord, 2003), a measure of expressive semantics and grammar, assesses a variety of relational concepts, including both the types of concepts measured by the TRC and also more complex concepts. In this study, we consider the performance of 9 – 11-year-olds with WS on this measure to address the question of whether relational language performance improves as children with WS get older. We also consider the relation between performance on the TRC and performance on the CELF-4 Formulated Sentences subtest for children for whom both measures were available.

Method

Participants—Participants were 29 children (16 girls, 13 boys) with genetically-confirmed WS aged 9.06 – 11.74 years (mean = 10.52 years, SD = 0.77). Some of the children were enrolled in a longitudinal study in which the CELF-4 Formulated Sentences subtest was administered annually. For these children, the data from their most recent assessment within the 9 – 11 year age range were included in order to examine the relational conceptual knowledge of older children. Ten of the participants had previously completed the TRC.

Measures—The CELF-4 Formulated Sentences subtest was designed to evaluate the ability to formulate complete semantically and grammatically correct sentences of increasing length and complexity. The subtest includes 28 items. For the first 24 items, the child is shown a picture and asked to use the word provided by the examiner (e.g., car, never, longest, although) in a sentence to describe the picture. The picture adds a contextual constraint to sentence formulation and allows for the determination of whether the child understands the word provided by the examiner. The last 4 items do not include pictures. The examiner provides a phrase and the child is asked to use it in a sentence. A scaled score (mean = 10, SD = 3) is provided. Norms are available for ages 5.0 – 21.99 years.

Procedure—The assessments were administered according to the standardized procedures.

Results

Performance relative to CELF-4 norms—Scaled scores for performance on the CELF-4 Formulated Sentences subtest are shown in Figure 2a. As indicated in the figure, most of the participants had considerable difficulty on this subtest; the modal scaled score was 1, the lowest possible. However, four participants scored at or above the mean for the general population.

Comprehension and grammatical use of CELF-4 Formulated Sentences target words as a function of word type—The CELF-4 Formulated Sentences includes a variety of types of target words, ranging from simple nouns to complex relational words. To better characterize the participants' performance on the types of words included in this subtest, we divided the 24 words for which the child was shown an accompanying picture into six types. Three (nouns, verbs, adverbs) did not measure relational concepts. The three remaining types measured relational concepts. The "TRC relational" type included words similar to those on the TRC (e.g., *longest*, *before*). The "simple relational" type included words whose correct use (given the pictures provided) would most likely involve linking two words or short phrases (e.g., *and*, *neither*). The "complex relational" type included words whose correct use would most likely involve linking simple sentences into a single complex sentence (e.g., *because*,

unless). We coded each response twice. Responses were first coded for whether the utterance produced indicated that the child understood the meaning of the target word, independent of the utterance's grammaticality. Responses were also coded for whether the target word was used correctly in a complete, grammatical sentence that was related to the picture (a sentence that would receive full credit based on the test authors' coding scheme). All utterances were coded by both authors, with the few disagreements resolved by consensus.

Coding results are shown in Figure 2b. As expected, virtually all the children produced utterances that indicated they understood the nouns and verbs included on the Formulated Sentences subtest, and most of the sentences produced for these words were grammatical. Understanding and use of adverbs in complete grammatical sentences was somewhat weaker. Performance on the relational word types was considerably weaker. Only 56% of responses to the TRC relational words (which the TRC norms make clear are expected to be mastered before age 8 years) indicated comprehension of the target term, with comprehension decreasing further for the simple relational words and further still for the complex relational words.

Relation between performance on CELF-4 Formulated Sentences subtest and TRC—Ten of the 29 children had completed the TRC as part of an earlier assessment. For these children, mean CA was 6.93 years ($SD = 0.73$ years) at completion of the TRC and 11.33 years ($SD = 0.83$ years) at completion of Formulated Sentences. A scatter plot of standard scores on the two assessments is shown in Figure 3a. Performance on the two tests was highly correlated [$r(8) = .87$], indicating strong continuity in performance on relational concepts over the age range studied in this project.

Many standardized tests have as their lowest standard score either a 1 (on a 1 – 19 scale with a mean of 10 and a standard deviation of 3) or a 55 (the equivalent score on a scale with a mean of 100 and a standard deviation of 15). The CELF-4 subtests are in the former group. In the sample of 10 children for whom both TRC and Formulated Sentences data were available, only 2 children scored at floor on Formulated Sentences. The TRC uses a T scale, with a mean of 50 and a standard deviation of 10. However, unlike most standardized assessments that use T scores, which have a floor of 20 (equivalent to the 1 and 55 above), the floor on the TRC is 0. To demonstrate the methodological importance of having a lower floor than is typically found on standardized assessments, we recoded the TRC data as if the floor were 55. As indicated in Figure 3b, under this scenario (which would actually be the case for comparisons of subtests for most standardized measures), 7 of the 10 children (the first and third circles for TRC standard score = 55 represent 2 children each) would have earned the lowest possible standard score. In this situation, the relation between TRC and CELF-4 standard scores has become nonlinear, effectively dividing into two clusters of scores, and a great deal of information regarding the relations between performance on the two assessments lost. Examination of the scatter plot made it clear that it was no longer appropriate to compute a Pearson correlation statistic.

Discussion

The results of this study clearly indicate that for most children with WS, the great difficulty during the early primary school years in comprehension of relatively simple relational words continues into the later school years and extends both to relational terms meant to link two words or short phrases and also to relational terms meant to link simple sentences into a single complex sentence. At the same time, there is considerable variability in comprehension of relational language, with some children performing at or close to age level on the Formulated Sentences subtest and showing excellent performance on our comprehension measure even as the modal performance for the WS group was at floor on Formulated Sentences and very low on our comprehension measure. For the subsample of children who had completed both the

TRC and then the Formulated Sentences subtest approximately 4½ years later, there was strong continuity in relative performance on relational concepts. We were able to quantify this continuity using a relatively small sample size, including many children who had considerable difficulty on the TRC, because of the very low floor for standard scores on this assessment. Had the floor been set at the usual level for developmental assessments, the linear relation would have been lost.

General Discussion

Vocabulary Profile in Williams Syndrome and Its Relation to Spatial Ability

Findings from previous research have clearly demonstrated that WS is associated with a specific cognitive profile, with relative strength in language and verbal short term memory and severe weakness in visuospatial construction. Previous studies also have shown that WS is associated with a pattern of strengths and weaknesses within language, with vocabulary ability stronger than grammatical ability and pragmatic ability more seriously limited. The present research demonstrates that WS also is associated with a pattern of strengths and weaknesses within vocabulary. In particular, receptive concrete vocabulary ability is a clear strength, with mean level of performance for a very large group of 5 – 7-year-olds at the bottom of the average range for the general population and 19% scoring at or above the 50th percentile for the general population. In contrast, receptive relational vocabulary is a clear weakness, with mean level of performance ~2 SD below that of receptive concrete vocabulary and 73% scoring at or below the 2nd percentile for the general population. This difficulty with relational/conceptual language is general, rather than specific to spatial concepts; when the WS group and the TD group were carefully matched for extent of relational vocabulary, the two groups did not differ on percentage of spatial concepts known. The extent of the difficulty children with WS have with relational language is underscored by the finding that performance on relational/conceptual language assessment averages only 3½ points higher than performance on assessment of visuospatial construction, the area of greatest weakness for individuals with WS. The difficulty also extends to more complex relational concepts, with the modal scaled score for 9 – 11-year-olds with WS on the CELF-4 Formulated Sentences subtest at floor (1) and examination of the sentences produced indicating very poor comprehension of the relational concepts tested. There was strong continuity in relative relational concept ability over a 4-year period.

Phillips et al. (2004) argued that the findings of their research coupled with those of Landau and Zukowski (2003) suggested that the difficulty individuals with WS have comprehending spatial terms is due to problems constructing mental models of verbal descriptions of space (cf. Johnson-Laird, 1983; Tversky, 1991). Phillips et al. also argued that difficulty in comprehending spatial terms may extend to difficulty in comprehending comparative adjectives because comprehension of these terms often is based on relational reasoning using spatial mental models of the comparison (e.g., De Soto, London, & Handel, 1965). This extension would explain Phillips et al.'s finding that for the WS group, performance on the only set of comparative adjectives (lighter/darker) included in the non-spatial items was significantly worse than performance on the remaining non-spatial items and was similar to performance on the most difficult spatial items. The authors note that to be confident that individuals with WS have difficulty with non-spatial comparative adjectives, data from performance on additional such pairs are needed. Finally, they argue that to be certain that the problem is with creation of spatial mental models, one must show that people with WS do not have a general impairment in relational meaning that extends beyond spatial and comparative relational concepts.

The data from Study 1 provide strong evidence that children with WS have considerable difficulty with comprehension of comparative adjectives (dimensional adjectives in TRC

parlance). This finding clearly is consistent with the hypothesis that individuals with WS have difficulty constructing spatial mental models. The data from all three studies show that the problem children with WS have with relational language extends well beyond spatial concepts and comparative/dimensional concepts to temporal concepts, quantitative concepts, and more abstract relational concepts such as the complex relational concepts tested by the CELF-4 Formulated Sentences subtest. Although Phillips et al. (2004) suggested that this result would indicate that the difficulty individuals with WS have with relational concepts cannot be attributed to the creation of spatial mental models, it is possible that a more general difficulty with relational concepts is exactly what would be predicted by difficulty constructing spatial mental models. Such a position is consistent with Walsh's (2003) A Theory of Magnitude (ATOM) framework (see also Gentner & Loewenstein, 2002 and Loewenstein & Gentner, 2005).

Walsh (2003) has argued that spatial, quantitative, and temporal processing are all controlled by a common magnitude system (ATOM). If he is correct, then it would be reasonable to expect that individuals with WS would have similar levels of difficulty with spatial, temporal, quantitative, and dimensional concepts, consistent with the findings of Study 1. Although studies of spatial mental models have not addressed the most complex types of relational concepts included in Study 3, it is possible that the construction of spatial mental models could facilitate comprehension of these terms as well. Walsh has further argued that the magnitude system that is common to spatial, quantitative, and temporal processing is located in the inferior parietal cortex. Meyer-Lindenberg et al. (2004; Meyer-Lindenberg, Mervis, & Berman, 2006) have identified a structural abnormality (reduced gray matter and sulcal depth) in this region in the brains of normal-IQ adults with WS and have shown that this abnormality in the intraparietal sulcus serves as a roadblock to dorsal stream information flow. The results of a path analysis based on data from fMRI studies comparing normal-IQ adults with WS to CA- and IQ-matched adults in the general population indicated that the only difference between the two groups was that the path from the intraparietal sulcus to the later dorsal stream region was significant for the control group but not for the WS group. This same abnormality has been argued to provide a basis for the severe difficulty individuals with WS have with visuospatial construction (Meyer-Lindenberg et al., 2004, 2006), providing further unification for the results of Studies 1 and 2.

Methodological Issues

Considerable lip service has been paid to the goal of closely matching the comparison groups to the target group(s) on the control variable(s) before comparing the groups on the variables of interest to the researchers. Nevertheless, in many cases the p values for between-group comparisons on the control variable(s) are within the range that Frick (1995) has argued indicates that the groups are clearly not matched ($p < .20$). In many other cases, the p values are in the ambiguous range (between .20 and .49). Studies in which $p \geq .50$ for the control variable comparison (the range Frick considers to indicate that the groups are matched) are becoming more common, but are still very much in the minority. The demonstration component of Study 1 provided a clear illustration of how the extent to which comparison and target groups are stringently matched on the control variable(s) impacts the conclusions that are drawn from comparisons on the variables that are the focus of the research. When the WS and TD groups were matched on relational vocabulary size at p levels $>.05$ but within the range that Frick considers clearly not matched, comparisons of performance on spatial concepts suggested that children with WS had specific difficulty with this type of relational concept. The same conclusion was reached when the groups were matched at p levels that were toward the lower end of Frick's ambiguous range. In contrast, when groups were very carefully matched, it became clear that children with WS do not have specific difficulty with spatial concepts relative to other types of relational concepts.

In Study 1, the WS group and the TD group differed in CA by less than 1 year. We have previously shown that even if the groups are well matched on the control variable(s), if the groups differ considerably in CA and/or if one or more groups includes a wide CA range, one cannot assume that the groups should perform at the same level on the target measures if the relation between the target variable and the control variable is the same for both groups. This is because development of many skills is nonlinear. For the same reason, use of statistical methods to account for CA differences is problematic. Examples of these difficulties and of one solution (standard score profiles) are provided in Mervis and Robinson (2003, 2005) and Mervis and Klein-Tasman (2004).

Researchers who study children with unusual disorders often have relatively easy access only to limited numbers of participants. In such cases, matching using the strategies employed in the present research may result in unacceptably small sample sizes. An alternative approach used by some researchers is to treat the matching variable as the covariate in an ANCOVA. This approach is appropriate if the groups included have similar CA ranges and distributions and two other conditions are met: First, the relation between the covariate and the target variable is approximately linear for both groups over the entire CA range of the participants. Second, the variance of residual dependent variable scores is approximately the same over the full range of covariate scores for each group. As mentioned above, the development of many of the abilities studied by language researchers is nonlinear and therefore would not meet these assumptions. (See Jarrold and Brock, 2004, for further discussion.)

The use of standard scores derived from well-normed assessments that appropriately measure the abilities on which the research is focused provides investigators with an excellent tool for identifying patterns of cognitive and language strengths and weaknesses. Unfortunately, in many cases these assessments are not yet available. Even when appropriate assessments are available, the norms may not have a low enough floor to accurately assess the children participating in the research. This is especially likely to be the case for assessments that use scaled scores on a 1 – 19 scale or T scores on a 20 – 80 scale. In these cases, the lowest standard score is only 3 standard deviations below the general population mean. In the present research, the majority of children with WS scored at floor on both the DAS Pattern Construction subtest and the CELF-4 Formulated Sentences subtest. The TRC is normed to 5 standard deviations below the general population mean, resulting in only 1 of 92 children with WS performing at floor and (since luckily most of the children for whom both TRC and CELF-4 Formulated Sentences data were available scored above floor on the CELF-4 despite its relatively high floor) allowing us to determine that the correlation between performance on an assessment of basic relational concepts and performance on an assessment measuring more complex relational concepts 4 years later was both linear and very strong for children with WS. However, when for demonstration purposes we assumed that the TRC also had a floor only 3 standard deviations below the general population mean, 7 of the 10 children now had the lowest possible standard score on the TRC, resulting in the linear nature of the relation being completely masked.

The problem of inadequate floors on standardized assessments is even more critical for syndromes for which the typical level of performance is lower than for WS, for example Down syndrome or males with Fragile X syndrome. Researchers studying these syndromes often resort to using AE scores in an attempt to differentiate among large numbers of participants who all earned the lowest possible standard score. Yet there are serious problems with AE scores (the median CA at which a particular raw score was obtained), including the fact that they are not on an interval scale yet are often treated as interval in statistical analyses; problems with AE scores are discussed in detail in Mervis (2004) and Mervis and Robinson (2005) as well as in the manuals for many standardized assessments (e.g., the CELF-4; Semel et al., 2003).

Test authors are becoming more sensitive to the problem of inadequate floors on their assessments. For example, the T scores for the revision of the DAS (DAS-II; Elliott, 2007) are normed to 4 standard deviations below the general population mean; the positive impact of this change for research on children with WS is apparent in that many fewer children are now scoring at floor on the Pattern Construction subtest (Mervis, unpublished data). The scales of the revision of the Vineland Adaptive Behavior Scales (VABS-2; Sparrow et al., 2005) are normed to almost 5 standard deviations below the general population mean, and the PPVT-4 (Dunn & Dunn, 2007) and the co-normed Expressive Vocabulary Test-2 (EVT-2; Williams, 2007) are both normed to more than 5 standard deviations below the general population mean. These changes should greatly increase the utility of well-conceived standardized assessments for research on children with language and/or intellectual disabilities. When comparing relative performance on standardized assessments with differing floor values, it is important to set the lowest possible standard score for all of the measures to the score corresponding to the floor on the measure with the highest floor. For example, when comparing standard scores on the TROG (floor = 55) to those on the PPVT (floor = 20) to determine if receptive vocabulary ability of a particular group differs significantly from receptive grammar ability, PPVT standard scores below 55 should be converted to 55. Without this type of adjustment, performance on the ability tested by the assessment with the higher floor would have an artificial advantage over performance on the ability tested by the assessment with the lower floor.

Conclusion

The results of the present research suggest that the relative strength in vocabulary ability that has been consistently found by researchers studying the language abilities of individuals with WS is actually a relative strength not in vocabulary ability in general but rather in concrete vocabulary ability. Conceptual/relational vocabulary ability is much more limited and is in fact at a similar level to that of the hallmark weakness in Williams syndrome, visuospatial construction. This weakness in relational language is not restricted to spatial concepts but rather extends to all types of relational concepts examined in these studies. Comparative studies involving groups of children who have other forms of developmental disabilities are needed to determine if this pattern is generally characteristic of children with intellectual or language disabilities or is restricted to children with particular syndromes or (unlikely) only to WS.

The finding that the spatial conceptual ability of children with WS was not more limited than other types of conceptual ability emerged only when groups were very carefully matched for overall level of relational language. This result provides additional support for the importance of adhering rigorously to methodological principles of matching when performing comparative studies of children with developmental disabilities. The availability of carefully designed standardized assessments that are normed to 4 or more standard deviations below the general population mean would greatly facilitate future progress in elucidating the language strengths and weaknesses of children with limited intellectual ability.

Acknowledgements

We thank the children and their families for their enthusiastic participation in these studies. We also thank the members of the Neurodevelopmental Sciences Laboratory for conducting some of the assessments, Joanie Robertson for database management, and Doris Kistler for statistical consultation. This research was supported by grant# R37 HD29957 from the National Institute of Child Health and Human Development and grant# R01 NS35102 from the National Institute of Neurological Disorders and Stroke.

References

- Bates, E. Early language development: How things come together and how they come apart; 1990, April; International Conference on Infant Studies; Montreal, Quebec.

- Bellugi U, Lichtenberger L, Jones W, Lai Z, St George M. I. The neurocognitive profile of Williams syndrome: a complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience* 2000;12 (Suppl 1):7–29. [PubMed: 10953231]
- Bellugi, U.; Marks, S.; Bihle, A.; Sabo, H. Dissociation between language and cognitive functions in Williams syndrome. In: Bishop, D.; Mogford, K., editors. *Language development in exceptional circumstances*. London: Churchill Livingstone; 1988. p. 177-189.
- Bennett C, LaVeck B, Sells CJ. The Williams elfin facies syndrome: The psychological profile as an aid in syndrome identification. *Pediatrics* 1978;61:303–306. [PubMed: 634688]
- Bishop, DVM. *Test for Reception of Grammar*. Manchester, UK: Chapel Press; 1989.
- Bishop, DVM. *Test for Reception of Grammar, version 2*. London: Psychological Corporation; 2003.
- Cassidy SB, Morris CA. Behavioral phenotypes in genetic syndromes: genetic clues to human behavior. *Advances in Pediatrics* 2002;49:59–86. [PubMed: 12214780]
- Cohen J. Things I have learned (so far). *American Psychologist* 1990;45:1304–1312.
- Csányi, FI. *Peabody Szókincs-Teszt*. Budapest, Hungary: Bárczi Gusztáv Gyógypedagógiai Főiskola; 1974.
- De Soto CB, London M, Handel S. Social reasoning and spatial paralogic. *Journal of Personality and Social Psychology* 1965;2:513–521. [PubMed: 5841356]
- Dunn, LM.; Dunn, DM. *Peabody Picture Vocabulary Test. 4*. Minneapolis, MN: Pearson Assessments; 2007.
- Dunn, LM.; Dunn, LM. *Peabody Picture Vocabulary Test – Revised*. Circle Pines, MN: American Guidance Service; 1981.
- Dunn, LE.; Dunn, LE. *Peabody Picture Vocabulary Test. 3*. Circle Pines, MN: American Guidance Service; 1997.
- Dunn, LM.; Dunn, LM.; Whetton, C.; Pintilie, D. *British Picture Vocabulary Scale*. Windsor: NFER-Nelson; 1982.
- Edmonston, NK.; Litchfield Thane, N. *TRC: Test of Relational Concepts*. Austin, TX: PRO-ED; 1988.
- Elliott, CD. *Differential Ability Scales*. San Antonio, TX: Psychological Corporation; 1990.
- Elliott, CD. *Differential Ability Scales. 2*. San Antonio, TX: Psychological Corporation; 2007.
- Frick R. Accepting the null hypothesis. *Memory and Cognition* 1995;23:132–138.
- Gentner, D.; Loewenstein, J. Relational language and relational thought. In: Amsel, E.; Byrnes, JP., editors. *Language, literacy, and cognitive development: The development and consequences of symbolic communication*. Hillsdale, NJ: Erlbaum; 2002. p. 87-120.
- Hillier LW, Fulton RS, Fulton LA, Graves TA, Pepin KH, Wagner-McPherson C, et al. The DNA sequence of chromosome 7. *Nature* 2003;424:157–164. [PubMed: 12853948]
- Jarrold C, Brock J. To match or not to match? Methodological issues in autism-related research. *Journal of Autism and Developmental Disorders* 2004;34:81–86. [PubMed: 15098961]
- Johnson-Laird, PN. *Mental models*. Cambridge, England: Cambridge University Press; 1983.
- Karmiloff-Smith A, Grant J, Berthoud I, Davies M, Howlin P, Udwin O. Language and Williams syndrome: How intact is “intact”? *Child Development* 1997;68:274–290.
- Kaufman, AS.; Kaufman, NL. *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Services; 1990.
- Kaufman, AS.; Kaufman, NL. *Kaufman Brief Intelligence Test. 2*. Circle Pines, MN: American Guidance Services; 2004.
- Klein-Tasman BP, Mervis CB. Distinctive personality characteristics of 8-, 9-, and 10-year-old children with Williams syndrome. *Developmental Neuropsychology* 2003;23:271–292.
- Klein-Tasman BP, Mervis CB, Lord C, Phillips KD. Socio-communicative deficits in young children with Williams syndrome: Performance on the Autism Diagnostic Observation Schedule. *Child Neuropsychology* 2007;13:444–467. [PubMed: 17805996]
- Landau, B.; Hoffman, JE.; Reiss, JE.; Dilks, DD.; Lakusta, L.; Chunyo, G. Specialization, breakdown, and sparing in spatial cognition: Lessons from Williams-Beuren syndrome. In: Morris, CA.; Lenhoff, HM.; Wang, PP., editors. *Williams-Beuren syndrome: Research, evaluation, and treatment*. Baltimore, MD: Johns Hopkins University Press; 2006. p. 207-236.

- Landau B, Zukowski A. Objects, motions, and paths: Spatial language in children with Williams syndrome. *Developmental Neuropsychology* 2003;23:107–139.
- Laws G, Bishop D. Pragmatic language impairment and social deficits in Williams syndrome: a comparison with Down's syndrome and specific language impairment. *International Journal of Language and Communication Disorders* 2004;39:45–64. [PubMed: 14660186]
- LeCouteur, A.; Lord, C.; Rutter, M. *The Autism Diagnostic Interview - Revised (ADI-R)*. Los Angeles, CA: Western Psychological Services; 2003.
- Leyfer OT, Woodruff-Borden J, Klein-Tasman BP, Fricke JS, Mervis CB. Prevalence of psychiatric disorders in 4 – 16-year-olds with Williams syndrome. *American Journal of Medical Genetics Part B* 2006;141B:615–622.
- Loewenstein J, Gentner D. Relational language and the development of relational mapping. *Cognitive Psychology* 2005;50:315–353. [PubMed: 15893523]
- Lord, C.; Rutter, M.; DiLavore, P. *Autism Diagnostic Observation Schedule Generic*. University of Chicago; 1998.
- Lukács, Á. *Language abilities in Williams syndrome*. Budapest, Hungary: Akadémiai Kiadó; 2005.
- Lukács, Á.; Pléh, C.; Racsmány, M. Language in Hungarian children with Williams syndrome. In: Bartke, S.; Siegmüller, J., editors. *Williams syndrome across languages*. Amsterdam: John Benjamins; 2004. p. 187–220.
- Mervis, CB. *Ecological approaches to cognition: Essays in honor of Ulric Neisser*. Mahwah, NJ: Erlbaum; 1999. *The Williams syndrome cognitive profile: Strengths, weaknesses, and interrelations among auditory short-term memory, language, and visuospatial constructive cognition*; p. 193–227.
- Mervis, CB. Cross-etiology comparisons of cognitive and language development. In: Rice, ML.; Warren, SF., editors. *Developmental language disorders: From phenotypes to etiologies*. Mahwah, NJ: Erlbaum; 2004. p. 153–186.
- Mervis, CB. Language abilities in Williams-Beuren syndrome. In: Morris, CA.; Lenhoff, HM.; Wang, PP., editors. *Williams-Beuren syndrome: Research, evaluation, and treatment*. Baltimore, MD: Johns Hopkins University Press; 2006. p. 159–206.
- Mervis, CB. Cognitive, language, and communicative development of children with deletion or duplication of the Williams syndrome region (7q11.23). *Symposium on Research on Child Language Disorders*; Madison, WI: 2007 Jun.
- Mervis CB, Becerra AM. Language and communicative development in Williams syndrome. *Mental Retardation and Developmental Disabilities Research Reviews* 2007;13:3–15. [PubMed: 17326109]
- Mervis, CB.; Kistler, DJ.; Peregrine, E.; Rowe, ML.; John, AE. Longitudinal assessment of intelligence in children and adolescents with Williams syndrome: A multilevel modeling analysis; 2006, July; *International Williams Syndrome Association Professional Conference*; Richmond, VA.
- Mervis C, Klein-Tasman B. Methodological issues in group-matching designs: Alpha levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders* 2004;34:7–17. [PubMed: 15098952]
- Mervis, CB.; Morris, CA. Williams syndrome. In: Mazzocco, MMM.; Ross, JL., editors. *Neurogenetic developmental disorders: Variation of manifestation in childhood*. Cambridge, MA: The MIT Press; 2007. p. 199–262.
- Mervis, C.; Robinson, B. Methodological issues in cross-group comparisons of language and/or cognitive development. In: Levy, Y.; Schaeffer, J., editors. *Language competence across populations: Toward a definition of specific language impairment*. Mahwah, NJ: Lawrence Erlbaum; 2003. p. 233–258.
- Mervis CB, Robinson BF. Designing measures for profiling and genotype/phenotype studies of individuals with genetic syndromes or developmental language disorders. *Applied Psycholinguistics* 2005;26:41–64.
- Mervis CB, Robinson BF, Bertrand J, Morris CA, Klein-Tasman BP, Armstrong SC. The Williams Syndrome Cognitive Profile. *Brain and Cognition* 2000;44:604–628. [PubMed: 11104544]
- Meyer-Lindenberg A, Kohn P, Mervis CB, Kippenhan JS, Olsen R, Morris CA, Berman KF. Neural basis of genetically determined visuospatial construction deficit in Williams syndrome. *Neuron* 2004;43:623–631. [PubMed: 15339645]

- Meyer-Lindenberg A, Mervis CB, Berman KF. Neural mechanisms in Williams syndrome: a unique window to genetic influences on cognition and behavior. *Nature Reviews: Neuroscience* 2006;7:380–393.
- Osborne, LR. The molecular basis of a multisystem disorder. In: Morris, CA.; Lenhoff, HM.; Wang, PP., editors. *Williams-Beuren syndrome: Research, evaluation, and treatment*. Baltimore, MD: Johns Hopkins University Press; 2006. p. 18-58.
- Paterson S. Language and number in Down syndrome: The complex developmental trajectory from infancy to adulthood. *Down Syndrome Research and Practice* 2001;7:79–86.
- Phillips CE, Jarrold C, Baddeley AD, Grant J, Karmiloff-Smith A. Comprehension of spatial language terms in Williams syndrome: Evidence for an interaction between domains of strength and weakness. *Cortex* 2004;40:85–101. [PubMed: 15070004]
- Philofsky A, Fidler DJ, Hepburn S. Pragmatic language profiles of school-age children with autism spectrum disorders and Williams syndrome. *American Journal of Speech-Language Pathology* 2007;16:368–380. [PubMed: 17971496]
- Rowe, ML.; Mervis, CB. Working memory in Williams syndrome. In: Alloway, TP.; Gathercole, SE., editors. *Working memory and neurodevelopmental conditions*. New York, NY: Psychology Press; 2006. p. 267-293.
- Semel, E.; Wiig, EH.; Secord, WA. *Clinical Evaluation of Language Fundamentals*. 4. San Antonio, TX: Harcourt Assessment Inc; 2003.
- Sigman M, Ruskin E. Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development* 1999;64Serial no. 256
- Sparrow, SS.; Cicchetti, DV.; Balla, DA. *Vineland Adaptive Behavior Scales*, 2nd ed., Survey Interview Form. 2. Minneapolis, MN: Pearson Assessments; 2005.
- Stojanovik V. Social interaction deficits and conversational inadequacy in Williams syndrome. *Journal of Neurolinguistics* 2006;19:157–173.
- Strømme P, Bjørnstad PG, Ramstad K. Prevalence estimation of Williams syndrome. *Journal of Child Neurology* 2002;17:269–271. [PubMed: 12088082]
- Tversky, B. Spatial mental models. In: Bower, GH., editor. *The psychology of learning and motivation: Advances in research and theory*. 27. San Diego, CA: Academic Press; 1991. p. 109-145.
- Walsh V. A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences* 2003;7:483–488. [PubMed: 14585444]
- Williams, KT. *Expressive Vocabulary Test*. 2. Minneapolis, MN: Pearson Assessments; 2007.

Figure 1. PPVT-III, TRC, and DAS Pattern Construction histogram standard score comparisons for the full sample of 92 children with WS (left column) and for the children with WS included in the equated sample of 31 children from Study 1 (right column).

Figure 2.

(a) Histogram of CELF-4 Formulated Sentences scaled scores for 29 9–11-year-olds with WS. (b) Percent correct for 29 9–11-year-olds with WS on the CELF-4 Formulated Sentences subtest as a function of word type. Solid bars indicate the percent of items for which the child's use of the target word demonstrated that he or she understood the meaning of the word. Striped bars indicate the percent of items for which the child used the word appropriately in a complete grammatically correct sentence. This percentage corresponds to the percent of items for which the child's response would be scored "2" according to the test authors' scoring guidelines for this subtest.

Figure 3.

(a) Scatter plot and line of best fit showing the relation between standard scores on the TRC and the CELF-4 Formulated Sentences subtest for 10 children with WS. Standard scores from both measures were converted to a common scale with mean = 100 and SD = 15. (b) Scatter plot and line of best fit between standard scores on TRC CELF-4 Formulated Sentences subtest after TRC standard scores <55 are converted to 55, so that the lowest possible standard score is the same on both measures. Despite the nonlinear distribution of scores (and therefore the inappropriateness of a linear regression analysis) the regression line yielded by such an analysis (shown in Panel b for illustration purposes) is similar to that in Panel a. This contrast highlights the importance of graphical examination of the data to determine if a planned statistical analysis is appropriate rather than simply computing the statistic and assuming the obtained result is valid.

Table 1

Descriptive statistics for WS and TD groups

		WS (<i>n</i> = 92)	TD (<i>n</i> = 72)	<i>t</i> -test <i>p</i> -value
CA	Mean (SD)	6.06 (.79)	5.38 (.83)	<i>p</i> < .001
	Range	5.00 – 7.95	4.01 – 6.97	
KBIT IQ	Mean (SD)	78.79 (14.39) ^{<i>I</i>}	103.07 (13.34) ^{<i>I</i>}	<i>p</i> < .001
	Range	44 – 112	70 – 136	
PPVT-III Raw Score	Mean (SD)	64.24 (17.58)	72.54 (21.07)	<i>p</i> = .007
	Range	29 – 109	14 – 112	
PPVT-III Std Score	Mean (SD)	86.73 (13.67)	101.61 (14.12)	<i>p</i> < .001
	Range	59 – 118	64 – 133	
TRC Raw Score	Mean (SD)	20.50 (10.60)	32.96 (12.93)	<i>p</i> < .001
	Range	0 – 45	4 – 56	
TRC Std. Score	Mean (SD)	55.79 (21.37)	91.71 (16.63)	<i>p</i> < .001
	Range	25 – 104	44 – 122	

^{*I*} Data for one child missing from analysis.

Table 2

TRC raw score and percent correct on spatial items for WS and TD groups matched for relational vocabulary size at various levels of stringency

Match criterion	WS	Main TD	p-value	WS	Sub-samples TD	p-value
(a) 4 – 45 raw scores						
TRC raw mean ⁿ (SD)	90 20.94 (10.28)	58 29.72 (10.76)	$p < .001$	31 20.94 (11.34)	31 29.87 (10.50)	$p = .002$
Spatial % mean (SD)	0.42 (0.21)	0.62 (0.22)	$p < .001$	0.41 (0.22)	0.63 (0.21)	$p = .001$
(b) 15 – 40 raw scores						
TRC raw mean ⁿ (SD)	53 26.58 (6.24)	40 28.70 (6.80)	$p = .123$	31 26.26 (6.45)	31 28.45 (6.84)	$p = .199$
Spatial % mean (SD)	0.52 (0.17)	0.61 (0.17)	$p = .015$	0.51 (0.15)	0.60 (0.18)	$p = .029$
(c) 20 – 40 raw scores						
TRC raw mean ⁿ (SD)	45 28.24 (5.18)	37 29.62 (6.18)	$p = .276$	31 28.26 (5.37)	31 29.68 (6.18)	$p = .338$
Spatial % mean (SD)	0.56 (0.15)	0.63 (0.16)	$p = .046$	0.57 (0.16)	0.63 (0.17)	$p = .129$
(d) Equated						
TRC raw mean ⁿ (SD)	31 28.39 (5.53)	31 28.45 (5.65)	$p = .964$	--	--	--
Spatial % mean (SD)	0.56 (0.14)	0.61 (0.16)	$p = .239$	--	--	--