Cross-sectional and longitudinal assessment of cognitive development in Williams syndrome

Short title: Cognitive development in Williams syndrome

Emily K. Farran[1], Harry R. M. Purser[2], Christopher Jarrold[3], Michael S. C. Thomas[4], Gaia Scerif[5], Vesna Stojanovik[6], Jo Van Herwegen[7]

[1]School of Psychology, University of Surrey. ORCID: 0000-0001-6409-0875

[2]Department of Psychology, Nottingham Trent University. ORCID: 0000-0003-3307-8421

[3]School of Psychological Science, University of Bristol. ORCID: 0000-0001-8662-0937

[4] Department of Psychological Sciences, Birkbeck, University of London. ORCID: 0000-0002-8231-6011

[5]Attention, Brain and Cognitive Development Group, Department of Experimental Psychology, University of Oxford, Oxford, UK. ORCID: 0000-0002-6371-8875

[6]School of Psychology and Clinical Language Sciences, University of Reading. ORCID: 0000-0001-6791-9968

[7] Department of Psychology and Human Development, Institute of Education, UCL's Faculty of Education and Society, London, UK. ORCID: 0000-0001-5316-1818

Address correspondence to:

e.farran@surrey.ac.uk

Professor Emily K. Farran

School of Psychology

University of Surrey

Guildford

Surrey. GU2 7XH

Funding: This research was supported by funding from the Williams Syndrome Foundation UK to Jo Van Herwegen and Michael Thomas. Legacy data was collected under multiple grants to the authors.

Conflict of interest: No conflict of interest is reported by the authors.

Ethics approval statement: Data for this study was obtained from the WiSDom database (https://blogs.ucl.ac.uk/wisdom/), a UK-based legacy database. The database was put together retrospectively using cognitive and behavioural data for individuals with WS from the last twenty-plus years from seven UK labs (labs are led by each of the authors and Annette Karmiloff-Smith, who is sadly no longer with us). Participant consent was provided at each lab when data was collected. Ethical approval for the WiSDom project was approved by UCL: Z6364106/2019/08/37

Author contribution statement: EF: Writing – Original Draft, Conceptualisation, Methodology, Formal analysis, Resources; HP: Conceptualisation, Methodology, Formal analysis, Data curation, Visualisation, Writing – Original Draft; CJ: Conceptualisation, Methodology, Resources, Writing – Review and Editing; MT: Conceptualisation, Methodology, Resources, Writing – Review and Editing; GS: Conceptualisation, Resources,

Writing – Review and Editing; VS: Resources, Writing – Review and Editing; JVH: Conceptualisation, Methodology, Data curation, Resources, Writing – Review and Editing

Data availability statement:

Data are available on request from the WisDom network: https://blogs.ucl.ac.uk/wisdom/

Acknowledgements: Thank you to the parents and individuals with WS and to the Williams Syndrome Foundation, UK. Thank you to lab members who contributed to the legacy data set.

Cross-sectional and longitudinal assessment of cognitive development in Williams syndrome

Research Highlights

- Cross-sectional and longitudinal developmental trajectories of verbal and non-verbal development, in the largest sample of individuals with Williams syndrome (WS) to-date, are presented.
- This research is unique due to large sample sizes, independence between cross-sectional and longitudinal samples, and a reliance on minimum three timepoints for longitudinal trajectories.
- Data support the WS cognitive profile of stronger verbal than non-verbal ability and a developmental model of delayed onset and a delayed rate.
- Cross-sectional and longitudinal data are discussed with reference to validating cross-sectional developmental patterns using longitudinal data and the importance of individual differences in understanding development.

Abstract

Williams syndrome (WS) is a rare genetic syndrome. As with all rare syndromes, obtaining adequately powered sample sizes is a challenge. Here we present legacy data from seven UK labs, enabling the characterisation of cross-sectional and longitudinal developmental trajectories of verbal and non-verbal development in the largest sample of individuals with WS to-date. In Study 1, we report cross-sectional data from between N=102 and N=209 children and adults with WS on measures of verbal and non-verbal ability. In Study 2, we report longitudinal data from N=17 to N=54 children and adults with WS who had been tested on at least three timepoints on these measures. Data support the WS characteristic cognitive profile of stronger verbal than non-verbal ability, and shallow developmental progression for both domains. Both cross-sectional and longitudinal data demonstrate steeper rates of development in the child participants than the adolescent and adults in our sample. Cross-sectional data indicate steeper development in verbal than non-verbal ability, and that individual differences in the discrepancy between verbal and non-verbal ability are largely accounted for by level of intellectual functioning. A diverging developmental discrepancy between verbal and non-verbal ability, whilst marginal, is not mirrored statistically in the longitudinal data. Cross-sectional and longitudinal data are discussed with reference to validating cross-sectional developmental patterns using longitudinal data and the importance of individual differences in understanding developmental progression.

Keywords: Williams syndrome, longitudinal data, verbal ability, non-verbal ability, cognitive development, individual differences

Introduction

Williams syndrome (WS) is a rare genetic syndrome with a prevalence of 1 in 7, 500 to 1 in 20, 000 live births (Morris et al., 1998; Strømme et al., 2002) and mild to moderate intellectual disability. Despite a mean IQ between 50 and 60 (Martens et al., 2008), WS is associated with an uneven cognitive profile, characterised at a group level by poorer non-verbal relative to verbal cognition (Martens et al., 2008). Furthermore, at the individual level, there are broad individual differences in the magnitude of the discrepancy between verbal and non-verbal ability in WS (Jarrold et al., 1998; Van Herwegen et al., 2011). In the current study, we investigate developmental progression of verbal and non-verbal ability, and the discrepancy in performance between these domains, in WS.

Most studies of WS have used cross-sectional data. While this is useful for determining cognitive profiles, it provides a snapshot of ability and fails to inform about developmental progression (Karmiloff-Smith, 1998). When cross-sectional data are used to plot developmental trajectories, it is not possible to differentiate individual differences from true change over time, and thus any developmental claims made using cross-sectional data ideally require validation from longitudinal data (Thomas et al., 2009). For rare genetic syndromes such as WS, sufficiently powered sample sizes of longitudinal data are difficult to obtain (Farran, 2021; Farran & Scerif, 2022). In the current study, we present cross-sectional and longitudinal data in the largest WS sample to-date, drawn from the WISDOM database (https://blogs.ucl.ac.uk/wisdom/). Below, we outline considerations related to reporting data from standardised measures, before reviewing existing longitudinal studies of the development of verbal and non-verbal ability in WS.

Reporting data from standardised assessments

Studies that have examined IQ or Standard Scores (SS) longitudinally (e.g., Fisher et al., 2016; Mervis & Pitts, 2015) ask questions related to stability over time. Stability in IQ or SSs of a WS group with increasing age reflects a similar relative position of the WS group

compared to the general population over time (i.e., the gap remains stable), whilst a decrease in IQ or SS over time reflects that the gap between the WS group and the general population widens over time. Other studies have used age-equivalence or raw scores and compared change over time in these variables to change in chonological age, to determine whether the rate of developmental progression is typical, faster or slower than expected for chonological age (e.g., Porter & Dodd, 2011; Jarrold et al., 2001).

For IQ and SS, floor effects are common and thus can reduce the sensitivity of the measure. For example, Farran et al. (2019) report T-scores for the Matrices subtest of the British Ability Scales III (BAS-3) (Elliott & Smith, 2011) for a sample of N=20 participants with WS, which range from 20 to 23. In fact, 18 of the 20 participants in this sample received a T-score of 20, despite a range of ability scores (the equivalent of raw scores) from 33 to 103. For longitudinal designs, floor performance means that a decline in SS is not possible to detect, and stability in SS might simply be an artefact of the individual remaining at floor.

In contrast to SSs, age equivalence scores provide a concrete indication of the absolute level of performance on a given measure, which can be directly compared across measures and time points. As noted by Jarrold et al. (2001), the underlying standardisation process that converts raw scores to age equivalent values necessarily also 'linearises' the resulting scores such that a year's developmental improvement is comparable at all ranges of the test.

Raw scores are the purest measure of performance. However, unlike age equivalence scores, raw scores are not directly comparable across tasks due to differences in scoring ranges. Furthermore, raw scores do not always progress linearly. For example, vocabulary development in younger children is much steeper than in older children. Thus, progression in raw scores in WS can be considered for single tasks only, and this must be within the context

of the nature of the growth curve of the typical population (i.e., that it might not be linear). These considerations have been taken into account in the current studies.

Longitudinal development of verbal and non-verbal ability in Williams syndrome

We now review studies of longitudinal development of verbal and non-verbal ability in WS (for a comparative summary, see Table 1). Jarrold an colleagues used age equivalence scores to determine the rates of development of receptive vocabulary (British Picture Vocabulary Scale; BPVS, Dunn et al., 1982) and the Pattern Construction (PC) subtest of the Differential Ability Scales (Elliott, 1990) cross-sectionally and longitudinally in WS. Cross-sectional data (N=16; 6 to 28 years) demonstrated significantly stronger verbal than non-verbal ability in WS at a group level (Jarrold et al., 1998). Longitudinal data was from N=15 of the same sample (6 timepoints over 40 months; Jarrold et al., 2001) (note, Jarrold et al. [1998] also presented a two timepoint subset of this data, with timepoints spaced by 8 months; see Table 1). In both cross-sectional and longitudinal datasets, the authors demonstrated shallower rates of development in age equivalence scores for non-verbal, relative to verbal ability. Thus, at a group level, the gap between verbal and non-verbal ability widened with increasing chonological age. Furthermore, using difference scores between verbal and non-verbal age equivalence scores, Jarrold et al. (1998) demonstrated that individual differences in the discrepancy between verbal and non-verbal ability can be explained by level of intellectual functioning (in this case using verbal ability as a proxy for intellectual functioning). That is, individuals with stronger intellectual functioning, have a larger discrepancy between verbal and non-verbal ability.

The pattern of steeper verbal than non-verbal progression was mirrored in cross-sectional data from Thomas et al. (2009) with N=28 individuals, also using BPVS and PC age equivalence scores. Thomas et al. (2009) additionally included a small longitudinal sample (N=4, 2 timepoints) to demonstrate the principle of longitudinal validation; whilst their cross-

sectional data predicted longitudinal change for BPVS as in Jarrold et al. (2001), it was less accurate for PC. The authors suggested this may have been due to PC floor effects. Thomas et al.'s (2009) longitudinal data, however, were simply for proof of principle, rather than formal interpretation.

Porter and Dodd (2011) used the Woodcock Johnson Test of Cognitive Ability – Revised (WJ-RCOG; Woodcock & Johnson, 1989) with individuals with WS (N=27; 5 to 45 years at timepoint 1, 2 timepoints spaced by 5 years). There was no differentiation in performance between verbal and non-verbal factors, SSs remained consistent across time and improvement in raw scores over time did not reflect any advantage of verbal subtests. In contrast to the studies above, this does not support the characteristic WS cognitive profile, and suggests consistency in cognitive profiles over time and similar rates of development for each cognitive factor. There was, however, a relationship between chonological age and change in SSs between timepoints for verbal ability only, thus differentiating verbal from non-verbal abilities. This reflected a decline in verbal SSs in younger participants (the gap between the WS group and the normative sample increased) and an increase in SSs in older participants (the gap between the WS group and the normative sample decreased). However, the authors point out that the WJ-RCOG verbal factor is confounded by contributions from executive function which could explain the lack of consistency with the studies above, and could account for the relationship between verbal ability and chonological age. Overall, therefore, comparison of this study to similar longitudinal studies is not straightforward due to a lack of comparability of measures.

Howlin et al. (2010) report cross-sectional (N=92, 19 to 55 years), and longitudinal cognitive data (N=47, 19 to 38 years at timepoint 1, 2 timepoints spaced by ten years) from individuals with WS. For both cross-sectional and longitudinal datasets, Verbal IQ was higher than Performance IQ (measured using Weschler test batteries (timepoint 1: Wechsler

Adult Intelligence Scale – Revised [WAIS-R]; timepoint 2: WAIS-III [Wechsler 1997]). This was supported at an individual level, with few participants showing the opposite profile. The authors report stable IQ over time in WS, from both datasets. In the cross-sectional data, performance on the Matrix reasoning subtest was the only task to show change, with higher SSs in the 40-year-old plus group than the 19- to 29-year-old group. In the longitudinal data, BPVS age equivalent scores increased over time. Other longitudinal changes were either within the 95% confidence interval of the test (PIQ) or confounded by the use of different versions of the test between timepoints (Expressive One Word Picture Vocabulary Test (EOWPVT; Brownall, 2000)). Thus, evidence for broad stability in IQ over time is consistent with Porter and Dodd (2011), and the discrepancy between verbal and non-verbal ability is consistent with Jarrold et al. (1998; 2001). With the exception that longitudinal progression was observed for BPVS scores only, there was little evidence that this gap widens with increasing age. This is unsurprising because all participants in this study were adults, i.e., the data did not capture the portion of the trajectory in which the verbal-nonverbal discrepancy emerged.

Fisch et al. (2012) used the Stanford-Binet IV (Thorndike et al., 1986) (N=17, 3 to 15 years of age at timepoint 1, 2 timepoints spaced by 2 years). They report a mean drop of 2.06 IQ points between timepoints (verbal and non-verbal IQ were not reported separately), and similarly to Porter and Dodd report the biggest longitudinal drop in IQ for the youngest participants. It should be noted, however, that there were large individual differences in difference scores between timepoints, with IQ differences ranging from -12 to +14 IQ points.

Mervis et al. (2012) used the Kaufman Brief Intelligence Test-2 (KBIT2; Kaufman & Kaufman, 2004) longitudinally with children with WS (N=40, 4 to 14 years at timepoint 1, 4 to 7 timepoints). They report no significant difference between verbal and nonverbal ability, and no significant change in SS over time for either verbal or non-verbal ability at a group

level, i.e., a lack of support for the WS cognitive profile. However, this might be explained by the substantial individual differences; large individual differences were reported for longitudinal intercepts for SS in both domains, and for rates of SS development (slopes) in the non-verbal domain. They also report a significant correlation between verbal and non-verbal SS intercepts at age 10 years (the age at which the model was centred), but not rates of development. This indicates that individuals with the lowest non-verbal ability at ten years were also those with the lowest verbal ability at this age, whilst the lack of association for rate of development is likely due to the large individual differences in slopes for non-verbal ability only. They also report a correlation between intercepts and slopes in the verbal domain, indicative of greater longitudinal progression in those individuals with higher intercepts at 10 years. Equivalent correlations are not reported for the non-verbal domain. This pattern for the verbal domain shows some parallel with Jarrold et al. (1998) who report that those with higher intellectual functioning show a larger verbal-nonverbal discrepancy, suggestive of steeper verbal progression for those individuals. Thus, whilst at a group level Mervis et al. (2012) demonstrate a lack of support for the WS cognitive profile, associational evidence appears to show some evidence that verbal ability is differentiated from non-verbal ability in their WS sample.

Mervis and Pitts (2015) report longitudinal development of vocabulary and IQ (Differential Ability Scales 2nd edition [DAS-II; Elliott, 2007]) in WS (N=76, 4 to 15 years at timepoint 1, two timepoints three years apart). At a group level, they report higher Verbal SS than Spatial SS, but that Non-verbal Reasoning SS was higher than both Verbal and Spatial SS. They also report a reduction in SS over time for vocabulary, Verbal SS and Non-verbal Reasoning SS (i.e., the gap between SS of the WS group and the normative sample increased), but not Spatial SS. This does not support the characteristic WS cognitive profile, or any developmental advantage for verbal development, and shows some similarity to Porter

and Dodd (2011). The authors also report no relationship between chonological age and change in SS between timepoints. However, when split into a younger (<7.5 years at timepoint 1) and an older group (>7.5 years at timepoint 1), the magnitude of absolute change was significantly higher for the younger group for all IQ measures. Whilst Mervis and Pitts (2015) report absolute change (as opposed to directional change), when coupled with the group decline in SS reported, this is broadly consistent with Fisch et al. (2012) and Porter and Dodd (2011).

Fisher et al. (2016) used the KBIT2 (Kaufman & Kaufman, 2004) with individuals with WS (N=52, 14 to 49 years at timepoint 1, two to seven timepoints, spaced by 2 to 9 years). Group means demonstrated the characteristic profile of stronger verbal than non-verbal IQ (verbal IQ: 77.9; non-verbal IQ: 70.8), although statistical comparison was not reported. Similar to Howlin et al. (2010) and Mervis et al. (2012), neither verbal IQ nor non-verbal IQ showed statistical change in SS over time.

Sauna-aho et al. (2019) used Wechsler test batteries (Wechsler Adult Intelligence Scale – Revised [WAIS-R; Wechsler, 1992]; Wechsler Intelligence Scale for Children – Revised [WISC-R; Wechsler, 1984]) with individuals with WS (N=25, 19 to 68 years at timepoint 1, N=18 at timepoint 2, spaced by 20 years). VIQ and PIQ were comparable, although no statistical comparison was reported. Change in IQ scores over time was within the 95% confidence intervals indicative of no change. The relationship between chonological age and IQ showed a quadratic function for both VIQ and PIQ. This reflected cross-sectional stability in VIQ from 19 to 40 years, followed by gradual decline, and improvement in PIQ between 19 and 50 years, followed by gradual decline.

In summary, studies of longitudinal development of verbal and non-verbal ability in WS have used varying measures, variables and age ranges. There is a broad consensus that verbal ability is higher than non-verbal ability in WS. Longitudinal rates of development

typically demonstrate similar rates for both verbal ability and non-verbal ability, with some studies showing a shallower rate of development in the non-verbal domain compared to the verbal domain (Jarrold et al., 2001). These differences across studies might relate to the purity of the measures employed and/or the level of intellectual functioning of the samples. Where measured, there is a consistent finding that for younger participants change over time indicates that the gap between individual with WS and their peers is increasing (Porter & Dodd, 2011; Fisch et al., 2012; Mervis & Pitts, 2015). This likely reflects the rapid neural and cognitive development in children, relative to adults. Sauno-aho et al. (2018) is the only study to demonstrate a quadratic function between IQ and age. This is due to the wide age range of participants (their oldest participant was 86 years), which encompassed an age range in which cognitive decline would be expected in the typical population.

Our study is distinguished from the preceding literature due to its large sample sizes, its conservative approach in ensuring independence in cross-sectional and longitudinal samples, and its reliance on greater numbers of timepoints for longitudinal trajectories, to reduce estimation error. We report data from the BPVS (Dunn et al., 1997; Dunn & Dunn, 2009) as a measure of verbal ability, whilst two measures of non-verbal ability are included: Ravens Coloured Progressive Matrices (RCPM; Raven, 1993), and the Pattern Construction subtest of the British Ability Scales (PC; Elliott et al., 2007; Elliot & Smith, 2011). In Study 1, we report cross-sectional data from N=209 individuals with WS for performance on BPVS, N=102 individuals with WS on RCPM and N=103 individuals with WS on PC. In Study 2, we report longitudinal data for individuals with WS who had been tested on at least three timepoints on BPVS (N=54), RCPM (N=41) and PC (N=17).

Study 1: Cross-sectional associations

Using the largest WS sample of its kind to-date, we tested models of (cross-sectional) development. If the rate of development in WS is slow, this predicts a shallow gradient of

broadly linear progression. If development in WS is characterised by delayed onset followed by developmental catch-up, this predicts a similar (albeit developmentally shifted), non-linear trajectory to that observed in typical development of relatively rapid development followed by a plateau in adulthood. We predicted a combination of these two models, that verbal and non-verbal progression in WS would be characterised by delayed onset and a delayed rate, followed by a plateau in adulthood. This would be reflected by a non-linear relationship between performance and CA. We predicted a non-linear trajectory because of the large age range of our sample, which ranges from children to adults. The rate of change might differ for children compared to adults, as exemplified by Porter and Dodd (2011), Fisch et al. (2012) and Mervis and Pitts (2015) because of differences in neural and cognitive development in children, relative to adults. In addition, we predicted that the effect size of this relationship would be negatively impacted by individual differences that are not wholly age-related, such as individual differences in intellectual functioning (Jarrold et al., 2001; Karmiloff-Smith, 1998; Mervis et al., 2012). We also investigated whether verbal abilities were superior to non-verbal abilities and, related to the point above, whether individual differences in the difference between verbal and non-verbal ability was associated with verbal ability (as a proxy for intellectual functioning), as observed by Jarrold et al. (1998). We predicted higher verbal than non-verbal ability and a larger magnitude of difference between verbal and non-verbal ability in those with higher verbal ability.

Method

Participants

Data were obtained from the WiSDom database, a database of retrospectively shared cognitive and behavioural data across seven WS labs in the UK (https://blogs.ucl.ac.uk/wisdom/). N=209 had completed the BPVS (version 2 or 3). BPVS version 1 had a different scoring procedure to BPVS2 and BPVS3, and so was deemed too

different to these latter versions, and participant data for BPVS 1 were not included. BPVS2 and BPVS3 had the same scoring procedure and were deemed comparable with no adjustment to scoring required. The final sample of N=209 ranged in age from 4 to 59 years, consisting of 88 females, 82 males, and 39 people for whom a record of sex was unavailable. N=102 had completed the RCPM (all versions are identical in content and scoring procedure), with an age range of 7 to 47 years: 31 females, 32 males, and 39 people for whom a record of sex was unavailable. N=102 had completed the PC test from the British Ability Scales 2 and 3, both of which have identical content and scoring procedure, ranging in age from 4 to 60 years consisting of 49 females, 39 males, and 14 people for whom a record of sex was unavailable. For all individuals, WS had been clinically or genetically confirmed.

Measures

The BPVS (Dunn et al., 1997; Dunn & Dunn, 2009) measures receptive vocabulary. For each trial, the participant hears a word and chooses which picture, from a set of 4 pictures, depicts the meaning of the word. In versions 2 and 3, trials are arranged in sets of 12, and the basal set is the lowest set in which the participant made no more than one error, and the ceiling set is the set in which the participant made 8 or more errors. Raw score is calculated as the ceiling item minus the number of errors.

The RCPM (Raven, 1993) is a measure of non-verbal reasoning. Across 36 trials, the participant is shown a pattern, or pattern sequence, with an element missing. They must choose, from a matrix of 6 pattern pieces, which one is the missing element to complete the pattern or pattern sequence. Raw score is the number of accurate answers out of 36.

PC is a visuo-spatial (non-verbal) subtest of the British Ability Scales (BAS II; Elliott et al., 2007; BAS III: Elliott & Smith, 2011). Participants are presented with 2D squares (earlier trials) or 3D cubes with faces of either one colour or of two colours divided along the

diagonal. Participants are asked to arrange the upper faces of the cubes (or the squares) to model either an example built by the experimenter or a 2D image in a stimulus booklet. Trials begin with a 2 by 1 design, with harder trials requiring 9 cubes in a 3 x 3 design. Age equivalence scores (calculated at the time of data collection) were used for this task, because the scoring procedure is such that raw scores are not meaningful without reference to the specific testing schedule in each testing session and this information was not available.

Results

Cross-sectional trajectories

The cross-sectional developmental trajectories of BPVS, RCPM and PC were established by first removing scores at floor (0 participants for BPVS, 0 for RCPM, 18 for PC), and then curve-fitting against chonological age (CA), selecting from linear, logarithmic, power, and quadratic models. Raw scores were used for BPVS and RCPM, but age equivalence score was used for PC as discussed in the method. A logarithmic model was significant and best-fitting for BPVS, $F(1, 207) = 189.042$, $p < .001$, $R^2 = .477$, AIC = 1918 (lower values of AIC indicate better model fit; for comparison, linear model AIC = 1950). The best fitting, but non-significant, model for RCPM was a power model , $F(1, 100) = 2.475$, $p = .119$, $R^2 = .024$, AIC = 57.6 (linear model AIC = 624), whereas a power model was both significant and best-fitting for PC, $F(1, 100) = 37.800$, $p < .001$, $R^2 = .274$, AIC = 22.5 (linear model AIC = 875), see Figure 1. Describing the curves, performance for both BPVS and PC increased at a constantly decreasing rate with advancing CA, whilst RCPM performance was not reliably related to CA (a 45-degree rotated correlation further confirmed this lack of relationship; $r(100) = -0.051$, $p = .609$; see Thomas et al., 2009).

Comparison of verbal and non-verbal ability

For individuals who were assessed at the same timepoint with both BPVS and RCPM (N=25), a repeated measures t-test was used to determine the difference in age equivalence scores for verbal and non-verbal ability respectively. Consistent with previous research, this demonstrated significantly stronger verbal than non-verbal ability in our sample, $t(24)=5.40$, $p < .001$, mean (SD) difference: 34.81 (32.22) months. A corresponding analysis for individuals who were assessed at the same timepoint with both BPVS and PC (N=61) also demonstrated significantly stronger verbal than non-verbal ability in our sample, $t(60)=10.07$, $p < .001$, mean (SD) difference: 39.25 (30.44) months.

Comparison of BPVS and RCPM for the full samples was carried out using a partially overlapping samples t-test in R (Derrick, 2018; R Core Team, 2021; see Derrick, Toher, & White, 2017, for an overview of the method). RCPM was used as the measure of non-verbal ability here on account of there being a higher volume of data for RCPM than for PC. This method enables two samples to be compared where those samples are only partially independent, i.e., where a proportion of participants are in both samples, and involves coding each datapoint as repeated measures or independent, with reference to the two variables. This allows one to use the full sample, but at the same time introduces the limitation of differences in within sample individual differences between the BPVS and RCPM samples. Consistent with the repeated measures t-test above, this demonstrated significantly higher age-equivalence for BPVS than for RCPM, $t(183.908) =5.230$, $p <.001$, mean difference: 15.70 months, equal variances not assumed.

We were also interested in individual differences in the magnitude of difference between verbal and non-verbal ability, and whether the magnitude of difference was associated with differences in intellectual functioning. Akin to Jarrold et al. (1998), for our repeated measures subsample (BPVS vs RCPM; N=25; BPVS vs. PC; N=61), we used verbal ability to account for differences in intellectual functioning across the sample, by plotting the

difference in verbal and non-verbal ability, against verbal ability. Figure 2a, 2b and 2c demonstrate verbal vs. non-verbal difference scores plotted against verbal ability for raw scores (adjusted for differences in the range of scores between measures; normalised raw score) and for age equivalence scores. These show significant linear relationships for BPVS vs RCPM difference scores: normalised raw score: $r=.742$, $N=25$, $p < .001$; age equivalence scores: $r= .895$, $N=25$, $p < .001$; and BPVS vs PC difference scores: age equivalence only (raw score unavailable): $r= .860$, $N=61$, $p < .001$.

Discussion

Consistent with the literature (e.g., Howlin et al., 2010; Jarrold et al., 1998; 2001), comparison of age equivalence scores demonstrated higher verbal than non-verbal ability for the full sample and the repeated measures sub-samples. This reflects the WS characteristic cognitive profile. The difference was very stark with over 15 months difference for the full sample and over 34 months difference for the repeated measures sample. The repeated measures sample, whilst smaller, does not have the confound of differences in within sample individual differences between the measures, which is a limitation of the full sample comparison. Taken together, we can be confident in the reliability of these findings.

The group level data above are important for confirming support for the WS cognitive profile, but the large individual differences in the discrepancy between verbal and non-verbal ability across the sample are notable. To address these individual differences, we replicated Jarrold et al. (1998) to demonstrate that individual differences in the discrepancy between verbal and non-verbal ability were largely accounted for by level of intellectual functioning. Using verbal ability as a proxy for intellectual functioning, we demonstrated that those with stronger verbal ability had a larger gap between verbal and non-verbal ability. This suggests that verbal ability (specifically receptive vocabulary) progresses at a steeper rate than non-verbal ability in our sample. Notably, verbal ability explains a large amount of these

individual differences in our sample, explaining 55% of the variance in raw score differences and 74% to 80% of the variance in age equivalence score. Nevertheless, as discussed below, progression with increased chonological age also supports the suggestion of steeper verbal than non-verbal progression in the child participants in our sample, i.e., many (but not all) of the more verbally able children are also older children.

To investigate progression with age on each measure, we plotted performance against chonological age. As predicted, scores on our three measures were not linearly related to chonological age, thus did not support a model of delayed rate of development. Instead we demonstrated non-linear relationships between chonological age and performance for BPVS and PC measures; these relationships broadly support our prediction of a model of delayed onset (performance at the intercept was low) and delayed rate (progression was shallow relative to normative progression), followed by a plateau. For BPVS, a significant logarithmic relationship was revealed (48% of variance explained). For PC, the relationship was fitted significantly by a power model (22% of variance explained). The pattern of little developmental change in adulthood observed for BPVS and PC (see Figures 1a and 1c) is consistent with Howlin et al. (2010), who report no developmental progression in BPVS scores in their adult WS sample, and with Porter and Dodd (2011), Fisch et al. (2012) and Mervis and Pitts (2015) who demonstrate differential patterns of development in their younger participants, relative to older participants, likely reflective of faster neural and cognitive development in children, compared to adulthood. However, as predicted, for both model fits a large amount of variance remained unexplained by chonological age. Note, whilst comparison to normed data is useful, a neurotypical comparison sample would enable a comparison of model fits to typical developmental trajectories. Furthermore, these are cross-sectional data and thus comparison to longitudinal studies is exercised with caution.

As shown in Figure 1b, the scatter plot for RCPM illustrates a cloud of data with no significant relationship with chonological age for the full age range. This contrasts with Howlin et al.'s (2010) finding of higher performance in their oldest group, 40 years plus, compared to their 19 to 29 years age group on a similar task, the matrix reasoning subtest of the WAIS-III. However, Howlin et al.'s (2010) finding is difficult to interpret and could simply reflect incidental differences in the range of intellectual functioning of their age bands. One would not necessarily predict progression in RCPM in the adult portion of our sample, but for the younger participants a lack of progression could reflect the low ceiling for non-verbal ability in WS coupled with large individual differences. This is consistent with our finding that the gap between verbal and non-verbal ability is stronger for those with higher verbal ability. However, the developmental pattern for RCPM differs from evidence of some developmental progression with age in our sample on the PC task, which is also a measure of non-verbal ability. Whilst error analysis of RCPM has demonstrated that it taps into similar non-verbal mechanisms in individuals with WS, compared to the typical population (Van Herwegen et al., 2011), PC performance is atypical in WS due to the construction demands, a particular weakness in WS (Farran & Jarrold, 2003). It is also a timed task (for all but the earliest items), which adds processing speed demands. Thus, although differences between RCPM and PC trajectories could reflect differences in the samples (participants for each measure were overlapping, but not identical), the difference is most likely related to differences in task approach. That is, construction draws on processes such as memory, executive functions, and fine motor skills. These demands likely present a more uniform performance limitation in WS, particularly in young children, and thus represent a strong developmental contributor to progression on this task.

In summary, in line with predictions, we have demonstrated the characteristic profile of stronger verbal than non-verbal ability in WS and shown evidence of developmental

progression in raw scores for both domains, with steeper development in verbal than non-verbal ability, particularly at the younger end of the range of ages in our sample. It is important to note, however, that whilst cross-sectional designs are useful for gathering large samples thus invoking sufficient statistical power and representativeness, conclusions related to change with age are confounded by between-participant individual differences and are of most value when considered in tandem with longitudinal data.

Study 2. Longitudinal trajectories

In Study 2, we investigated the longitudinal trajectories of verbal ability and non-verbal ability, using the same three tasks: BPVS, RCPM and PC. In contrast to previous studies, we classified longitudinal data as a minimum of three timepoints. Three timepoints provide a more interpretable insight into longitudinal trajectories. With just two timepoints one is unable to index the reliability of the rate of change seen between these points (as two datapoints are necessarily connected by a straight line), which can introduce larger estimation error of gradients than when more timepoints are included. Furthermore, with more than two timepoints one can examine whether the rate of change remains constant over two or more intervals, with some consequent indication of the reliability of that rate of change. For BPVS and RCPM, cross-sectional datapoints from Study 1 were excluded from Study 2, in order that the two datasets were independent. This enabled comparison of cross-sectional and longitudinal findings, whilst avoiding the confound of overlapping datasets. These two elements of the design drastically reduced our longitudinal sample size but increased scientific rigour.

We were interested in the developmental patterns of verbal and non-verbal ability, predicting that the longitudinal data would validate the cross-sectional data, i.e., that group level and within-participant longitudinal rates of development would be shallower for non-

verbal ability than for verbal ability. We were also interested in whether participants' changes in score over time (i.e., the gradient of their longitudinal trajectory) would be related to chonological age. This was carried out for replication purposes (e.g., for comparison with Porter & Dodd, 2011 and Mervis & Pitts, 2015) and for comparison with study 1. We predicted that chonological age would not account for a large amount of variance in this analysis. However, in line with study 1 data and our hypothesis of developmental catch-up in WS, we predicted that, broadly, we would observe a model of delayed onset and delayed rate, and larger developmental change (steeper gradients) for children, compared to adults due to faster neural and cognitive changes in children compared to adults.

Method

The same measures were used as in Study 1. Longitudinal data were collated for BPVS and RCPM for the subsets of the Study 1 samples of individuals with at least three testing records ~6 months apart or more, with resulting subsamples of 41 and 53 individuals, respectively. In addition, cross-sectional data from the analyses in Study 1 were excluded in order that the two studies consisted of independent samples. For PC, there were very few repeat testing records across individuals; there were 17 individuals with at least three testing records, but data common with the cross-sectional analyses were used for the first time-point

Results

To test for any inconsistency of gradients between T1-T2 and T2-T3, repeated-measures t-tests were conducted and revealed no significant difference between the gradients for BPVS, $t(101.93) = 0.530$, $p = .597$, mean difference = 0.145 (T2-T3 > T1-T2, ns), equal variances assumed, RCPM, $t(80) = 0.393$, $p = .696$, mean difference = 0.035 (T2-T3 > T1-T2, ns), equal variances not assumed, or PC, $t(17.18) = 1.161$, $p = .262$, mean difference = 6.34

(T2-T3 < T1-T2, ns). Thus, the rate of change in our sample remained constant over the timepoints measured, indicative of reliability of the longitudinal gradients in our sample.

Longitudinal trajectories

BPVS

Individual participant linear trajectories are shown in Figures 3a-3c. A two-level mixed model was used to evaluate the relationship between BPVS raw scores and CA. Mixed models allow for both differing numbers of observations and differing ranges of covariates (e.g., CA) across participant; such models also account for the 'random' dependencies of observations within each participant rather than treating them as measurement error, as is the case in traditional non-hierarchical linear models. At Level 1, linear change of BPVS raw score with CA was modelled, with intercept allowed to vary randomly at Level 2 to model differences in this parameter across participants. Gradient was treated as a fixed rather than random effect in this model, given that possible developmental change was the primary subject of investigation.

Three individuals had negative coefficients (6%), 50 had positive ones (94%). Pseudo-*R*-square model fit was 0.83. There was a significant increase of 0.23 raw score points with each month of CA, $t(230) = 12.447$, $p < .001$ (95% CI = 0.19, 0.26). This compares to an increase of 0.85 points per month in the typical population (based on a linear line-of-best-fit for the BPVS age equivalence norms data, although note that this applies to norms data for 3 years 9 months to 15 years 9 months only).

RCPM

For RCPM, a model was constructed that was identical in structure to that above, with RCPM raw score instead of BPVS. Fifteen individuals (37%) had negative coefficients, 26 (63%) had positive ones. To better interpret the large number of negative coefficients, we

determined how many of these were likely to be statistical error vs. true decline. Of the fifteen individuals, 7 had scores with overlapping confidence intervals between the highest and lowest scores (the RCPM manual states that ±2 raw score points are within the 97.5 Confidence Interval) and thus are not considered to represent true decline. The remaining 8 individuals had a broader range of scores; however, even for this subgroup, not all showed consistent decline across their set of scores. The overall model fit was evaluated by deriving a pseudo-*R*-square value from the correlation between the fitted and the observed values, resulting in a figure of 0.74 using the MuMin package (Bartoń, 2022). There was a modest but significant increase of 0.008 raw score points with each month of CA, $t(196) = 2.105$, $p = .036$ (95% CI = 0.00052, 0.016). This compares to an increase of 0.27 points per month in the typical population (based on a linear line-of-best-fit for RCPM age equivalence norms, although note that this applies to norms data for 5 years 6 months to 11 years 6 months only).

PC

Finally, a corresponding model was constructed for PC. Age equivalence score was used as the dependent variable, for the same reasons as in Study 1. All 17 participants had positive coefficients (Pseudo-*R*-square model fit was 0.71). There was a significant increase of 0.12 age equivalence score points with each month of CA, $t(52) = 5.027$, $p < .001$ (95% CI = 0.069, 0.16) (compared to one age equivalence month for each month of CA in the typical population).

Comparison of verbal and non-verbal longitudinal gradients

We compared longitudinal gradients for BPVS and RCPM using a partially overlapping samples t-test, i.e., each data point was coded as repeated measures or independent, with reference to the two variables. This allows one to use data from the full longitudinal sample. In order to compare like-for-like between tests we used age equivalence scores for this

analysis; gradients were extracted from the relationship between age-equivalence (Y-axis) and CA (X-axis): BPVS (M=0.204, SD=0.444); RCPM (M=0.070, SD=0.260). Despite a substantial mean difference, the partially overlapping samples t-test revealed no significant difference between the longitudinal gradients for BPVS and RCPM: $t(46.232) = 1.850$, $p = .071$, mean difference = 0.135 (BPVS>RCPM, ns), equal variances not assumed. Mean difference is an effect size measure, computed by imputing the mean value for all 'missing' values (where there is no repeated measure). The corresponding analysis comparing gradients of BPVS (M=0.204, SD=0.444) and PC (M=0.263, SD=0.299) showed no significant difference between them: $t(32.092) = 0.594$, $p = .557$, mean difference = 0.059 (PC>BPVS, ns), equal variances not assumed.

Relationships between Chonological age and cognitive change

To determine whether the patterns of longitudinal development above were associated with the age of the participants, we investigated the relationship between the slope function of their timepoints and their chonological age (i.e., gradient of raw score [Y-axis] against CA [X-axis]). We used mean age across their testing points as our CA variable. The relationships were determined by curve-fitting, selecting from linear and quadratic models. For both BPVS and RCPM, the best fitting model was a quadratic model: BPVS, $F(2, 50)=6.196$, $p =. 004$, $R^2=.206$; RCPM, $F(2, 38)=5.292$, $p= .009$, $R^2=.218$. For PC, a linear relationship was the best model, but this was not significant, $F(1, 15)= 2.415$, $p= .141$, $R^2=.139$ (Figure 4).

Discussion

Study 2 tracked verbal and non-verbal ability across a minimum of three timepoints in a large sample of individuals with WS. Longitudinal trajectories of verbal and non-verbal ability demonstrated significant longitudinal progression for each task, BPVS, RCPM and PC, at a group level. The significant progression, albeit relatively shallow, for all three tasks,

contrasts to the cross-sectional data, where progression with age was observed for BPVS and PC only. Thus, whilst this study shows some validation of cross-sectional developmental patterns, it also demonstrates the value of longitudinal data in understanding development and suggests that, in the cross-sectional RCPM data, large individual differences were masking evidence of cross-sectional developmental progression.

At an individual level, positive longitudinal developmental slopes were observed for 94% of individuals for BPVS and for 100% of individuals for PC, but only for 63% of individuals for RCPM (although, note different sample sizes per task). Thus, whilst at a group level, participants progressed in RCPM over time, there was heterogeneity at an individual level. This does have some similarities to the cross-sectional data, where no relationship with age was observed for RCPM. This heterogeneity was not, however, reflected in the PC task, also a measure of non-verbal ability. As discussed in Study 1, perhaps this difference between these two non-verbal tasks reflects the construction demands (and relatedly, motor, processing speed and/or executive demands) in the PC task. A decline in RCPM score in a subgroup of our sample is difficult to interpret. Inspection of Figure 3b demonstrates that the negative slopes for RCPM are not age related. For some the decline was within the confidence interval of the measure and may not represent true decline. Others showed a spiky trajectory across their timepoints, which could indicate the influence of environmental factors such as fatigue or time of day. We cannot rule out, however, that for some, RCPM showed true decline across timepoints.

Developmental progression between verbal and non-verbal ability was compared, considering the full sample (i.e., combining repeated measures and independent samples data). Contrary to predictions and to the cross-sectional data, whilst the difference in slope values was in the predicted direction (BPVS > RCPM), this trend for a difference in longitudinal progression between BPVS and RCPM in WS was not significant. This could

indicate that the developmental timeline of individual trajectories (a minimum of three timepoints) was not sufficient for verbal and non-verbal trajectories to diverge, but could also reflect that the difference in cross-sectional slopes between verbal and non-verbal ability reported in Study 1 is influenced by individual differences. This longitudinal pattern in our data does not support that observed by Jarrold et al. (2001) who demonstrated shallower longitudinal progression for nonverbal ability (measured using PC) compared to verbal ability (measured using BPVS). Whilst both studies had a similar age composition, it is possible that the difference across studies reflects differences in intellectual functioning (our sample is less able, with a maximum BPVS score of 204 compared to ~230 for Jarrold et al., 2001), the length of developmental timelines of the sample (minimum of 3 timepoints in our sample vs. 6 timepoints in Jarrold et al., 2001), or differences in sample size (our sample is larger for BPVS and so less impacted by individual differences). The pattern observed in the current study is consistent with Porter and Dodd (2011), who report improvement over time in raw scores for their sample of children and adults, but no specific advantage for verbal subtests.

We were also interested in whether participants' changes in score over time would be related to chonological age. Based on the literature, we predicted larger developmental change for children than for adults. This was largely supported; the relationship between chronological age and developmental progression in BPVS and RCPM was best explained by a quadratic function. Observation of Figure 4 indicates that the steepest gradients on these measures were amongst the youngest participants, and that developmental progression then reduced with age until early adulthood when rates of progression reached a plateau, i.e., adults showed consistent rates of relatively shallow progression over time. No significant relationship was observed for PC, but this likely reflects a lack of power. This supports our hypothesis of a period of developmental catch-up in the children in our sample. This could be

an impact of the late onset of many verbal and non-verbal competencies (e.g., Laing et al., 2002), which would be observed as steeper developmental progression in younger participants as their neural and cognitive capabilities are progressing at speed, relative to older participants. This pattern is consistent with Fisch et al. (2012) and Mervis and Pitts (2015) who report higher levels of change in their younger participants, whilst Porter and Dodd (2015) observed this pattern in the verbal domain only. It is important to note, however, that whilst the relationship between chronological age and slope value was significant, for both BPVS and RCPM, the proportion of variance explained by this relationship was 21% and 22% respectively, leaving a large amount of variance unexplained.

General Discussion

In the current paper, we present cross-sectional and longitudinal data in the largest WS sample to-date, drawn from the WISDOM database (https://blogs.ucl.ac.uk/wisdom/). Ours is the most rigorous work in this area for two reasons. First, our cross-sectional and longitudinal BPVS and RCPM samples are independent and thus any similarities or differences are not confounded by shared data between the studies. Second, the current paper reports longitudinal data with a minimum of three longitudinal datapoints for each individual. Our work is not, however, without limitations. These are legacy data and thus we have a mix of repeated measures and independent samples across our three tasks, and a range in the number of developmental timepoints within our sample. Whilst this has been accounted for statistically, future studies could employ a repeated measures design and a fixed number of timepoints. Furthermore, our measures were not from the same testing battery which makes comparison across them less reliable than if they had been from the same testing battery with common normative data.

Our cross-sectional data demonstrated support for the WS cognitive profile of stronger verbal than nonverbal abilities and support a developmental model of non-linear

development in WS, rather than linearly slow development. This is also mirrored in the analyses of the relationship between longitudinal change over time and chronological age. This suggests that development in WS is characterised by delayed onset followed by a delayed rate of developmental catch-up, i.e., a non-linear trajectory of relatively rapid development followed by a plateau in adulthood, similar to that observed in typical development (albeit developmentally shifted and shallower). These findings inform parents and educators of predicted developmental patterns in WS, thus enabling them to broadly determine developmental potential and tailor approaches accordingly. For example, focussing more support on weaker areas of cognition in early development, and focussing more support on compensation strategies, by drawing on relative strengths, in later development are potential strategies afforded by the non-linear trajectories reported here (see also Van Herwegen et al., 2019).

The cross-sectional data suggest that individual differences in intellectual functioning largely explain the discrepancy between verbal and non-verbal ability, whilst the longitudinal data suggest that the discrepancy between verbal and non-verbal ability remains constant with increasing age for the age range in our sample. For the longitudinal data, although not significant, there was a trend towards diverging developmental trajectories. It could be that the time windows available across timepoints for the longitudinal data analysis were not spread over a sufficiently long time to capture the process of developmental divergence. Also, first timepoint of longitudinal data was used for study 1 and not included in study 2. Thus, the study 2 longitudinal data do not include the timepoint where development is likely at its fastest. These arguments could explain why there was not significant evidence that the discrepancy increased with increasing age for the range of developmental timelines, abilities and ages of our sample. Future research which includes larger samples and longer timelines,

and which start at earlier points in development is needed to fully answer questions of developmental divergence.

In summary, our data support the WS cognitive profile of stronger verbal than non-verbal ability, but also demonstrate that for both domains, when compared to TD norms, longitudinal developmental progression is delayed and shallower than in the typical population. This suggests that the gap between individuals with WS and their typically developing peers widens over time; this relates to the finding that children with WS are likely to move from mainstream schools to more specialist educational settings in the transition from primary to secondary school (Van Herwegen et al., 2018). Consistent with previous research, model fits to the cross-sectional and longitudinal data demonstrate steeper rates of development in the child participants in our sample than the adolescent and adults in our sample. One inconsistency between cross-sectional and longitudinal data relates to the rates of development for verbal and non-verbal domains. The cross-sectional data demonstrated that verbal ability was associated with difference scores between verbal and non-verbal ability, suggestive of a steeper rate of development for verbal abilities than nonverbal abilities. The longitudinal data, however, do not mirror this finding and instead demonstrate no statistical difference between the longitudinal slope values for verbal vs. non-verbal ability in our large sample. The mixed findings for a difference in rate of development is also observed in the literature with some studies reporting a verbal advantage in rate of development (Jarrold et al., 2001) and others not (Mervis & Pitts, 2015; Porter & Dodd, 2011), and thus invites further research. The current series of studies complements and builds upon previous knowledge of developmental progression in verbal and non-verbal domains in WS and is one of the largest comparisons of cross-sectional and longitudinal data to date. Whilst most of our cross-sectional findings were validated longitudinally, where they were not, this enabled us to differentiate between effects of individual differences vs. longitudinal

development, and therefore to quantify a confound that has long existed in the cross-sectional literature.

References

Bartoń, K. (2022). *MuMIn: Multi-Model Inference. R package version 1.46.0.* https://CRAN.R-project.org/package=MuMIn

Brownall, R. (2000) *Expressive One-Word Picture Vocabulary Test*. Academic Therapy Publications, Novato, CA.

Derrick, B. (2018). Partiallyoverlapping: Partially Overlapping Samples Tests. R package version 2.0. URL https://CRAN.R-project.org/package=Partiallyoverlapping

Derrick, B., Toher, D., & White, P. (2017). How to compare the means of two samples that include paired observations and independent observations: A companion to Derrick, Russ, Toher and White (2017). *The Quantitative Methods for Psychology*, 13(2), 55-65. https://doi.org/10.20982/tqmp.13.2.p120

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale, Third Edition*. London: GL Assessment.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale, Second Edition*. NFER-Nelson.

Dunn, L. M., Dunn, L. M., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Scale*. Windsor, U.K.: NFER-Nelson.

Elliot, C. D. (1990). *Differential Ability Scales*. New York: The Psychological Corporation

Elliott, C. D., & Smith, P. (2011). *British Ability Scales—Third Edition (BAS-3)*. London: GL Assessment.

Elliott, C.D. (2007). *Differential Ability Scales, 2nd edition*. San Antonio, TX: Psychological Corporation. https://doi.org/10.1037/t15074-000

Farran, E. K., & Scerif, G. (2022). Genetic syndromes, neuroconstuctivism, and replicable research; challenges and future directions. *Infant and Child Development*. e2307. https://doi.org/10.1002/icd.2307

Farran, E.K. (2021). What can neurodevelopmental disorders tell us about developmental pathways? Realising neuroconstructivist principles now and in the future. In Mareschal, D., Knowland, V., & Thomas, M. S. (Eds). *Taking development seriously: a Festschrift for Annette Karmiloff-Smith. Neuroconstructivism and the multi-disciplinary approach to understanding the emergence of mind*. Routledge. https://doi.org/10.4324/9780429445590-14-14

Farran, E.K., Bowler, A., Karmiloff-Smith, A., D'Souza, H., Mayall, L., Hill, E.L. (2019). Cross-domain associations between motor ability, independent exploration and large-scale spatial navigation; Attention Deficit Hyperactivity Disorder, Williams syndrome and typical development. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2019.00225

Farran, E.K. & Jarrold, C. (2003). Visuo-spatial cognition in Williams syndrome; Reviewing and accounting for the strengths and weaknesses in performance. *Developmental Neuropsychology, 23,* 175-202. https://doi.org/10.1207/S15326942DN231&2_8

Fisch, G. S., Carpenter, N., Howard-Peebles, P. N., Holden, J. J., Tarleton, J., Simensen, R., & Battaglia, A. (2012). Developmental trajectories in syndromes with intellectual disability, with a focus on Wolf-Hirschhorn and its cognitive–behavioral profile. *American Journal on Intellectual and Developmental Disabilities*, *117*(2), 167-179. https://doi.org/10.1352/1944-7558-117.2.167

Fisher, M., Lense, M., & Dykens, E. (2016). Longitudinal trajectories of intellectual and adaptive functioning in adolescents and adults with Williams syndrome. *Journal of Intellectual Disability Research*, *60*(10), 920-932. https://doi.org/10.1111/jir.12303

Howlin, P., Elison, S., Udwin, O., & Stinton, C. (2010). Cognitive, Linguistic and Adaptive Functioning in Williams Syndrome: Trajectories from Early to Middle Adulthood. *Journal of Applied Research in Intellectual Disabilities*, *23*, 322-336. https://doi.org/10.1111/j.1468-3148.2009.00536.x

Jarrold, C., Baddeley, A. D., & Hewes, A. K. (1998). Verbal and Nonverbal Abilities in the Williams Syndrome Phenotype: Evidence for Diverging Development Trajectories. *Journal of Child Psychology and Psychiatry*, *39*, 511-523. https://doi.org/10.1017/S0021963098002443

Jarrold, C., Baddeley, A. D., Hewes, A. K., & Phillips, C. (2001). A longitudinal assessment of diverging verbal and non-verbal abilities in the Williams syndrome phenotype. *Cortex*, *37*, 423–431. https://doi.org/10.1016/S0010-9452(08)70583-5

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive neuroscience*, *2*, 389-398. https://doi.org/10.1016/S1364-6613(98)01230-3

Kaufman, A.S., Kaufman, N.L. (2004). Kaufman Brief Intelligence Test, 2nd edition. Circle Pines, MN: American Guidance Services.

Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Research Review: Williams syndrome: a critical review of the cognitive, behavioral, and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry*, *49*(6), 576-608. https://doi.org/10.1111/j.1469-7610.2008.01887.x

Mervis, C.B., Robinson, B.F., Bertrand, J., Morris, C.A., Klein-Tasman, B.P., & Armstrong. S.C. (2000). The Williams syndrome cognitive profile. *Brain and cognition,* 44, 604-628. https://doi.org/10.1006/brcg.2000.1232

Mervis, C. B., & Pitts, C. H. (2015). Children with Williams syndrome: Developmental trajectories for intellectual abilities, vocabulary abilities, and adaptive behavior.

American Journal of Medical Genetics Part C: Seminars in Medical Genetics. https://doi.org/10.1002/ajmg.c.31436

Mervis, C. B., Kistler, D. J., John, A. E., & Morris, C. A. (2012). Longitudinal assessment of intellectual abilities of children with Williams syndrome: multilevel modeling of performance on the Kaufman Brief Intelligence Test—Second Edition. *American Journal on Intellectual and Developmental Disabilities*, *117*(2), 134-155. https://doi.org/10.1352/1944-7558-117.2.134

Morris, C. A., Demsey, S. A., Leonard, C. O., Dilts, C., & Blackburn, B. L. (1988). Natural history of Williams syndrome: Physical characteristics. *Journal of Pediatrics*, *113*, 318-326. https://doi.org/10.1016/S0022-3476(88)80272-5

Porter, M., & Dodd, H. (2011). A longitudinal study of cognitive abilities in Williams syndrome. *Developmental Neuropsychology*, *36*(2), 255-272. https://doi.org/10.1080/87565641.2010.549872

Raven, J. C. (1993). *Coloured progressive matrices*. Information Press Ltd.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sauna-aho, O., Bjelogrlic-Laakso, N., Sirén, A., Kangasmäki, V., & Arvio, M. (2019). Cognition in adults with Williams syndrome—A 20-year follow-up study. *Molecular genetics & genomic medicine*, *7*(6), e695. https://doi.org/10.1002/mgg3.695

Strømme, P., Bjømstad, P. G., & Ramstad, K. (2002). Prevalence estimation of Williams syndrome. *Journal of child neurology*, *17*(4), 269-271. https://doi.org/10.1177/088307380201700406

Thomas, M. S. C., Annaz, D., Ansari, D., Scerif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders.

Journal of Speech, Language, and Hearing Research, 52, 336-358. https://doi.org/10.1044/1092-4388(2009/07-0144)

Thorndike, R.L., Hagen, E.P., Sattler, J.M. (1986). *The Stanford-Binet Intelligence Scale, 4th edition*. Chicago, IL: Riverside Publishing.

Van Herwegen, J., Ashworth, A., & Palikara, O. (2018). Parental views on special educational needs provision: cross-syndrome comparisons in Williams Syndrome, Down Syndrome, and Autism Spectrum Disorders. Research in Developmental Disability, 80, 102-111. https://doi.org/10.1016/j.ridd.2018.06.014

Van Herwegen, J., Farran, E. K., & Annaz, D. (2011). Item and error analysis on Raven's Colored Progressive Matrices in Williams syndrome. *Research in Developmental Disabilities*, 32, 93–99. https://doi.org/10.1016/j.ridd.2010.09.005

Van Herwegen, J., Ranzato, E., Karmiloff-Smith, A., & Simms, V. (2019). Eye movement patterns and approximate number sense task performance in Williams syndrome and Down syndrome: a developmental perspective. *Journal of Autism and Developmental Disabilities* 49(10),4030-4038. *doi: 10.1007/s10803-019-04110-0*

Van Herwegen, J., Rundblad, G., Davelaar, E.J., & Annaz, D. (2011). Variability and standardised test profiles in typically developing children and children with Williams syndrome. *British Journal of Developmental Psychology,* 29, 883-894. https://doi.org/10.1111/j.2044-835X.2010.02015.x

Wechsler, D. (1997) *Manual for the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)*. The Psychological Corporation, San Antonio, TX. https://doi.org/10.1037/t49755-000

Wechsler, D. (1984). *WISC-R: Wechler intelligence test for children (Finnish edition)*. Helsinki, Finland: Psykologien Kustannus Oy.

Wechsler, D. (1992). *WAIS-R: Wechsler Adult Intelligence Scale—Revised (Finnish edition)*. Helsinki, Finland: Psykologien Kustannus Oy.

Woodcock, R. W., & Johnson, M. B. (Eds.). (1989). *Woodcock-Johnson Psycho-educational Battery–Revised*: Itasca, IL: Riverside Publishing.

Table 1: Summary of longitudinal studies of verbal and non-verbal development in individuals with WS

Study author and year in alphabetical order		**N**	**Age range at timepoint 1**	**Design (N of timepoints)**	**Measures**	**DVs**	**Verbal vs. Non-verbal**	
							slope value	**group level**
Fisch et al. (2012)		17	3 to 15 years	longitudinal (2)	Stanford-Binet IV	IQ		
Fisher et al. (2016)		52	14 to 49 years	longitudinal (2 to 7)	KBIT2	IQ	V=NV	V>NV
Howlin et al. (2010)	Study 1	92	19 to 55 years	cross-sectional (1)	BPVS EOWPVT WAIS-R	SS AE	V=NV	V>NV (19 to 29 years) V=NV (30 years+)
	Study 2	47	19 to 38 years	longitudinal (2)	BPVS EOWPVT WAIS-R/ WAIS III	SS AE	V=NV	V>NV
Jarrold et al., 1998	Study 1	16	6 to 28 years	cross-sectional (1)	BPVS, PC	AE	V > NV	V > NV
	Study 2	16	6 to 28 years	longitudinal (2)	BPVS, PC	AE	longitudinal: V = NV cross-sectional: V>NV	V>NV
Jarrold et al., 2001		15	6 to 28 years	longitudinal (6)	BPVS, PC	AE	V > NV	V>NV
Mervis et al. (2012)		40	4 to 14 years	longitudinal (4 to 7)	KBIT2	SS	V=NV	V=NV
Mervis and Pitts (2015)		76	4 to 15 years	longitudinal (2)	DAS-II	SS	Spatial >V=NV	NV>V>Spatial
Porter and Dodd (2011)		27	5 to 45 years	longitudinal (2)	WJ-RCOG	SS raw score z-score	V=NV	V=NV

Sauni-aho et al. (2018)		18	19 to 68 years	longitudinal (2)	WAIS-R WISC-R	IQ	V=NV	V>NV
Thomas et al. (2019)	Example 1	28	5 to 12 years	cross-sectional (1)	BPVS, PC	AE	V >NV	
	Example 2	4	5 to 12 years	longitudinal (2)	BPVS, PC	AE	V=NV	

Key: dark grey shading: not applicable; light grey shading: based on means (statistical analyses not conducted). AE: age equivalence scores; BPVS: British Picture Vocabulary Scale; DAS-II: Differential Ability Scales 2nd edition; EOWPVT: Expressive One Word Picture Vocabulary Test; KBIT2: Kaufman Brief Intelligence Test-2; PC: Pattern Construction subtest of the Differential Ability Scales; SS: Standard Scores; V: Verbal; NV: Non-verbal; WAIS-R III: Weschler Adult Intelligence Scale 3rd edition; WAIS-R: Weschler Adult Intelligence Scale – Revised; WISC-R: Wechsler Intelligence Scale for Children – Revised; WJ-RCOG: Woodcock Johnson Test of Cognitive Ability – Revised.

Figure 1a. Cross-sectional developmental trajectory of BPVS raw score against chronological age. N = 209.

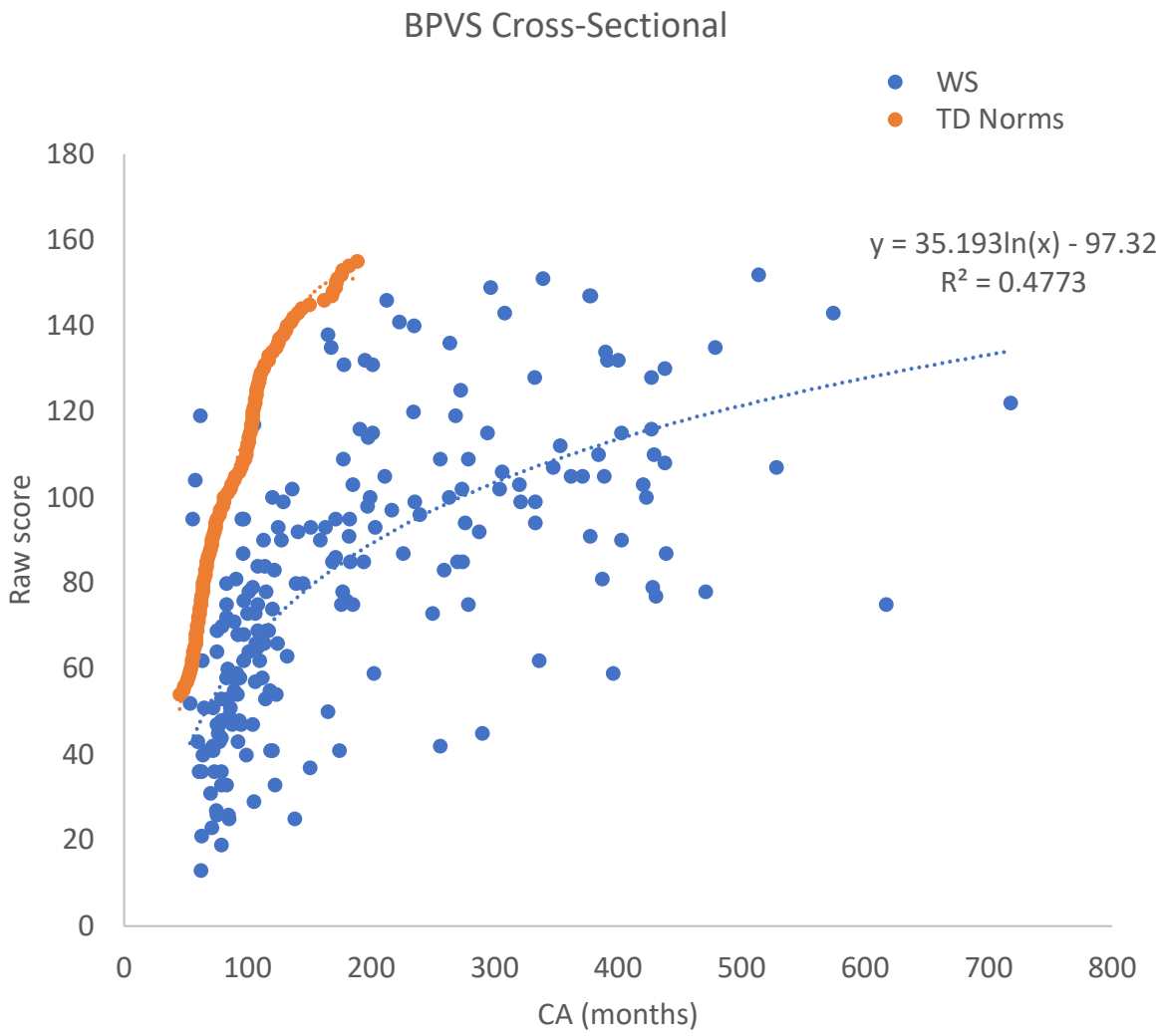

Figure 1b. Cross-sectional developmental trajectory of RCPM raw score against chronological age. N = 102.

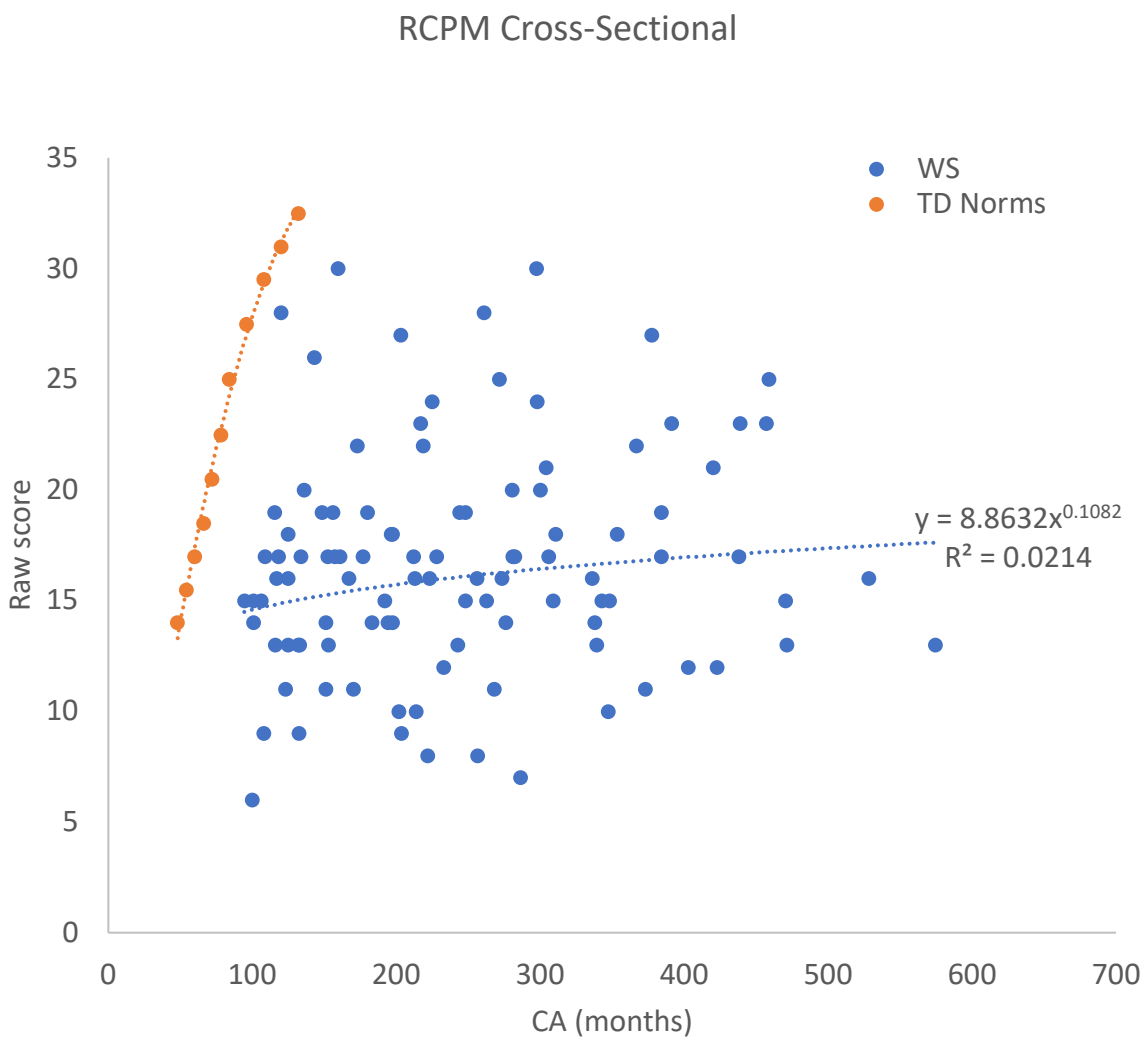

Figure 1c. Cross-sectional developmental trajectory of PC age-equivalent against chronological age. N = 102.

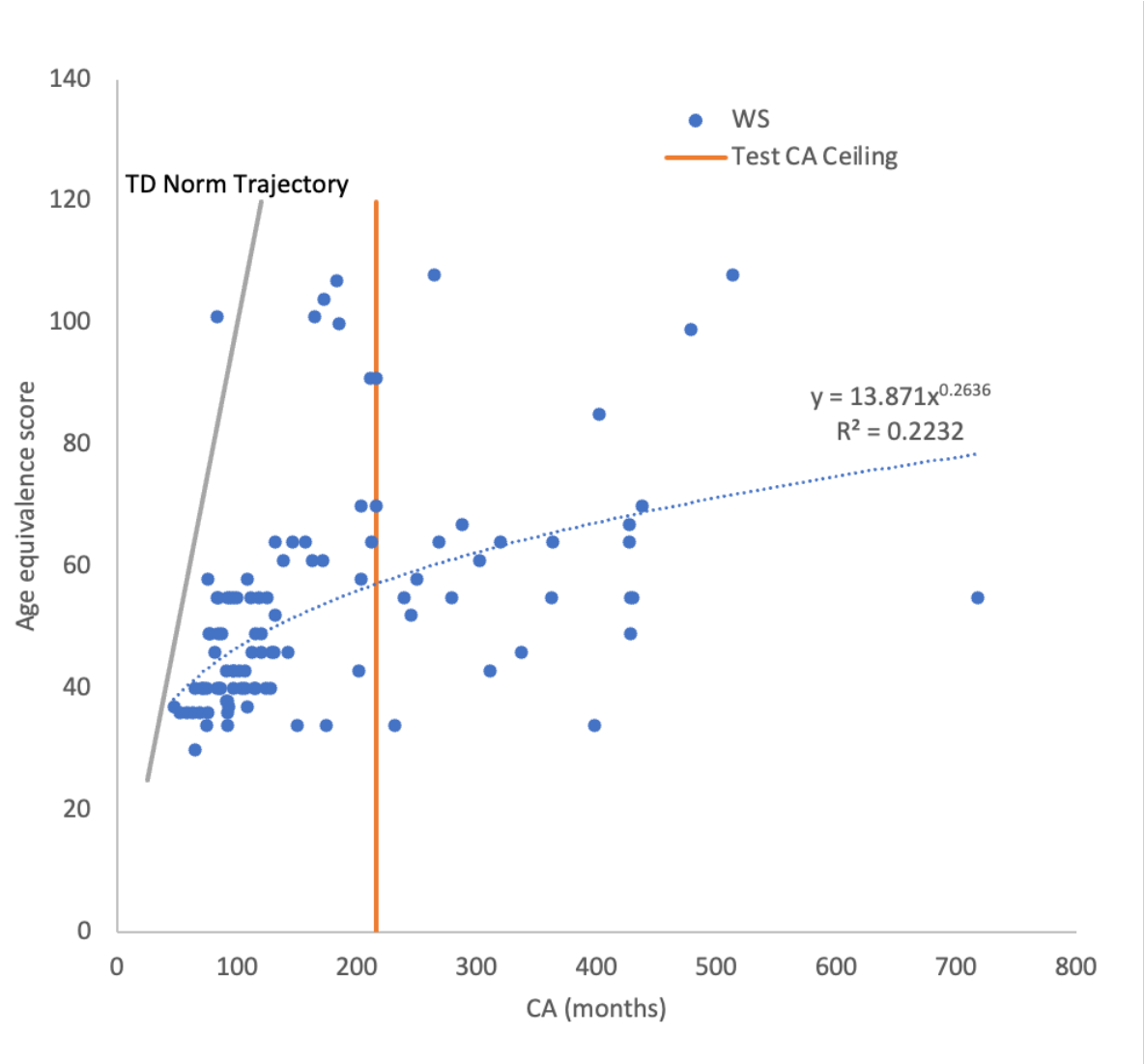

Figure 2a: Developmental change in the difference between verbal and non-verbal ability (RCPM), plotted using normalised raw score

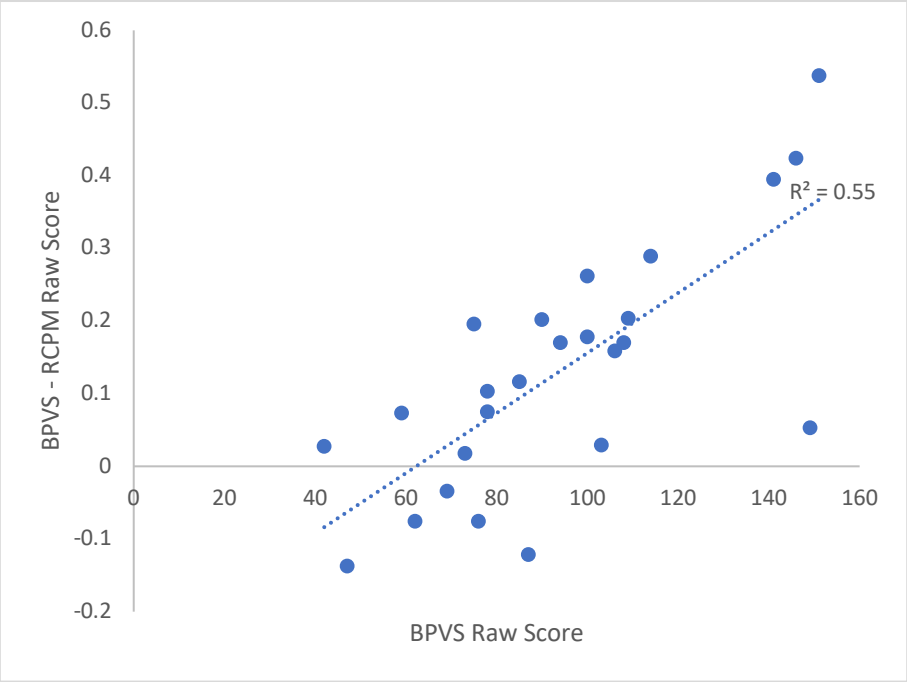

Figure 2b. Developmental change in the difference between verbal and non-verbal ability (RCPM), plotted using age equivalence score

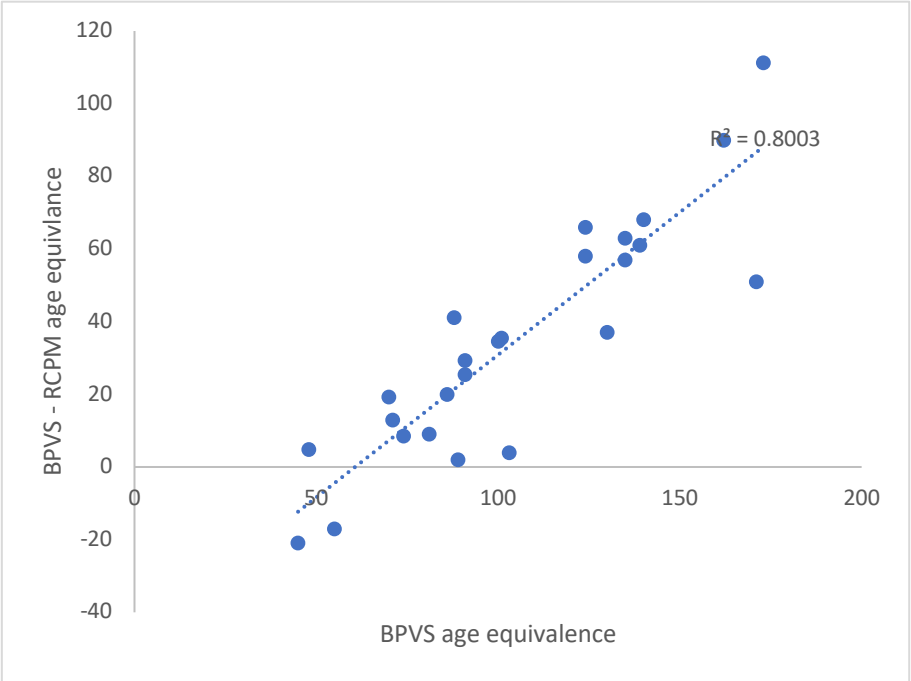

Figure 2c. Developmental change in the difference between verbal and non-verbal ability (PC), plotted using age equivalence score

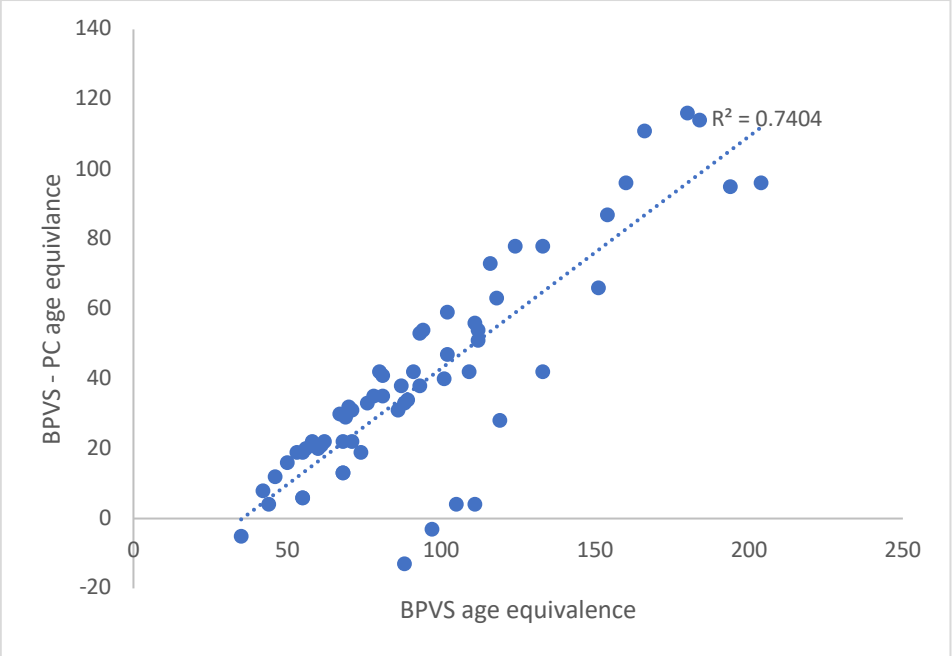

Figure 3a. Linear longitudinal trajectories in BPVS raw scores

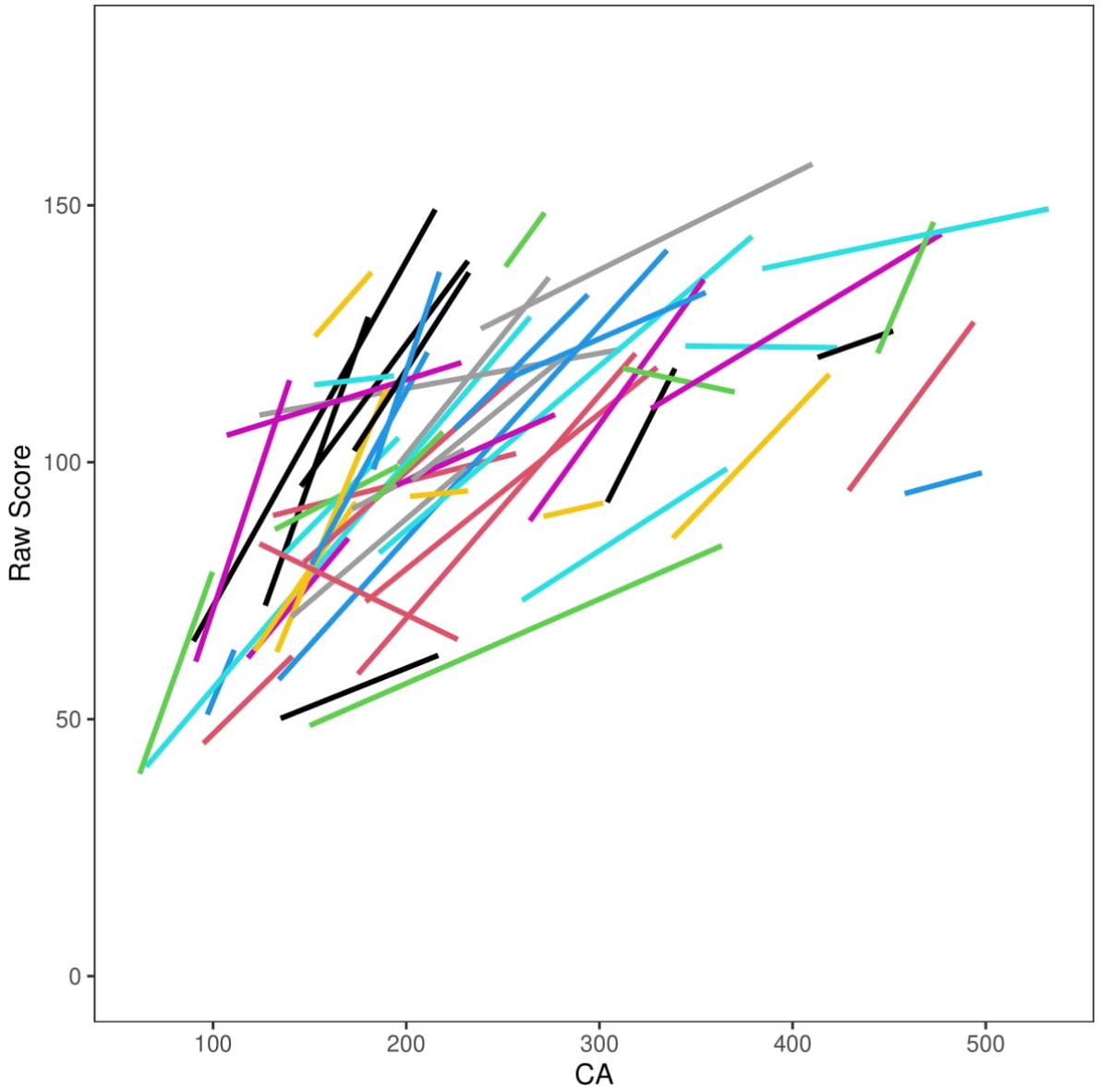

Figure 3b. Linear longitudinal trajectories in RCPM raw scores

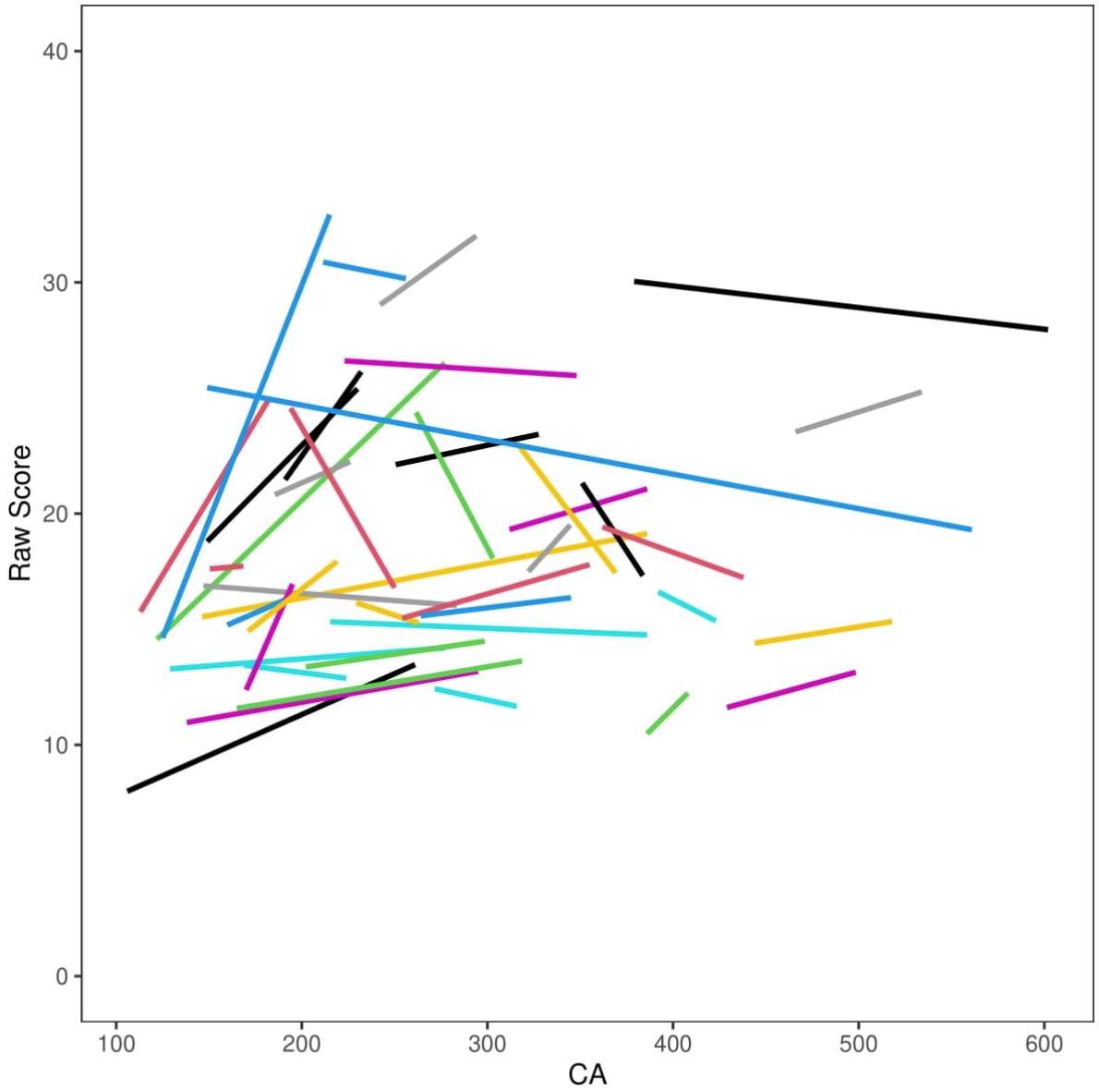

Figure 3c. Linear longitudinal trajectories in PC age equivalence

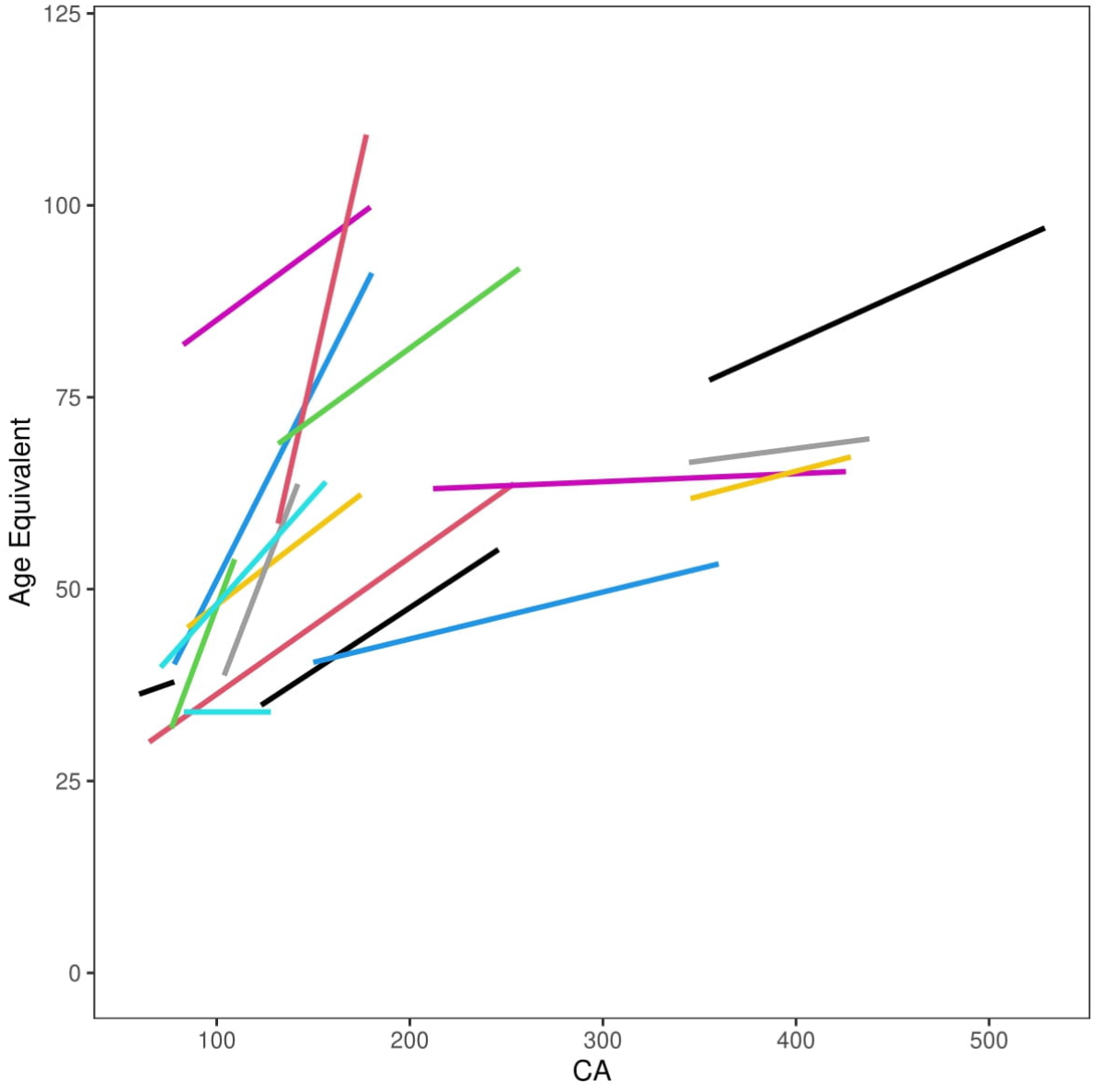

Figure 4: The relationship between chronological age and developmental progression gradients.