Research in Developmental Disabilities xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Research in Developmental Disabilities

Item and error analysis on Raven's Coloured Progressive Matrices in Williams Syndrome[☆]

Jo Van Herwegen[a,*], Emily Farran[b], Dagmara Annaz[a]

[a] School of Health and Social Sciences, Middlesex University, The Burroughs, Town Hall, Hendon NW4 4BT, UK
[b] Department of Psychology and Human Development, Institute of Education, 20 Bedford Way, London WC1H 0AL, UK

ARTICLE INFO

Article history:
Received 27 August 2010
Accepted 7 September 2010

Keywords:
Williams Syndrome
Raven's Coloured Progressive Matrices
Types of errors
Item analysis

ABSTRACT

Raven's Coloured Progressive Matrices (RCPM) is a standardised test that is commonly used to obtain a non-verbal reasoning score for children. As the RCPM involves the matching of a target to a pattern it is also considered to be a visuo-spatial perception task. RCPM is therefore frequently used in studies in Williams Syndrome (WS), in order to match WS participants to a control group or as a single measure to predict performance on a test-condition in developmental trajectory analyses. However, little is known about the performance of participants with WS on the RCPM. The current study compared the type of errors and the difficulty of each item for 53 participants with WS to 53 typically developing children who were individually matched on the total raw score for RCPM. Results showed that the participants with WS made the same proportion of error types and that the proportion of error types changed similarly to those of typically developing controls over development. Furthermore, the differential item difficulty between the two groups was highly similar. It is therefore argued that, although participants with WS are delayed on RCPM, their performance is not atypical which suggests that RCPM performance is supported by typical mechanisms. The RCPM is therefore a useful tool to match WS to control groups or to construct developmental trajectories.

1. Introduction

Williams Syndrome (WS) is a rare developmental disorder with a prevalence of approximately 1 in 20 000 live births (Morris, Demsey, Leonard, Dilts, & Blackburn, 1988) and is caused by the hemizygous deletion of 28 genes on chromosome 7 (Donnai & Karmiloff-Smith, 2000; Tassabehji, 2003). WS is characterised by a fractionated cognitive profile with relatively good face-processing and language abilities in contrast to poor performance on tasks such as drawing, visuo-spatial abilities, memory and number processing in the context of an overall general intelligence score within the mild to moderately impaired range (Martens, Wilson, & Reutens, 2008; Mervis et al., 2000). Furthermore, studies investigating the cognitive processes underlying behavioural outcomes have shown that individuals with WS use different strategies compared to controls. For example, studies examining face processing abilities in WS have found that people with WS tend to look more at individual features compared to controls (Annaz, Karmiloff-Smith, Johnson, & Thomas, 2009; Deruelle, Mancine, Livet, Cassé-

☆ This research was supported by grants from the Williams Syndrome Foundation, UK, the Nuffield Foundation (DA and JVH, 2008), British Academy and Autour des Williams, France.
* Corresponding author. Permanent address: Department of Psychology, Kingston University, Penrhyn Road, Kingston Upon Thames, Surrey KT1 2EE, UK. Tel.: +44 020 8547 2803; fax: +44 0208 417 2292.
E-mail address: j.vanherwegen@kingston.ac.uk (J. Van Herwegen).

0891-4222/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ridd.2010.09.005

Perrot, & de Schonen, 1999; Karmiloff-Smith et al., 2004). This local bias has also been claimed to explain the weaker visuo-spatial abilities reported in people with WS (Abreu, French, Annaz, Thomas, & de Schonen, 2005; Mervis et al., 2000). Yet, other studies have shown that the bias to process elements locally rather than globally is only present in the drawing abilities but not in visual perception or identification of patterns tasks (see Farran & Jarrold, 2003; Farran, Jarrold, & Gathercole, 2003 for a discussion).

Due to the fact that overall cognitive development is impaired in individuals with WS, studies (e.g. Farran, Jarrold, & Gathercole, 2001; Riby & Hancock, 2009) have often matched participants with WS to control groups based upon performance levels on a specific task, such as Raven's Coloured Progressive Matrices (RCPM; Raven, Court, & Raven, 1990). RCPM is a standardised test which, although described as a measure of fluid intelligence (Woliver & Sacks, 1986), involves visual matching of a target to a pattern and is thus a measure of specific visuo-spatial ability (Gunn & Jarrold, 2004). Previous studies investigating RCPM scores in children with WS have found that performance scores are well below what is expected for their chronological age. For example, Brock, Jarrold, Farran, Laws, and Riby (2007) found that for 17 out of 41 children with WS between the ages of 6 and 17 years old, no mental age (MA) equivalent score could be calculated and that they obtained raw scores lower than 15 points on the RCPM. The remaining 24 participants had a mean MA of 7 years and 5 months (SD 1.5) while their mean chronological age (CA) was 12 years old (SD = 3.1) (Brock et al., 2007). However, little is known about how the performance of individuals with WS compares to typically developing (TD) controls on individual RCPM items and how their performance can be characterised i.e., delayed, atypical or both. First of all, an analysis of the types of errors made by participants with WS can provide useful information about the processing strategies employed during task completion. For example, the type of errors made by participants with WS can inform us whether they show a bias to process local elements rather than the entire pattern. Secondly, it is possible for individuals with a developmental disorder to obtain the same raw score or mental age equivalent score without passing necessarily the same items as the level of difficulty of each item might be different across groups (Facon & Nuchadee, 2010). Thus, it is unclear whether the level of difficulty of RCPM items is the same among individuals with WS and typical controls. Without such knowledge it is difficult to draw any conclusions from comparison studies in WS which have matched groups based upon RCPM performance or from studies which have used performance on a single MA measure to predict performance on a test-condition and make comparisons to the TD-group (see Thomas et al., 2009 for a discussion). The current study aimed therefore to investigate whether the types of errors made by a large group of participants with WS differed from TD controls matched for the overall number of errors made, and how error patterns changed with increasing CA in both groups. This would provide further information whether the underlying processes used by participants with WS are typical or atypical. Lastly, this study investigated the functioning of individual items of RCPM in participants with WS and TD controls and examined whether the same items caused difficulties in both groups.

2. Methods

2.1. Participants

Fifty-three participants with Williams Syndrome between the ages of 5;0 and 41;04 years old (mean = 18;03, SD = 9;10) completed the RCPM test. Performance of participants with WS was individually matched to that of TD children, based upon the total raw score plus or minus 1. This ensured that different types of error patterns would not be confounded by different levels of performance ability between the two groups. This matching approach resulted in both groups having very similar mean raw scores and standard deviations (WS: mean: 19.34, SD = 4.91; TD: mean: 19.68, SD = 5.08). A one-way ANOVA with group as between factor showed that there was no significant difference for overall performance on the RCPM between the two groups ($F(1,105) = 0.122$, $p = 0.727$). The TD children had an average chronological age of 5 years and 8 months (SD = 1;03 range: 3;04–9;04 years). Participant and parental consent was obtained for all participants and they were all informed that they could withdraw from the study at any time.

2.2. Materials and methods

Raven's Coloured Progressive Matrices (Raven et al., 1990) is a standardised test which includes 36 items which are divided into three sub-tests of 12 items each. In each item the participant is presented with a coloured pattern of which one part is missing and the participant is asked to select the missing part out of six options. In each sub-test the items are ordered in increasing difficulty. There is no set time limit to complete items and all participants were encouraged to complete all 36 items. Testing took place on a one-to-one basis in a quiet room either at the participant's home or school and was carried out according to the Manual's instructions.

2.3. Data analyses

2.3.1. Error type analysis

In accordance with Raven et al. (1990), each error was coded using one of the following four categories: (1) Difference, (2) Inadequate Individuation, (3) Repetition of the Pattern, and (4) Incomplete Correlation. 'Difference' errors are types of errors in which the participant selects an option that either has no pattern of any kind or has no direct relation to the target pattern.

Please cite this article in press as: Van Herwegen, J., et al. Item and error analysis on Raven's Coloured Progressive Matrices in Williams Syndrome. *Research in Developmental Disabilities* (2010), doi:10.1016/j.ridd.2010.09.005

In 'Inadequate Individuation' errors the participant is unable to combine all the features of a figure and thus selects an option resulting in irrelevancies, distortion of the pattern, or selects an incomplete pattern altogether. 'Repetition of the Pattern' errors occur when the participant selects an option which is merely a copy of part of the pattern immediately above or beside the target gap pattern. Finally, 'Incomplete Correlation' errors include those errors in which the participant correctly identified half of the pattern but the option chosen is incorrectly orientated or does not entirely complete the pattern.

For each participant the number of errors per type was divided by the total amount of errors made by that participant. This ensured that the type of errors made would not be confounded with the total number of errors made.

In the first instance a group-comparison was carried out in order to establish whether the WS group made each type of error more or less often compared to the TD group. Furthermore, the type of errors are considered in terms of the three different sub-tests because the problems were not organised so that there would be an equal amount of the error types across the sub-tests. For example, sub-test Ab includes the majority of problems in which the discrete figures could be apprehended as parts of an organised "whole" or as an individual entity (Raven et al., 1990). It is therefore possible that the WS group make the same total number of error types but that they are distributed differently across the three sub-tests compared to the TD group. Secondly, the relationship between chronological age and type of error was investigated using developmental trajectories (Thomas et al., 2009). This allowed investigation of whether error patterns in participants with WS change in a similar way to the changes observed in the TD group.

2.3.2. Item analysis

In line with Facon and Nuchadee (2010) a statistical analysis based on the transformed item difficulties method was employed due to its simplicity and the fact that it requires a moderate number of subjects compared to more sophisticated methods to investigate differential item functioning. For the delta-plot analysis, the proportion of participants (p) who answered the item correctly was calculated per group separately. The complement of each p-score (i.e., $1 - p$) was converted into a z-score so that the easiest items would receive lowest z-scores and most difficult items the highest z-scores. Next, delta-scores were calculated based on the z-scores using the formula: $\Delta = 4z + 13$ (see Facon & Nuchadee, 2010 for a discussion). Subsequently, the delta scores of the two groups were plotted on a bivariate graph. If the difficulty of the items is similar between groups the correlation should be close to 1. For each item, the perpendicular distance from the data point for that item to the line of best fit for the distribution of items is calculated as D. This allows one to compare the item's difficulty between the two groups. The greater the distance (D), the greater the differential functioning of the item. In line with Facon and Nuchadee (2010), a D greater than 1.5 Δ units indicates a significant differential functioning for that item for one group compared to the other group.

3. Results

3.1. Type of error analyses

Fig. 1 shows the distribution of the types of errors made per group across the four different types of errors. A repeated-measures ANOVA analysis comparing the proportion of errors per error type, did not show a significant interaction between group and type of error ($F(3,102) = 1.775$, $p = 0.157$). This revealed that, as a group, the proportions of each error type made by WS participants did not differ from those of the TD group.

However, comparison of the number of errors made per sub-test showed that there was a significant difference in both the TD group ($F(1,52) = 330.017$, $p < 0.001$, $\eta_p^2 = 0.864$) as well as in the WS group ($F(1,52) = 573.666$, $p < 0.001$, $\eta_p^2 = 0.917$). As shown in Fig. 2, participants in both groups made the least errors sub-test A and most errors in sub-test B.

3.2. The relationship between type of errors made and chronological age

As the WS group included participants from a wide age-range while the TD control participants were much younger, developmental trajectories were constructed for each type of error per group. This allowed investigation of how the number of each type of error changed with increasing CA. Fig. 3 shows the relationship between CA (in months) and RCPM total performance raw score for the TD and WS groups. For the TD group ($F(1,52) = 36.572$, $p < 0.001$, $\eta_p^2 = 0.42$) and the WS group ($F(1,52) = 11.232$, $p = 0.002$, $\eta_p^2 = 0.18$) there was a clear relationship between increasing CA and overall performance. No direct comparison of the developmental trajectories could be carried out as the WS participants were much older compared to the TD group ($t(104) = -9.279$, $p < 0.001$). Still, as CA accounted for 42% of the variance in RCPM score for the TD group, but only 18% of the variance in RCPM score for the WS group (see Fig. 3), the relationship between CA and RCPM score was relatively weaker for the WS group.

Due to the fact that performance improved with increasing CA but that this relationship is different between the two groups, it is possible that the distribution of types of errors made by participants changes with increasing CA, and also that these changes in the pattern of errors with age differs between the two groups. In order to investigate this possibility, developmental trajectories were constructed per group for each of the different types of errors separately. As can be seen in Fig. 4, for the TD group the number of 'Difference' and 'Inadequate Individuation' errors decreased significantly with increasing CA (Difference errors: $F(1,52) = 8.228$, $p = 0.006$, $\eta_p^2 = 0.14$; Inadequate Individuation errors: $F(1,52) = 22.590$, $p < 0.001$, $\eta_p^2 = 0.31$), while the number of 'Repetition of the pattern' errors increased with increasing CA ($F(1,52) = 22.727$,

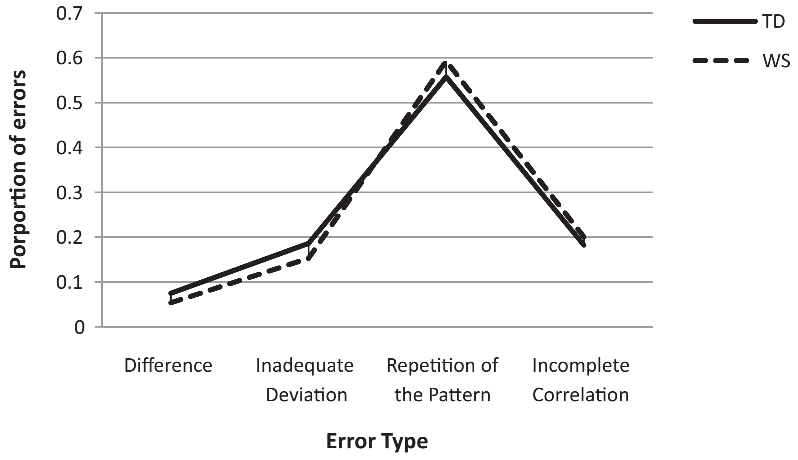

Fig. 1. Proportion of each type of error made by each group.

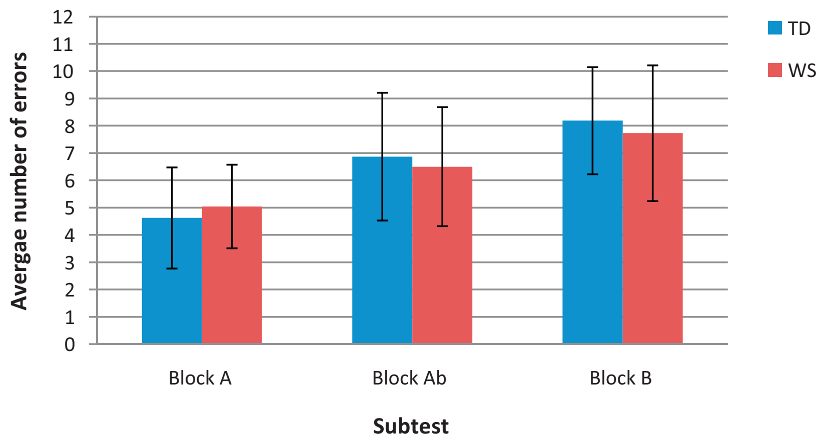

Fig. 2. Total number of errors made per sub-test for TD and WS group.

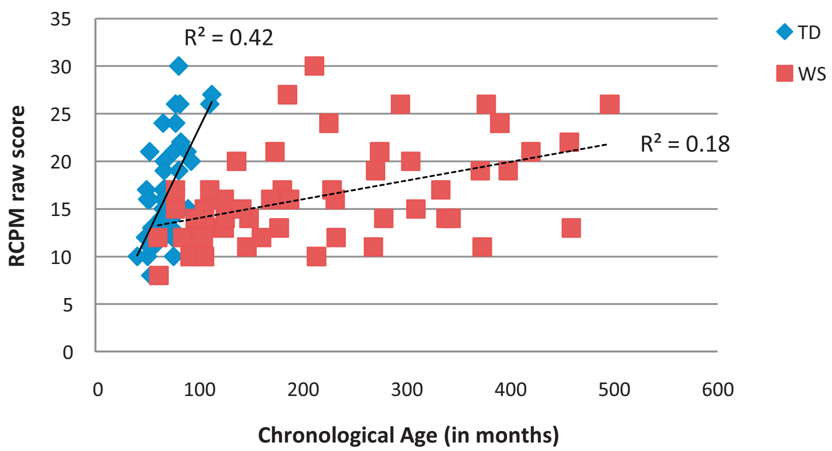

Fig. 3. The relationship between raw score on RCPM and chronological age (in months) for the TD and WS group.

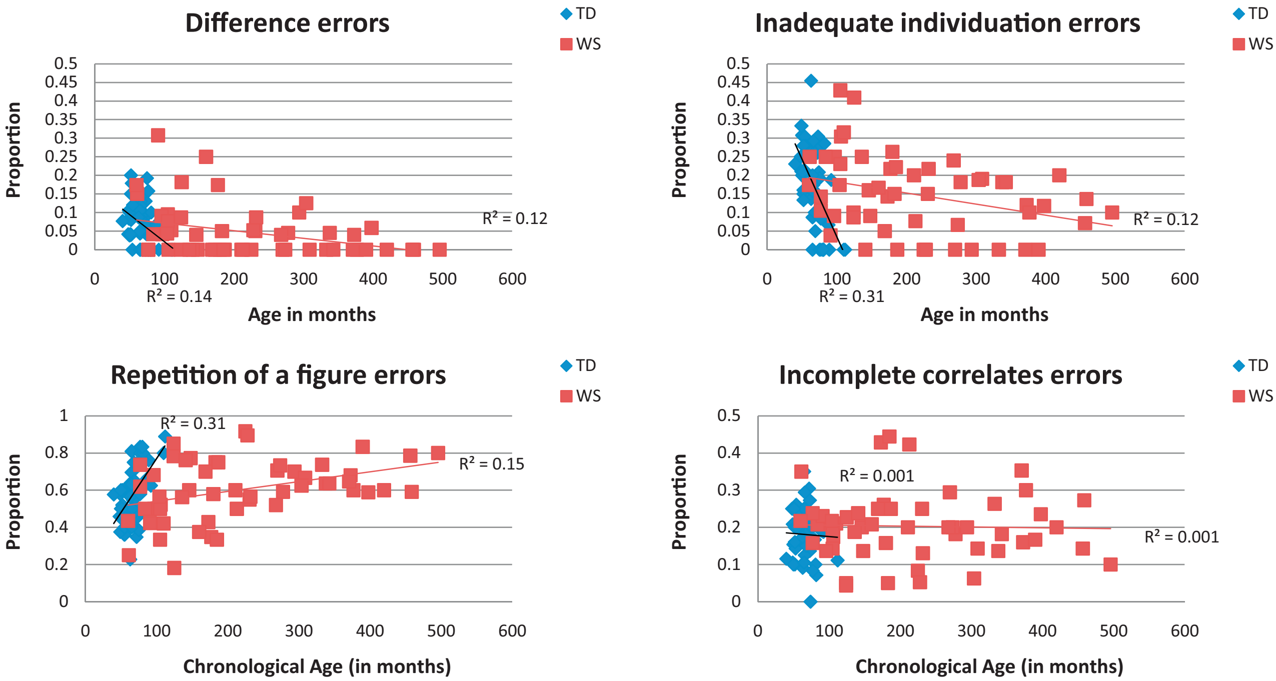

Fig. 4. The relationship between chronological age and the proportion of each type of error for TD and WS group.

$p < 0.001$, $\eta_p^2 = 0.31$). There was no significant relationship between the number of ‘Incomplete Correlation’ errors and CA ($F(1,52) = 0.071$, $p = 0.792$, $\eta_p^2 < 0.01$). Examination of how the proportion of each type of error changes with increasing CA in participants with WS showed exactly the same patterns as observed in the TD group. The amount of ‘Difference’ and ‘Inadequate Individuation’ errors also decreased significantly with increasing CA (Difference errors: $F(1,52) = 7.190$, $p = 0.01$, $\eta_p^2 = 0.12$; $F(1,52) = 6.719$, $p = 0.012$, $\eta_p^2 = 0.12$). The number of ‘Repetition of the patttern’ errors increased significantly with increasing CA ($F(1,52) = 8.894$, $p = 0.004$, $\eta_p^2 = 0.15$) but there was no relationship between the number of ‘Incomplete correlation’ errors and CA ($F(1,52) = 0.046$, $p = 0.832$, $\eta_p^2 < 0.01$) in this group.

3.3. Item analyses

Fig. 5 shows the proportion of participants who passed the item and shows that the proportion of correct reponses between the two groups is very similar.

Indeed, as shown in Fig. 6, the delta scores obtained from each item correlated highly between the two groups (*rho* = 0.89, $N = 36$, $p < 0.01$). In addition, only for three items (sub-test A item 8: −3.37, sub-test Ab item 4: 2.32 and sub-test B item 10:

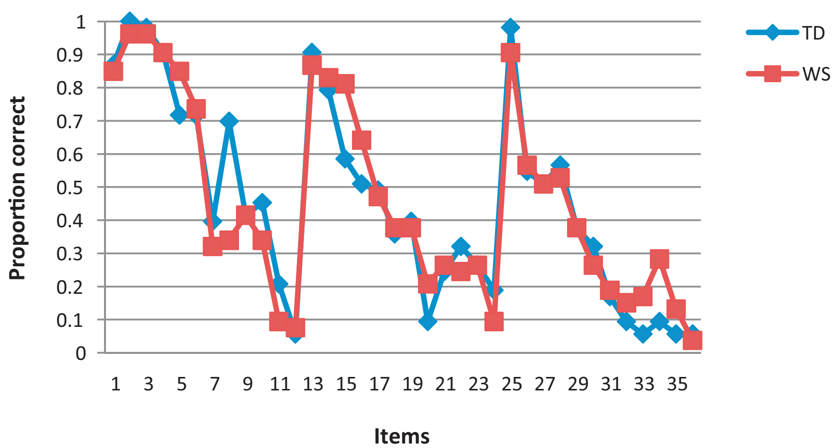

Fig. 5. Proportion of participants who answered the item correctly for the WS and TD group.

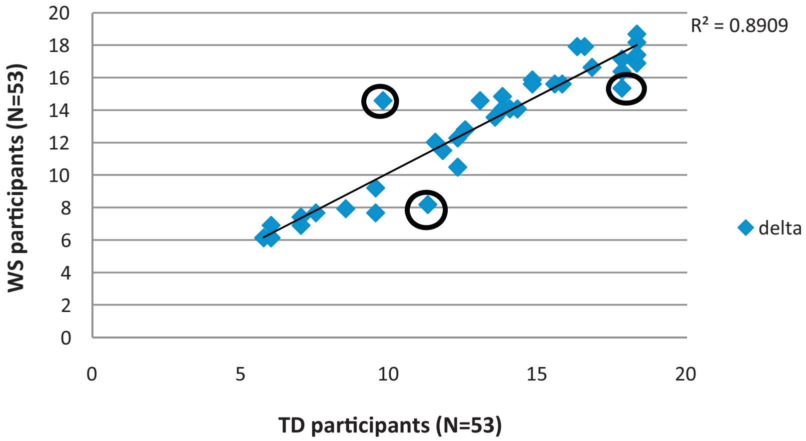

Fig. 6. Bivariate distribution of Δ scores for the TD and WS group. Circled plots correspond to D scores that exceeded the 1.5 Δ threshold.

1.59) the D-values were larger than the 1.5 threshold and thus there were very few differential functioning items between the two groups.

4. Discussion

The Raven's Coloured Progressive Matrices (Raven et al., 1990) is a widely used standardised task with TD and developmental disorders populations in order to obtain a measure of non-verbal ability. Studies very often use raw score or a derived non-verbal mental age equivalent score to either match participants with WS to other clinical or control groups or to build developmental trajectories. The aim of the current study was to assess the type of errors made by a large group of individuals with WS in order to characterise their performance on RCPM as delayed, atypical or both delayed and atypical. The results of the current study show that, although participants with WS are developmentally delayed on RCPM, they make the same types of errors as TD children who were individually matched for the number of errors made. Furthermore, the patterns of errors in the WS group developed in a similar way as in the TD group with the number of 'Difference' and 'Inadequate individuation' errors decreasing and the number of 'repetition of the pattern' errors increasing in older participants compared to younger ones. These patterns of errors are similar to those reported in TD children by Gunn and Jarrold (2004). However, in contrast to Gunn and Jarrold, the current study did not find an increase in incomplete errors with age. One difference between the two studies is that, although the TD children had a similar CA, the mean RCPM raw scores for TD children was much higher in the current study (19.68, SD = 5.08) compared to the mean of TD children in Gunn and Jarrold (2004) (13.20, SD = 1.86). Finally, analysis of the items' difficulty using delta-plots showed that there was very little differential functioning between the two groups. There was only a difference between the two groups for three items. Two of these items (sub-test Ab item 3 and sub-test B item 10) were answered correctly by more WS participants compared to the TD group and thus these items appear to be easier for the WS than TD group. This difference might be explained by the fact that the participants with WS were much older compared to the TD participants. For only one item (sub-test A item 8) did more TD participants pass compared to WS ($D > 1.5$). Closer inspection of performance on the items in sub-test A showed that it was not the case that this item was more difficult for the WS participants compared to the other items. However, this difference between the groups was probably caused by the fact that many more TD participants answered this item correctly compared to previous and succeeding items. Inspection of the item itself showed that item 8 of sub-test A is very similar to item 7, with exception of the fact that the figure now includes a double rather than a single stripe. One possible explanation therefore might be that the fact that item 7 precedes item 8 facilitated TD children to answer item 8 correctly while this was not the case in WS children. Thus, the similarity between the two items might explain why more TD children, but not WS children, answered item 8 correctly. However, future studies investigating visual rule learning and facilitation effects in WS are required in order to provide more conclusive explanations.

The current findings suggest that, first of all, performance on RCPM in participants with WS is delayed, in that they obtain a similar raw score to TD participants who are much younger, but do not show an atypical pattern of performance. Secondly, RCPM is a reliable measure to obtain a non-verbal functioning score in WS and can be used with confidence in clinical practice as well as in research studies when matching WS to other participants groups or constructing developmental trajectories. Finally, it can be concluded that, although participants with WS might use different strategies compared to TD participants in face recognition or visuo-spatial construction tasks, they do not use different strategies when solving visuo-perceptual problems such as those included in the RCPM. This finding supports the view by Farran and Jarrold (2003) that, whether participants with WS exhibit a local processing style relates to the specific demands of the tasks, i.e., a local bias is observed on visuo-spatial construction, but not perceptual tasks (with exception of face recognition tasks).

Please cite this article in press as: Van Herwegen, J., et al. Item and error analysis on Raven's Coloured Progressive Matrices in Williams Syndrome. *Research in Developmental Disabilities* (2010), doi:10.1016/j.ridd.2010.09.005

Although the results in the current study seem rather clear-cut further studies are required. First, while we report results from a very large sample given the prevalence of WS, the sample size nevertheless restricted the method of item analysis that could be employed. Similarly, as mentioned above, even though the method chosen was the most suitable for small sample sizes, it is possible that the differences observed for three of the items were an artefact of the small sample size rather than true group differences. Secondly, examination of item and error patterns would provide further information about the underlying abilities required to solve the problems using cross-syndrome developmental trajectories comparisons included in the RCPM. For example, participants with autism are reported to have a local processing bias but it has been suggested that this bias is perceptual rather than constructional (see Farran & Jarrold, 2003 for a discussion). It can therefore be predicted that if the RCPM is mainly a perceptual task, participants diagnosed with autism would show different patterns of errors compared to those in the current study.

Acknowledgements

We thank the Williams Syndrome Foundation, UK and all of the participants with Williams Syndrome who took part in this study. We are grateful to all the schools and typically developing children for their participation.

References

Abreu, A. M., French, R. M., Annaz, D., Thomas, M., & de Schonen, S. (2005). A «visual conflict» hypothesis for global-local visual deficits in Williams Syndrome: Simulations and data. In *Proceedings of the 27th annual cognitive science society* (pp. 45–50). Mahwah, NJ: LEA.
Annaz, D., Karmiloff-Smith, A., Johnson, M. H., & Thomas, M. S. C. (2009). A cross-syndrome study of the development of holistic face recognition in children with autism, Down syndrome and Williams syndrome. *Journal of Experimental Child Psychology, 102*, 456–486.
Brock, J., Jarrold, C., Farran, E. K., Laws, G., & Riby, D. M. (2007). Do children with Williams Syndrome really have good vocabulary knowledge? Methods for comparing cognitive and linguistic abilities in developmental disorders. *Clinical Linguistics & Phonetics, 21*(9), 673–688.
Deruelle, C., Mancini, J., Livet, M. O., Cassé-Perrot, C., & de Schonen, S. (1999). Configural and local processing of faces in children with Williams syndrome. *Brain and Cognition, 41*, 276–298.
Donnai, D., & Karmiloff-Smith, A. (2000). Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics, 97*, 164–171.
Facon, B., & Nuchadee, M.-L. (2010). An item analysis of Raven's colored progressive matrices among participants with Down syndrome. *Research in Developmental Disabilities, 31*(1), 243–249.
Farran, E. K., & Jarrold, C. (2003). Visuospatial cognition in Williams Syndrome: Reviewing and accounting for the strengths and weaknesses in performance. *Developmental Neuropsychology, 23*(1 & 2), 173–200.
Farran, E. K., Jarrold, C., & Gathercole, S. E. (2001). Block design performance in the Williams syndrome phenotype: A problem with mental imagery? *Journal of Child Psychology and Psychiatry, 42*(6), 719–728.
Farran, E. K., Jarrold, C., & Gathercole, S. E. (2003). Divided attention, selective attention and drawing: Processing preferences in Williams syndrome are dependent on the tasks administered. *Neuropsychologia, 41*, 676–687.
Gunn, D. M., & Jarrold, C. (2004). Raven's matrices performance in Down syndrome: Evidence for unusual errors. *Research in Developmental Disabilities, 25*, 443–457.
Karmiloff-Smith, A., Thomas, M., Annaz, D., Humphreys, K., Ewing, S., Brace, N., et al. (2004). Exploring the Williams syndrome face-processing debate: The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry, 45*(7), 1258–1274.
Martens, M. A., Wilson, S. J., & Reutens, D. C. (2008). Research review: Williams Syndrome: A critical review of the cognitive, behavioural and neuroanatomical phenotype. *Journal of Child Psychology and Psychiatry, 49*(6), 576–608.
Mervis, C. B, Robinson, B. F., Bertrand, J., Morris, C. A., Klein-Tasman, B. P., & Amstrong, S. C. (2000). The Williams syndrome cognitive profile. *Brain and Cognition, 44*, 604–628.
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*(2), 318–326.
Raven, J. C., Court, J. H., & Raven, J. C. (1990). *Manual for Raven's progressive matrices and vocabulary scales—section 2: Coloured progressive matrices*. Oxford: Oxford Psychologists Press.
Riby, D. M., & Hancock, P. J. B. (2009). Do faces capture the attention of individuals with Williams syndrome or Autism? Evidence from tracking eye movements. *Journal of Autism and Developmental Disorders, 39*, 421–431.
Tassabehji, M. (2003). Williams-Beuren syndrome: A challenge for genotype-phenotype correlations. *Human Molecular Genetics, 12*(2), 229–237.
Thomas, M. S. C., Annaz, D., Ansari, D., Serif, G., Jarrold, C., & Karmiloff-Smith, A. (2009). Using developmental trajectories to understand developmental disorders. *Journal of Speech, Language and Hearing Research, 52*, 336–358.
Woliver, R. E., & Sacks, S. D. (1986). Intelligence and primary aptitudes: Test design and tests available. In R. B. Cattell & R. C. Johnson (Eds.), *Functional psychological testing: Principles and instruments* (pp. 166–188). New York: Brunner/Mazel.

Please cite this article in press as: Van Herwegen, J., et al. Item and error analysis on Raven's Coloured Progressive Matrices in Williams Syndrome. *Research in Developmental Disabilities* (2010), doi:10.1016/j.ridd.2010.09.005