Moss, J., Magiati, I., **Charman, T.** & Howlin, P. (2008). Stability of the Autism Diagnostic Interview – Revised from pre-school to elementary school age in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 1081-1091.

Stability of the Autism Diagnostic Interview – Revised from pre-school to elementary school age in children with autism spectrum disorders

Jo Moss [1], Iliana Magiati [2], Tony Charman[3] & Patricia Howlin [1]

[1] St George's, University of London

[2] Institute of Psychiatry, King's College London

[3] UCL Institute of Child Health

Suggested running head: Stability of the ADI-R.

Abstract

This study examined the stability of scores on the ADI-R from pre-school to elementary school age in children with autism spectrum disorders (ASD). Participants were 35 children who, at T1, all had a clinical diagnosis of ASD. On initial assessment (mean age 3.5 years; SD 0.6), all met ADI-R algorithm criteria for autism. ADI-R assessments were repeated at follow up (FU; mean age 10.5 years; SD 0.8). Changes in ADI-R total, domain and ADI-R algorithm item scores were assessed. Twenty-eight children continued to score above the ADI-R cut-off for autism at FU, although significant decreases in ADI-R domain and item scores were also found. In conclusion, while classification of children according to ADI-R criteria, generally remained stable between pre-school and elementary school age, many children demonstrated significant improvements in symptom severity.

Key words: Autism Diagnostic Interview-Revised, autism spectrum disorder, autism, longitudinal study, symptom severity.

Stability of the Autism Diagnostic Interview – Revised from pre-school to elementary school age in children with autism spectrum disorders

The issue of diagnostic stability in the field of autism spectrum disorders (ASD) has become an increasing focus of research and is of particular relevance for studies of prevalence and intervention. Several studies confirm that pre-school diagnoses of autism made by *expert* clinicians remain relatively stable 2-22 years after initial diagnosis (Billstedt, Gillberg & Gillberg, 2005; Cox, e*t al.*, 1999; Eaves & Ho, 2004; McGovern & Sigman, 2005; Moore & Goodson, 2003; Lord, 1995; Stone *et al.*, 1999).

Other studies have explored the stability of autism/ASD diagnoses based on specific diagnostic instruments. Although the ADI-R is among the best validated of diagnostic interview measures, follow-up studies indicate that ADI-R based diagnoses made early in life may be less stable than clinical judgement (see Charman & Baird, 2002; Chawarski, Klin, Paul & Volkmar, 2007). For example, follow-up studies by Lord and colleagues (2006) reported that clinical judgement at age 2 was a more accurate predictor of diagnosis at age 9 than an ADI-R at 2 years. Lord *et al.* (2006) found the ADI-R to be initially more inclusive than clinical judgement for both autism and ASD criteria based on Risi *et al.*'s ASD criteria (2006). Lord *et al.* (2006) noted that the use of standardised assessments such as the ADI-R *alongside* clinical judgement was likely to result in more stable diagnosis than either ADI-R or clinical judgement alone. Moreover, Charman *et al.* (2005) in a study of 26 children monitored the numbers meeting ADI-R criteria for autism at 4 time points from 2 to 7 years. Six participants changed

ADI-R algorithm status at least twice during this period; another 8 changed algorithm status once, indicating that ADI-R classifications of autism may not remain stable over time.

Understanding the developmental trajectory of autism spectrum *symptomatology* over time also provides further insight into the stability of ASD diagnoses. Comparisons of ADI-R current scores with retrospective ratings for scores at 4-5 years (the age used for rating diagnostic algorithm items) generally show steady improvements in symptomatology with age (Boelte & Poustka, 2000; Fecteau, Mottron, Berthiaume & Burack, 2003; Piven, Harper, Palmer & Arndt, 1996; Seltzer *et al.*, 2003; Shattuck *et al.*, 2007). There are clearly limitations to retrospective studies of this kind, but the prospective study by Charman *et al.* (2005) also showed significant improvements in domain scores from 2 to 7 years, although patterns of change differed across domains and age groups (see also Cox *et al.,* 1999). However, the variance in ADI-R scores also increased with age and although many children did improve, others showed an increase in symptom severity. Starr, Szatmari, Bryson and Zwaigenbaum (2003) analysed change over 2 years in 68 high functioning children originally assessed on the ADI-R at a mean age of 5.8 years. They found significant improvements in each domain, with only a few items indicating a worsening of behaviour.

The current prospective study employed a sample of 35 children with an independent diagnosis of autism/ASD prior to participation in the study. Participants were originally assessed on the ADI-R at a mean age of 3.5 years and then again at a mean age of 10.5 years. Standardised assessments of cognitive and language level were available at both time points and participants had a broad range of intellectual ability (T1 IQ <20 to >130). The study focused on the ADI-R with respect to the following issues: a. stability of the ADI-R at three levels: meeting algorithm criteria, domain and item level

scores for the group as a whole and at an individual level; b. exploration of T1 factors that were associated with ADI-R outcomes at FU.

In line with previous studies (e.g. Bolte and Poustka, 2000; Fecteau *et al.*, 2003; Piven *et al.*, 1996; Seltzer *et al.*, 2003; Starr *et al.*, 2003), it was predicted that ADI-R scores would decrease between the ages of 3 and 10 years, indicating an improvement in symptom severity over time.

Method

Recruitment

Initial recruitment was conducted between July 1998 and April 2000 (T1). Parents of children with a diagnosis of autism or ASD were contacted via the following sources: Local Educational Authorities, diagnostic and treatment centres and specialist schools across greater London and the UK National Autistic Society (NAS). In September 2005, all families included at T1 were invited to participate in the follow-up study (FU).

Participants were included in the current study if they met the following criteria[1] at T1: a. chronological age (CA) 22-54 months; b. enrolled in specialist pre-school provisions (either home-based behavioural programmes or autism specific nurseries); c. independent professional diagnosis of autism or ASD prior to participation *and* meeting algorithm criteria for autism on the ADI-R; d. no additional major medical diagnoses; e. English the main language spoken at home. Additionally, all children included in the current study were those who had received an ADI-R assessment at T1 *and* FU.

Participants

A total of 75 children was assessed between July 1998 and April 2000. Six to 8 years later, 56 participants (72.7% return rate) agreed to take part in the FU study, 35 of whom met all of the inclusion

criteria. Of the 21 participants not included in the study, 12 did not have ADI-R data both at T1 and FU, 4 did not meet criteria for autism on the ADI-R at T1 and 5 did not have a confirmed independent diagnosis of autism/ASD prior to participation in the study. There were no significant differences between those included in the current study and the 21 children excluded from this sample.

Table 1 describes participant characteristics at T1 and FU. Thirty-two participants were male; 24 were of white origin, 4 of Black British, 3 of Asian British and 4 of mixed race origin. Twenty-nine had an independent clinical diagnosis of autism and 6 a diagnosis of ASD.

(Insert Table 1 about here)

Measures

Primary measure- Autism spectrum symptomatology: The Autism Diagnostic Interview- Revised (ADI-R; Rutter, Le Couteur & Lord, 2003) was used to assess ASD symptomatology. The ADI-R is a semi-structured interview focusing on three domains: Communication (Verbal- VC and Non Verbal- NVC), Reciprocal Social Interaction (RSI) and Repetitive Behaviour and Stereotyped Patterns (RBSP). Each item is scored from 0 (no impairment) to 2/3 (very severe delay/ deviance). In order to meet ADI-R criteria for autism, a child needs to score above cut-off in all three domains and to present with developmental concerns prior to the age of 3 years. ADI-R cut-off scores are 10 for RSI, 7 NVC, 8 for VC and 3 for RBSP. The suggested broader cut off criteria for ASD proposed by Risi *et al.* (2006) were employed where participants no longer met the algorithm criteria of the ADI-R at FU. According to Risi *et al.* (2006), individuals who meet the criteria for autism on the RSI *and* on either the NVC/VC domain or the RBSP domain can be considered to meet ASD criteria on the ADI-R. Only those children using regular, functional phrase speech were rated on the VC domain (T1 N = 7; FU N=20); all participants were assessed on the NVC domain at T1 and FU. In order to ensure that scores were

comparable at both time points, total ADI-R scores were computed using only the NVC domain. This total ADI-R score (i.e. ADI-R total =RSI + NVC + RBSP) was used to examine *overall* changes in symptomatology from T1 to FU.

Typically, on the ADI-R the informant provides details both on current symptoms and behaviour at 4-5 years. For the present study information was collected on the children's *current* behaviour at both T1 and FU in order to facilitate comparisons across the two time points. Four questions on the ADI-R vary according to the age of the child ('play with peers', 'friendships, 'circumscribed interests' and 'imaginative play with peers') and were only administered when chronologically appropriate. Thus, they were not administered at both T1 and FU for all individuals and to ensure comparability across assessments these items were not included in the item level comparisons.

Cognitive ability: Four assessment tools were employed: the Merrill-Palmer Scale of Mental Tests (MP; Stutsman, 1948), the Bayley Scales of Infant Development (Bayley, 1993), the Wechsler Pre-school and Primary Scale of Intelligence (WPPSI; Revised and III version; Wechsler, 1990; 2003) and the Wechsler Intelligence Scale for Children (WISC-IV; Wechsler, 2004). Due to changes in age and ability between T1 and FU, it was not always appropriate to use the same test at both time points, although every effort was made to use the same instrument over time as far as possible.

Selection of the appropriate assessment was based on chronological age, developmental level and language ability. To avoid repeated, separate comparisons of T1 and FU cognitive scores obtained from different tests and in order to utilise the most relevant IQ data, a "best test" IQ and Mental Age (MA) score was created for each child using the most age appropriate/ best standardised test available at each time point according to the following hierarchy: WISC > WPPSI (higher level) > Bayley> MP>

WPPSI (lower level) (all full scale scores). At T1 the best available test was the WPPSI for one participant, the Bayley for 17 and the MP for 17 participants. At FU the best available test was the WISC for 10 participants, the WPPSI (higher) for 6 and the MP for 19 participants.

Adaptive behaviour: Adaptive behaviour was assessed using the Vineland Adaptive Behavior Scales (VABS, Survey form; Sparrow, Balla &Chicchetti, 1984). Standard and age equivalent scores were obtained on the Vineland Adaptive Behavior Composite score (VABS ABC) and each of the three domains (communication, daily living skills, socialization).

Language: Language comprehension was assessed using the British Picture Vocabulary Scales – 2nd Edition (BPVS; Dunn, Dunn, Whetton & Burley, 1997). Expressive language was assessed using the Expressive One Word Picture Vocabulary Test (EOWPVT; Gardener, 1990; Brownell, 2000). Since most children had very limited speech at T1, standard and age equivalent scores on these tests were highly skewed towards basal and thus raw scores were employed in the analyses reported here.

Reliability: Inter-rater reliability was obtained for 23 assessments (5 WISC, 4 WPPSI, 5 MP, 5 BPVS and 4 EOWPVT) for 12 randomly selected participants. Video recordings of the assessments were scored independently by two observers. Intra-class correlation coefficients for total raw scores between the two observers were very high ($r = 1.0$, $p < .001$).

Inter-rater reliability was obtained on ADI-R scores for 6 randomly selected participants (3 verbal, 3 nonverbal). Two examiners, blind to each other's scores, rated the algorithm items of the ADI-R independently using detailed written notes made during the interview. Inter-rater reliability for

algorithm cut-off scores for autism, domain and total level raw scores was good (Kappa = 1.0 for all cut off scores and Intraclass Correlation Coefficients for domain and total level scores were >.86).

Procedure: T1 and FU assessments and interviews were carried out by three experienced research psychologists (SD, T1 assessments only; JM, FU assessments only; IM, T1 & FU assessments). All T1 assessments were conducted at home; most FU child assessments were administered at school but 5 children were assessed at home at parental request. All parent interviews were conducted with the main caregiver at both T1 and FU. Child assessments were conducted within 8 weeks following the parent interviews.

Data analysis: The stability of ADI-R scores from T1 to FU was assessed on three levels: algorithm cut-off, domain and item level scores. Nonparametric analyses were employed where data were not normally distributed. McNemar tests were conducted to investigate changes in ADI-R algorithm status; paired t-tests were conducted on each domain and the total ADI-R score. Wilcoxon Matched-Pairs Signed-Ranks tests were conducted to assess the stability of the ADI-R at item level. T1 variables associated with ADI-R scores at FU were identified using Pearson correlational analyses. Paired t-tests or Mann-Whitney U tests (as appropriate) were conducted in order to explore any T1 differences between individuals whose ADI-R classification changed from T1 to FU and those whose classification remained stable and between those individuals who showed the most and the least improvement in ADI-R scores. A conservative p value of $\leq .01$ was employed in all analyses.

Results

Stability of ADI-R diagnostic criteria

All 35 participants scored above the algorithm cut-off for autism on all three domains of the ADI-R at T1. At FU 28 participants (80%) scored above the cut off for autism on all three domains; 32 (91.4%)

scored above the cut off for autism on the RSI domain; 15 (42.9%) scored above the cut off on the VC domain; 28 (80%) scored above the cut off on the NVC domain and 30 (85.7%) scored above the cut off on the RBSP domain. McNemar tests revealed no statistically significant differences with regard to the number of participants meeting algorithm criteria for autism on any of the domains or on the total score between T1 and FU.

Seven children (20%) who met ADI-R criteria for autism at T1 no longer met algorithm criteria at FU. Two of these met Risi *et al.*'s (2006) broader criteria for ASD on the ADI-R; 5 failed to meet *either* the autism criteria or the Risi *et al.* (2006) broader ASD criteria on the ADI-R at FU (see Measures section). Among the 7 children who no longer met full ADI-R criteria, 3 failed to meet criteria on one domain (RBSP domain), 3 did not meet criteria on 2 domains (two on RSI and VC domains, one on NVC and RBSP domains) and one did not meet criteria on any domain.

Further analysis was conducted to explore differences between those children whose ADI-R algorithm classification remained stable and those who no longer met criteria for autism at FU (see Table 2; Mann Whitney U tests employed due to small and unequal N in groups). Those children with less stable classifications had significantly higher VABS ABC age equivalent scores (U = 33.5, N = 34, p = .01), VABS ABC standard score (U = 37.0, N = 35, p = .01), VABS communication age equivalent (U=35.5, N=35, p= .01) and VABS daily living age equivalent scores (U=36.5, N = 35, p = .01) at T1. No other significant differences were identified.

(Insert Table 2 about here)

Domain and total scores

Mean group scores for RSI and NVC domains and for total ADI-R score decreased significantly between T1 and FU (see Table 3). The decrease in the VC score did not reach significance (likely due to low N) and no significant difference was identified on the RBSP domain.

(Insert Table 3 about here)

In order to examine the extent of change shown by *individual* children, ADI-R change (FU-T1) scores at domain and total score level were divided into quartiles (see Figure 1 for details of quartile ranges in each domain and the total ADI-R score). Although some children made 'moderate' to 'major' improvements in each of the domains assessed, many showed no or only 'minor' change.

(Insert Figure 1 about here)

Item level scores

Stability of ADI-R item level scores between T1 and FU was assessed within each domain (RSI, NVC, VC and RBSP). Typically scores of 2 and 3 are collapsed in the algorithm output for the ADI-R. For the purposes of this analysis, the distinction between scores of 2 and 3 was retained in order to ensure that any change over time could be identified. Analysis on verbal items was conducted only for those individuals who were verbal at initial assessment (N= 7). Figure 2 illustrates the mean item level scores of items in the RSI, NVC, VC and RBSP domains of the ADI-R at T1 and FU[2].

(Insert Figure 2 about here)

Significant decreases in symptom severity were identified on 4 of the 13 items of the RSI domain (interest in other children, response to children's approaches, use of other's body to communicate, inappropriate facial expression). A significant decrease was also identified on 3 of the 7 items of the NVC domain (gestures, nodding and social play). There were no significant changes in the VC domain. A significant decrease was identified in one of the 7 items of the repetitive behaviour domain (repetitive use of objects).

T1 variables associated with ADI-R scores at follow-up

Table 4 describes Pearson correlation coefficients between the following T1 variables: ADI-R score, IQ, ADI-R overall language level (according to Q19 of the ADI-R), VABS communication, daily living and socialization standard scores and outcome on ADI-R total and domain scores at FU. Correlation coefficients indicated that communication (VABS communication score and overall level of language), socialization skills, IQ and symptom severity at T1 all demonstrate a strong association with outcome on the ADI-R at FU.

(Insert Table 4 about here)

In order to examine the extent of change shown by individual children, ADI-R change (FU-T1) scores at domain and total score level were divided into quartiles. Participants whose change scores were ≥ 75th or ≤ 25th quartile were considered to have made the most and least progress respectively. T1 receptive and expressive language skills were significantly better in those who made the most progress between T1 and FU. No other significant differences were identified (See Table 5).

(Table 5 about here)

Discussion

The purpose of the current study was prospectively to assess the stability of ASD symptomatology, according to the ADI-R, from pre-school to elementary school in a group of children with a wide range of intellectual and verbal ability.

No statistically significant changes in the number of participants meeting autism criteria on the ADI-R between T1 and FU were identified at domain or total score level. All of the 35 children selected for participation met full ADI-R criteria for autism at T1, including criteria for abnormality before 36 months. At FU, 28 children met full ADI-R criteria for autism based on their current behaviour. Since all the children had also received an independent diagnosis of ASD from a qualified clinician prior to inclusion in the study, this stability in ADI-R classification also reflects the overall reliability of early clinical diagnosis (see also Stone *et al.*, 1999; Eaves & Ho, 2004; Moore & Goodson, 2003). However, in common with other recent studies that have examined stability of clinical diagnosis (Sutera *et al*., 2007; Turner & Stone, 2007); some children who initially met ADI-R criteria for autism later on only met criteria for broader ASD or fell below even the ASD threshold. Although clinical diagnosis was not re-ascertained in the current study, either by the research team or the children's local clinicians at FU, all children still presented with behaviours associated with ASD and most showed significant impairments and delays in their development and functioning. .

The stability of ADI-R classification was greater in the present study than in that of Charman *et al*. (2005). Charman *et al.* reported that between the ages of 2 and 7 years, 14 children out of the 26 participants changed ADI-R classification at least once, 5 changed status twice and one child (with a clear clinical diagnosis of autism) was given a different ADI-R classification on each of the 3 assessments. Charman *et al.* (2005) suggested that these changes might be attributable to other factors

such as repetition effects and differences in administration/scoring of the ADI-R by different examiners over the 4 time points. The present sample was older when first seen (3.5 vs. 2.5 years in the Charman study) with some participants as old as 4 years at T1, thus greater stability might be expected. However, an examination of ADI-R trajectories in children from 3 to 7 years in the Charman *et al*. study still indicated considerably more change, with 11 out of that sample of 26, changing from ADI-R autism-positive ratings at age 3 to "non-autistic" ratings at age 7. One child moved from "non-autistic" to "autistic" over the same period. Although differences in the age range between the Charman *et al.* study and the current sample may account for some of the differences in overall stability, the reasons for this apparent discrepancy are not entirely clear and indeed, given the wide spread of intellectual and verbal ability in the present sample (initial IQ 16-130+; only 9 verbal participants at T1) one might have expected considerably less stability than was actually recorded. The methodological factors outlined in the Charman *et al.* study are unlikely to have played a role in the current study given the longer time period between assessments, the small number of interviewers involved and the fact that inter-rater reliability for ADI-R scoring was very high at FU.

At the domain level, however, the findings generally parallel those of Charman *et al.* (2005) and a number of other studies (Boelte & Poutska 2000; Fecteau *et al.*, 2003; Piven *et al.*, 1996; Seltzer *et al*, 2003; Shattuck *et al.*, 2007; Starr *et al.*, 2003) indicating improvements in symptom severity over time, especially on scores in the RSI and NVC domains of the ADI-R. Total ADI-R scores also decreased significantly. These results indicate that such improvements in the severity of autism-specific behaviours can occur in children who fall within the moderate-severe range of intellectual impairment as well as those of higher verbal and cognitive ability.

The relative improvements found in the different domains of the ADI-R have been reported in several other studies, although profiles of change seem to be more variable in retrospective studies or those involving shorter follow-up periods (Moore & Goodson, 2003; Seltzer *et al.*, 2003; Shattuck *et al.*, 2007). On the whole, improvements in the RBSP domain tend to be less marked than improvements in the RSI and NVC/VC domains (Charman *et al.*, 2005; Fecteau *et al.*, 2003; Piven *et al.*, 1996). Whilst measurement issues may play some role here (there are fewer items in the ADI-R RBSP domain than in the RSI, NVC and VC domains; see also Piven *et al.*, 1996), the consistency of the reported pattern of change across the three domains suggests that this may be more than just an artefact of measurement and that the developmental trajectory of each of the components of the triad of impairments may well differ.

Few studies to date have considered the stability of ADI-R symptom severity beyond the algorithm and domain levels. In the current study, changes in the severity of specific symptoms were also considered. Whilst findings should be interpreted with caution due to the use of multiple comparisons, the results highlight several behaviours that showed significant improvement over time. The symptoms showing most change in the RSI domain were: 'interest in other children', 'response to other children', 'appropriate facial expression' and 'use of other's body to communicate'. Within the NVC domain, 'gestures', 'nodding' and 'social play' were the specific behaviours that improved. Improvement in 'repetitive use of objects' was identified within the RBSP domain. No significant improvements were identified within the VC domain although this is likely to reflect the small number of participants who were verbal at T1. Fecteau *et al.* (2003) and Starr *et al.* (2003) also reported improvements in similar items of the ADI-R although Cox *et al.* (1999) noted improvements in different items (gaze, quality of social overtures, pointing and headshaking) between the ages of 20 and 42 months, suggesting that patterns of change may vary according to age group.

Investigation of the specific factors associated with changes in symptom severity over time highlighted several T1 variables that were associated with outcome at FU. Correlational analyses indicated that T1 communication skills, IQ and symptom severity were highly correlated with outcome at FU. At an individual level, comparison of those children who no longer met ADI-R criteria for autism at FU to those whose ADI-R classification remained stable indicated that early adaptive behaviour skills may also be an important factor in predicting later symptom severity. Comparison of individuals who made the most and least progress on the ADI-R between T1 and FU also indicated that language skills at pre-school age are likely to be associated with outcome at elementary school age.

In summary the findings from the current study suggest that children who have better language skills, less severe autism symptomatology and higher adaptive behaviour skills at the pre-school age are more likely to show improvements in symptom severity at elementary school age. This is consistent with the findings of several previous studies (Charman *et al.*, 2005; Coplan & Jawad, 2005; Eaves & Ho, 2004; McGovern & Sigman, 2005). Interestingly, it should be noted that in this sample, improvements in symptom severity were not notably associated with increases in IQ. In fact, IQ scores generally remained stable or decreased between T1 and FU. However, MA scores did show an increase between T1 and FU and therefore the improvements in symptom severity could, in part, be accounted for by general developmental progress. This might be especially the case for items included in the ADI-R algorithm that measure early communicative developmental abilities (e.g. gestures, nodding etc).

Methodological limitations

Whilst few studies to date have systematically assessed long term change in such young children with a wide range of intellectual disability, the present sample may consist of a greater proportion of children functioning within the moderate to severe range of disability than would be representative of the wider

autism population (cf. Edelson, 2006). Thus, at T1 30% of participants had IQ scores of ≤50. Furthermore, since the inclusion criteria also required them to have received a formal ASD diagnosis prior to initial assessment at mean age 3.5 years, it is likely that the sample may include a higher proportion of children with more severe difficulties than is typical of the general autism population. (Many children in the UK do not receive a confirmed diagnosis of autism until after this age; Howlin & Moore, 1997).

It might also be argued that analysis using ratings of 0-3 on the ADI-R increases the chances of finding significant change over time (changes in scores from 3 to 2 on the ADI-R algorithm are considered to be less significant than change from 2 to 1 or 1 to 0; Starr *et al.*, 2003). Nevertheless, our results are consistent both with those of Fecteau *et al.* (2003), who used the same scoring method and those of Starr *et al.* (2003), who employed the more conservative method of collapsing scores 2 and 3. Given the relative severity of intellectual ability in the current sample compared to the Starr *et al.* (2003) and Fecteau *et al.* (2003) studies, the consistency with which improvements in autistic symptomatology have been observed over time suggests that this profile of change is not an artefact of measurement but is likely to reliably represent the typical course of development of the disorder.

Summary

The findings of the current study suggest that ADI-R classification of autism at a pre-school age remains stable throughout the elementary school years, although some children initially meeting criteria for autism went on to meet ASD criteria or fell below this threshold at FU. However, the severity of autism-specific symptoms improves over time. As found in previous research, improvements are more likely to occur within the areas of communication and social interaction than in the repetitive behaviour domain. Overall, children who make the most improvements in symptom

severity are likely to be those who demonstrate less severe autism symptomatology, better language and adaptive behaviour skills at pre-school age. Although the number of children who demonstrate changes in ADI-R classification is small, identifying the variables that characterise those children who *do* change over time may be important in understanding the factors underlying later outcome in autism symptom severity in the wider autism population.

References

Bayley, N. (1993). *Bayley Scales of Infant Development – Second Edition*. San Antonio, TX: Psychological Corporation.

Billstedt, E., Gillberg, C., & Gillberg, C. (2005). Autism after adolescence: Population-based 13 to 22 year follow-up study of 120 individuals with autism diagnosed in childhood. *Journal of Autism and Developmental Disorders, 35,* 351-360.

Boelte, S & Poustka, F. (2000). Diagnosis of autism: The connection between current and historical information. *Autism: International Journal of Research and Practice, 4,* 382-390.

Brownell, R. (2000). *Expressive One Word Picture Vocabulary Test – 2000 Edition.* Academic Therapy Publications, Novato, California.

Charman, T & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2 to 3 year old children. *Journal of Child Psychology and Psychiatry, 43,* 289-305.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry, 46,* 500-513.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology and Psychiatry, 48,* 128-138.

Coplan, J. & Jawad, A. (2005). Modelling clinical outcome of children with autistic spectrum disorders. *Pediatrics, 116,* 117-122.

Cox, A., Charman, T., Baron-Cohen, S., Drew, A., Klein, K., Baird, G., Swettenham, J. & Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry, 40,* 719-732.

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

Eaves, L. & Ho, H. (2004). The very early identification of autism: Outcome to age 4 1/2 - 5. *Journal of Autism and Developmental Disorders, 34,* 367-378.

Edelson, MG. (2006) Are the majority of children with autism mentally retarded? A systematic evaluation of the data. *Focus on Autism & Other Developmental Disabilities*, *21,* 66-83

Fecteau, S., Mottron, L., Berthiaume, C., & Burack, J. (2003). Developmental changes of autistic symptoms. *Autism: International Journal of Research and Practice 7,* 255-268.

Gardner, M.F. (1990). *Expressive One-Word Picture Vocabulary Test (Revised)*. Academic Therapy Publications, Novato, California.

Howlin, P. & Moore, A. (1997). Diagnosis in Autism: A survey of over 1,200 parents. *Autism: International Journal of Research and Practice, 2,* 135-162.

Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry, 36,* 1365-1382.

Lord, C., Risi, S., DiLavore, P., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry, 63,* 694-701.

Magiati, I., Charman, T., & Howlin, P. (2007). A two-year follow-up study of early intensive home-based behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry, 48*, 803-812.

McGovern, C. & Sigman, M. (2005). Continuity and change from early childhood to adolescence in autism. *Journal of Child Psychology and Psychiatry, 46,* 401-408.

Moore, V. & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? *Autism: International Journal of Research and Practice, 7,* 47-63.

Piven, J., Harper, J., Palmer, P., & Arndt, S. (1996). Course of behavioral change in autism: A retrospective study of high-IQ adolescents and adults. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 523-529.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E., Leventhal, B. & Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *American Academy of Child and Adolescent Psychiatry, 45,* 1094-1103.

Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism Diagnostic Interview- Revised (ADI-R).* Los Angeles: Western Psychological Services.

Seltzer, M., Krauss, M., Shattuck, P., Orsmond, G., Swe, A., & Lord, C. (2003). The symptoms of autism spectrum disorders in adolescence and adulthood. *Journal of Autism and Developmental Disorders, 33,* 565-581.

Shattuck, P., Seltzer, M., Greenberg, J., Orsmond, G., Bolt, D., Kring, S. Lounds, J. & Lord, C. (2007). Change in autism symptoms and maladaptive behaviors in adolescents and adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders,* 37, 1735-1747.

Sparrow, S.S., Balla, D. & Cicchetti, D.V. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.

Starr, E., Szatmari, P., Bryson, S., & Zwaigenbaum, L. (2003). Stability and change among high-functioning children with pervasive developmental disorders: A 2 year outcome study. *Journal of Autism and Developmental Disorders, 33,* 15-22.

Stone, W., Lee, E., Ashford, L., Brissie, J., Hepburn, S., Coonrod, E. & Weiss, B. (1999). Can autism be accurately diagnosed in children under 3 years? *Journal of Child Psychology and Psychiatry, 40,* 219-226.

Stutsman, R. (1948). *Merrill-Palmer Scale of Mental Tests*. Los Angeles: Western Psychological Services.

Sutera, S., Pandey, J., Esser, E.L., Rosenthal, M.A., Wilson, L.B., Barton, M., Green, J., Hodgson, S., Robins, D.L., Dumont-Mathieu, T. & Fein, D. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37,* 98-107.

Turner, L.M. & Stone, W.L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry, 48,* 793-802.

Wechsler, D. (1990). *Wechsler Pre-school and Primary Scale of Intelligence-Revised UK Edition* (WPPSI-R UK). Sidcup, Kent: Psychological Corporation.

Wechsler, D. (2004). *Wechsler Pre-school and Primary Scale of Intelligence – Third Edition UK* (WPPSI-III UK). Sidcup, Kent: Psychological Corporation.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children – Fourth UK Edition* (WISC-IV UK). Sidcup, Kent: Psychological Corporation.

Author Note

Dr. Jo Moss: Department of Community Health Sciences, St. George's University of London, UK

Dr. Iliana Magiati: Department of Psychology, Institute of Psychiatry, Kings College London, UK

Professor Tony Charman: Institute of Child Health, University College London, UK

Professor Patricia Howlin: Department of Community Health Sciences, St. George's University of London, UK

Professor Patricia Howlin and Dr. Jo Moss are now at the Department of Psychology, Institute of Psychiatry, Kings College London. Dr Iliana Magiati is currently working as a Clinical Psychologist at the Community Child Psychology Service and the Child Development Team, Islington Primary Care NHS Trust.

This study was conducted as part of Iliana Magiati's clinical doctoral thesis. We are extremely grateful to the children, families, teachers and schools who participated in the study. We wish to acknowledge Dr. Samantha Day (SD) for completing many intake assessments and Dr. Judy Gould for her advice in the early stages of the project. The intake stage of the study was funded by Action Research.

Address for correspondence:

Dr. Jo Moss

Department of Psychology

Institute of Psychiatry

Henry Wellcome Building, PO 77

De Crespigny Park

London, SE5 8AF

Telephone: +44 (0)20 7848 0417

E-mail: jo.moss@iop.kcl.ac.uk

Footnotes

[1] Participants in the current study were selected from a sample of children who were recruited as part of a longitudinal study investigating the effectiveness of early intervention in pre-school children enrolled in either eclectic nurseries or home-based Early Intensive Behavioural Intervention (EIBI) programmes (Magiati, Charman and Howlin, 2007), thus there is some overlap here with the participants involved in that study. The current study was a longer follow-up study of these participants focusing on stability and change of autism-specific symptomatology and behaviour over time.

[2] Mean scores are presented in the graphs for ease of reference, but analyses were conducted using non-parametric techniques.

Table 1: Participant characteristics at T1 and FU assessments (N=35)

		T1	FU
		Mean (SD)/ Median *Range*	Mean (SD)/ Median *Range*
Mean Chronological Age (years)		3.5 (.6) *2.3-4.5*	10.5 (.8) *9.1-12.1*
Mean IQ[1]		68.9 (33.0) *16.0-137.5*	52.2 (22.0) *15.9-124.0*
Mean MA[1] (months)		29.4 (16.1) *5.0-56.0*	67.9 (32.0) *19.0-168.4*
Median Receptive language[2]	*Raw Score*	0.0 *0.0-32.0*	33.0 *0.0-117.0*
	Standard score	40.0 *40.0-94.0*	40.0 *40.0-117.0*
	N scoring at Basal	20	9
Median Expressive language[3]	*Raw Score*	0.0 *0.0-39.0*	27.0 *0.0-119.0*
	Standard Score	68.0 *49.0-120.0*	55.0 *55.0-121.0*
	N scoring at Basal	25	13
Mean VABS ABC [4]	*Age Equivalence (months)*	19.9 (5.8) *12.0-33.0*	44.5 (26.8) *13.0-110.0*
	Standard Score	58.4 (5.8) *50.0-72.0*	37.8 (17.4) *20.0-86.0*
Mean VABS communication	*Age Equivalence (months)*	15.6 (6.1) *8.0-33.0*	47.1 (35.8) *11.0-135.0*
	Standard Score	59.3 (7.7) *47.0-78.0*	41.4 (20.5) *20.0-101.0*
Mean VABS daily living skills	*Age Equivalence (months)*	21.0 (6.1) *11.0-39.0*	47.2 (24.4) *17.0-110.0*
	Standard Score	62.8 (6.7) *55.0-89.0*	33.2 (19.1) *20.0-84.0*
Mean VABS socialisation	*Age Equivalence (months)*	13.7 (4.9) *6.0-29.0*	35.7 (25.2) *10.0-116.0*
	Standard Score	59.1 (5.2) *52.0-71.0*	48.8 (14.5) *28.0-97.0*

Verbal ability (functional speech)[5]	*N phrases or more*	7	20
	N at least five words	8	6
	N Fewer than five words	20	9

[1] For details of IQ assessments see measures section below; [2] British Picture Vocabulary Scales (Dunn *et al.*, 1997) - median values reported due to data being non-normally distributed; [3] Expressive One Word Picture Vocabulary Test (Gardener, 1990; Brownell, 2000) -median values reported due to data being non-normally distributed.; [4] Vineland Adaptive Behavior Scales – Survey Form (Sparrow, Balla & Chicchetti, 1984); [5] Data based on the current language item (Q 19) of the ADI-R.

Table 2: T1 characteristics of participants who scored above (N=28) and below (N=7) the cut off for autism on the ADI-R at FU.

T1 Variable		**Above cut off FU (N=28)** Median[1]/*Range*	**Below cut off FU(N=7)** Median[1]/*Range*	**U**	***p***
IQ[2]		56.3 *(16.0-137.0)*	97.9 *(46.0-121.4)*	64.5	.17
MA[2] (months)		20.5 *(5.0-56.0)*	46.0 *16.0-51.0*	59.5	.11
Receptive language[3]	*Raw Score*	0.0 *(0.0-22.0)*	17.0 *(0.0-32.0)*	66.5	.14
Expressive language[4]	*Raw Score*	0.0 *(0.0-18.0)*	7.0 *(0.0-39.0)*	58.0	.04
VABS ABC[5]	*Age Equivalence*	17.0 *(12.0-33.0)*	24.00 *(19.00-32.00)*	33.5	.01
	Standard Score	56.00 *(50.0-68.00)*	63.0 *(61.00-72.0)*	37.0	.01
VABS Communication	*Age Equivalence*	13.0 *(8.0-33.0)*	18.0 *(14.0-28.0)*	35.5	.01
	Standard Score	56.0 *(47.0-77.0)*	61.0 *(56.0-78.0)*	44.5	.03
VABS daily living skills	*Age Equivalence*	18.0 *(11.0-34.0)*	26.0 *(19.0-39.0)*	36.5	.01
	Standard Score	61.0 *(55.0-70.0)*	68.0 *(59.0-89.0)*	49.5	.05
VABS socialisation.	*Age Equivalence*	12.0 *(6.0-29.0)*	16.0 *(12.0-24.0)*	41.0	.02
	Standard Score	57.0 *(52.0-70.0)*	61.0 *(56.0-71.0)*	47.0	.04
ADI-R total score[6]		38.0 *(26.0-46,0)*	36.0 *(18.0-43.0)*	57.5	.09

[1] Median values are reported due to small and uneven N across the two groups; [2] For details of IQ assessments see measures section; [3] British Picture Vocabulary Scales (Dunn *et al.*, 1997); [4] Expressive One Word Picture Vocabulary Test (Gardener, 1990; Brownell, 2000); [5] Vineland Adaptive Behavior Scales – Survey Form (Sparrow, Balla & Chicchetti, 1984); [6] ADI-R total score = RSI + NVC + RBSP. Only NVC score was used in order to include both verbal and nonverbal participants.

Table 3: Domain and total scores on the ADI-R at T1 and FU (N= 35)

ADI-R Domain	**T1** Mean (SD)*Range*	**FU** Mean (SD)*Range*	**N**	**T**	***p***

RSI	21.1 (4.5)	18.4 (5.9)	35	2.85	.01
	10.0-28.0	*3.0-28.0*			
NVC*	10.7 (2.9)	8.9 (3.5)	35	4.19	<.001
	3.0-14.0	*1.0-14.0*			
VC	13.3 (3.7)	9.1(4.4)	7	1.79	.12
	8.0-18.0	*4.0-19.0*			
RBSP	4.7 (1.3)	4.5(1.9)	35	.56	.58
	3.0-9.0	*2.0-10.0*			
Total Score**	36.4 (6.9)	31.8 (10.0)	35	3.36	.002
	17.0-46.0	*6.0-46.0*			

* For the purposes of analysis, all participants received a score on the NVC domain regardless of whether or not they were verbal.

** Total score = RSI + NVC + RBSP. Only NVC score was used in order to include both verbal and nonverbal participants. RSI cut-off = 10, NVC cut-off = 7; VC cut-off = 8 and RBSP cut-off = 3.

Table 4: Pearson Correlation between T1 predictor variables and outcome on the ADI-R at FU

T1variables	**FU ADI-R scores**			
	Total Score	**RSI score**	**NVC score**	**RBSP score**
ADI-R total score	**.59*****	**.55****	**.62*****	.26
RSI score	**.48****	**.48****	**.52****	.08
NVC score	**.68*****	**.62*****	**.69*****	**.42***
RBSP score	-.10	-.14	-.12	.14
Overall level of language	**.72*****	**.72*****	**.67*****	.33
IQ	**-.63*****	**-.63*****	**-.67*****	-.15
VABS ABC Communication (standard score)	**-.55*****	**-.59*****	**-.42***	-.31
VABS Daily Living Skills (standard score)	**-.37***	**-.38***	-.26	-.30
VABS Socialization (standard score)	**-.55****	**-.59*****	**-.48****	-.18

***<.001; **<.01; * <.05

Table 5: Characteristics of participants who made the most (N=8) and least (N=11) progress on ADI-R scores from T1 to FU.

T1 variables	**Most progress (N=8)**	**Least progress (N=11)**	**U**	***p***
	Median[1]	**Median**[1]		
	Range	***Range***		
Chronological Age (years)	3.5 *2.6-4.0*	3.42 *2.8-4.2*	42.0	.90
Mental Age (months)[2]	46.0 *5.0-52.0*	19.0 *14.0-43.0*	32.0	.32
IQ[2]	97.9 *16.0-121.4*	50.0 *34.0-97.7*	28.5	.20

Receptive language[3] (raw score)	14.0 *0.0-32.0*	0.0 *0.0-5.0*	18.00	.01
Expressive language[4] (raw score)	0.5 *0.0-39.0*	0.0 *0.0-0.0*	22.00	.01
VABS[5] ABC score (ss)	61.0 *52.0-72.0*	55.0 *50.0-66.0*	23.5	.09
VABS[5] communication (ss)	61.0 *54.0-78.0*	57.0 *47.0-61.0*	26.0	.13
VABS[5] daily living skills (ss)	63.0 *59.0-89.0*	62.0 *55.0-77.0*	35.5	.48
VABS[5] socialisation (ss)	59.0 *53.0-71.0*	58.0 *53.0-64.0*	28.0	.19
ADI-R total score[6]	38.0 *18.0-46.0*	36.0 *26.0-43.0*	37.0	.56

[1]Median values reported due to small N in groups; 2 For details of IQ assessments see measures section; [3] British Picture Vocabulary Scales (Dunn *et al.*, 1997); [4]Expressive One Word Picture Vocabulary Test (Gardener, 1990; Brownell, 2000); [5] Vineland Adaptive Behavior Scales – Survey Form (Sparrow, Balla & Chicchetti, 1984); [6]ADI-R total score = RSI + NVC + RBSP. Only NVC score was used in order to include both verbal and nonverbal participants.

Figure Caption Sheet

Figure 1: Number of children demonstrating the most and least improvements on ADI-R domain and total scores and ADI-R change score quartiles.

Figure 2: Mean item scores in the RSI, NVC, VC and RBSP domains of the ADI-R at T1 and FU

Figure 1 Top

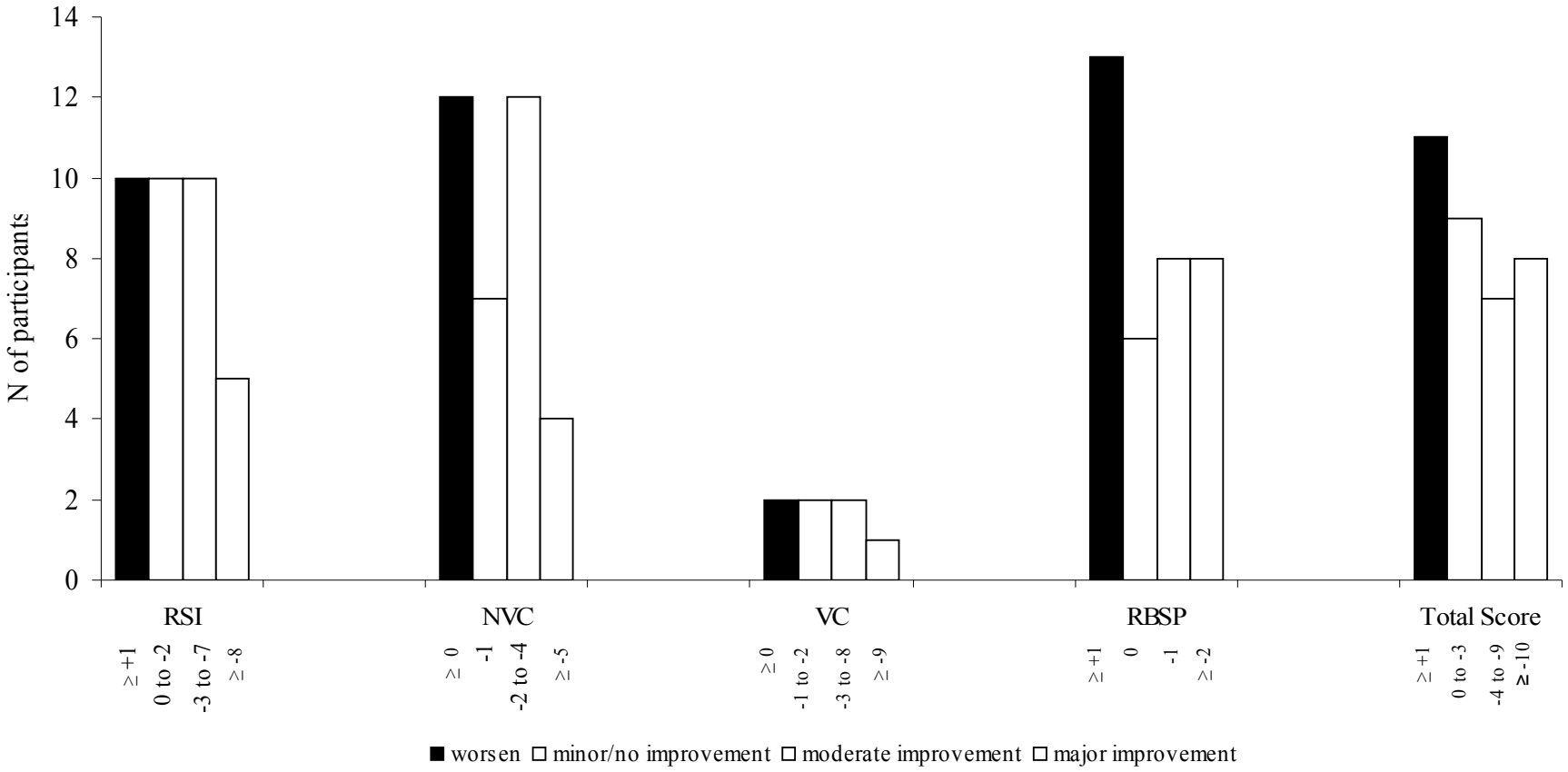
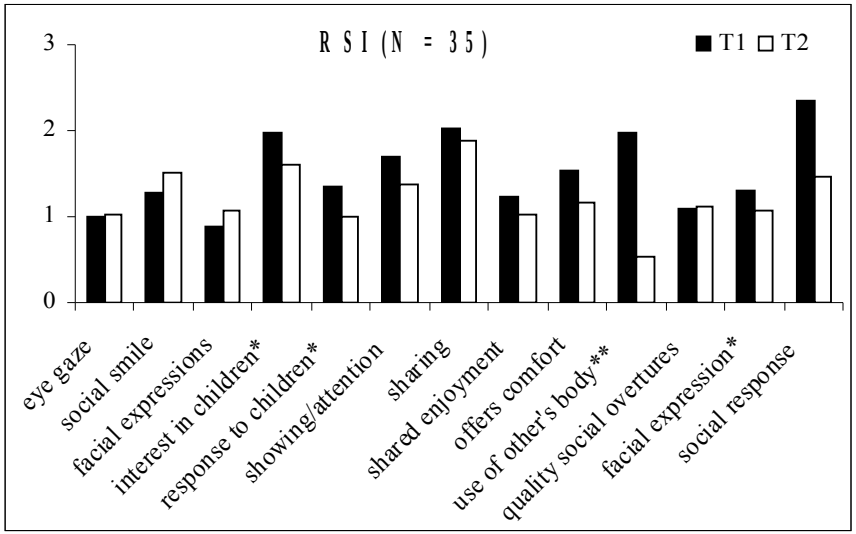
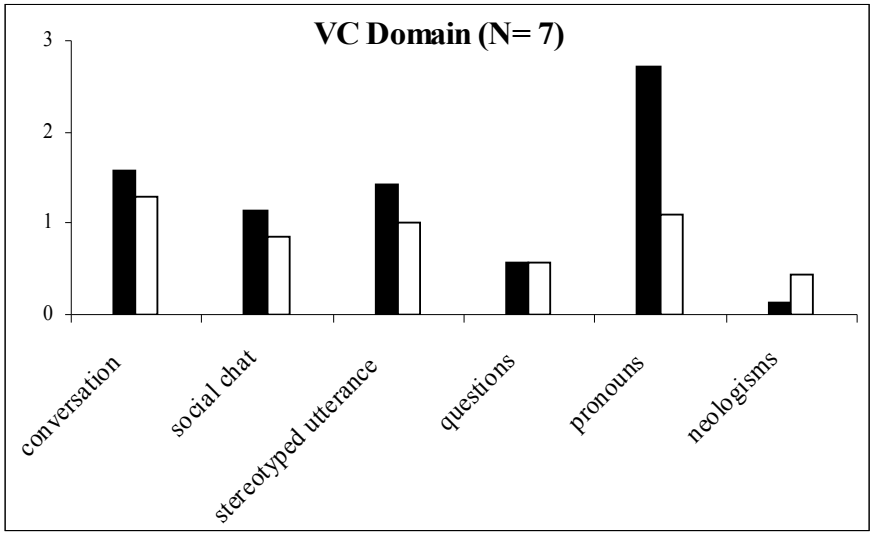
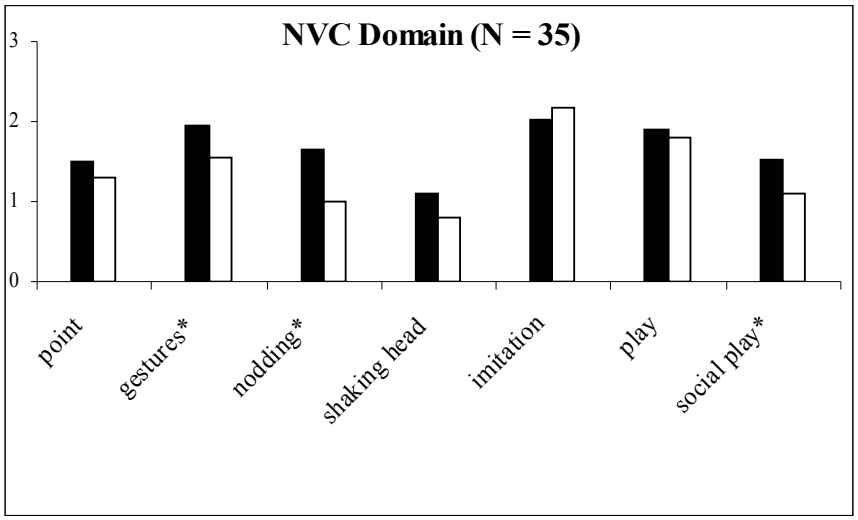
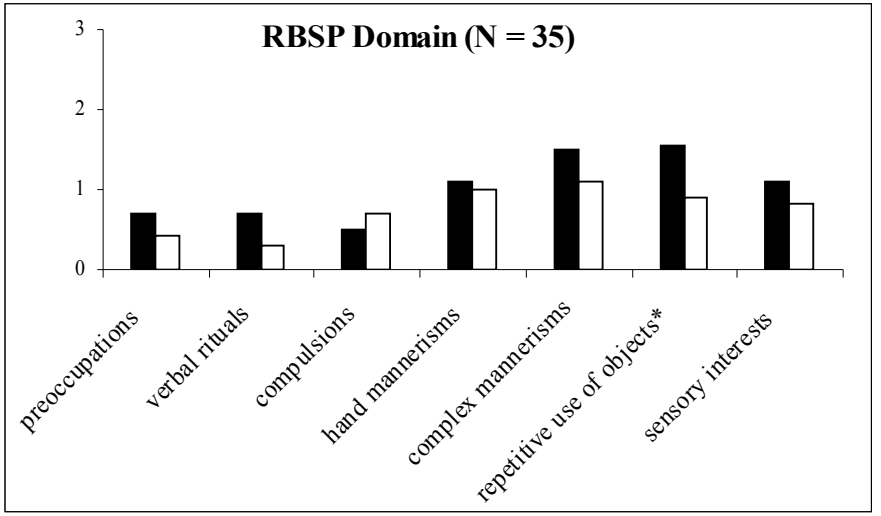


Figure 2 Top


RSI (N = 35)
T1 T2
3
2
1
0
eye gaze
social smile
facial expressions
interest in children*
response to children*
showing/attention
sharing
shared enjoyment
offers comfort
use of other's body**
quality social overtures
facial expression*
social response


VC Domain (N= 7)
3
2
1
0
conversation
social chat
stereotyped utterance
questions
pronouns
neologisms


NVC Domain (N = 35)
3
2
1
0
point
gestures*
nodding*
shaking head
imitation
play
social play*


RBSP Domain (N = 35)
3
2
1
0
preoccupations
verbal rituals
compulsions
hand mannerisms
complex mannerisms
repetitive use of objects*
sensory interests

* p≤ .01
** *p* <.001