Pasco, G., Gordon, R. K., Howlin, P. & **Charman, T.** (2008). The Classroom Observation Schedule to Measure Intentional Communication (COSMIC): An observational measure of the intentional communication of children with autism in an unstructured classroom setting. *Journal of Autism and Developmental Disorders, 38,* 1807-1818.

Running head: Observing intentional communication in the classroom

The Classroom Observation Schedule to Measure Intentional Communication (COSMIC): An observational measure of the intentional communication of children with autism in an unstructured classroom setting

Greg Pasco, Rosanna K Gordon, Patricia Howlin, Tony Charman,
Department of Community Health Science, St George's University of London

Abstract

The Classroom Observation Schedule to Measure Intentional Communication (COSMIC) was devised to provide ecologically valid outcome measures for a communication-focused intervention trial. Ninety-one children with autism spectrum disorder aged 6 years 10 months (SD 16 months) were videoed during their everyday snack, teaching and free play activities. Inter-rater reliability was high and relevant items showed significant associations with comparable items from concurrent Autism Diagnostic Observation Schedule – Generic (Lord et al., 2000) assessments. In a subsample of 28 children initial differences in rates of initiations, initiated speech/vocalisation and commenting were predictive of language and communication competence 15 months later. Results suggest that the use of observational measures of intentional communication in natural settings is a valuable assessment strategy for research and clinical practice.

Key Words: Observation; Ecological validity; Intentional communication; Classroom

Impairment in communication is a core feature of autism spectrum disorder (ASD) (American Psychiatric Association, 2000; World Health Organisation, 1993) and the development of certain early social communication skills has been found to be an important prognostic indicator for later language and social development (Bruinsma, Koegel, & Koegel, 2004; Charman et al., 2003; Mundy, Sigman, & Kasari, 1990; Sigman & Ruskin, 1999). However, assessing the language and communication abilities of young and/or severely delayed children with autism can be problematic, and in some cases assessment of language skills using standardised tests may be unachievable (Charman, Drew, Baird, & Baird, 2003). Many of these children do not use speech as a means of expressive communication, (Tager-Flusberg, Paul, & Lord, 2005) and very few formal assessments of communication abilities are standardised below a 2-year age equivalent (Charman, 2004; Marans, 1997), making them potentially unsuitable. Furthermore, the formal context of an assessment may not enable children with ASD to demonstrate language competence, as their skills may be highly situation specific, or best observed in more natural settings (Charman et al., 2003; Wetherby, Schuler, & Prizant, 1997). Guidelines based on expert opinion (e.g. National Initiative for Autism: Screening and Assessment [NIASA], 2003; National Research Council [NRC], 2001) recommend that standardised tests should constitute only a part of the assessment of the communicative strengths and needs of a child with autism.

A number of research-based assessments of children's early social communication skills has been developed to provide profiles of the social communication skills of children with ASD and to measure changes in communication skills following intervention. The *Behavior Sample* of the Communication and Symbolic Behavior Scales – Developmental Profile (CSBS-DP: Wetherby & Prizant, 2002) and the Early Social Communication Scales (ESCS: Mundy, Hogan,

& Doehring, 1996) are two of the best known and mostly widely used (Wetherby, 2006). Other instruments have been described and evaluated in the literature, including the Communicative Intention Inventory (Coggins & Carpenter, 1981), the Prelinguistic Communication Assessment (PCA: Stone, Ousley, Yoder, Hogan, & Hepburn, 1997) and the Social Communication Assessment for Toddlers with Autism (SCATA: Drew, Baird, Taylor, Milne, & Charman, 2007). Although primarily designed as a diagnostic instrument, the Autism Diagnostic Observation Schedule – Generic (ADOS-G: Lord et al., 2000) has also been used to provide measures of specific aspects of social communication and interaction (e.g. Aldred, Green, & Adams, 2004; McConachie, Randle, Hammal, & Le Couteur, 2005). The tasks included in these assessments contain specific 'presses' or 'communicative temptations' designed to elicit child responses, particularly in relation to functional communication skills such as joint attention and requesting, and the nonverbal behaviours, such as gaze switching and gestures, that may accompany them. The psychometric properties, validity and reliability of these instruments have been reported elsewhere (Paul, 2005; Wetherby, 2006).

The procedures involved in these structured observational assessments are designed to be "specially-created analogues of [natural] settings" (Harris, Belchic, Blum, & Celiberti, 1994, p.128). They are widely used because observation in a natural unstructured setting can be very time consuming, and a child may not display particular behaviours even over an extended observation period (Wetherby & Prizant, 2005). However, whilst these instruments offer an efficient method for ascertaining information relating to pertinent communication skills and deficits of children with ASD, they do not provide direct measures of children's communicative behaviour in natural everyday settings or in relation to their regular communication partners. Guidelines (NIASA, 2003; NRC 2001) suggest that a comprehensive ecologically based

assessment strategy should consider children's language in natural contexts such as the home, classroom or wider community. Bishop (1998) advises that researchers and clinicians using observational procedures need to appraise the representativeness of the samples of behaviour observed and recognise that micro-analytic approaches can be time consuming and often suffer from poor inter-rater reliability.

Watson, Lord, Schaffer, and Schopler (1987) developed a schedule designed to measure the spontaneous communication of children and adults with autism in everyday settings. Abrahamsen and Mitchell (1990) and Stone and Caro-Martinez (1990) have used this to rate the intentional communication of children with autism in educational and other settings and report acceptable levels of inter-rater reliability. Murdoch, Cost, and Tieso (2007) report high levels of inter-rater reliability and content validity for the Social-Communication Assessment Tool (S-CAT), an observational measure that incorporates four aspects of social communication behaviour occurring during interactions with typically-developing peers. Three studies (Hwang & Hughes, 2000; McHale, Simeonsson, Marcus, & Olley, 1980; Stahmer & Ingersoll, 2004) report the use of specifically-developed coding schemes to measure multiple aspects (e.g. communicative form, function and communication partner) of the social communication skills of children with autism in everyday classroom settings. In each case adequate levels of inter-observer agreement were reported.

Howlin, Gordon, Pasco, Wade, and Charman (2007) conducted a randomised controlled trial of Picture Exchange Communication System (PECS: Frost & Bondy, 2002) training for teachers of children with autism. This was an effectiveness trial, assessing the benefits of PECS under 'real world' conditions (Flay et al., 2005) and involved a large (N = 84), well-characterised sample of school-age children with autism, observed 3 times over a period of 2 years. The

primary outcome measures were required to be ecologically valid, based on observations of communication skills in everyday settings. As all the participants were nonverbal or in the early stages of language development, particular measures were required of their initiations (verbal and nonverbal), use of picture symbols and speech. Existing coding schemes, such as that described by Watson et al. (1987), were considered for this purpose, but none were found to cover the range of codes and aspects of communication required for the study, without substantial amendment. Howlin et al. (2007) detected immediate treatment effects for rates of observed initiations and symbol use, but not for rates of speech. These immediate treatment effects did not generalize to dependent variables based on ADOS-G domain scores or outcome measures derived from standardised assessments of vocabulary development.

The purpose of the present report is to describe the content, reliability and validity of the schedule devised to provide these outcome measures – the Classroom Observation Schedule to Measure Intentional Communication (COSMIC). Data on reliability and concurrent validity are based entirely on baseline, pre-treatment, observations and assessments, and are therefore not affected by the intervention delivered as part of the study. The data on predictive validity relate only to the children in the control group who did not receive treatment as part of the Effectiveness of PECS study (N = 28) in order for the results to be 'uncontaminated' by any effects of the intervention.

Method

Development of COSMIC

Specific aspects of social communication were selected, based on the requirements of the intervention study and items included in other assessments of early social communication (e.g. Carr & Felce, 2006; Mundy et al., 1996; Watson et al., 1997; Wetherby & Prizant, 2002;

Wetherby & Prutting, 1984). Operational definitions (see Appendix) were developed for each item, and videos of children's communicative behaviour in classroom settings rated in order to ascertain the robustness of these definitions, develop inter-rater agreement and modify the content of COSMIC where appropriate. The final list of items included in COSMIC is shown in Table 1. The Record Form used for recording children's communicative interactions is shown in the Appendix.

Participants

Ninety-one children aged between 4 and 11 years participated in this study. All were nonverbal or at a 1-word level of expressive communication and had a formal clinical diagnosis of autism. Eighty-three of the children were participants in the Effectiveness of PECS study (Howlin et al., 2007).

Videoing and Assessment

The children were videoed for 15 minutes during which their usual snack time took place, and for an additional 15 minutes during other classroom activities (e.g. one-to-one teaching, group teaching and free play). Teachers were asked to carry on as if the researcher were not present during the recording, as the aim was to observe children during their familiar everyday activities. Children were administered module 1 of the ADOS-G (Lord et al., 2000), the *Visual Reception* and *Fine Motor* scales of the Mullen Scales of Early Learning (Mullen, 1995), the Expressive One Word Picture Vocabulary Test (EOWPVT: Brownell, 2000) and the British Picture Vocabulary Scales (BPVS: Dunn, Dunn, Whetton, & Burley, 1997), where achievable, to provide standardised measures of autism diagnosis and symptomatology, nonverbal cognitive ability, and expressive and receptive vocabulary, respectively.

Inter-rater Reliability

Analysis of inter-rater reliability was based on recordings of 40 randomly-selected sessions, equally balanced between snack and non-snack contexts, representing 22% of all baseline sessions. Twenty-four of these sessions were independently rated by 2 of the authors (GP & KG), and the remaining 16 were rated by one of these authors and one of 2 additional raters. These additional raters were naïve to the overall purpose of the study, and had no previous contact with or knowledge of the participants or involvement in the original intervention study. These 2 naïve raters received approximately 3 hours of training using videos of participant children, as well as a copy of the coding definitions (see Appendix).

Concurrent Validity

Concurrent validity was examined by comparing rates of COSMIC items from baseline observations of children with scores on items from what is considered to be the 'gold standard' observational measure used to assist in the diagnosis of autism – the ADOS-G. Items for comparison were selected on the basis that items or combinations of items from the ADOS-G appear to measure similar aspects of communicative behaviour as the items from COSMIC. Items selected for comparison were: (1) rates of *Initiated Speech/Vocalisation* from COSMIC and ADOS-G item A1 *Overall Level of Non-Echoed Language*; (2) rates of *Echolalia* from COSMIC and ADOS-G item A4 *Echolalia*; (3) rates of *Gesture/Pointing* from COSMIC and ADOS-G items A7 *Pointing* and A8 *Gestures*; and (4) rates of *Comment* from COSMIC and ADOS-G item B10 *Spontaneous Initiation of Joint Attention*.

The sessions upon which the COSMIC data were based comprised unstructured everyday classroom-based activities, whereas ADOS-G scores were based on children's behaviours during a semi-structured interaction with an unfamiliar examiner. Some ADOS-G items are scored with reference to a specific task or press, whereas others are based on a summary of the child's

behaviour across the assessment. Of the ADOS-G items selected for these comparisons the majority are summary items. For example, scoring for the *Pointing* item on the ADOS-G may include examples of unsolicited spontaneous pointing, spontaneous pointing in response to specific materials (e.g. if the child spontaneously points to bubbles fired from a bubble gun by the examiner) and pointing deliberately elicited by a task (e.g. the examiner holds up 2 items to see whether the child points to one of them as a request). Scores for the *Spontaneous Initiation of Joint Attention* item are based upon examples of joint attention bids by the child that are primarily responses to specific elements of the ADOS-G assessment, such as the bubbles or a remote control toy being operated, but may also be entirely spontaneous on the child's part.

Predictive Validity

Many children scored zero on the EOWPVT and BPVS at both baseline and follow-up. This meant that scores from these standardised vocabulary assessments were not useful as appropriate measures of language competence at outcome for the investigation of predictive validity. An ADOS-based measure of Language and Communication Competence (ALCC) score was therefore created as a measure of expressive language ability. The 4 items from the ADOS-G *Communication* domain that assess 'normative' aspects of communicative behaviour, as opposed to stereotyped or atypical symptomatology, were summed to provide the ALCC score[1]. As ADOS-G items are measures of severity, scores were reversed so that higher ALCC totals relate to higher levels of language and communication competence. Predictive validity was assessed on the basis of scores from 28 children selected from the non-treatment arm of the Effectiveness of PECS study by examining associations between baseline rates of specific COSMIC items with ALCC scores from ADOS-G assessments conducted 15 months later.

Baseline items selected for the assessment of predictive validity were: *Initiation*; *Initiated Speech/Vocalisation*; and *Comment*. The hypothesis that these items might be predictive of increases in language competence at outcome was based largely on findings from previous research. Thus, rates of *Initiated Speech/Vocalisation* can be considered as broad indicators of a child's overall level of language and communication ability; the amount of commenting and/or joint attention ability demonstrated by children with autism has been reported to be a good predictor of language outcome (e.g. Drew et al., 2007; Sigman & Ruskin, 1999; Siller & Sigman, 2002); the rate of initiation was a primary outcome measure in the Effectiveness of PECS study, based on an assumption about the importance of spontaneous communication for children with autism, and initial rates of initiations have also been found to be predictive of later language ability (Drew et al., 2007).

Results

Participants

Age and sex. The mean age of children at baseline was 6 years 10 months (SD = 16 months). The youngest child was 3:11 and the oldest was 10:02. Ten of the 91 participants were girls.

Autism diagnosis. All the children had clinical diagnoses of autism made by local clinicians. Seventy-nine children scored above the ADOS-G diagnostic cut-off for *autism*, the remaining 12 scored above the *autism spectrum* cut-off.

Nonverbal mental age and DQ. Many of the children in this study were older than the age for which the Mullen is standardised (68 months), so neither standardised scores nor nonverbal IQ scores are presented. Age equivalent scores from the *Visual Reception* and *Fine Motor* scales of the Mullen were added together and divided by 2 in order to provide a nonverbal mental age

(NVMA) equivalent. The mean NVMA equivalent was 25.4 (SD = 9.6) months. The NVMA was multiplied by 100 and divided by chronological age in order to provide a nonverbal developmental quotient (NVDQ) score. The mean NVDQ score was 31.8 (SD = 12.7).

Expressive and receptive vocabulary. The distributions of scores for EOWPVT and BPVS were highly negatively skewed with large proportions of zero scores, so median and interquartile scores are presented. The median EOWPVT raw score was 0 (IQR 0 – 9), with 61 children scoring 0 on this assessment. The median BPVS raw score was 1 (IQR 0 – 11), with 38 children scoring 0. As so many of the children scored 0 on these assessments, neither standard scores nor age equivalents were calculated.

Rates of COSMIC Items

Even though several items from different categories may be recorded simultaneously within a single interaction (e.g. *Communication partner*, *Intentionality*, *Form* and *Function* of communicative interaction) – see Record Form in the Appendix – the data presented here are for each item separately. The median rate (per minute) and interquartile range of each item from baseline observations are presented in Table 1. Medians are presented as the distributions of nearly all items were highly negatively skewed, some items having a large proportion of zero rates. The numbers of children observed to display each behaviour at least once are also presented along with correlations (Pearson's r) between rates of each item and ADOS-G Diagnostic total and NVDQ scores from concurrent assessments. These correlations demonstrate the relationships between each behaviour rated using COSMIC and the degree of autistic symptomatology and nonverbal cognitive ability. Parametric statistics are presented here because the broad pattern of results is very similar to that based on nonparametric correlations, and because partial correlations can be performed more easily using Pearson's statistic (see

Predictive validity below). As multiple correlations were performed, results were tested at the 0.1% level of significance. Relatively strong negative associations between rates of *Correct response*, *Speech* and *Gesture/Pointing* and ADOS-G Diagnostic totals were found, along with strong positive associations between rates of *Speech* and NVDQ scores.

[place Table 1 about here]

Inter-rater Reliability

Inter-rater reliability was calculated using intra-class correlation coefficients (ICC), based on the frequency for each item observed by each rater in each session. The ICC and p-values and the mean total frequency are shown in Table 2. ICC values were 0.78 (all $p < .001$) or above for all but one item. The one item with a lower ICC value, *Request social routine* (ICC = 0.59, $p < .01$), had a mean observed frequency of 0.71 (SD = 1.43) per 15 minute session, and had the lowest variance of all items. Landis and Koch (1977) characterise values of reliability coefficients between 0.61 and 0.80 as "substantial" and those above 0.80 as "almost perfect".

[place Table 2 about here]

Concurrent Validity

Initiated Speech/Vocalisation. ADOS-G item A1 (*Overall Level of Non-Echoed Language*) has 5 possible scores, ranging from "No words or word approximations" to "Regular use of utterances with two or more words". As ADOS-G item scores are severity scores, with higher values relating to greater severity or autistic symptomatology, scores for this and subsequent items were reverse coded, so that higher scores relate to better language ability. Rates of *Initiated Speech/Vocalisation* and ADOS-G speech category scores were both significantly correlated with concurrent NVDQ scores (Pearson's $r = 0.29$, $p < .01$, and $r = 0.51$, $p < .001$, respectively). In order to control for the underlying influence of children's nonverbal cognitive

abilities Pearson's correlations between rates of *Initiated Speech/Vocalisation* and the ADOS-G speech category scores were performed, with NVDQ scores partialled out. There was a moderate and highly significant association between *Initiated Speech/Vocalisation* and ADOS-G speech category ($r = 0.58$, $p < .001$, $df = 88$).

Echolalia. *Echolalia* is indicated in COSMIC by the combination of *Passive compliance* and *Speech* in an interaction. ADOS-G item A4 (*Immediate Echolalia*) has 4 codes relating to the use of immediate echolalia during the assessment, ranging from "Rarely or never repeats others' speech" to "Speech largely consists of immediate echolalia". There is also a category for children whose spoken language is too limited to make a judgement. For the purposes of this analysis, scores relating to this latter category were treated as equivalent to the "Rarely or never" category. Both COSMIC and ADOS-G echolalia variables were significantly associated with NVDQ scores ($r = 0.34$ and 0.48, respectively, both $p < .001$). Pearson's correlations between the COSMIC and ADOS-G variables for echolalia, with NVDQ partialled out, showed a moderate and significant association ($r = 0.34$, $p < .01$, $df = 88$).

Gesture/Pointing. The 2 relevant items from the ADOS-G, A7 (*Pointing*) and A8 (*Gestures*), both have a range of scores relating to the complete absence or regular and flexible use of the behaviour in question. For the purpose of this analysis, scores from both items were added together. Rates of *Gesture/Pointing* and combined ADOS-G *Pointing* and *Gestures* scores were significantly correlated with NVDQ ($r = 0.32$, $p < .01$, and $r = 0.43$, $p < .01$, respectively). Pearson's correlations between rates of *Gesture/Pointing* from COSMIC and combined ADOS-G *Pointing* and *Gestures* scores, with NVDQ partialled out, showed a small but significant correlation ($r = 0.26$, $p < .05$, $df = 88$).

Comment. ADOS-G item B10 (*Spontaneous Initiation of Joint Attention*) contains 3 codes relating to the flexible use of eye contact to reference an object to share interest, partially referencing an object, or the lack of any observed referencing. Rates of *Comment* from COSMIC and ADOS-G joint attention scores were significantly correlated with NVDQ ($r = 0.29$, and $r = 0.32$, both $p < .01$). Rates of *Comment* from COSMIC and the ADOS-G joint attention score were not significantly associated whether or not NVDQ was partialled out ($r = 0.12$, n/s, and $r = 0.03$, n/s, $df = 88$).

Predictive Validity

The mean ALCC score for the 28 children in the Effectiveness of PECS study *No Treatment Group* was 3.4 (SD = 2.7) at baseline and 4.1 (SD = 3.0) at follow-up. A paired samples t-test showed that follow-up scores were significantly higher than at baseline ($t = -2.07$, $p < .05$, $df = 27$), indicating an increase in language and communication competence, as observed during the ADOS-G assessments, over time (mean = 15.3 months (SD = 0.74)).

Table 3 shows a correlation matrix of rates of the COSMIC items *Initiation*, *Initiated Speech/Vocalisation* and *Comment*, ALCC scores and NVDQ scores at baseline, and partial correlations between these COSMIC items and baseline ALCC scores, controlling for NVDQ. Baseline rates of each of the COSMIC variables were significantly associated with ALCC but not NVDQ scores at baseline. There were also significant correlations between ALCC scores and NVDQ scores. When NVDQ scores were partialled out, only *Initiated Speech/Vocalisation* was significantly correlated with ALCC score.

[place Table 3 about here]

ALCC scores at baseline and follow-up were significantly associated ($r = 0.78$, $p < .001$). Further correlations between the rates of baseline COSMIC variables and ALCC follow-up

scores were carried out, with baseline NVDQ scores partialled out. All 3 COSMIC variables were significantly associated with ALCC follow-up scores (See Table 4). The size of the correlations indicates that initial differences in rates of *Initiation*, *Initiated Speech/Vocalisation* and *Comment* contributed to approximately 18%, 38% and 17%, respectively, of the variance in follow-up ALCC scores. For the purpose of comparison, the same analysis was carried out with scores for initiation of and response to joint attention from the baseline ADOS-G assessments. When initial ALCC scores were partialled out, neither ADOS-G joint attention measure was significantly associuated with ALCC follow-up scores (r = 0.10 & 0.19, respectively, both n/s).

[place Table 4 about here]

Discussion

This paper reports on the systematic observation and coding of the communicative behaviour of children with autism who are nonverbal or in the early stages of language development. Data were collected in unstructured classroom settings and generally the frequency of communicative behaviours was low. Of all the items included in COSMIC only interactions with teachers occurred more frequently than once per minute and only *Correct response* and *Action* were observed to occur more than once every 2 minutes. Only about a third of the children interacted with another child in 30 minutes, and just 23 of the 91 children used their communication skills for the purpose of commenting. Children's rates of speech and use of gestures and pointing, but not use of picture symbols, were significantly associated with the degree of autistic symptomatology and nonverbal cognitive ability. The rate of responding correctly to the prompts and questions of others was negatively associated with the degree of autistic symptomatology.

In relation to the independent coding of videoed observations acceptable levels of inter-rater reliability were achieved for all but one item – the rarely-occurring *Request social routine* item. Concurrent validity was primarily investigated via comparison with scores from a semi-structured investigator-led assessment with specific item definitions. There was a significant degree of association for 3 of the 4 comparisons between items from COSMIC and those from the ADOS-G. These associations covered items that may index broad underlying communication competence (i.e. spontaneous speech and/or vocalisation) as well as those measuring very particular aspects of communicative behaviour (i.e. echolalia). However, these correlations were not so strong as to suggest that the use of an observational measure of social communication is redundant. This point is underlined by the fact that immediate treatment effects in the Effectiveness of PECS study (Howlin et al., 2007) were detected only for measures from COSMIC.

It may not be surprising that differences in these children's initial levels of spontaneous speech and vocalisation were predictive of levels of communicative competence at follow-up, but it is interesting that initial rates of commenting were also associated with outcome. The importance of measures of early declarative and joint attention abilities as predictors of later language skill is well established (Charman, 2003). Several longitudinal studies (e.g. Drew et al., 2007; Mundy et al., 1990; Sigman & Ruskin, 1999) employing fine-grained definitions of joint attention have also found initial levels of joint attention ability, particularly *responsive* joint attention, to be predictive of later gains in either expressive or receptive language ability. The present study demonstrates that broader measures of declarative function, based on observations of children in natural contexts, can also provide predictors of gains in communicative competence, albeit over a relatively short period. Of further interest is the fact that it was child

initiated comments that predicted their subsequent competence in *expressive* language. Moreover, for the same group of children, neither the initiation of nor response to joint attention, measured via the baseline ADOS-G assessments, predicted language competence scores at follow-up when initial levels of language competence were partialled out.

COSMIC was developed to provide ecologically valid outcome measures for the Effectiveness of PECS study. In that study immediate treatment effects were detected for dependent variables based on COSMIC (rates of *Initiations* and *PECS use*) but not for rates of *Speech/Vocalisation* and outcomes based on the ADOS-G, EOWPVT or BPVS. Pre- to post-intervention changes in communicative behaviours appeared to be clinically meaningful for children in the intervention groups: from 15 to 26 *Initiations* per hour and from 12 to 40 interactions involving *PECS use* per hour. This pattern of results suggests that the COSMIC variables, based on direct observations of children in everyday classroom settings, provided sensitive measures of those aspects of communicative behaviour targeted by PECS intervention. In contrast, standardised measures of language (EOWPVT, BPVS) did not prove useful for assessing the abilities of the children who participated in the Effectiveness of PECS study, all of whom were primarily nonverbal and of low cognitive ability.

There is a number of limitations to the present study. Firstly, we have no data regarding the stability of the behaviours rated using COSMIC. That is to say, each child was videoed only once at each time point. In order to ascertain the stability of each variable we would have needed to film each child on several occasions at each time point during each assessment period. This would have enabled us to determine how 'representative' a child's observed behaviour was of their 'typical' behaviour. Bruckner, Yoder, and McWilliam (2006) outline a procedure, known as a Generalizability study, to calculate the optimum number of observations required for a specific

variable, which takes into account stability across observations, as well as the agreement between raters for that variable. A prospective Generalizability study investigating the stability of COSMIC items is currently in progress.

Our investigation of inter-rater reliability also has limitations. Due to resource constraints, 2 of the raters were involved in the recruitment, randomisation, assessment and videoing procedures of the Effectiveness of PECS study and were therefore not blinded to child characteristics. However, the integrity of the findings relating to reliability were partially preserved by recruiting 2 additional raters who were unfamiliar with the participants and the specific purpose of the overall study. The percentage of all observations selected to be rated independently by both raters (22%) was an arbitrary figure. Walter, Eliasziw, and Donner (1998) provide a means for calculating the number of raters and observations required in reliability studies, based on prior or pilot data indicating the approximate reliability coefficient, and a given minimum acceptable level of reliability. Assuming a minimum acceptable reliability coefficient of 0.60, 2 raters and 35 observations would have been an optimal configuration for those variables with ICC values at or above 0.80 (13 of the 16 variables reported here). For items with ICC values below 0.70, suitable reliability estimates would require more raters, more observations or a lower level of acceptable reliability. For example, for a variable with an ICC value in the region of 0.60, and given an acceptable minimum reliability coefficient of 0.50, approximately 128 observations would need to be coded by 5 raters. If the acceptable minimum reliability coefficient was 0.40, then just 35 observations could be coded by 5 raters.

Only 4 items or combinations of items from COSMIC were investigated for concurrent validity. This was largely because concurrent validity could only be ascertained in relation to items from the ADOS-G that tapped the same concept or aspect of behaviour. All appropriate

comparisons between items from each measure were investigated. With regard to the investigation of predictive validity, the number of children whose data were included (N = 28) might be considered a small sample. Inclusion of the remaining children who were assessed at final follow-up would have resulted in more robust statistical values, but then the findings may have been affected by the fact that these additional children had received intervention as part of the Effectiveness of PECS study in the interim period. Furthermore, the period from initial to follow-up observation, approximately 15 months, does not constitute a particularly extensive period to assess change in language skills. Finally, our measure of language competence, the ALCC, is derived from scores of an assessment based on a semi-structured procedure, and the validity of using this variable as an index of language ability has not been established.

In conclusion, the present study demonstrates that the use of an ecologically valid measure of children's early social communication skills in a classroom setting can provide important information to supplement data obtained from semi-structured and/or standardised assessments of children's language. Adequate levels of reliability can be achieved using such a measure, and the training requirements for raters that are relatively modest compared to the commitment required to become a reliable administrator and rater of an instrument such as the ADOS-G. Given the desirability of measuring the intentional communication of children with autism in familiar everyday settings and with their regular communication partners, the use of observational coding schemes such as COSMIC for research and clinical purposes should be further explored and encouraged.

Word count: 4,888 excl Refs

References

Abrahamsen, E. P., & Mitchell, J. R. (1990). Communication and sensorimotor functioning in children with autism. *Journal of Autism and Developmental Disorders, 20*(1), 75-85.

Aldred, C., Green, J., & Adams, C. (2004). A new social communication intervention for children with autism: pilot randomised controlled treatment study suggesting effectiveness. *Journal of Child Psychology Psychiatry, 45*(8), 1420-1430.

American Psychiatric Association. (2000). *DSM-IV-TR: Diagnostic and Statistical Manual of Mental Disorders (4th ed.) - Text Revision*. Washington, DC: APA.

Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry, 39*, 879-891.

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

Bruckner, C. T., Yoder, P. J., & McWilliam, R. A. (2006). Generalizability and decision studies: An example using conversational language samples. *Journal of Early Intervention, 28*(2), 139-153.

Bruinsma, Y., Koegel, R. L., & Koegel, L. K. (2004). Joint attention and children with autism: A review of the literature. *Mental Retardation and Developmental Disabilities Research Reviews, 10*(3), 169-175.

Carr, D., & Felce, J. (2007). The effects of PECS teaching to phase III on the communicative interactions between children with autism and their teachers. *Journal of Autism and Developmental Disorders, 37*(4), 724-737.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society of London, Series B - Biological Sciences, 358*(1430), 315-324.

Charman, T. (2004). Matching preschool children with autism spectrum disorders and comparison children for language ability: Methodological challenges. *Journal of Autism and Developmental Disorders, 34*(1), 59-64.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Drew, A., & Cox, A. (2003). Predicting language outcome in infants with autism and pervasive developmental disorder. *International Journal of Language and Communication Disorders, 38*(3), 265-285.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*(1), 213-236.

Coggins, T. E., & Carpenter, R. L. (1981). The Communicative Intention Inventory - a System for Observing and Coding Children's Early Intentional Communication. *Applied Psycholinguistics, 2*(3), 235-251.

Drew, A., Baird, G., Taylor, E., Milne, E., & Charman, T. (2007). The Social Communication Assessment for Toddlers with Autism (SCATA): An instrument to measure the frequency, form and function of communication in toddlers with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37*(4), 648-666.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale - 2nd edition*. Windsor NFER-Nelson.

Flay, B. R., Biglan, A., Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., et al. (2005). Standards of evidence: criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151-175.

Frost, L., & Bondy, A. (2002). *The Picture Exchange Communication System Training Manual - 2nd edition*. Newark, DE: Pyramid Educational Products, Inc.

Harris, S. L., Belchic, J., Blum, L., & Celiberti, D. (1994). Behavioral assessment of autistic disorder. In J. L. Matson (Ed.), *Autism in Children and Adults: Etiology, Assessment, and Intervention* (pp. 127-146). Belmont, CA: Brooks/Cole Publishing Co.

Howlin, P., Gordon, R. K., Pasco, G., Wade, A., & Charman, T. (2007). The effectiveness of Picture Exchange Communication System (PECS) training for teachers of children with autism: a pragmatic, group randomised controlled trial. *Journal of Child Psychology and Psychiatry, 48*(5), 473-481.

Hwang, B., & Hughes, C. (2000). The effects of social interactive training on early social communicative skills of children with autism. *Journal of Autism and Developmental Disorders, 30*(4), 331-343.

Landis, J. R., & Koch, G. G. (1977). Measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Marans, W. D. (1997). Communication assessment. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of Autism and Pervasive Developmental Disorders (2nd ed.)* (pp. 427-447). New York, NY: John Wiley & Sons.

McConachie, H., Randle, V., Hammal, D., & Le Couteur, A. (2005). A controlled trial of a training course for parents of children with suspected autism spectrum disorder. *Journal of Pediatrics, 147*(3), 335-340.

McHale, S. M., Simeonsson, R. J., Marcus, L. M., & Olley, J. G. (1980). The social and symbolic quality of autistic children's communication. *Journal of Autism and Developmental Disorders, 10*(3), 299-310.

Mullen, E. (1995). *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Services.

Mundy, P., Hogan, A. E., & Doehring, P. (1996). *A preliminary manual for the abridged Early Social Communication Scales.* Miami, FL: University of Miami.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115-128.

Murdock, L. C., Cost, H. C., & Tieso, C. (2007). Measurement of social communication skills of children with autism spectrum disorders during interactions with typical peers. *Focus on Autism and Other Developmental Disabilities, 22*(3), 160-172.

National Initiative for Autism: Screening and Assessment. (2003). *National Autism Plan for Children*. London: National Autistic Society.

National Research Council. (2001). *Educating Children with Autism. Committee on Educational Interventions for Children with Autism. Division of Behavioral and Social Sciences and Education.* Washington, DC: National Academy Press.

Paul, R. (2005). Assessing communication in autism spectrum disorder. In F. R. Volkmar, R. Paul, A. Klin & D. J. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders (3rd ed.)* (pp. 799-816). Hoboken, NJ: John Wiley & Sons.

Sigman, M., & Ruskin, E. (1999). Continuity and change in the social competence of children with autism, Down syndrome, and developmental delays. *Monographs of the Society for Research in Child Development, 64*(1), 1-114.

Siller, M., & Sigman, M. (2002). The behaviors of parents of children with autism predict the subsequent development of their children's communication. *Journal of Autism and Developmental Disorders, 32*(2), 77-89.

Stahmer, A. C., & Ingersoll, B. (2004). Inclusive programming for toddlers with autism spectrum disorders: Outcomes from the children's toddler school. *Journal of Positive Behavior Interventions, 6*(2), 67-82.

Stone, W. L., & Caro-Martinez, L. M. (1990). Naturalistic observations of spontaneous communication in autistic children. *Journal of Autism and Developmental Disorders, 20*(4), 437-453.

Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal communication in two- and three-year-old children with autism. *Journal of Autism and Developmental Disorders, 27*(6), 677-696.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin & D. J. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders (3rd ed.)* (pp. 335-364). Hoboken, NJ: John Wiley & Sons.

Walter, S. D., Eliasziw, M., & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Statistics in Medicine, 17*(1), 101-110.

Watson, L., Lord, C., Schaffer, B., & Schopler, E. (1987). *Teaching Spontaneous Communication to Autistic and Developmentally Handicapped Children*. Austin, TX: Pro-ed.

Wetherby, A. M. (2006). Understanding and measuring social communication in children with autism spectrum disorders. In T. Charman & W. Stone (Eds.), *Social and Communication Development in Autism Spectrum Disorders: Early Identification, Diagnosis, and Intervention* (pp. 3-34). New York, NY: Guilford Press.

Wetherby, A. M., & Prizant, B. M. (2002). *Communication and Symbolic Behavior Scales - Developmental Profile. First Normed Edition.* Baltimore, MD: Paul H. Brookes.

Wetherby, A. M., & Prizant, B. M. (2005). Enhancing language and communication development in autism: Assessment and intervention guidelines. In D. Zager (Ed.), *Autistic Spectrum Disorders: Identification, Education and Treatment (3rd ed.)* (pp. 327-365). Mahwah, NJ: Lawrence Erlbaum Associates.

Wetherby, A. M., & Prutting, C. A. (1984). Profiles of communicative and cognitive social abilities in autistic children. *Journal of Speech and Hearing Research, 27*(3), 364-377.

Wetherby, A. M., Schuler, A. L., & Prizant, B. M. (1997). Enhancing language and communication development: Theoretical foundations. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of Autism and Pervasive Developmental Disorders (2nd ed.)* (pp. 513-538). New York, NY: John Wiley & Sons.

World Health Organisation. (1993). *ICD-10: International Classification of Diseases - 10th revision: Classification of mental and behavioural disorders.* Geneva: WHO.

Appendix

COSMIC
Classroom Observation Record Form

Child study number: **Observation/visit:**

Observer: **Date:**

Communicative function		**Spontaneity & Intentionality**		**Form**	
Q	request object	**I**	initiation	**P**	picture/symbol
W	request social routine	**C**	correct response	**S**	speech
D	comment, shared attention	**U**	unwanted/inappropriate	**V**	vocalisation
T	refusal/protest	**N**	non-interactive	**M**	signing
		Z	no response	**G**	gesture/pointing
		O	passive compliance	**X**	action

Context N.B. record onset/offset times		**Communication partner**	
K	snack	**1**	teacher/teaching assistant
A	group teaching/table activity	**2**	other adult/researcher
B	one-to-one session	**3**	other child
F	free play/unstructured		

	Context	**Partner**	**Function**	**Spon & Intent**	**Form (I)**	**Form (II)**	***Notes***
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							

Definitions of COSMIC items

Spontaneity & Intentionality	
Initiation	▫ Code initiation when the child spontaneously initiates an interaction. Initiation should not be coded when the communication partner clearly prompts the interaction verbally, physically, or otherwise ▫ If the child is using PECS or another symbol system, a communicative act is coded as an initiation if at least three seconds have elapsed since the adult placed a symbol on the PECS book, touched or moved the book as a prompt for the child to use a symbol ▫ Also code initiation when the child's response is an elaboration, contradiction or correction to the communication partner – e.g. The adult says, "There's your coat" and the child responds, "That's not my coat: *this* is my coat" (pointing to a different coat) Notes: ▫ In order for a child's initiation to be coded as such, it requires the presence, attention and/or proximity of an intended or potential communication partner as the recipient of that initiation. A child's communication must be directed to at least one other person, who is either looking towards the child, or close enough to hear or otherwise be aware of the communicative attempt ▫ Self-directed speech or vocalisations are not coded unless they are clearly part of an interaction with another person ▫ With the exception of PECS use, no specific time gap is specified to distinguish coded initiations from correct responses. The observer must make a judgement based upon the context of the communicative exchange whether or not the child's communicative act has been prompted by the communication partner ▫ In relation to initiations using PECS or other symbol systems the 'three second rule' does not apply if the child is clearly unaware of the teacher's replacement of a symbol or movement of the PECS book – i.e. if the child's back is turned to the book, or s/he is obviously distracted by the activity in which s/he is engaged
Correct response	▫ Code correct response when a child responds *appropriately* or complies with an instruction or prompt (e.g. the child sits after being told to "sit down") ▫ This code should be used if the content of the child's response is incorrect, but is appropriate to the context (i.e. during a work session the teacher instructs the child to find the *blue* marble, but the child picks up a *red* marble) Note: For children using PECS or other symbol systems this code should be used when the teacher replaces a symbol or places their book in front of them, and the child's next use of the symbol is within three seconds (unless the child is engaged in an activity or is unaware of the replacement, in which case code initiation)
Unwanted/ Inappropriate	▫ This code is used when a child's communicative act is clearly inappropriate, aggressive and/or disruptive Behaviours coded as unwanted/inappropriate may include throwing objects, hitting others, pushing furniture, loud screaming or *active* resistance to a request or instruction – e.g. falling to the floor in response to an instruction to sit
Non-interactive	▫ Non-interactive is used when the child responds to an approach by withdrawing, avoiding further interaction, or responding in a non-meaningful or stereotyped manner

	▫ This code may also be used to classify an *approach* by the child that is clearly not *interpersonal*, where for example they are attempting to take something from another person without looking at them or otherwise interacting with them Interactions coded as non-interactive do not require further classification in terms of form or communicative function
No response	▫ This code is used when the child does not respond in any way to a request, approach or prompt ▫ If no response is expected or necessary (e.g. the teacher says "good sitting", or otherwise comments on the child's actions) then no code should be given Interactions coded as no response do not require further classification in terms of form or communicative function
Passive compliance	▫ This code is primarily used when a child is being physically prompted (e.g. being led by the hand towards their seat, or completing a task via a hand-over-hand prompt) ▫ Use of this code requires that the child is *engaged* in the process, and should not be used if the child is being dragged, or if hand-over-hand prompting is being used, but the child is not attending to the task at hand Passive compliance is also used to classify examples of immediate echolalia, and should be paired with a code of speech. However if the repeated word is a response to a specific instruction this should be coded as correct response (e.g. the teacher hold up a biscuit and says, "Say biscuit. Biscuit", and the child repeats "biscuit")
Form	
Picture/Symbol	▫ The form of an interaction should be coded as picture/symbol if it involves the child giving or pointing to a symbol, picture, photograph, object of reference or other symbolic representation of an object, food item or activity Notes: ▫ If a child is communicating using PECS or another symbol-based system, and points to a symbol as part of the interaction the *form* should be coded as picture/symbol rather than gesture/pointing ▫ Manipulation of symbols that does not involve interaction with another person should not be coded – e.g. if a child has been prompted "Check your schedule" and then takes a symbol from his/her timetable and puts it in a 'finished' pocket ▫ If a child using PECS constructs a 'sentence strip' using a number of symbols, the interaction is coded when the strip is *handed over* to the communication partner, or *pointed to* communicatively
Speech	▫ Single words, short phrases and whole sentences should be coded as a single example of speech ▫ This may include word approximations and speech of poor intelligibility Notes: ▫ Where a child has used several forms of communication simultaneously, two of these forms can be coded. If speech or vocalisation is one of the observed forms, this should always be coded
Vocalisation	▫ Sounds that do not appear to have a speech-like quality, but that are being produced for apparently communicative purposes, should be coded as vocalisation. This may include crying, moaning or wailing, if used with some apparent communicative intent Note: ▫ Where the child's vocalisation is not intelligible as a recognised word, the ob-

	server should make a judgement based on perceived intonation and phonemic quality, as well as on the response of the communicative partner (who may be familiar with the child's idiosyncratic speech sounds, or aware of the immediate communicative context) ▫ If the distinction between speech and vocalisation is still ambiguous, code as speech
Signing	▫ Use this code for any use of sign language Note: If there is ambiguity as to whether a child's action is a formal sign or an iconic/descriptive gesture, code as gesture
Gesture/Pointing	▫ This code includes head nodding and shaking, pointing, descriptive, demonstrative or instrumental gestures
Action	▫ This code covers a range of behaviours, including sitting down, reaching, walking, putting a toy in a box, etc Any manipulation of symbols, pictures or photographs should be coded as picture/symbol
Communicative function	
Request for object/ snack	▫ Use this code for any communicative act where an object, toy, snack item, etc., is requested, whether this is spontaneously initiated by the child or prompted by an adult ▫ In cases where a request has been made, and the communication partner asks for a repeat or rehearsal of the request, do not assign this code for subsequent requests
Request for social routine	▫ This code is used when the child makes a request for a game or activity that is clearly social or interpersonal in nature – such as tickling, hugging or other informal social routines This code may be used when the request is for a formal game or activity, for a game of chess, for example, but not where the child is simply requesting that the adult facilitates an activity that will not involve them, such as switching the computer on, or reaching a toy that is on a high shelf
Comment	▫ Comment is coded when a child initiates joint attention verbally or non-verbally or spontaneously refers to an event, object or action. This may include a description of a picture, object or event ▫ If a child is clearly making a request for the object in question, albeit indirectly, code as a request for object Codes of comment should always occur in conjunction with a code of initiation, and a response to a direct question (e.g. "What can you see in this picture?") should not be coded as a comment
Refusal/Protest	▫ This code may used be used to classify a range of behaviours from appropriate refusal to inappropriate screaming as a protest. The specific nature of the behaviour will be made clearer by the accompanying codes for the interaction

Author Note

Greg Pasco, St George's, University of London, UK. Rosanna K Gordon, St George's, University of London, UK. Patricia Howlin, St George's, University of London, UK. Tony Charman, UCL Institute of Child Health, London, UK.

Greg Pasco is now at Autism Research Centre, University of Cambridge, UK. Rosanna K Gordon is now at Royal College of Psychiatrists' Research and Training Unit, London, UK. Patricia Howlin is now at Institute of Psychiatry, Kings College London, UK.

This study was funded by the Three Guineas Trust. The authors would like to acknowledge the children, their parents and the various professionals working with them who participated in this study. We would also like to express our gratitude to Philippa Moss and Anita Arora, who contributed to the investigation of inter-rater reliability. Some of the data presented in this article are based on work carried out as part of the doctoral thesis of the first author.

Correspondence concerning this article should be addressed to Greg Pasco, Autism Research Centre, University of Cambridge, Douglas House, 18b Trumpington Road, Cambridge CB2 8AH, UK; Tel: 44 (0)1223 746155; e-mail gp310@medschl.cam.ac.uk.

Footnotes

[1] ADOS module 1 *Communication* domain items contributing to ALCC score: A1 *Overall Level of Non-Echoed Language*; A2 *Frequency of Vocalization Directed to Others*; A7 *Pointing*; A8 *Gestures*

Table 1

Median rates per minute, interquartile range, number of children observed to display each behaviour and correlations with ADOS Diagnostic total scores and NVDQ scores at baseline

COSMIC Category COSMIC item	Median	Interquartile range	No. of children observed to display behaviour	Correlation with ADOS Diagnostic total	Correlation with NVDQ score
Communication Partner					
Teacher	1.77	1.27 – 2.53	91	- 0.16	0.01
Other adult	0.00	0.00 – 0.00	11	- 0.25	0.22
Other child	0.00	0.00 – 0.03	33	- 0.20	0.24
Spontaneity & Intentionality					
Initiation	0.20	0.07 – 0.40	73	- 0.25	0.25
Correct response	0.80	0.33 – 1.17	90	- 0.43 *	0.21
Unwanted/Inappropriate	0.00	0.00 – 0.13	42	0.07	- 0.13
Noninteractive	0.17	0.07 – 0.30	81	0.22	- 0.27
No response	0.07	0.03 – 0.17	77	0.07	- 0.03
Passive compliance	0.27	0.13 – 0.47	90	0.32	- 0.25
Form					
Picture/Symbol	0.20	0.00 – 0.90	80	0.10	- 0.08
Speech	0.03	0.00 – 0.80	46	- 0.40 *	0.42 *
Vocalisation	0.00	0.00 – 0.13	45	0.15	0.04
Signing	0.00	0.00 – 0.00	19	- 0.12	0.03
Gesture/Pointing	0.03	0.00 – 0.20	52	- 0.41 *	0.32
Action	0.70	0.43 – 1.03	91	0.04	- 0.16
Communicative function					
Request object/snack	0.17	0.07 – 0.37	83	0.01	- 0.04
Request social routine	0.00	0.00 – 0.03	25	- 0.18	0.08
Comment	0.00	0.00 – 0.03	23	- 0.22	0.29
Refusal/Protest	0.00	0.00 – 0.13	50	0.03	- 0.06

N = 91

* $p < .001$

Table 2

Inter-rater reliability: Intra-class correlations (ICC), p-values and mean total frequencies for each COSMIC item

COSMIC item	ICC	p	Mean frequency (SD) per 15-minute session
Spontaneity & Intentionality			
Initiation	0.80	< .001	3.70 (4.92)
Correct response	0.97	< .001	12.14 (12.76)
Unwanted/Inappropriate	0.88	< .001	1.75 (5.18)
Noninteractive	0.78	< .001	3.15 (4.74)
No response	0.78	< .001	1.75 (2.17)
Passive compliance	0.96	< .001	3.84 (3.53)
Form			
Picture/Symbol	0.82	< .001	2.14 (4.39)
Speech	0.97	< .001	5.23 (9.23)
Vocalisation	0.88	< .001	1.66 (4.98)
Signing	0.96	< .001	0.63 (1.80)
Gesture/Pointing	0.91	< .001	1.64 (3.11)
Action	0.91	< .001	13.53 (11.16)
Communicative function			
Request object/snack	0.94	< .001	3.55 (5.21)
Request social routine	0.59	< .01	0.71 (1.43)
Comment	0.80	< .001	0.53 (1.80)
Refusal/Protest	0.91	< .001	2.05 (5.46)

N = 40

Table 3

Correlations between COSMIC items, ADOS Language & Communication Competence (ALCC) scores and Nonverbal Developmental Quotient (NVDQ) scores at baseline and partial correlations between COSMIC items and ALCC scores, controlling for NVDQ scores (Pearson's r)

	Bivariate correlations (*Partial correlations controlling for NVDQ shown in parentheses*)			
	Initiated Speech/ Vocalisation	Comment	ALCC score at baseline	NVDQ score
Initiation	0.88 *** (*0.87* ***)	0.62 *** (*0.59* **)	-0.39 * (-*0.33*)	0.22
Initiated Speech/ Vocalisation		0.76 *** (*0.73* ***)	-0.55 ** (-*0.50* **)	0.28
Comment			-0.38 * (-*0.23*)	0.36
ALCC score at baseline				0.57 **

N = 28 df = 25 * p < .05 ** p < .01 *** p < .001

Table 4

Correlations between rates of COSMIC items at baseline and ADOS Language & Communication Competence (ALCC) follow-up scores, and with baseline Non-verbal Developmental Quotient (NVDQ) scores partialled out (Pearson's r)

	ALCC score at follow-up	ALCC score at follow-up with NVDQ partialled out
Initiation	-0.45 *	-0.42 *
Initiated Speech/ Vocalisation	-0.62 ***	-0.62 **
Comment	-0.55 **	-0.45 *

N = 28 df = 25 * $p < .05$ ** $p < .01$ *** $p < .001$