Early predictors of language and social communication impairments at 9-11 years: A follow-up study of early-referred children

Shula Chiat and Penny Roy

Language & Communication Science

City University London

Abstract

Purpose: To evaluate hypotheses that early sociocognition will predict later social communication and early phonology will predict later morphosyntax in clinically referred preschoolers.

Method: Participants were 108 children aged 9-11 years who had been referred to clinical services with concerns about language at 2½-3½ years. Predictors at Time 1 (T1) were measures of sociocognition, word/nonword repetition, and receptive language. Outcome measures at Time 3 (T3) included a social communication questionnaire completed by parents, and tests of nonword repetition, morphosyntax, and receptive language.

Results: Group- and case-level analyses revealed early sociocognition to be the strongest predictor of social communication problems, which by T3 affected almost a third of the sample. At group level, early phonology, which was a significant problem for the majority of children at T1, was a weak predictor of morphosyntax at T3. However, at case level, the majority of children with poor morphosyntax and nonword repetition at outcome had had very low repetition scores at T1.

Conclusions: In early language referrals, it is important to identify and address sociocognitive problems, a considerable risk for later social communication and ASD. The majority of early referred children had phonological problems, often severe, but these require further investigation to determine their longer-term significance for language.

It is a commonplace that children attain developmental milestones at different ages, posing a challenge for early identification of problems: at what point does late or slow attainment become a sign of long-term problems and impairment? In the case of language, the challenges are manifold. Since language is a complex system for mapping meaning intentions onto phonological structures, the child's ability to process meaning intentions, sound structures, and relations between these is critical (Chiat, 2001). Since children acquire the particular language to which they are exposed, language input is also critical. Hence, slow emergence of language can arise from a multiplicity of child-internal and/or external factors (Desmarais et al., 2008). It is therefore unsurprising that, as Brown pointed out back in 1973, onset and rate of early language development vary widely (Flax et al., 2009) and that many 'late talkers' and children referred to clinical services with concerns about language turn out to be 'late bloomers'. In follow-up studies recruiting children identified as late talkers (at 2 years) or by clinical referral (at 4 years), using different initial and outcome measures, the findings are remarkably similar: with each year from 2 to 5 years, roughly half the children with delay move into the normal range (Bishop & Edmundson, 1987; Dale et al., 2003; Paul, 1996; Rescorla, Dahlsgaard, & Roberts, 2000; Whitehurst & Fishel, 1994). Nevertheless, at a group level, significant correlations are found across time, indicating that children with language delay are at risk of weak language skills. While important and useful in their own right, these robust research findings offer little guidance to clinicians responsible for assessing and making decisions about individual children. How does the clinician distinguish the 'late bloomer' from the preschooler at risk of language impairment?

A growing body of research has attempted to address this well-recognised problem by identifying potential predictors of outcome and evaluating their validity at a case rather than group level. Again, studies vary in age of recruitment, sampling (population, late talkers, clinically referred children), initial and outcome measures (parent report of vocabulary and/or grammar, measures of utterance length or complexity, receptive and expressive language, narrative). Again, outcomes are largely consistent, considering the variation in participants and methodologies. In line with group level results, prediction is poor particularly for children with identified risk: while normal performance is a relatively reliable predictor of normal outcome, the 'normalisation' observed in group studies is reflected in weak prediction at case level. In an evaluation of parent report measures (covering children's vocabulary, grammar, reference to past and future, and nonverbal ability) as predictors of four-year-old outcome for two-year-olds with early language delay, Dale et al. (2003) conclude that 'the accuracy of prediction from 2-year measures was too poor to be of practical utility in discriminating persistent and transient difficulties' (p.555). Inclusion of further factors (parental concern and ear infections) had little effect on results.

In the light of these outcomes, it is not surprising that research findings play a very indirect role in clinical decision-making. According to reports and analysis of clinicians' judgements in both the USA and UK, decisions are based on a combination of informal observations of verbal and nonverbal behaviours and clinical evaluation of performance on standardised measures (Emanuel, Chiat, & Roy, 2007; Roulstone, 2001; Records & Tomblin, 1994). Clinicians work on the assumption that wide-ranging social, cognitive and linguistic skills are important foundations for 'typical' language and communication outcomes and potential evidence of longer-term risk, and draw on the breadth of their knowledge and experience to evaluate these.

The study of predictors that we report takes up this assumption, turning attention from measures of language itself to measures of very early skills that are known to underpin the development of language and communication. Recall our initial observation that language acquisition depends on the child's ability to process meaning intentions, sound structures, and relations between these. Taking up the first of these abilities, we hypothesise that children who have early difficulties engaging with other people and recognising their intentions will have difficulties understanding the purpose of communication and determining the meaning behind people's words, affecting their development of language and its use. We refer to skills in engaging with other people and inferring intentions behind their actions and words as *sociocognitive skills* and our hypothesis as the *sociocognitive hypothesis*. The skills we target have been investigated in studies of language development in typically developing children (Brooks & Meltzoff, 2008; Carpenter, Nagell, & Tomasello, 1998) and children with ASD (Charman et al., 2005; Toth, Munson, Meltzoff, & Dawson, 2006), but rarely in children with language delay. One exception is the research of Thal and colleagues (Thal, Tobias, & Morrison, 1991; Thal & Tobias, 1992) who investigated use of communicative gestures in late talkers and found an association with receptive language abilities. In line with our approach, Desmarais et al. (2010) draw attention to the heterogeneity of children with early vocabulary delay and highlight the occurrence of difficulties with comprehension and engagement in communication as well as expressive language.

Taking up the second of the key abilities identified above, we hypothesise that children who have early difficulties processing and storing the sound patterns of language will have difficulty with lexical phonology and morphosyntax, affecting their acquisition of words and sentence structures. We refer to this as the *phonological hypothesis*. The phonological skills that we consider have been studied extensively in school-age children (Conti-Ramsden, Botting & Faragher, 2003; Ellis-Weismer et al., 2000; Gathercole, 2006). In contrast, they have received little attention in studies of early language delay and early referred children, though some studies have highlighted the rate of *speech* difficulties in these groups (Bishop and Edmundson, 1987; Pharr, Ratner, & Rescorla, 2000; Rescorla & Ratner, 1996).

Our approach and hypotheses are in line with the conclusion Desmarais et al. (2008) draw from their systematic review of the literature on late-talking toddlers: 'future studies must include children from a clinical sample and document the various communication variables involved in language delay that have to do with both expressive and receptive skills…to identify subgroups of children with a vocabulary delay at age 2 years with distinct profiles…' (p.385). Our predictions are also in line with profiles of older children with a diagnosis of language impairment. It is well recognised that children meeting criteria for Specific Language Impairment (SLI) form a heterogeneous group. While some conform to the 'typical SLI' profile with hallmark deficits in language and particularly morphosyntactic structure, others present with deficits in pragmatics and social communication with or without structural deficits (Bishop, 1998; Bishop et al., 2000; Botting & Conti-Ramsden, 1999). Although standard exclusion criteria for SLI rule out diagnoses of ASD and other developmental disorders, many children have concomitant difficulties that make differential diagnosis problematic, and overlaps between SLI and ASD have been the subject of considerable debate (Bishop, 2000; Bishop & Norbury, 2002; Botting & Conti-Ramsden, 1999; Leyfer et al., 2008; Williams, Botting, & Boucher, 2008). Based on our hypotheses, a key goal of our study is to identify predictors of these different profiles of language and social communication. If we can uncover precursors of these different outcomes, this would not only be helpful for early clinical decision-making and intervention, but might throw more light on the heterogeneous language profiles and trajectories observed in school-age children.

In order to investigate our hypothesised predictors, we assessed a large sample of children referred to speech and language therapy services at 2½-3½ years (T1). This sample is different from samples of late talkers who are selected on the basis of very limited expressive language at 2 years rather than clinical referral (for example, Paul, 1996; Rescorla, Dahlsgaard, & Roberts, 2000). Since children in our sample are older (up to 3½ years at referral) and their language has been of sufficient concern to prompt a referral for clinical assessment, as a group they are likely to have more persistent and severe difficulties than children in late talker samples. Nevertheless, as with late talker samples, a notable proportion of young clinically referred children ‘catch up’ with peers within one to two years of referral, and outcomes for those who do not will vary (Bishop & Edmundson, 1987). We would argue that this is a very important population in which to investigate clinical prediction, since it is the population whom clinicians are asked to assess and diagnose, and about whom clinical decisions must be made.

The cohort we recruited were assessed again at 4-5 years (T2), roughly 18 months after referral and initial assessment. Outcomes were broadly in line with our predictions (Chiat & Roy, 2008). Our measure of early receptive language, the auditory section of the Preschool Language Scales 3 (UK) (Boucher & Lewis, 1997), emerged as the best *all-round* predictor of language and communication outcomes. This replicates research revealing receptive language to be the strongest predictor of language outcomes, with important implications for early clinical assessment (Desmarais et al, 2008; Flax et al., 2009). However, predictions from receptive language measures are for broad outcome; they do not differentiate between deficits in language structure and social communication observed in older children. In line with our predictions of specific outcomes, at a group level we found that:

- Early sociocognition was the strongest predictor of later social communication.
- Early phonology was the strongest predictor of later morphosyntax.

This paper reports a further follow-up study (T3), when the children were age 9-11 years and approaching the crucial and challenging transition to secondary school. The goals of this follow-up were:

1. to investigate social communication, morphosyntax and receptive language outcomes seven years after the children were first assessed;
2. to determine whether T1 Auditory PLS continued to be the strongest all-round predictor of outcomes at T3;

3. to evaluate our sociognitive and phonological hypotheses against longer-term evidence through group- and case-level analyses of relations between T1 sociocognition and T3 social communication, and between T1 phonology and T3 morphosyntax.

Methods

Participants

The original T1 sample comprised 181 children (75.1% boys) referred to speech and language therapy services in the Greater London area. Criteria for inclusion were age 2;6-3;6 at time of referral; reason for referral was concern about language development (not speech); no report of congenital problems, hearing loss, oro-motor difficulties, and no diagnosis of autism or ASD; no concerns about nonverbal ability; English as first/main language.

Attrition at T2 was relatively low, with 163 children (74.8% boys) participating. Following ethical approval for our T3 study, letters were sent to parents and children inviting them to participate in the T3 follow-up. Children were only included if they and their parents gave informed consent. Within the resources available, we were able to reach just under two-thirds of the T2 sample. Four of these children were excluded from analyses on the grounds that they had nonverbal IQ scores below the 5^{th} percentile at T3 and at both earlier time points. Of the remaining 108 children, 94.4% had nonverbal IQ scores in the normal range (≥85) at T3, leaving six children who did not. These children were included because they had scores close to normal at one or more time points, and all had scores ≥80 at T3. The final sample had a mean age of 10;5, with SD 6.91 months. Parent questionnaires were returned for 93 of these children.

Given the level of attrition, the T3 sample cannot be assumed to be fully representative of the original sample. Comparison between the T1 children who did and did not participate in T3 revealed significant differences in T1 nonverbal IQ (means of 92.29 vs. 87.85) and T1 receptive language (means of 88.61 vs. 83.64), but T1 measures of expressive language, sociocognition and phonology did not differ significantly. The observed differences indicate that as a group the ‘retained’ sample were less severely impaired at the time of referral, but were equally at risk on our very early processing measures. Since T3 analyses are based on the performance of the T3 sample for both T1 and T3 measures, the sample provides a valid basis for evaluating hypothesised relations between these, albeit with reduced power. Where data are missing due to non-compliance, numbers of participants included in the analysis are specified.

This study was approved by the Research Ethics Committee of the School of Health Sciences at City University London.

Assessments

Nonverbal abilities were measured using the nonverbal subscales of the British Abilities Scales II (Elliott, 1996) appropriate for children’s age: at T1, the nonverbal composite comprising Picture Similarities and Block Building subscales; at T3, the average of Pattern Construction and Matrices.

Time 1

Three T1 predictors were evaluated: receptive language, sociocognition, and phonological processing and memory. The receptive language and phonological measures yield continuous scores. For case-level analyses using categorical scores, ROC analyses were applied to continuous scores and revealed that a cut-off of -1.5 SD produced the best levels of predictiveness. Our measure of sociocognition yields a composite score and details of cut-offs for this measure are presented below.

Receptive language was measured using the auditory subscales of the Preschool Language Scale – 3(UK) (PLS; Boucher & Lewis, 1997).

Sociocognition was measured using our novel Early Sociocognitive Battery (ESB). This assesses three sets of sociocognitive skills. In the *social responsiveness subtest*, the tester acts out a sequence of scenarios in which she expresses six feelings such as hurt and surprise, and children are scored for looks to her face. In the *joint attention subtest*, the tester shows the child a set of six plastic eggs and opens these to reveal an object inside; she then looks (and if necessary points) at a matched object to the side, front or back of the child. Children are scored for alternating gaze between tester and object and for following tester's gaze or pointing gesture. Finally, in the *symbolic understanding subtest*, the tester presents the child with six gestures, then six miniature objects, and finally six uses of pretend objects, and asks the child to choose one out of six objects that matches ('is the best one for') each demonstrated gesture, miniature, or pretend object.

In the case of the joint attention and symbolic subtests, we used data from typically developing samples to determine cut-offs for 'low' performance (below -1.5 SD) and 'borderline' performance (between -1.5 and -1 SD) for each age band. In the absence of normative data for social responsiveness, the third and last of the sociocognitive subtests, we derived cut-offs from the distribution of the clinic sample itself which showed no age effects. For the composite on the ESB, performance was classified as 'low' if children scored in the low category on at least two of the three sociocognitive tests; 'normal' if they scored in the normal category on at least two; and borderline in all other cases.

Full information on the protocol for the ESB, criteria for scoring, and cut-offs for categories of performance are available in Chiat and Roy (2006) and at: http://www.city.ac.uk/health/research/research-areas/lcs/veps-very-early-processing-skills/veps-assessments.

Phonological processing and memory were measured using the Preschool Repetition Test (PSRep; Seeff-Gabriel et al., 2008). This is a standardised test for children aged 2-6 years comprising 18 real words followed by 18 nonwords which children are asked to copy. Score is for total number of items repeated correctly. Performance on the PSRep was unexpectedly poor, with almost half the sample scoring at least 2 SD below the mean, substantially more than the proportion attaining such low scores on our other T1 measures, and many times more than the 2.5% found in the standardisation sample. This in part reflects the rapid development of repetition skills in the majority of children (evident in standardisation results). We give further consideration to our unexpected findings on PSRep when we evaluate the implications of our T3 findings for the phonological hypothesis (see Discussion below).

Time 3

Assessment at T3 targeted a range of language, literacy, social communication and attention skills. In this paper, we focus on the assessments that are directly relevant to evaluation of the sociocognitive and phonological hypotheses, measuring language and social communication outcomes at T3. Prediction of literacy and literacy-related outcomes and the role of attention will be evaluated and reported in future papers.

Broad language abilities were measured using the two receptive subscales of the Clinical Evaluation of Language Fundamentals – Fourth Edition UK (CELF-4(UK); Semel, Wiig, & Secord, 2006) which make up the score for receptive language. A cut-off of -1 SD was used for categorical analyses, since the aim was to identify children at risk of comprehension difficulties in the middle school years.

Social communication was assessed using parent ratings on the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005). This questionnaire is designed to assess autistic impairment on a quantitative scale across a range of severity, rather than provide an all-or-nothing diagnosis. It comprises 65 items and is standardised for the age range 4-18 years. Given its aims and scope, the SRS is well suited to evaluation of social communication outcomes predicted by our sociocognitive hypothesis. The full set of items yield a Total score which is transformed into a T-score (mean 50, SD 10). T-scores ≥76 are in the 'severe' range, and are strongly associated with a clinically diagnosable ASD, while T-scores of 60-75 are classed as 'mild to moderate', and are typical for children with mild or 'high-functioning' ASD. The SRS has proven reliability, and the cut-offs for 'severe' and ‘mild to moderate’ have been validated against expert clinician diagnoses of autism spectrum conditions. For our categorical analyses, we used a score of ≥63 as the criterion for our low social communication category. This was identified as the optimal cut-off from ROC analysis based on the clinical diagnosis of social communication problems and/or ASD in our sample at T3 according to parental report.

Phonological processing and memory were measured using the Children’s Test of Nonword Repetition (CNRep; Gathercole & Baddeley, 1996). This is only standardised up to age 8;11, so the cut-off for low performance was based on the distribution of raw scores in our sample. Scores below the first tertile in our sample were taken as ‘low’ (see justification for cut-offs in next section).

Morphosyntax was measured using a battery of three assessments, two newly developed for this study, and one developed by Conti-Ramsden et al. (2011). Together these yielded four scores which were used to derive a categorical composite for Morphosyntax.

The School Age Sentence Imitation Test-English 32 (SASIT-E32)

This test was created as part of a Cooperation in Science and Technology project set up to develop new tests in a range of languages for assessment of bilingual children (see http://www.bi-sli.org/). Children are asked to repeat 32 sentences targeting a range of morphosyntactic and syntactic structures including complex auxiliary structures, passives, object wh- questions, sentence complements, temporal and conditional sentence adjuncts, and object relative clauses. Sentences were audio-recorded and presented in a fixed randomised order on a laptop.

This test yields three accuracy scores: number of sentences fully correct (all words repeated in correct order); number of content words correct (maximum 121) whether or not correctly inflected; and number of function words correct (maximum 176). The Morphosyntax composite used in this study included the two word scores (content and function) on the grounds that these provide an acute probe of children’s knowledge of sentence structure and grammatical morphemes (see Seeff-Gabriel, Chiat, & Dodd, 2010; Polišenská, Chiat, & Roy, submitted). Inflections are a further potential source of information about morphosyntactic abilities, and well recognised as a source of difficulty for children with SLI. However, since English uses a diverse range of function words which are well represented in our target sentences, and these are easier to detect and score than inflections (which often constitute a single word-final segment), we judged the function word score to be an adequate measure of grammatical morpheme production. Interrater reliability for scoring of content words, function words and whole sentences based on 10% of the sample was high, with Cronbach's alphas of 1, 0.98 and 1 respectively.

Grammatical judgement

This novel assessment was based on an experimental task developed by McDonald (2008). Materials comprised 16 grammatically correct sentences exemplifying a range of structures and 16 ill-formed sentences derived from the correct version by omitting a preposition or determiner (e.g. Those boys play same game every day), omitting a plural, tense or aspect marker (e.g. Here are four cup on the table), or inverting the main verb rather than auxiliary verb to form a question (e.g. Drives the teacher a fancy car?). The 32 items were divided into two sets, A and B, such that each set contained equal numbers of correct and incorrect items, and only one version (either correct or incorrect) of each sentence. Items were presented on a laptop allowing touchscreen responses. This test yields a total score for number of sentences judged correctly (maximum=32), the third of the scores making up the Morphosyntax composite.

Full protocols for the above two tests are again available at: http://www.city.ac.uk/health/research/research-areas/lcs/veps-very-early-processing-skills/veps-assessments.

Past tense elicitation: PPT20 (Conti-Ramsden et al., 2011)

This is a short version of the past tense elicitation task developed by Marchman, Wulfeck and Ellis Weismer (1999). Children are presented with a picture of an activity accompanied by a sequence of sentences such as ‘This boy is walking. He walks every day. Yesterday he…?’. They are expected to complete the final sentence using the verb from the lead-in sentences (in this case ‘walk’) but in the past tense. While target items and pictures were those used by Conti-Ramsden, minor modifications were made to keep the presentation as simple and consistent as possible, with all lead-in sentences following the format ‘The *noun* is *verb*-ing. Every day he/she/it *verb*-s. Yesterday he/she/it…?’. This test yields a total score for number of correctly produced past tense forms of verbs (maximum=20), the fourth and final score making up the Morphosyntax composite.

Each test included two practice items to familiarise children with the task and provide feedback explaining why their response was right/wrong.

Normative data are not yet available for the newly created SASIT-E32 and Grammatical Judgement tasks, and given the changes in presentation of the PPT20, it was not appropriate to apply the normative scores provided in Conti-Ramsden (2011). As with the CNRep (see above), cut-offs for low performance were based on the distribution of raw scores. These were heavily skewed towards the top end, with a third of the children scoring at ceiling on every measure (see Results). This is to be expected for basic morphosyntactic tasks on which typically developing children score at or close to ceiling (as found by Conti-Ramsden et al., 2011). Histograms revealed a significant drop at the first tertile, and on these grounds, scores below the first tertile cut-off were taken as ‘low’.

The construction of our Morphosyntax composite, combining content word and function word scores from SASIT-E32, Grammatical Judgement total, and Past Tense total, is supported by high correlations ranging from 0.56 (between SASIT-E32 function words and Grammatical Judgement scores) to 0.81 (between SASIT-E32 content and function word scores). Performance on the morphosyntactic battery was classified as 'low' if children scored in the low category on at least two of the four measures; 'normal' if they scored in the normal category on at least three; and borderline in all other cases. A cut-off between 'low' and 'borderline/normal' was used for categorical analyses.

Procedure

Following receipt of consent forms, parents were sent the social communication questionnaire (SRS). They were also sent a questionnaire covering background medical educational information about their child. This included questions about clinical diagnoses and speech and language therapy received since T2.

Children were assessed in a quiet room at school or at home according to parental preference, in two sessions each lasting 1½-2 hours. Assessments were administered by three qualified Speech and Language Therapists who had not participated in previous phases of the research and were blind to the performance of children at T1 and T2. The order of assessments was fixed, and was designed to vary activities through each session. In addition, sets A and B of the grammatical judgement task (see above) were split between the two sessions. All tasks were scored online, apart from SASIT-E32, which was audiorecorded for checking of online scoring and interrater reliability.

Results

Table 1 shows the mean, standard deviation, minimum, and maximum standard scores on T1 and T3 measures apart from the ESB, together with the proportion of scores in the 'low' category according to cut-offs used for all measures.

Performance of T3 sample at T1

At T1, the mean score for receptive language was just within the normal range, and just over a third of the sample scored in the low category (below -1.5 SD). Turning to our early measures of sociocognition and phonology, about a quarter of the sample fell in the low category on the ESB, compared with over two-thirds on the PSRep.

INSERT TABLE 1 ABOUT HERE

Children's binary categorical scores on the ESB and PSRep (low vs borderline/normal) were used to determine whether children in the T3 sample had been low on both, one or neither at T1. As shown in Figure 1, just over two-fifths of the children performed above the cut-off on both; just under a third were low on phonology only, compared with about one twentieth low on sociocognition only; and a fifth were low on both measures. As pointed out in our T2 follow-up study (Chiat & Roy, 2008), the small proportion of children with a 'pure' sociocognitive profile may be due to the exclusion of children diagnosed with ASD, and also the likelihood that children who have fluent speech are not referred to speech and language therapy services even if their social interaction or comprehension are limited.

INSERT FIGURE 1 ABOUT HERE

Performance of T3 sample at T3

Results at T3 are shown in Table 1. As this table shows, the group mean for receptive language was in the low average range, with just under a third of the sample scoring in the low category (below -1 SD). For social communication, taking our cut-off of T-score ≥63 (based on ROC analysis as reported above), almost one third (30.1%) of the sample fell in our low category. For our measure of word-level phonology, the CNRep, the mean score was low, equivalent to a standard score of 75 for the *oldest* group in the CNRep standardisation sample (8;0-8;11) who were *younger* than children in our T3 sample. In contrast, mean scores on our four morphosyntactic measures were close to ceiling (see 'Assessments' section above). Taking our cut-off of two scores in the bottom tertile on at least two measures, just over one third fell in the low category for Morphosyntax.

Children's categorical scores on the SRS, the Morphosyntax Battery and Receptive CELF were used to explore language and social communication profiles at T3. Figure 2 shows the distribution of children who were in the low category on both SRS and language measures (Morphosyntax in 2a, Receptive CELF in 2b), on SRS only, language only, or neither. Together, these pie charts illustrate that just over half the sample fell in our 'normal' category on both the SRS and at least one language measure, with the rest fairly evenly divided between children who were in the low category on both (17.6%, 16.1%), on SRS only (13.2%, 14%), and language only (17.6%, 15.1%).

INSERT FIGURE 2 ABOUT HERE

We also considered clinical outcome at T3, taking *receipt of clinical services* as the outcome measure in terms of whether or not children had on-going contact with Speech and Language Therapy services. This serves at the same time as a validation of the T3 outcome measures we adopted. Given limited clinical and educational resources, by 9-11 years only children who have severe difficulties in mainstream classrooms are likely to receive clinical support/intervention (though this will to some extent vary according to local policy and provisions). According to parent report, 15 children (15%) were currently receiving speech and language therapy (SLT) (see Table 1). As expected, there was a significant association between children receiving SLT and those with SRS and/or receptive language problems at T3 ($\chi^2(3)=35.93$, $p<.001$). Of the 54.8% of children with no problems (see Figure 2b), only one child was receiving ongoing SLT. Moreover, this child had a receptive language score of 85, the threshold for receptive language problems. Of those with problems, three-fifths with combined problems compared with about one fifth with either SRS problems only or receptive problems only were receiving SLT. The mean receptive language scores and SRS scores for children with problems and receiving SLT differed significantly from the mean scores for children with problems but not receiving SLT (Receptive CELF: $mean_{receiving\ SLT}=71.71$, SD=12.69; $mean_{not\ receiving\ SLT}=86.04$, SD=14.7; $t(40)=-3.11$, $p=.003$; and SRS: $mean_{receiving\ SLT}=76.14$, SD=18.47; $mean_{not\ receiving\ SLT}=64.36$, SD=17.22; $t(40)=2.04$, $p=.05$). As expected, then, children currently receiving clinical services were likely to have more pervasive and severe difficulties than those who were not.

Evaluation of predictors for outcomes at group level using continuous and composite scores

The heterogeneity of language and social communication profiles at and between time points was as we expected, and provides evidence to evaluate our hypotheses about early predictors of language and social communication outcomes. In order to evaluate the contribution of performance on T1 measures to T3 outcomes at a group level, logisitic regression analyses were conducted. Preliminary correlational analyses were carried out (see Table 2), and only those predictors that were significantly correlated with outcome were entered in the regression analysis. All three predictors were significantly correlated with Morphosyntax outcomes. However, only Auditory PLS and the ESB correlated with Receptive CELF and SRS, so PSRep was not included in regression analyses for these outcomes. As can be seen in Table 2, the association between Auditory PLS and ESB at T1 was significant, and both showed moderate and significant correlations with SRS at T3 (marginally but not significantly higher for SRS). Interestingly, correlations between Auditory PLS and ESB at T1 and Receiving SLT at T3, *seven years later*, were at least as large as *concurrent* correlations between language and social communication measures and Receiving SLT at T3.

INSERT TABLE 2 ABOUT HERE

As our main goal was to assess the relative strength of our T1 predictors rather than determine the most efficient model, we ran logistic regression analyses entering all our T1 predictors simultaneously (Table 3a). For these analyses, we adopted the cut-offs for our T3 binary outcome measures shown in Table 1. Two of our three T1 predictors, Auditory PLS and PSRep standard scores, were continuous. For the ESB, which yields a composite categorical score, we created dummy variables for low and borderline performance, taking normal as the reference group.

Table 3 reports the beta values (B), their standard errors (SE) and significance values, together with the odd ratios (OR) and their 95% confidence intervals (CI) for the T1 predictors entered; R^2 (Nagelkerke) and goodness-of-fit statistics for the total models are also included. Hosmer-Lemeshow tests confirmed that each full model we present was a ‘good fit’ of the data (nonsignificant χ^2 p values ranged from .29 to .68). Table 3a shows the full models for our four outcomes (Receptive CELF, SRS, CNRep, and Morphosyntax) with simultaneous entry of predictors. In the case of SRS outcome, both the ESB and Auditory PLS added significantly to the amount of change explained by the model (together accounting for over a third of the variance); in line with the sociocognitive hypothesis, the ESB added more than Auditory PLS. In the case of CNRep outcome, the PSRep was the only predictor to add significantly to the amount of change explained by the model, in line with our phonological hypothesis. However, contrary to this hypothesis and findings at T2, early phonology was not uniquely predictive of Morphosyntax at T3. In this case, the total model was much more predictive than any one T1 predictor taken on its own.

INSERT TABLE 3 ABOUT HERE

Following these regression analyses evaluating our sociocognitive and phonological hypotheses, we also considered predictiveness of T1 Auditory PLS, ESB and PSRep for our T3 clinical outcome measure Receiving SLT. Logistic regression with simultaneous entry revealed that T1 Auditory PLS and ESB were significant predictors of Receiving SLT for the sample as a whole, and together explained just under half of the variance (Table 3b). As can be seen, low ESB added significantly to the model, as did auditory PLS, though to a lesser extent.

Evaluation of predictor profiles for outcome profiles at case level using categorical scores

In order to determine the strength of these predictors at a case level, and compare prediction of 'low' vs 'normal' outcomes, six case-level measures were calculated: sensitivity (proportion of children with 'low' outcome who were 'low' on the predictor); specificity (proportion of children who were 'normal' on outcome who were 'normal' on the predictor); positive likelihood ratio (LR+, the ratio of the odds of a positive test result at T1 in individuals affected compared to those unaffected at T3, interpreted as small (2-4), moderate (5-10), or large (>10)); negative likelihood ratio (LR-, the ratio of the odds of a negative test result at T1 in those affected compared to those unaffected at T3, interpreted as small (0.2-0.5), moderate (0.1-0.2), or large (<0.1)); and positive and negative predictive value (PPV, the proportion of children with positive results who do have the disorder, and NPV, the proportion with negative results who do not).

Results are presented in Table 4. As this shows, Auditory PLS attained a similar level and balance of strengths across the four outcome measures, achieving acceptable levels of specificity (0.77-0.84), but weak sensitivity (0.5-0.57). Children with poor outcomes at T3 were two-to-three times more likely to have had low Auditory PLS scores at T1 compared with children without such problems at T3 (the likelihood being slightly higher for SRS and Receiving SLT than for Receptive CELF and Morphosyntax oucomes). The strength of the ESB as a specific predictor of SRS outcome seen in the regression analysis was most evident in high specificity (0.89) and LR+ of 5.31, indicating that a low ESB score was over five times more likely in children with poor SRS outcomes compared with those with good SRS outcomes. Positive and negative predictive values were both relatively high (0.7 and 0.83 respectively). However, ESB proved relatively weak on sensitivity (0.57) and LR- (0.48). A similar picture emerges when we take current receipt of clinical services as the outcome criterion for caseness, with both the Auditory PLS and the ESB faring well. Most strikingly, the ESB achieved good levels of sensitivity (.8) and specificity (.86), yielding moderate or close-to-moderate LR+ (5.67) and LR- (.23).

The group-level significance of the PSRep was apparent in acceptable levels of sensitivity for CNRep (0.91) and Morphosyntax (0.82), and in the case of CNRep, good LR- levels (0.22) and negative predictive values (0.91). Hence, children with poor nonword repetition and morphosyntactic outcomes were likely to have had problems with PSRep, and good PSRep indicates relatively low risk for later difficulties with nonword repetition. However, these strengths came at the expense of very weak specificity (0.41 and 0.38 respectively) indicating a high rate of false positives, and correspondingly poor levels of LR+. Thus, the predictive strength of the PSRep complements the strengths of the other two predictors which lay in their specificity.

INSERT TABLE 4 ABOUT HERE

Discussion

Our follow-up study was motivated by two hypotheses predicting relations between early profiles of phonological and sociocognitive skills and later profiles of language and social communication. Investigation of performance at 9-11 years, seven years after children were first referred, provides a strong test of these hypotheses. Starting with T3 outcomes, our sample shows the heterogeneous profiles that we expected and that underpinned our predictions. The notable proportion who had social communication problems at T3 (almost a third according to our cut-off of 63 on the SRS) may seem surprising, given that none of the sample were diagnosed with ASD at T1. However, this outcome corroborates evidence of social communication deficits in school-age children with SLI, and the overlap between SLI and ASD (Bishop & Norbury, 2002; Botting & Conti-Ramsden, 1999; Leyfer et al., 2008), which was a key motivation for the sociocognitive hypothesis. A similar proportion of the sample, almost a third, had language problems according to measures of receptive language or morphosyntax. Close to half of those with social communication problems and half of those with language problems were impaired on both, making up a seventh of the sample in total.

How effective were our early measures in predicting these T3 outcomes, and what light do our findings throw on the nature of school-age deficits?

Early sociocognition as a predictor of social communication outcomes

In line with the sociocognitive hypothesis and findings at T2 (Chiat & Roy, 2008), our Early Sociocognitive Battery emerged as the strongest predictor of social communication outcome. It achieved the best levels of specificity (0.89), and good PPV and NPV, with over two-thirds of positive results being true positives and four-fifths of negative results being true negatives. Since the ESB measures very basic skills in social responsiveness, joint attention and understanding of intentions that normally emerge prior to verbal communication and develop through the first and second years of life (Carpenter et al., 1998), our findings suggest that some of the children had primary deficits in sociocognition that contributed to, and may even be responsible for, their later social communication problems. No previous study has investigated this range of early sociocognitive skills in children referred with early language delay, so ours is the first to report the prevalence of sociocognitive deficits in this population and to identify links to later social communication problems (with or without wider language deficits). The inferences we have drawn from raise the possibility that some of these children may have ASD that was not diagnosed at an earlier age, perhaps because symptoms were mild and/or language difficulties were more conspicuous. This possibility invites further investigation.

In line with our original predictions (Chiat & Roy, 2008), the ESB was also a relatively strong predictor of language outcomes, reaching comparable levels of specificity to receptive language at case level. Poor ESB performance was more than twice as likely in children with poor receptive language and morphosyntax at T3 compared with children with no identifiable language problem. This outcome is consistent with previous evidence that another measure of sociocognition, gesture, is significantly associated with receptive language abilities (Thal, Tobias, & Morrison, 1991; Thal & Tobias, 1992), and with recent findings on the predictiveness of the Language Use Inventory (LUI; O'Neill, 2009; Pesco & O'Neill, 2012), a new parental questionnaire about children's realisation of pragmatic functions, for language and pragmatic outcomes. Research on late talkers and early referred children has until recently focused on deficits in vocabulary and morphosyntax and we would argue that more attention needs to be given to early sociocognitive deficits and to early interventions addressing these problems (see also Weiss & Theadore, 2011).

However, interpretation of the ESB must take account of findings on sensitivity and LR-, which fell well short of acceptable levels of prediction, missing just under half the children with social communication problems at T3. It could be that these children had sociocognitive difficulties that were not detected by the ESB but might have been evident in assessments of other sociocognitive behaviours (e.g. nonverbal imitation), or that stemmed from other problems (e.g. attention deficits). A further possibility is that social communication problems were late-emerging, unrelated to earlier difficulties, in at least some children.

Early phonology as a predictor of phonological and morphosyntax outcomes

In line with the phonological hypothesis, PSRep emerged as a signficant predictor of nonword repetition at T3. In contrast, and contrary to the phonological hypothesis, early PSRep proved a weak predictor of morphosyntactic outcome. Case-level analysis of categorical performance revealed the source of this weakness: while PSRep was the most likely of our measures to identify children with later deficits in morphosyntax and nonword repetition (yielding a sensitivity of 0.82 and 0.91 respectively), this came at the expense of low specificity indicating an unacceptably high rate of false positives. Our finding that early repetition had been a significant predictor of morphosyntax at T2 (Chiat & Roy, 2008) suggests that the phonological skills required for the PSRep are important for laying down basic phrasal structure of the language, but are far from sufficient for later morphosyntactic development. The contribution of phonology is further indicated by our finding that PSRep at T1 was significantly predictive of nonword repetition at T3 which was in turn highly correlated with morphosyntax at T3 (in line with well-established evidence of concurrent relations between nonword repetition and language and the status of nonword repetition as a potential marker of SLI in school-age children: see Gathercole, 2006 and Conti- Ramsden, Botting, & Faragher, 2001). It is nonetheless evident from our longitudinal results that our school-age morphosyntactic tasks (for example our School-Age Sentence Imitation Test eliciting complex structures such as sentence complements, temporal clauses, relative clauses) drew on semantic-syntactic aspects of sentence processing that go beyond those needed to establish the basic phrasal structures elicited in our preschool sentence imitation test.

Our findings on the extent to which PSRep over-predicted problems at both T2 and T3 have led us to re-consider the reasons for difficulties with this simple task of repeating words and nonwords. It is striking that the rate and severity of impairment on the PSRep were notably higher than on any other T1 measure: a full half of the sample scored below -2 SD, and two-thirds fell into our low category (cut-off <-1.5 SD). This, together with the relatively low risk associated with poor PSRep performance, points to the possibility that low PSRep scores may in many cases reflect relatively peripheral (even if severe) difficulties in phonological production, as opposed to difficulties with phonological processing and memory which the PSRep aimed to identify. Scoring of the PSRep makes allowances for systematic substitutions to take account of speech difficulties (see Seeff-Gabriel, Chiat & Roy, 2008), but severe delay in the development of speech planning/programming may give rise to limitations or errors in speech output that cannot be distinguished from errors due to deficits in phonological processing and memory. Our suggestion that phonological production problems may be a key factor in our T1 results on

PSRep is supported by previous evidence on speech production in late talkers. Pharr, Ratner and Rescorla (2000) and Rescorla and Ratner (1996) reported strikingly limited consonant and vowel inventories at 2 years and syllable structures at 3 years in late talkers, pointing to constraints on speech production that could be responsible for limitations in these children's expressive language. This evidence has received surprisingly little attention in the late-talker literature.

If, as we propose, very poor PSRep is in some cases due to severe limitations in the production of lexical phonology that impact on children's expressive language, this could account for early clinical referral of some children in our sample. Severe limitations in speech production that disrupt intelligibility and communication are highly conspicuous and a cause of concern for parents (who might fail to notice less conspicuous but ultimately more concerning comprehension problems). Our suggestion that very poor performance may be due to relatively peripheral though potentially persistent phonological production difficulties invites further investigation of input as well as output speech abilities, and how these relate to receptive and expressive language, in children with very poor repetition performance.

Turning to prediction of clinical outcome, children receiving clinical services at T3 were likely to have more pervasive and/or severe problems on our T3 measures. T1 Auditory PLS and ESB both contributed significantly to this clinical outcome, together accounting just under half of the variance. Not surprisingly, the proportion of variance explained was higher than for any of our T3 outcome measures taken separately, and case-level prediction reached acceptable levels.

Limitations

Sample attrition reduced the size and in some respects the representativeness of our T3 sample, impacting on the power of our T3 analyses. Nevertheless, the sample was sufficiently large to conduct relevant analyses, and significant associations were found.

Deciding on cut-offs for impairment is necessary to evaluate clinical predictiveness at a case level, but is always problematic. The measures we used were motivated by our hypotheses, were not always standardised, and yielded different types of score and composite scores. Decisions on cut-offs were made on the basis of ROC analyses where these were applicable, but where normative data were not available and/or measures produced composite scores, cut-offs were determined from distributions of performance in our sample. Clearly, results based on categorical performance that we report depend on the cut-offs selected, which must be taken into account in considering their implications.

Finally, we have pointed out above potential precursors of language and social communication that we did not assess; no doubt we have overlooked others that may play an important role.

Conclusion

Notwithstanding the limitations that we have identified, our investigation of the longer term implications of children's performance on the PSRep and ESB has thrown new light on the heterogeneity of early language referrals and their outcomes. The ESB revealed that a notable proportion of children had sociocognitive difficulties when first assessed although no child had a diagnosis of ASD at this age. These difficulties were significantly associated with long term social communication problems, which affected a surprisingly high number of children by 9-11 years. Given the importance of sociocognition for understanding communication and discovering meanings expressed by language, such early sociocognitive difficulties were expected to contribute to impairments in receptive language, with repercussions for language production and use. Indeed, the all-round predictiveness of early sociocognition for outcomes at 9-11 years was close to the predictiveness of early receptive language performance found at both T2 and T3. However, given the under-identification of later social communication problems, we conclude that the ESB, though valuable, is by no means sufficient, and there is clearly a need to investigate other possible predictors of deficits in social communication.

Our findings on morphosyntax outcomes highlight the multifactorial nature of these skills by middle childhood. Counter to our hypothesis, early phonology as measured by the PSRep was not a specific predictor of morphosyntax as it had been at 4-5 years; all our predictors played a role, but none uniquely. Nonetheless, case-level analyses showed that most children with problems in morphosyntax and nonword repetition at T3 had had problems with PSRep at T1. In the many cases where poor PSRep was not associated with longer-term problems, we suggest this might reflect serious difficulties with speech production that may have impacted on children's language production in the short-term.

Based on the results of our longer-term follow-up study, seven years after children's first referral, we conclude that our measures of early skills make a useful contribution to early assessment and identification of early needs. The ESB, in particular, enables systematic measurement of a range of sociocognitive skills and indicates risk for social communication difficulties that are increasingly recognized both clinically and in research. Further research investigating the nature of the widespread phonological problems we observed on the PSRep, and how input and/or output phonological difficulties impact on the development of lexical phonology and morphosyntax, will clarify further the implications of poor repetition abilities for longer term outcome and hence for early intervention. More generally, we conclude that tracking the outcomes of deficits in early skills known to underpin language development offers a promising way forward for understanding the heterogeneous outcomes observed in early referred children.

Acknowledgements

This research was funded by the Economic and Social Research Council, Award No. RES-000-23-0019 for the initial study, and Award No. RES-000-22-4093 for the follow-up. Many thanks to our research assistants Joanna Piper, Liz Elliman and Mia Travlos (follow-up study), and Talia Barry, Sophie Edgington, Renia Kaperoni, Luisa Martinez, Louise Occomore and Sharonne Williams (initial study). Thanks also to parents, teachers and children for their participation. We are grateful to anonymous reviewers for their valuable comments on an earlier version of this paper.

References

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.

Bishop, D.V M. (2000). Pragmatic language impairment: A correlate of SLI, a distinct subgroup, or part of the autistic continuum? In D.V.M. Bishop & L.B. Leonard (eds.), *Speech and Language Impairments in Children: Causes, Characteristics, Intervention and Outcome* (pp. 99-113). Hove, UK: Psychology Press.

Bishop, D.V.M., Chan, J., Adams, C., Hartley, J., & Weir, F. (2000). Evidence of disproportionate pragmatic difficulties in a subset of children with specific language impairment. *Development and Psychopathology, 12,* 177-199.

Bishop, D.V.M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156-173.

Bishop, D.V.M. & Norbury, C.F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: A study using standardized diagnostic instruments. *Journal of Child Psychology and Psychiatry, 43,* 917-929.

Botting, N. & Conti-Ramsden, G. (1999). Pragmatic language impairment without autism: The children in question. *Autism, 3,* 371-396.

Boucher, J., & Lewis, V. (1997). *Pre-school Language Scale-3 (UK Adaptation)*. London: Harcourt Brace & Company Books.

Brooks, R., & Meltzoff, A.N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language, 35,* 207-220.

Brown, R. (1973). *A first language: The early stages*. Cambridge, Mass.: Harvard University Press.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63,* (Serial no. 255).

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.

Chiat, S. (2001). Mapping theories of developmental language impairment: Premises, predictions and evidence. *Language and Cognitive Processes, 16,* 113-42.

Chiat, S., & Roy, P. (2008). Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry, 49,* 635-645.

Constantino, J.N., & Gruber, C.P. *Social Responsiveness Scale (SRS)*. Torrance, CA: Western Psychological Services.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, 42,* 741-748.

Conti-Ramsden, G., Durkin, K., Simkin, Z., Lum, J.A.G., & Marchman, V. (2011). The PTT-20: UK normative data for 5- to 11-year-olds on a 20-item past-tense task. *International Journal of Language and Communication Disorders, 56,* 243-248.

Dale, P.S., Price, T.S., Bishop, D.V.M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46,* 544-560.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking-toddlers. *International Journal of Language and Communication Disorders, 43,* 361-389.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., Rouleau, N. (2010). Three profiles of language abilities in toddlers with an expressive vocabulary delay: Variations on a theme. *Journal of Speech, Language, and Hearing Research, 53,* 699-709.

DSM-5 Development. Retrieved August 6, 2012, from http://www.dsm5.org/Pages/Default.aspx

Elliot, C.D. (1996). *BAS II: British Ability Scales*. Second Edition. Windsor, Berkshire: NFER-Nelson.

Emanuel, R., Chiat, S., & Roy, P. (2007). Evaluation of the clinical decisions made for 2-year-olds referred for speech and language therapy: A follow-up study. *International Journal of Speech and Language Disorders, 42*, 1-15.

Flax, J.F., Realpe-Bonilla, T., Roesler, C., Choudhury, N., & Benasich, A. (2009). *Journal of Learning Disabilities, 42,* 61-75.

Gathercole, S.E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27,* 513-543.

Gathercole, S.E., & Baddeley, A.D. (1996). *The Children's Test of Nonword Repetition (CNRep)*. London: Psychological Corporation.

Leyfer, O.T., Tager-Flusberg, H., Dowd, M., Tomblin, J.B., & Folstein, S.E. (2008). Overlap between autism and Specific Language Impairment: Comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule scores. *Autism Research, 1,* 284-296.

Marchman, V.A., Wulfeck, B., & Ellis Weismer, S. (1999). Morphological productivity in children with normal language and SLI: A study of the English past tense. *Journal of Speech, Language, and Hearing Research, 42,* 206-219.

McDonald, J.L. (2008). Grammaticality judgments in children: The role of age, working memory and phonological ability. *Journal of Child Language, 35,* 247-268.

O'Neill, D.K. (2009). Language Use Inventory: An assessment of young children's pragmatic language development for 18- to 47-month-old children [Manual].Waterloo, Ontario, Canada: Knowledge in Development.

Paul, R. (1996). Clinical implications of the natural history of slow expressive language development. *American Journal of Speech-Language Pathology, 5,* 5-20.

Pesco, D., & O'Neill, D.K. (2012). Predicting later language outcomes from the Language Use Inventory. *Journal of Speech, Language and Hearing Research, 55*, 421-434.

Pharr, A.B., Ratner, N.B., & Rescorla, L. (2000). Syllable structure development of toddlers with expressive specific language impairment. *Applied Psycholinguistics, 21,* 429-449.

Polisenska, K., Chiat, S., & Roy, P. (submitted). The contribution of linguistic knowledge to immediate verbal recall in children.

Records, N.L., & Tomblin, J.B. (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech and Hearing Research, 37*, 144-156.

Rescorla, L., Dahlsgaard, K., & Roberts, J. (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3;0 and 4;0. *Journal of Child Language, 27*, 643-64.

Rescorla, L., & Ratner, N.B. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research, 39*, 153-165.

Roulstone, S. (2001). Consensus and variation between speech and language therapists in the assessment and selection of pre-school children for intervention: A body of knowledge or idiosyncratic decisions? *International Journal of Language and Communication Disorders, 36*, 329-348.

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language & Communication Disorders, 45,* 691-702.

Seeff-Gabriel, B., Chiat, S., & Roy, P. (2008). *Early Repetition Battery*. London: Harcourt Assessment, Ltd.

Semel, E., Wiig, E.H., & Secord, W.A. (2006). *Clinical Evaluation of Language Fundamentals.* 4th edition. London: The Psychological Corporation.

SPSS (2007). Version 14.0. London: SPSS Inc.

Thal, D., & Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. *Journal of Speech and Hearing Research, 35,* 1281-1289.

Thal, D., & Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers: A 1-year follow-up. *Journal of Speech and Hearing Research, 34,* 604-612.

Toth, K., Munson, J., Meltzoff, A.N., & Dawson, G. (2006). Early predictors of communication development in young children with autism spectrum disorder: Joint attention, imitation, and toy play. *Journal of Autism and Developmental Disorders, 36,* 993-1005.

Weiss, A.L., & Theadore, G. (2011). Involving parents in teaching social communication skills to young children. *Topics in Language Disorders, 31,* 195-209.

Whitehurst, G.J., & Fischel, J.E. (1994). Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology and Psychiatry, 35,* 613-648.

Wiliams, D., Botting, N., & Boucher, J. (2008). *Language in autism and specific language impairment: Where are the links? Psychological Bulletin, 134,* 944-963.

Table 1: Mean, SD, minimum and maximum standard scores and proportion of children in low category for all T1 predictors and T3 outcomesT1 predictors (cut-off for low category)

	N	*Mean*	*SD*	*Min*	*Max*	*% (no.) in 'low' category*
Auditory PLS (<-1.5 SD)	106	86.42	17.40	50	127	31.1 (33)
ESB: Sociocognition (low on at least 2/3 tests)	108	-	-	-	-	25.9 (28)
PSRep: Phonology (<-1.5 SD)	108	71.48	16.85	50	130	69.4 (75)
T3 outcomes (cut-off for low category)						
Receptive CELF (<-1 SD)	108	91.77	17.56	55	131	30.6 (33)
SRS (T-score ≥ 63)	93	55.72	17.42	35	111	30.1 (28)
CNRep (<25/40)	108	25.89	5.8	9	36	31.5 (34)
SASIT content word (<119/121)	107	118.09	4.56	95	121	-
SASIT function word (<170/176)	107	168.63	14.07	47	176	-
Grammatical Judgement (<30/32)	108	29.42	3.28	15	32	-
Past Tense Test 20 (<16/20)	108	16.35	4.23	0	20	
Morphosyntax Battery (low on at least 2/4 measures)	106	-	-	-		35.8 (38)
Currently receiving SLT	100					15 (15)

Table 2: Correlations between T1 predictor variables and T3 outcome variables (shaded area of table)

	T1 predictors			*T3 outcomes*			
	ESB	PSRep	Recep CELF	SRS	CNRep	Morpho-syntax	Receiving SLT
T1 predictors							
Aud PLS	.56***	.49***	.36***	.43***	.26**	35***	.46***
ESB		.37***	.32**	.52***	.19*	.29**	.55***
PSRep			.09	.17	.39***	.28**	.21*
T3 outcomes							
Recep CELF	.			.32**	.29**	.39***	.46***
SRS	.	.			.22*	.31**	.48***
CNRep						.39***	.26*
Morphosyntax							.33**

$^{+}$T1 ESB and T3 outcomes are categorical scores

***p<.001; **p<.01; *p=borderline (.054, .053)

Table 3: Logistic regression analyses

(a) T1 predictors for specific T3 outcomes, simultaneous entry

T1 predictors				*95% CI OR*	
	B	*SE*	*OR*	*Lower*	*Upper*
	T3: Receptive language CELF				
Intercept	-3.82**	1.31			
Auditory PLS	.04*	.02	1.04	1.00	1.07
ESB low	.89	.61	2.44	.75	5.63
ESB borderline	1.12	.63	3.07	.9	10.51
	$X^2(3)=18.17$, p<.001, Nagelkerke R Square=22.3%				
	T3: Social Responsiveness Scale (SRS)				
Intercept	-4.67**	1.59			
Auditory PLS	.04*	.02	1.04	1.0	1.08
ESB low	2.04**	.67	7.7	2.08	28.5
ESB borderline	.71	.72	2.03	.5	8.31
	$X^2(3)=28.24$, p<.001, Nagelkerke R Square=37.6%				
	T3: Children's Test of Nonword Repetition				
Intercept	-3.72*	1.54			
Auditory PLS	.02	.02	1.02	.98	1.05
ESB low	-.1	.64	.9	.26	3.15
ESB borderline	-1.12	.76	.33	.07	1.45
PSRep	.06**	.02	1.06	1.02	1.11

	$X^2(4)=21.96$, $p<.001$, Nagelkerke R Square=26.2%

	T3: Morphosyntax				
Intercept	-4.72	1.44**			
Auditory PLS	.02	.02	1.03	.99	1.07
ESB low	.57	.6	1.76	.55	5.66
ESB borderline	1	.62	2.72	.81	9.1
PSRep	.02	.02	1.03	.99	1.0
	$X^2(4)=18.73$, $p=.001$, Nagelkerke R Square=22.6%				

(b) T1 predictors of T3 clinical status, simultaneous entry

	T3: Currently receiving SLT				
T1 predictors				*95% CI OR*	
	B	*SE*	*OR*	*Lower*	*Upper*
Intercept	-6.34**	2.44			
Auditory PLS	.09*	.04	1.09	1.02	1.17
ESB low	3.15**	1.16	23.24	2.4	225.05
ESB borderline	1.37	1.3	3.95	.31	50.4
	$X^2(3)=36.58$, $p<.001$, Nagelkerke R Square=54.2%				

Table 4: Sensitivity, specificity, positive and negative likelihood ratios, and positive and negative predictive values for T1 predictors and T3 outcomes

T1 predictor	*Sensitivity*	*Specificity*	*LR+*	*LR-*	*PPV*	*NPV*
	T3: Receptive language CELF					
Auditory PLS	.5	.77	2.18	.65	.48	.78
ESB	.42	.81	2.27	.71	.5	.76
	T3: Social responsiveness scale (SRS)					
Auditory PLS	.57	.84	3.6	.51	.62	.82
ESB	.57	.89	5.31	.48	.7	.83
	T3: Children's Nonword Repetition Test					
Auditory PLS	.5	.78	2.25	.64	.52	.77
ESB	.41	.81	2.18	.73	.5	.75
PSRep	.91	.41	1.53	.22	.41	.91
	T3: Morphosyntax					
Auditory PLS	.51	.79	2.46	.61	.58	.75
ESB	.4	.81	2.07	.75	.54	.71
PSRep	.82	.38	1.32	.48	.42	.79
	T3: Currently receiving SLT					
Auditory PLS	.8	.8	3.91	.25	.41	.96
ESB	.8	.86	5.67	.23	.5	.96
PSRep	.87	.31	1.25	.44	.18	.93

Figure 1: Distribution of ESB and PSRep profiles at T1

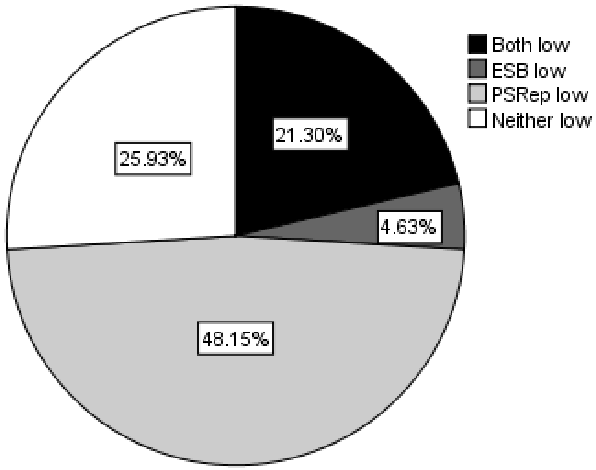

Figure 2: Distribution of SRS and (a) Morphosyntax, (b) Receptive CELF at T3

(a) Distribution of SRS and Morphosyntax (n=91)	(b) Distribution of SRS and Receptive CELF (n=93)