What our hands tell us:

A 2 year follow-up investigating outcomes in subgroups of children with language delay

Hilary O'Neill, Corresponding Author, Enable Ireland Early Services, Airbridge House, Dunshaughlin Business park, Dunshaughlin, Co. Meath, Ireland (honeill@enableireland.ie)

Carol-Anne Murphy, University of Limerick, Ireland

Shula Chiat, City, University of London, United Kingdom

Keywords: developmental language disorder, gesture, language delay, late talkers, outcomes, specific language impairment, symbolic comprehension.

Running head: Two-year follow-up, predicting outcomes of early language delay

Abstract

Purpose: This study followed up children identified with expressive language delay (ELD) or receptive-expressive language delay (R/ELD) at 2 years of age, time one (T1), in order to identify their language profiles at 4-5 years, time two (T2), and explore relationships to T1 language, gesture use and symbolic comprehension.

Method: Nineteen of 22 children were seen at follow-up (9/10 from R/ELD group; 10/12 from ELD group). T1 measures assessed receptive and expressive language, gesture use and symbolic comprehension. At T2 we assessed receptive and expressive language, sentence repetition, and expressive phonology.

Results: Outcomes for the R/ELD group were significantly poorer, with all children continuing to have delay in receptive and/or expressive language compared to just 20% of the ELD group. Expressive phonology delay was common in both groups. T1 receptive language showed the most pervasive correlations with T2 language measures, but categorical performance on all three T1 measures correctly predicted language outcomes in 16-17 of the 19 children.

Conclusion: Findings add to evidence that receptive language is a strong predictor of outcomes. Gesture use and symbolic comprehension are also strong predictors and clinically valuable as part of play-based assessments with implications for theoretical understanding and intervention planning.

In this paper we report on a follow-up of children with language delay two years after their initial clinical identification at 2-3 years with receptive/expressive (R/ELD) or expressive-only language delay (ELD). Our first study (O'Neill & Chiat, 2015) found performance on two measures of nonverbal sociocognitive skills, gesture use and symbolic comprehension, were significantly associated with receptive but not expressive language, with the R/ELD group scoring significantly below the ELD group on both. These findings supported interactionist theories proposing that new language forms build on social, motor and cognitive precursors, (Ambridge & Lieven, 2011), and were in line with evidence supporting such theories. Our follow-up study aimed to investigate the later language profiles of the R/ELD and ELD groups and to identify whether our early measures of gesture use and symbolic comprehension differentiated and might help with prognosis of longer-term outcomes.

Gestures are defined as actions used to intentionally communicate, expressed by the hands, facial expressions or body movements (Iverson & Thal, 1998). Deictic gestures appear early in development (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979) and can be split into two categories: contact actions, such as giving something to a person or pushing a hand away; and distal actions, which require no contact with another such as pointing/showing (McLean, McLean, Brady, & Etter, 1991). Representational gestures, sometimes termed symbolic or iconic, appear around 12 months of age (Acredolo & Goodwyn, 1988) and are used to represent an object or action. Symbolic comprehension refers to an understanding that symbols such as gestures, pictures, or objects can represent a meaning (Tomasello, Striano, & Rochat, 1999).

Gesture use and language development

The relationship between gesture use and language development has been attested across studies of children who have typically developing language and those in clinical groups. Bates and Dick (2002) reviewed the evidence demonstrating common underlying neural correlates for gesture use and language development and the co-emergence of gesture use and language milestones in early development. In typically developing children, the first symbolic gestures appear before and during the one word stage of development and express similar

meanings to words, for example, turning arms and palms over to symbolise ‘all gone’ (Acredolo & Goodwyn, 1988). The production of gesture-word combinations (e.g., pointing to a chair and saying ‘mommy’ thus requesting the mother to sit) accompanies or slightly precedes the transition from the single-word to two-word stage (Capirci, Iverson, Pizzuto, & Volterra, 2008). The frequency and diversity of children’s early gestures has been found to predict later vocabulary size and even more specifically the lexical items that will emerge (Acredolo & Goodwyn, 1988; Iverson & Goldin-Meadow, 2005). Accordingly, research in typically developing children not only points toward an association between early language and gesture, but suggests that later language milestones can be predicted by examining early gesture use (Namy & Waxman, 1998; Rowe & Goldin-Meadow, 2009).

Turning to clinical evidence, in children with Autism Spectrum Disorder (ASD), Ozcaliskan, Adamson and Dimitrova, (2016) found that deictic gesture at 30 months predicted vocabulary development a year later, and Yoder, Watson and Lambert (2015), that the use of intentional communication acts including non-conventional and conventional gestures, measured at 12-16 months, was related to receptive and expressive language growth at 2-4 years. Likewise, cross-sectional studies of children with language delay have reported significant correlations between gesture and profiles of receptive and expressive language (O'Neill & Chiat, 2015; Thal, Tobias, & Morrison, 1991; Thal & Tobias, 1992; Thal & Tobias, 1994), with different uses of gesture sequences and single gesture types related to single word production versus comprehension across a series of studies by Thal and colleagues (Thal, Marchman, & Tomblin, 2013).

Gesture use and prognosis

The prognostic potential of gesture invites further exploration given the well-established clinical challenge of differentiating transient from persistent impairment in late talkers, identified by limited expressive language at 2 years (Scheffner Hammer et al., 2017; Paul, Murray, Clancy, & Andrews, 1997; Rescorla, 2002), and in children presenting with early receptive and/or expressive language delay in the absence of a known cause (Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008). Variability in the persistence of early delay is

evident across epidemiological studies (Reilly et al., 2010) and studies of clinical groups (Bishop & Edmundson, 1987; Chiat & Roy, 2013). While 20% of two-year-old children may be identified with levels of expressive language that are below expectations (Zubrick, Taylor, Rice, & Slegers, 2007), as many as 60% (Dale, Price, Bishop, & Plomin, 2003) to 75% (Scheffner Hammer et al., 2017) will no longer meet the criteria for delay by 4-5 years. Samples of preschool children referred to clinical services are older (4 years in Bishop & Edmundson, 1987; 2½-4 years in Chiat & Roy, 2008) and might therefore be expected to have more severe and/or pervasive problems than those identified as late talkers (e.g. Rescorla, Dahlsgaard, & Roberts 1997; Scheffner Hammer et al., 2017). Follow-up of these clinic samples found that while 50% of children continued to have difficulties, up to half performed in the normal range 1-2 years later.

Population studies researching predictors of longer-term risk have largely focused on environmental, family and birth factors, with language abilities receiving limited consideration and, to our knowledge, none investigating nonverbal sociocognitive abilities as potential predictors. The finding that low socioeconomic status is one of the strongest predictors of late-talking and persisting language delay has important implications for intervention policies and provisions, however, its contribution to clinical prognosis and decision-making for individual clinically-referred children is less clear. Conversely, while separation of receptive language from expressive language may not be informative in population studies due to strong correlations between them (Scheffner Hammer et al., 2017), there is increasing evidence that delay in receptive language and on measures that go beyond language may be important indicators of risk at case level. According to Bishop, Snowling, Thompson, Greenhalgh, & CATALISE (2016), children with receptive language difficulties and poor use of gesture (as well as a family history of impairment) have poorer prognosis and less likelihood of recovery from early difficulties than children identified with expressive difficulties only. Desmarais et al.'s (2008) review of the literature on late-talking toddlers also highlighted the importance of assessing language comprehension, while a follow-up of study of nine late talkers (Thal et al.,

1991) found that the four late talkers who had not caught up were those with poorer language comprehension and gesture at first assessment one year earlier.

Other sociocognitive skills and prognosis

Along with receptive language, Chiat and Roy (2008, 2013) drew attention to the importance of sociocognitive skills – skills in early social engagement and understanding – for language development, and the possibility that these may be better indicators of long-term outcome than language itself. To assess sociocognitive skills, they administered the Early Sociocognitive Battery (ESB; Chiat & Roy, 2008) which comprises three subtests. The first, Social Responsiveness, scores children for the number and duration of looks to the assessor's face while she is expressing six feelings or emotional responses to an event, for example fear or surprise. The second subtest, Joint Attention, scores children for the number of times they alternate gaze between assessor and object, and follow the assessor's eye-gaze or finger-point, given 6 opportunities for each. The third subtest, measuring symbolic comprehension, was included in the present study and is described below. Follow-up of children referred with concerns about language at 2½-4 years (Chiat & Roy, 2013; Roy & Chiat, 2014) found that low performance on the ESB was the strongest predictor of social communication problems and contact with speech and language services seven years later, though receptive language was also predictive.

The present study

The few studies that have included measures of gesture use and other sociocognitive skills suggest that these may help with prognosis and may be informative for clinical decision-making regarding appropriate and timely provision of intervention particularly to those children with poor prognosis for recovery. However, few longitudinal studies have considered outcomes for children with early receptive versus expressive language delay and compared the predictive value of early language and nonverbal sociocognitive skills for these groups. The purpose of the current study was to add to the evidence base by following up the 2-3 year old children who participated in our cross-sectional study (O'Neill & Chiat, 2015) when they were 4-5 years and due to commence primary school.

Participants in the initial study were 22 monolingual English speaking 2-3-year-olds (mean 28.9 months, range 24-35 months) referred to local speech and language therapy clinics with concerns about language development, no history of hearing loss or repeated ear infections, and no identified learning disabilities, behavioural disturbances, neurological impairments or social/emotional impairments. Children were classified as presenting with receptive and expressive language delay (R/ELD, n = 10) or expressive language delay only (ELD, n = 12) using a cut-off score of -1 SD (standard score ≤ 85) on the relevant subscales of the PLS-3 (UK). Consistent with Zimmerman, Steiner, Pond, Boucher and Lewis (1997), we identified children with language scores falling below the average range. The study focused on children's use of gesture, i.e. their production of nonverbal symbols to refer to objects, actions and events. In addition, we assessed children's symbolic comprehension, i.e. their understanding that nonverbal symbols such as gestures or objects can represent a meaning, on the grounds that both draw on an underlying capacity for symbolic representation (Namy & Waxman, 1998).

Gesture use was assessed with two sections of the Communication and Symbolic Behaviour Scale (CSBS; Wetherby & Prizant, 2003). The Communicative Temptations section consists of eight structured situations that provide opportunities for children to communicate using gestures or vocalizations to comment or request continuation of the activity (e.g., wind-up toys, bubbles). The Sharing Books section allows the child to choose and examine a book, while the examiner shows interest in what the child points out or comments on. Two gesture scores were calculated: a distal gesture score which includes gestures made where the child's hand does not touch the object or person (e.g. pointing, reaching), or any symbolic gestures made by the hands (e.g. depictive gestures used to represent a function) and a conventional gesture score which is a measure of variety of gestures used socially (e.g. nodding and shaking the head, waving). These two scores were combined to yield a total frequency gesture score for each child.

Symbolic comprehension was assessed using the symbolic comprehension subtest of the ESB (Chiat & Roy, 2008), where gestures, miniatures, and pretend objects are used to

indicate which of six objects the child should roll down a chute. In the gestural condition, the tester mimes an action related to each target object (hammer, comb, toothbrush, bottle, sock and scissors); in the miniature condition, she holds up a miniature version of the target object (teddy, brush, book, shoe, spoon and t-shirt); and in the substitute object condition, she uses the substitute object as if it were the target object (cup used as a hat, banana as a telephone, stick as a crayon, shell as a plate, apple as a ball, brick as soap). For each item, the tester says 'Can you find the…?' presenting the gesture, miniature or substitute object, and then uses a sweeping hand gesture across the six objects saying 'Which is the best one? Which one goes best with this?'. The task therefore involves minimal verbal instruction. With six items in each condition, the maximum score is 18.

Our initial study found scores for both sociocognitive measures to be significantly associated with receptive language scores, and the ELD and R/ELD groups differed significantly with almost no overlap between groups. These findings prompted the current investigation of longer-term outcomes to determine whether our measures of early nonverbal sociocognitive skills would predict language abilities across time and would provide distinct and useful information to guide clinical decision-making and intervention.

Our follow-up study set out to address the following questions:

- Do children with receptive expressive language versus expressive-only language delay at 2-3 years show significant differences in language outcome two years later?
- Are T1 language, gesture and symbolic comprehension scores significantly correlated with T2 language scores?
- How accurately do measures of language, gesture and symbolic comprehension at T1 identify children with persisting receptive expressive vs expressive-only language delay at T2?

In addition, following our finding that better receptive language was associated with better gesture use (also reported by Thal et al., 1991) and with better symbolic comprehension, we explored the possibility that, in children with age-appropriate receptive language, expressive

language may be held up by severe delay in expressive phonology that affects lexical production, rather than a delay in lexical acquisition itself (Chiat & Roy, 2013). Expressive phonology is difficult to assess at 2-3 years and has not, to our knowledge, been investigated as a predictor of resolved vs persisting language delay. However, analyses of late talker data have identified limited consonant and vowel inventories and syllable structures at 2-3 years and have also suggested that constraints on speech production could be the reason for late talking (Pharr et al., 2000; Rescorla & Ratner, 1996). If children show a delay in expressive phonology at T2, presumably their expressive phonology was even more limited at T1, and this could account for delays in expressive language which resolve once expressive phonology is sufficient to produce recognizable words.

Method

Participants and Procedure

Of the 22 children (16 boys, 6 girls) assessed at time 1 (T1), 19 (14 boys, 5 girls) were available at time 2 (T2), 9/10 from R/ELD group and 10/12 from ELD group. Three families did not respond to the invitation to participate at follow-up. At T2, children's mean age was 4;9 (mean in months = 57, SD = 3.3, min 51, max 61). As at T1, there was a significant age difference between the two groups, with the R/ELD group (mean 58.5 months) older than the ELD group (mean 55.2 months) (Mann Whitney U = 18, p = .028). The T2 study was approved by the Health Service Executive Regional Ethics Committee, and all parents of participants provided informed written consent. Parents were also asked about their child's current educational status and receipt of speech and language therapy between T1 and follow-up. In the two-year period, all children from the R/ELD group received speech and language therapy averaging between 6 and 30 sessions. Six of 8 children in the ELD group received speech and language therapy averaging between 6 and 18 sessions.

Each participant was seen individually in a quiet clinic room accompanied by one or both parents, for sessions lasting 45-60 minutes. Assessments were administered in a set order over 2-3 sessions and all children completed the full assessment battery apart from one to whom the Sentence Imitation Test (SIT) was not administered due to fatigue. Assessments

were run by speech and language therapy students unaware of the children's original assessment status. They were trained in assessment administration and attended a further training day with the first author to ensure accuracy in assessment administration and scoring.

Assessments

Assessments carried out T1 and T2 are listed in Table 1.

INSERT TABLE 1 ABOUT HERE

For further details of T1 assessments, see O'Neill and Chiat (2015). At T2, we assessed language using the Auditory Comprehension and Expressive Communication subscales of the PLS-4 UK (Zimmerman, Pond, & Steiner, 2009) and classified children as having persisting delay in receptive and/or expressive language (PLD) if they scored ≤ -1.25 standard deviations below the mean (standard score ≤81) on the relevant subscales. This is consistent with international epidemiological studies (Reilly et al., 2010) and the finding that children with scores at this level have persistent language delay on longitudinal follow-up (Tomblin, 2008). Children were also assessed on a Sentence Imitation Test (SIT), from the Early Repetition Battery (Seeff-Gabriel, Chiat, & Roy, 2008), as a measure of expressive morphosyntax. The SIT consists of 27 sentences and yields several standard scores. In this study, children's responses were scored for whole sentence correct (all words in the target sentence repeated in the correct order, with no additions; maximum score = 27). Finally, for reasons explained above, we included an assessment of children's expressive phonology, the phonology subtest of the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd, Hua, Crosbie, Holm, & Ozanne, 2002) which has norms for Irish children. This provides standard scores with a population mean of 10 and standard deviation of 3. A score below 7 (< -1 SD) represents a phonological difficulty. Nonverbal ability was assessed using Raven's Coloured Progressive Matrices (RCPM; Raven, 2008).

Reliability

At T1, performances on the CSBS and ESB were scored by the lead researcher from video recordings and three of the 22 video recordings (13.6%) were randomly selected for reliability

analysis by a second registered speech and language therapist who had previous experience in scoring these assessments and was unaware of the children's classification. There was a high level of agreement for both by each rater (92.6 % for CSBS and 100% for ESB). The high rate of agreement on the latter reflects the clear-cut scoring criteria. At T2, two speech and language therapy students were present during administration of all standardized assessments. All assessments were scored live with scoring checked following each assessment. The DEAP assessment was both transcribed live and audio-recorded for later accuracy.

All assessments at both time points, bar the ESB, are standardized assessments with high levels of reliability reported in the test manuals.

Results

Table 2 presents descriptive data on T1 language assessments and T2 assessments according to T1 subgroup (ELD vs R/ELD).

INSERT TABLE 2 ABOUT HERE

Due to small numbers of participants in each group, nonparametric analyses were conducted using Mann-Whitney U Test for group comparisons and Wilcoxon Signed Rank Test for within-group comparisons across time. On our measure of nonverbal IQ at T2, the Raven's, all children scored above the cut-off of 70 for learning disability, with 15 in the normal range (≥ 85) (7/9 from the R/ELD group and 8/10 from the ELD group), and no significant difference between T1 ELD and R/ELD groups ($U = 35.5$, $p= 0.434$).

Language outcomes at T2

Comparing receptive language outcomes for the two groups, Table 2 shows the mean score for Auditory PLS increased in the R/ELD group but decreased in the ELD group. Although these changes across time were not significant (R/ELD group: $z= -1.68$, $p = 0.093$; ELD group: $z = -1.785$, $p = 0.74$), they narrowed the gap between groups observed at T1 and the group difference fell just short of significance at T2 ($U = 22$, $p=0.06$). On the other hand mean scores for expressive language, which did not differentiate groups at T1, showed a substantial increase in the ELD group from 73.9 to 90.7 ($z= -2.805$, $p = 0.005$), but no change in the R/ELD

group, resulting in a significant group difference at T2 (U = 0.06, p=0.002). Figure 1, showing the groups' mean scores for Auditory and Expressive PLS at T1 and T2, provides a graphic illustration of these changes in group language profiles. The two groups also differed significantly on the T2 Sentence Imitation Test (U = 13.5, p=0.01), with the ELD group gaining higher scores than the R/ELD group.

INSERT FIGURE 1 ABOUT HERE

Looking at language outcomes categorically, 8/10 children in the ELD group had receptive and expressive scores above our cut-off of -1.25 (SS > 81), of whom six fell in the average range (≥ 85) on both. The two remaining children scored below the cut-off for receptive language at T2. In contrast, all nine children in the R/ELD group scored below the cut-off of -1.25 in one modality (3 receptive only; 3 expressive only) or both (3). Thus, children in the R/ELD group had persisting delay, but varied in the scope of delay (see Discussion). A chi square comparing the distribution of outcomes (above vs not above the cut-off on both receptive and expressive language at T2) in the T1 ELD and R/ELD groups confirmed that these were significantly different (Fisher's exact p=0.001). The SIT showed a similar though non significant group difference, with just 2/8 children in the R/ELD group in the normal range at T2, (SIT data missing for one child), compared with 7/10 in the ELD group (Fisher's exact p=0.07); scores of the remaining children fell below -1.25 SD.

In contrast to findings for sentence level assessments, the groups did not differ on our measure of expressive phonology, the DEAP (U = 43.5, p=0.895). Using the DEAP cut-off for impairment of -1 SD (standard score <7) percent consonants correct, the majority of children in both groups had phonological impairment (R/ELD 6/9; ELD 7/10). Strikingly, though, phonological impairment on its own was confined to the ELD group (5/10 children); in the R/ELD group, it only occurred with persisting language delay.

Relations between gesture, symbolic comprehension and language at T1 and language profiles at T2

A key aim of our study was to explore whether early measures that related to language profiles at T1 would also relate to language profiles at T2, and might therefore support clinical

prediction and decision-making. To examine relations between T1 measures of gesture, symbolic comprehension and receptive and expressive language and T2 measures of receptive and expressive language and sentence repetition, we conducted Spearman's correlational analyses. Results revealed that T1 receptive language was significantly correlated with all three T2 language outcome measures, while T1 expressive language was only significantly correlated with sentence repetition (see Table 3). This is consistent with the findings from our group comparison and the considerable change in expressive language for the T1 ELD group displayed in Figure 1 above. Turning to our T1 nonverbal sociocognitive measures, gesture use was significantly correlated with expressive language and sentence repetition at T2, while symbolic comprehension was significantly correlated only with expressive language.

INSERT TABLE 3 ABOUT HERE

Based on the correlational results, our investigation of T1 measures as potential predictors focused on T1 language, gesture and symbolic comprehension. Since sample size precluded the use of regression analyses to investigate the relative contribution of T1 measures to T2 language outcomes, we considered how accurately classification of children on our T1 predictor measures (receptive language, gesture use, and symbolic comprehension) identified their classification on T2 measures of receptive and expressive language. We classified children at T2 as recovered with or without expressive phonology delay (recovered group, n = 8), or as having persisting delay in receptive and/or expressive language (PLD group, n = 11). To determine optimal cut-offs for T1 gesture and symbolic comprehension, i.e. cut-offs on gesture and symbolic comprehension that best distinguished children at T2 who did versus did not have a persisting language delay, we plotted children's scores on T1 gesture and symbolic comprehension against their T2 language classification. As Figure 2 shows, the optimal cut-off for gesture was 25: all 8 children in the recovered group scored above and 9/11 in the PLD group below this cut-off (Fisher's exact p=0.001). Interestingly, this cut-off for gesture yielded exactly the same division of children as T1 receptive language, the R/ELD group scoring 14-23 and the ELD group 26-61 for gesture use.

Turning to symbolic comprehension, the optimal cut-off score was five: 6/8 children in the recovered group scored above and 10/11 in the PLD group scored at or below this cut-off (Fisher's exact p=0.024). Thus, our three predictor measures achieved high and almost identical accuracy in classifying children (17/19 children classified correctly by measures of gesture use and receptive language, and 16/19 by symbolic comprehension). There were negligible differences in sensitivity (numbers of children with PLD correctly identified by the predictor measure: 8/8 for gesture/receptive language, 6/8 for symbolic comprehension) and specificity (numbers of recovered children correctly identified by the predictor measure: 9/11 for gesture/receptive language, 10/11 for symbolic comprehension).

INSERT FIGURE 2 AND TABLE 4 ABOUT HERE

Discussion

Our first research question in this small-scale follow-up study of children with receptive expressive versus expressive-only language delay at 2-3 years addressed their language outcomes at 4-5 years. Significant group differences on formal language measures of receptive and expressive language were found between the two groups. Consistent with previous studies, early receptive delay conferred a greater likelihood of persisting delay in the late pre-school/school-entry years (Chiat & Roy, 2008; Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2010; Thal et al., 1991; Scheffner Hammer et al., 2017). All nine children in the R/ELD group had persisting delay in receptive and/or expressive language at follow-up, compared with only two of the 10 children in the ELD group. However, group profiles had changed. The gap in receptive language that distinguished the groups at T1 had narrowed and fell just short of significance, but there was now a striking divergence in expressive language. While the R/ELD group showed no change on the expressive subscale of PLS-4, the ELD group's mean standard score increased from 73.6 to 90.7 and was now in the low average range. This is in line with previous reports that late talkers who 'catch up' typically perform at the low end of the normal range (Rescorla, 2002).

Within the R/ELD and ELD groups, however, language profiles became less consistent. Variation was most notable in the R/ELD group. One third of these children still had both receptive and expressive scores below our T2 cut-off of -1.25 SD; one third were below the cut-off on receptive but not expressive language; and one third on expressive but not expressive language. T2 profiles of the ELD group were more consistent, but unexpectedly, there was a slight decline in receptive language (from mean score of 97.5 to 92.4) and two children now scored below our -1.25 cut-off. This unexpected decline in receptive language is, however, in line with findings that late-talkers are at risk of weaknesses in language and other cognitive and social domains (Capone Singleton, 2018).

These findings add to evidence of different outcomes for early receptive vs expressive delay and the more pervasive difficulties experienced by children with receptive language impairment (Law, Campbell, Roulstone, Adams, & Boyle, 2008), while the heterogeneous profiles within groups are consistent with the reported failure to establish stable subgroups within language impairment (Bishop, Snowling, Thompson, Greenhalgh, & CATALISE-2 consortium, 2017). A number of extraneous factors could contribute to instability of individual profiles and heterogeneity of outcomes for the two groups in our study. Apart from incidental factors that might affect a child's performance on a particular day, we administered different versions of the PLS at T1 (PLS-3) and T2 (PLS-4), in line with requirements to use the most up-to-date version of a standardised test (Jakubowitz & Schill, 2008). Changes in group profile could therefore reflect changes in assessment demands. Demands also change with age and stage of assessment. Most notably, later sections of the receptive subscale of PLS-4 include phonological awareness tasks. Many children in the R/ELD group reached ceiling before these sections, but in the case of the ELD group, phonological awareness might account for the decline in the ELD group's receptive language scores and the puzzling finding that two children now met criteria for receptive language delay. Further assessment would be needed to confirm whether this is the case, or whether some children with ELD do not keep up with the growing language comprehension of their peers'. Another potential factor affecting language outcomes is receipt of intervention between T1 and T2. However, the T1 ELD group,

which showed more consistent catch-up with peers, had received considerably less intervention than the R/ELD group in the intervening time.

The most consistent and notable change in profiles was the increase in the ELD's expressive language. We had proposed that early expressive-only language delay might arise from a severe delay in expressive phonology and resolve once children's expressive phonology has advanced sufficiently to produce recognizable words (see introduction above; also Chiat & Roy, 2013). This proposal was motivated by our finding that use of gesture – which obviates speech production – was consistently higher in the ELD than the R/ELD group at T1, and by previous suggestions (Thal & Tobias, 1992) that children with expressive delay make use of gesture to circumvent their difficulties with verbal communication and are more likely to catch up with peers than children with receptive delay. We have no direct evidence of expressive phonology at T1 since assessment is problematic with minimally verbal children, but the majority of children fell in the impaired range on DEAP, our measure of expressive phonology at T2, and the ELD group did not differ from the R/ELD group on this measure. Assuming some catch-up in expressive phonology over the intervening period, it is likely that children's expressive phonology was even more delayed at T1 so this could have been responsible for their limited expressive language. If children had more problems beyond expressive phonology, for example difficulties in recognizing, storing and/or accessing words, we might expect less dramatic catch-up. However, further investigation of expressive phonology in children with expressive-only delay is needed to substantiate this interpretation.

Our second question addressed relations between our T1 measures of gesture use and symbolic comprehension as well as language, and our T2 measures of receptive and expressive language and sentence repetition. At T1, concurrent correlations between symbolic comprehension, gesture and receptive language had all been significant, while expressive language did not significantly correlate with any other measure. Relations with T2 measures diverged: T1 receptive language was significantly correlated with all T2 language measures, and most strongly with T2 expressive language. However, T1 gesture only significantly correlated with language production tasks (expressive language and sentence

repetition), and T1 symbolic comprehension only with expressive language. The finding that neither gesture nor symbolic comprehension correlated with receptive language was contrary to expectations, but measurement issues may again have played a role (see comments on the receptive subscale of PLS-4 above).

We had planned to distinguish receptive and expressive language outcomes in addressing our final question, how accurately T1 measures identified persisting delay. Given the heterogeneity of language profiles we found at T2, and mixed correlations between T1 measures and specific measures of language at T2, we classified children with persisting delay (PLD) whether this affected receptive or expressive language. Using this classification, T1 receptive language, symbolic comprehension and gesture use all achieved similarly high levels of accuracy, correctly accounting for 16-17 of the 19 children at follow-up.

Theoretical and clinical implications

Our study set out to investigate the prognostic potential of gesture use and symbolic comprehension and found that both measures correlated with later expressive language ability, however, early receptive language was most strongly correlated with all language outcomes and was the most accurate in identifying persisting delay. Our findings that strengths in early gesture use is associated with later 'catch up' in expressive language is in line with those from previous studies (Thal et al., 1991). We might infer that the rates of catch-up reported for late talkers and children with language delay in many studies reflect their focus on expressive language, with little attention to receptive language and the longer-term risk of receptive delay (Desmarais et al., 2008). This is not to overlook the risk of late talkers and children with expressive delay remaining at the low end of the average range for verbal skills and needing extra support with expressive phonology, phonological awareness, literacy, and increased language demands of school (Rescorla, 2002, Capone Singleton, 2018). It also leaves open the possibility that a small proportion of children with ELD have deeper problems with a poorer prognosis.

While similar levels of prediction were achieved by receptive language, gesture use and symbolic comprehension, our nonverbal sociocognitive measures did not improve

identification of ELD children who have persisting delay over assessment of receptive language. We would argue that these measures nevertheless make a valuable contribution to early clinical assessment and theoretical understanding of early language delay. Firstly, while traditional language assessments can identify delay in comparison to peers, they are limited in their ability to provide information on a child's functional communication skills, their communicative intent, how they communicate and the frequency, type and variety of means used by the child to compensate for their communication delay. These nonverbal sociocognitive measures have distinct potential in clinical assessment because they require understanding of communicative intentions, and in the case of gesture use, a desire to communicate, but they do not require words to convey meaning. These assessment tools not only provide a window into a child's underlying preverbal abilities but can provide valuable information to aid intervention (see below). Since children in the R/ELD group all had low scores on our nonverbal sociocognitive tasks, we infer that they had difficulties with communicative intentions whether conveyed verbally or nonverbally. Though these measures did not relate to receptive language at T2 as they had at T1, they may in the longer term throw light on changes in receptive and expressive language that we observed at T2, particularly in the R/ELD group. In this case, we predict they will have longer-term difficulties with communication (social communication problems) even if their receptive and/or expressive language delay resolves (see Roy & Chiat, 2014, for further discussion and evidence). Conversely, children in the ELD group who made good use of gesture, and in most cases gained higher scores for symbolic comprehension, are inferred to have good understanding of communication and development of communicative intent, implying that their delay in language stems from processing of the forms of language or linking these to meanings (Chiat, 2001). We have already argued that many of these children may only be delayed in their ability to *produce* words.

Secondly, our assessments of gesture use and symbolic comprehension are both child centred and incorporate engaging play-based activities in a structured but flexible format. The CSBS allows for naturalistic assessment of spontaneous communication but at the same time,

as a standardized assessment, provides norms. Clinically, traditional formal assessment with 2-year-olds is frequently problematic due to attention levels and variable interest in assessment materials and as a result cautious interpretation is required. Given that both our nonverbal sociocognitive measures readily engage 2-year-olds and matched the receptive language measure in identifying persisting delay, our findings build on previous studies that these assessments are a valuable clinical tool in gathering clinically useful information in a natural and child-focussed way (Thal et al., 1991; Crais, Watson, & Baranek, 2009). Finally, these measures may provide functional, clinically relevant information to aid the planning of intervention programmes in line with research by Meyers, McBride and Peterson (1996) which revealed that professionals found play-based assessments resulted in more useful information that could be directly translated into intervention.

Limitations, further research and conclusions

The limitations of our study are evident. Our sample was small, and subgroups even smaller, so results and interpretations must be treated with caution. While attempts were made to match the ELD and R/ELD groups at T1, there was a slight but significant age difference between our subgroups (though the age advantage of the R/ELD group did not confer an advantage at either time point). Other confounding variables such as socioeconomic status and family history of language delay were not examined and nonverbal cognitive skills were only assessed at T2 (with no difference between the two groups). At T2 only 19/22 children were available for follow up. It is possible that inclusion of the three children not available (one with R/ELD, two with ELD) could have affected the overall group outcomes. However, this is unlikely given the strong trend in those who did return. Additionally, we used different versions of the PLS across T1 (PLS 3) and T2 (PLS 4). Language comprehension and production are complex processes involving a range of skills, and receptive and expressive language assessments vary in the skills they target. Furthermore Consequently, studies with large samples and carefully targeted assessments are needed to unpack the contribution of early skills to later outcomes.

Nonetheless, our study adds to increasing evidence that children with expressive-only language delay are at relatively low risk of longer-term language delay, though likely to remain at the low end of the average range, and that children with delay in receptive language are at substantially higher risk. It supports the possibility that expressive-only delay may in many cases arise from delay in expressive phonology, but highlights different possible sources of delay in expressive as well as receptive language. We have argued that gesture use and symbolic comprehension, which rely on understanding and use of nonverbal symbols and were found to be linked to early receptive delay, are informative about children's functional communication, may guide intervention to support this, and may throw light on the nature of early language delays and their variable outcomes.

This study only followed the children up to their late preschool years. It is important to investigate the longer-term predictive value of gesture use for receptive and expressive language in the primary school years. In addition, large-scale longitudinal studies that include these measures of nonverbal sociocognitive skills, measures of expressive phonology, as well as receptive and expressive language are needed to clarify their contribution to outcomes.

REFERENCES

Acredolo, L., & Goodwyn, S. (1988). Symbolic Gesturing in Normal Infants. *Child Development, 59*(2), 450-466.

Ambridge, B., & Lieven, E. (2011). *Child Language Acquisition: Contrasting Theoretical Approaches*. Cambridge, UK: CambridgeUniversity Press.

Bates, E., Benigni, L., Bretherton, I., Camaioni, L., & Volterra, V. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic press.

Bates, E., & Dick, F. (2002). Language, gesture, and the developing brain. *Developmental Psychobiology, 40*(3), 293-310.

Bishop, D.V.M., & Edmundson, A. (1987). Language- Impaired 4 year olds. Distinguishing Transient from Persisting Impairment. *Journal of Speech and Hearing Disorders*, *52,* 156-173.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE-2 consortium. (2017). Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology & Psychiatry*. doi:10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE, C. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLoS One, 11*(7), e0158753. doi:10.1371/journal.pone.0158753

Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (2008). Gestures and words during the transition to two-word speech. *Journal of Child Language, 23*(3), 645-673.

Capone Singleton, N. (2018). Late Talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics of North America*, *65*(1), 13-29.

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*(6), 635-645.

Chiat, S., & Roy, P. (2013). Early Predictors of Language and Social Communication Impairments at Ages 9–11 Years: A Follow-Up Study of Early-Referred Children. *Journal of Speech, Language, and Hearing Research, 56*(6), 1824-1836.

Crais, E.R., Watson, L.R., & Baranek, G.T. (2009). Use of gesture development in profiling children's prelinguistic communication skills. *American Journal of Speech-Language Pathology, 18,* 95-108.

Dale, P. S., Price, T. S., Bishop, D. V., & 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*(3), 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 & Communication Disorders, 43*(4), 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.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology (DEAP)* London: The Psychological Corporation.

Fenson, L., Dale, P., Reznick, J., Thal, D., Bates, E., Hartung, J., . . . Reilly, J. (1993). MacArthur-Bates Communicative Developmental Inventories. *San Diego: Singular*.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture Paves the Way for Language Development. *Psychological Science, 16*(5), 367-371.

Iverson, J. M., & Thal, D. J. (1998). Communicative transitions: There's more to the hand than meets the eye. In A. Wetherby, S. Warren, & J. Reichle (Eds.), *Translations in prelinguistic communication* (pp. 59-86). Baltimore: Paul H. Brookes.

Jakubowitz, M., & Schill, M. J. (2008). Ethical Implications of Using Outdated Standardized Tests. *SIG 16 Perspectives on School-Based Issues, 9*(2), 79-83.

Law, J., Campbell, C., Roulstone, S., Adams, C., & Boyle, J. (2008). Mapping practice onto theory: the speech and language practitioner's construction of receptive language impairment. *International Journal of Language & Communication Disorders, 43*(3), 245- 263.

McLean, J. E., McLean, L. K. S., Brady, N. C., & Etter, R. (1991). Communication profiles of two types of gesture using nonverbal persons with severe to profound mental retardation. *Journal of Speech and Hearing Research*, *34,* 294-308.

Meyers, C.L., McBride, S.L., & Peterson, C. A. (1996). Transdisciplinary play-based assessment in early childhood special education: An examination of social validity. *Topics in Early Childhood Special Education*, *16,* 102-126.

Namy, L., & Waxman, S. (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference *Child Development, 69*, 295-308.

O'Neill, H., & Chiat, S. (2015). What Our Hands Say: Exploring Gesture Use in Subgroups of Children With Language Delay. *Journal of Speech, Language, and Hearing Research, 58*(4), 1319-1325.

Ozcaliskan, S., Adamson, L. B., & Dimitrova, N. (2016). Early deictic but not other gestures predict later vocabulary in both typical development and autism. *Autism, 20*(6), 754-763.

Paul, R., Murray, C., Clancy, K., & Andrews, D. (1997). Reading and metaphonological outcomes in late talkers. *Journal of Speech, Language & Hearing Research, 40*(5), 1037-1047.

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

Raven, J. (2008). Raven's e Educational: Coloured progressive matrices (CPM): London: Pearson Assessment.

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., . . . Bretherton, L. (2010). Predicting language outcomes at 4 years of age: findings from Early Language in Victoria Study. *Pediatrics, 126*(6), e1530-e1537.

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language and Hearing Research, 45 (2), 360-371*.

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, 643-664.

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.

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in Early Gesture Explain SES Disparities in Child Vocabulary Size at School Entry. *Science, 323*(5916), 951-953.

Roy, P., & Chiat, S. (2014). Developmental pathways of language and social communication problems in 9–11 year olds: Unpicking the heterogeneity. *Research in Developmental Disabilities, 35,* 2534-2546.

Scheffner Hammer, C., Morgan, P., Farkas, G., Hillemeier, M., Bittetti, D., & Maczuga, S. (2017). Late Talkers: A Population-Based Study of Risk Factors and School Readiness Consequences. *Journal of Speech, Language and Hearing Research*, *60,* 607-626.

Seeff-Gabriel, B. K., Chiat, S., & Roy, P. (2008). *Early Repetition Battery*: Pearson.

Thal, D., J., 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.

Thal, D. J., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continued delay. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 169-201). Baltimore, MD: Brookes.

Thal, D. J., & Tobias, S. (1992). Communicative gestures in children with delayed onset of oral expressive vocabulary. *Journal of Speech, Language, and Hearing Research, 35*(6), 1281-1289.

Thal, D. J., & Tobias, S. (1994). Relationship between language and gesture in normally developing and late talking toddlers. *Journal of Speech and Hearing Research, 37*, 147-170.

Tomasello, M., Striano, T. & Rochat, P. (1999). Do young children use objects as symbols? *British Journal of Developmental Psychology*, 17, 563-584.

Tomblin, B. (2008). Validating diagnostic standards for specific language impairment using adolescent outcomes. In C. F. Norbury, B. Tomblin, & D. Bishop (Eds.), *Understanding developmental language disorders: From theory to practice* (pp. 93-114). Hove: Psychology Press.

Wetherby, A. M., & Prizant, B. M. (2003). *CSBS manual: communication and symbolic behavior scales*: Brookes Pub.

Yoder, P., Watson, L. R., & Lambert, W. (2015). Value-added predictors of expressive and receptive language growth in initially nonverbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 45*, 1254-1270.

Zimmerman, I., Pond, R. E., & Steiner, S. (2009). *Preschool Language Scale-Fourth Edition (PLS-4UK)* Oxford: Pearson.

Zimmerman, I., Steiner, V. G., Pond, R. E., Boucher, J., & Lewis, V. (1997). *Preschool Language Scale-3 (UK)*. London: Harcourt Brace.

Zubrick, S. R., Taylor, C. L., Rice, M. L., & Slegers, D. W. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research, 50*(6), 1562-1592.

Table 1: Time 1 and 2 assessments

Time 1 Assessments	**Time 2 Assessments**
Preschool Language Scale-Third Edition (PLS 3UK) (Zimmerman et al., 1997)	Preschool Language Scale-Fourth Edition (PLS 4UK) (Zimmerman, Pond, & Steiner, 2009)

MacArthur-Bates Communicative Development Inventory: Words and Gestures (CDI-WG) (Fenson et al., 1993)	Sentence Imitation Test (SIT) from Early Repetition Battery (ERB) (Seeff-Gabriel et al., 2008)
Communication and Symbolic Behaviour Scales (CSBS) (Wetherby & Prizant, 2003)	Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al., 2002)
Early Sociocognitive Battery (ESB) (Chiat & Roy, 2008)	Raven's Coloured Progressive Matrices (Raven, 2008)

Table 2: Mean, standard deviation (SD), median, minimum and maximum standard scores for PLS at T1 and T2 and age, Raven's, DEAP, and Sentence Imitation Test at T2, according to group

	ELD (N=10, 7 boys)		**R/ELD (N=9, 7 boys)[1]**	
	Mean (SD)	**Range**	**Mean (SD)**	**Range**

	Median		**Median**	
Age (months)	55.2 (2.2) 55	51-58	58.5 (3.6) 60	51-61
Raven's	99.4 (16.2) 100	75-135	93.3 (9.3) 92.5	80-106
T1 PLS-3 (UK) Auditory Comprehension	97.5 (6.5) 97	86-112	68.7 (5.5) 68	59-78
T2 PLS-4 (UK) Auditory Comprehension	92.4 (14.1) 93	71-124	78.6 (13.1) 80	55-96
T1 PLS-3 (UK) Expressive Communication	73.6 (5.3) 73	65-82	69.9 (5.4) 71	61-82
T2 PLS-4 (UK) Expressive Communication	90.7 (6.29) 90	83-102	70.4 (14.1) 67.5	54-89
Sentence Imitation Test	89.7 (16.2) 90.5	66-114	69.8 (15.3) 66.5	54-97
DEAP percent consonants correct*	5.3 (2.09) 3.5	3-11	5.3 (3.6) 3.5	3-13

*Standard scores on DEAP have population mean of 10 and SD of 3

[1] One child did not complete the Sentence Imitation Test

Table 3: Correlations between language (PLS-3 Auditory Comprehension and Expressive Communication), symbolic comprehension and gesture use at T1 and language (PLS-4 Auditory Comprehension and Expressive Communication, and Sentence Imitation Test) at T2

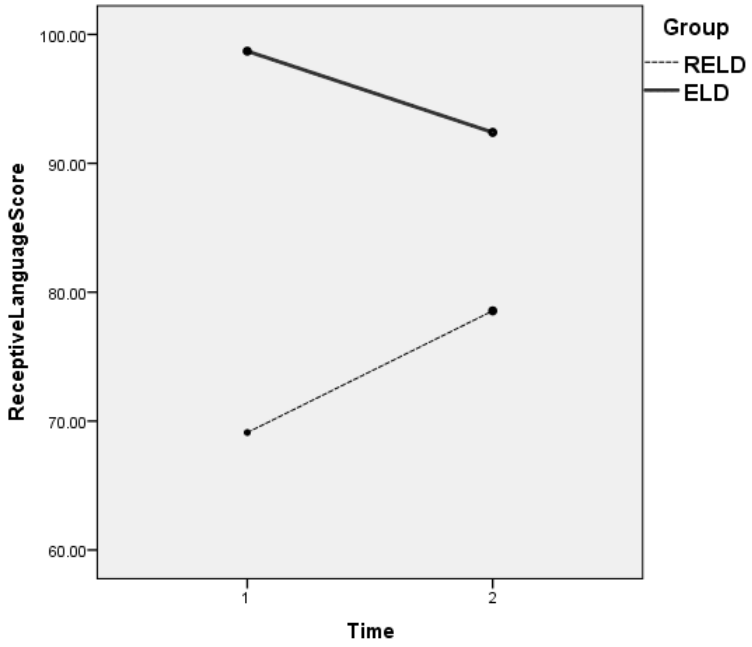
	Time 2

	Receptive Language	Expressive Language	Sentence Repetition
Time 1			
Receptive language	.492*	.718**	.496*
Expressive language	.299	.447	.782**
Symbolic comprehension	.342	.675**	.371
Gesture use	.110	.611**	.534*

** Correlation significant at the 0.01 level (2-tailed)

* Correlation significant at the 0.05 level (2-tailed)

Figure 1: Means for receptive and expressive language of R/ELD and ELD groups over time


Group
RELD
ELD
ReceptiveLanguageScore
100.00
90.00
80.00
70.00
60.00
1
2
Time

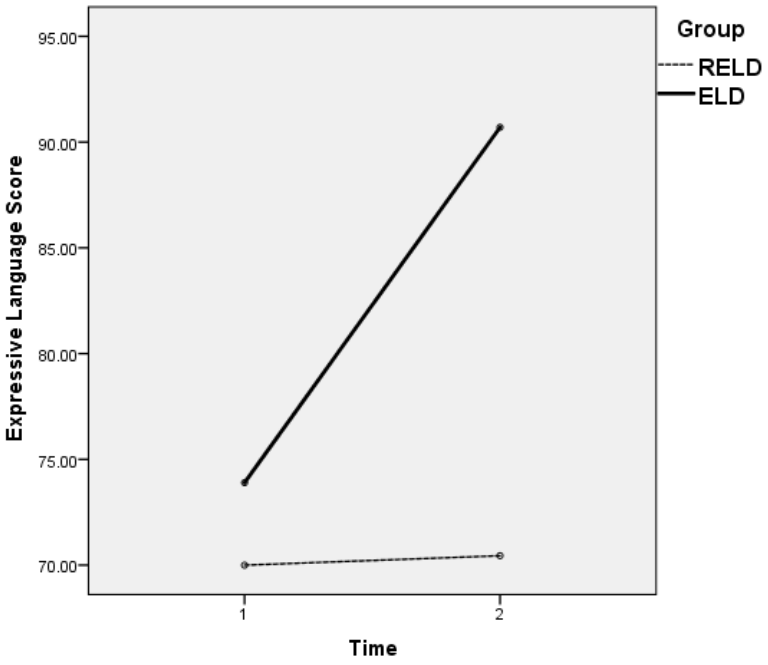
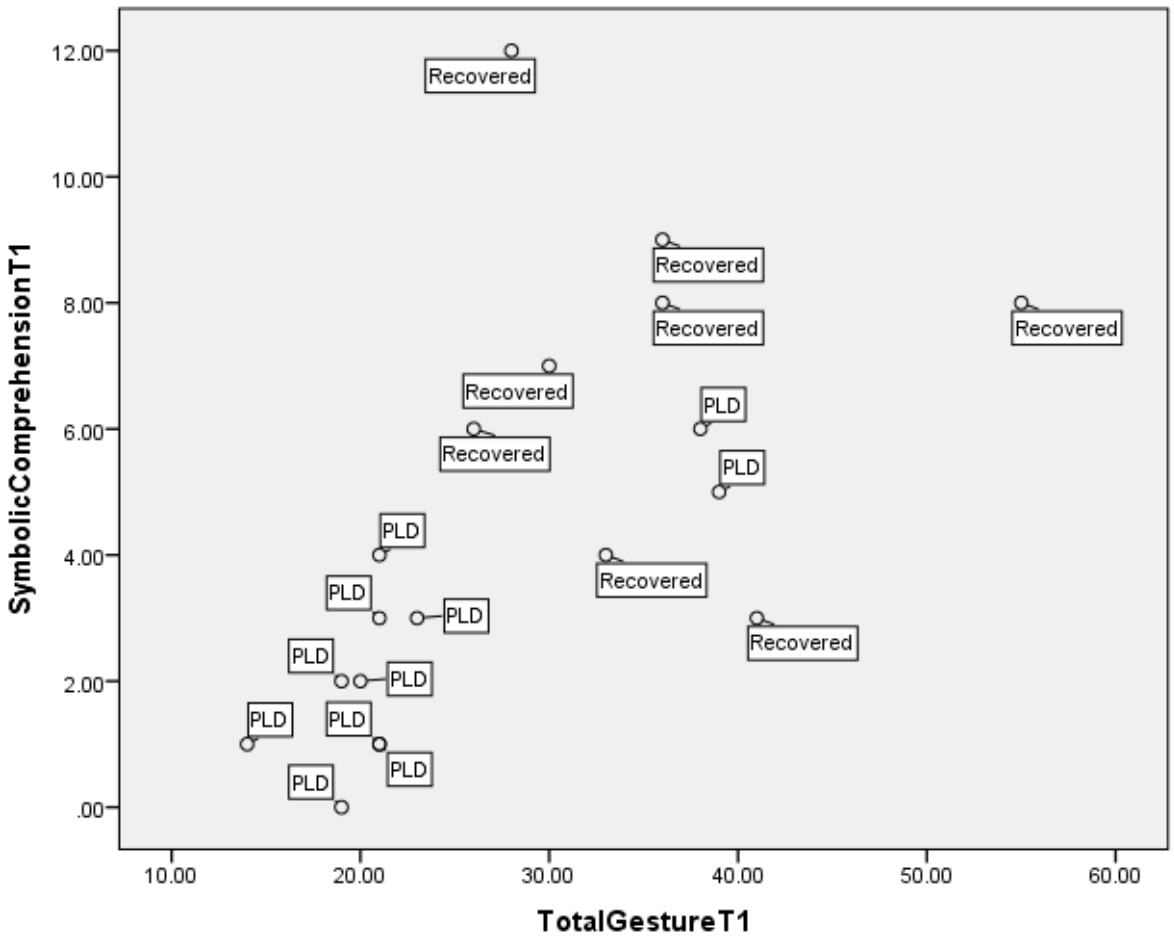

Group
RELD
ELD
Expressive Language Score
95.00
90.00
85.00
80.00
75.00
70.00
1
2
Time

Figure 2: T2 language outcomes[1] according to T1 scores for gesture and symbolic comprehension



[1]PLD: persisting language delay; Recovered: above the cut-off on receptive and expressive language measures (but may present with expressive phonology below mean).