# Grammaticality Judgements of Children with and without Language Delay

**Matthew Saxton, Julie Dockrell,
Eleri Bevan and Jo van Herwegen**

**Institute of Education, University of London**

## Abstract
We examined the grammatical intuitions of children both with and without language delay, assessed via a task presented on computer. We targeted three grammatical structures often reported as compromised in children with language impairments (copula, articles and auxiliaries). 26 children (8 girls) with language delay were recruited (mean age 4;10, range 3;8–6;0). These children met the standard criteria for Specific Language Impairment and underwent an intervention focusing on the three targets. The intervention supplied negative evidence for target omissions, according to the precepts of the Direct Contrast hypothesis (Saxton, 1997). Each child received 20-minute therapeutic sessions daily over a six week period. Children with language delay were tested at four points: pre-intervention, mid-intervention, immediately post-intervention and again six months later. To provide a basis for comparison, we also recruited a group of 116 typically developing children (62 girls) (mean age 5;9, range 3;1-7;9). The grammaticality judgement task yielded two measures: (1) judgement of correctness; and (2) reaction time. Although clear differences were found between typical and atypical children, it was clear that even among the oldest typical children, none performed at ceiling. We also found significant improvements in the performance of language delayed children after the intervention.

## 1 Introduction

Surprisingly little research has been conducted on children's knowledge of grammar in the school years (for a notable exception, see Nippold, 2007). This general lack of interest is based on the assumption that there is actually very little to investigate, in the sense that the mental grammar of a five- or six-year-old is taken to be, to all intents and purposes, equivalent to that of an adult. By the time the child reaches school, therefore, the main business of language acquisition is believed, by many, to be complete. This view is promulgated, for the main part, by researchers in the nativist tradition. Thus, Lightfoot (2005, p.50) suggests that "children attain a fairly rich system of linguistic knowledge by five or six years". On this view, grammar has been taken care of by the age of five, reducing language acquisition from that point onwards to a matter of simply expanding one's vocabulary. However, the pervasiveness of this view is matched only by the lack of empirical evidence to support it.

Our own interest in the grammatical knowledge of school-age children rose out of our work on children with language delay (LD). We developed a therapeutic intervention for children with language delay, most of whom were of school age (see


Child Language Seminar 2007
University of Reading

below). Our intervention was targeted on three grammatical structures: copula, auxiliary verbs and articles. Typically, interventions for LD children tend to focus exclusively on expressive language as a measure of intervention effects (e.g. Leonard, Camarata, Pawlowska, Brown & Camarata, 2007). It is a moot point as to whether receptive measures are needed. The aim of an intervention might be taken as simply to improve the child's speech output to the extent that grammatical forms are more closely aligned, in both frequency and quality, with those of children with typical language (TL). At the same time, though, speech output might improve without any appreciable change in the child's knowledge or understanding of grammar. The basis for any improvements might therefore be attributed to non-linguistic causes, such as increasing powers of attention and imitation. Alternatively, the intervention may have influenced the development of the child's knowledge of grammar. Improvements in the child's speech output would then be ascribed to a more mature system of grammatical knowledge. This latter outcome is perhaps more desirable since the end result would be a mental grammar more closely allied to that of TL children. Generally, it is apparent that improvements in the child's speech output do not necessarily equate with increases in grammatical knowledge. For this reason, we decided to include receptive measures of our three therapeutic targets, in order to gauge more thoroughly the effects of our intervention.

Intervention outcomes for children with Language Delay can also be assessed by comparing LD performance, both pre- and post-intervention, with that of Typical Language children. In fact, one might argue that such comparisons are essential. Simply finding that LD children have impaired grammatical intuitions would not, in and of itself, be especially informative. If TL children, matched for age with the LD sample, also exhibit impaired performance, then no delay or disorder could be imputed to the Language Delay group for the particular targets under scrutiny. As noted above, though, surpassingly little data on the grammatical intuitions of typical school-aged children is available. In particular, we know of no previous studies examining TL children's intuitions about our target grammatical morphemes: articles, copula, and auxiliaries. A central aim of the work reported here, therefore, was to gather baseline data from TL children. To this end, we decided to include Typical Language children from a fairly broad age range (3 to 7 years), so that any developmental trends could be more clearly discerned.

### *1.1 Grammaticality Judgements*

A standard method for tapping into children's knowledge of grammar is to elicit their judgements on key sentences, both grammatical and ungrammatical. The ability to make such judgements is dependent on the child's metalinguistic skills and these develop only gradually. It has been argued that metalinguistic capacity with respect to grammar emerges at about seven years (Gombert, 1992). But Gombert's conception of what he calls *metasyntactic ability* carries with it the notion of conscious reflection on grammar. The interest here, though, is on a less sophisticated ability, namely the ability to distinguish sentences on their basis of their grammaticality. This latter capacity does not necessarily require conscious analytic abilities, though it does demand sensitivity to grammatical form. it is not surprising, therefore, that some studies report children making reliable judgements much earlier than seven years. For example, Crain & Nakayama's youngest participants were 3;2, while Gleitman, Gleitman & Shipley (1972) elicited judgements from children aged 2;6. In this latter study, children were asked to say if sentences like *Bring me the ball* or **Ball me the bring* were either "good" or "silly". The two-year-olds in this study considered grammatical imperatives to be "good", while rejecting about half of ungrammatical imperatives as "silly".

Unfortunately, it is not entirely clear that the basis for child responses was sentence grammaticality. The overall accuracy score for these children is very roughly 75%, a figure which is probably significantly higher than chance. However, Gombert (1992) argues that we should distinguish between responses to grammatical versus ungrammatical sentences. This follows from the view that responses to grammatical sentences cannot be taken as a reliable index of child sensitivity to grammaticality. For Gombert (1992, p.41), grammatical sentences comprise , "sound sequences ..... with which the children are customarily surrounded", that is, they are familiar. Ungrammatical strings, on the other hand, will not be familiar and might be rejected on those grounds alone.

On this interpretation, the overall success rate of 75% noted by Gleitman et al. (1972) needs to be broken down into its component parts: roughly 100% correct for grammatical sentences; and 50% correct for ungrammatical sentences. the former figure *could* reflect a familiarity effect, while the latter figure indicates chance responding and, for Gombert, confirms his view that familiarity, not grammaticality, underpins child responses: children are reduced to chance when faced with unfamiliar sentences. An alternative, and equally plausible, basis for responding is semantic plausibility: Ungrammatical sentences are likely to be relatively difficult to assign a meaning to and are thus more likely to be rejected (de Villiers & de Villiers, 1974). One needs, therefore, to take into account the fact that child responding may depend on several factors, only one of which is, potentially, grammatical sensitivity.

Gombert provides a salutary reminder that responses to grammatical versus ungrammatical sentences should be distinguished in analysis. Unfortunately, the two kinds of sentence are often conflated (e.g., Johnson & Newport, 1989; van der Lely & Ullman, 1996). Gombert's point on this issue is pertinent. However, his argument for ignoring responses to grammatical sentences on the grounds of their familiarity is less convincing. One problem is that there is no special reason for believing that children will, in fact, be familiar with the grammatical sentences presented to them. Children do not hear all possible grammatical sentences: their limited experience with language is confounded by the infinite number of possible sentences that might be produced. If one then persists with the argument that children are prone to accept 'familiar' sentences, then one would have to couch the argument in terms of familiarity with sentence *structure*, rather than familiarity with precise sequences of words. And if that is the case, then one could, in fact, impute child responses to grammatical sensitivity (rather than familiarity, or meaning, or some other factor).

In the present study, there are two further reasons to believe that judgements may well be based on grammatical knowledge: (1) the age of our participants; and (2) the kind of errors being examined. With regard to age, we recruited Typical Language children between the ages of 3 and 7 years. Our children with Language Delay, meanwhile, were aged 3 to 6 years. A conservative view of the age at which children can make judgements on grammaticality would be four years (e.g., Bowey, 2005). Since the majority of children in our study lie above this age threshold, we can be more certain that they will be judging sentences on their grammatical merits. Second, our target structures are grammatical morphemes. Morpheme omission errors rarely have a significant impact on the semantic interpretability of a sentence. For example, the meaning of **He big man* is transparent, even without the copula (of course, many languages, including Russian, do not have the copula at all, and Russians get by just fine without it). Therefore, if a child rejects the sentence **He big man*, it is more likely to be on the basis of linguistic form rather than meaning.

*1.2 Intervention*

Our intervention is based on work in first language acquisition demonstrating the facilitative effects of corrective input (e.g., Saxton, Backley & Gallaway, 2005; Strapp, Bleakney, Helmick & Tonkovich, 2008). Specifically, the focus is on cases where the adult models a correct grammatical form directly contingent on a child error:

(1) Child: *He wiped him.*
[ reflexive context ]
Adult: *He wiped himself.*

(2) Child: All by her own.
Adult: All by herself.

The child errors in (1) and (2) can be characterised as errors of commission. But the concept of error can be extended to include cases where children omit obligatory morphemes. Such omission errors are, of course, prevalent in the early stages of typical language development (roughly from 18 months to three years), but they are also characteristic of older children who experience significant language delay (e.g., Bishop, 1997). Examples of corrective input in these cases involve the adult supplying the missing morpheme directly following the child omission:

(3) Child: We made cake.
Adult: Yes, we made a cake.

(4) Child: He tall man.
Adult: He's a tall man.

The Direct Contrast hypothesis predicts that the direct juxtaposition of child error and correct alternative has the power to function as corrective input (Saxton, 1997). In this context, then, it is predicted that the adult model signals not only that a given form is correct, it simultaneously functions to signal the rejection of the child form. Grammatical forms can be modelled in many contexts, with immediate contingency on a child error being but one of them. Non-contrastive modelling of grammatical forms is described here as *positive evidence*, with contrastive models (as in examples 1-4) falling under the heading of *negative evidence.* Children in our intervention were exposed to intensive modelling of our target structures in the form of negative evidence. We predicted that, following the intervention, the intuitions of LD children would more closely approximate those of their TL peers with respect to copulas, articles and auxiliaries.

*1.3 Developmental Trajectories*

The standard method for establishing a developmental disorder in a given population is to make comparisons with children of the same chronological age (CA) and also with children of the same mental age (MA). If the target group exhibits an impairment relative to the CA-matched group, but *not* to the MA-matched group, then a developmental delay is inferred. If the target group is impaired relative to both CA-matched *and* MA-matched groups, then a disorder is indicated (Hodapp, Burack & Zigler, 1990). One disadvantage with this approach is that the developmental pathways of typical and atypical groups are obscured. This follows from the use of age as a dependent variable in both CA and MA groups and the deployment of cross-sectional comparisons. There is then, an inherently non-developmental stance in the matching approach. An alternative approach seeks to derive functions relating performance to age for both typical and atypical groups and then compare the functions (using regression analyses) to see if they differ. This so-called trajectory approach provides a much clearer emphasis on change over time (e.g., Karmiloff-Smith, Thomas, Annaz, Humphreys, Ewing, Brace, Duuren, Pike, Grice & Campbell, 2004; Thomas, Annaz, Ansari, Scerif, Jarrold & Karmiloff-Smith, under review). The trajectories approach requires that children are recruited across a reasonable age range in both typical and atypical groups. This approach allows one to distinguish two aspects of development, with regard to a target behaviour: onset and rate. A difference in onset can be detected by comparing the intercepts for the regression lines for the two groups. A difference in rate of development is determined by comparing the gradients of the two regression lines. One can therefore establish at least three distinct developmental patterns using the trajectories approach: (1) delayed onset; (2) slower developmental rate; and (3) both delayed onset and slower growth rate.

## 2 Method

### *2.1 Participants*

Two groups of children were drawn from nurseries, primary schools and language support units in the South East of England. All participants were monolingual English speakers. None of the children in either group held diagnoses of auditory, attentional, behavioural or neuromotor disorders.

*Group One.* 30 children with Language Delay (LD) were recruited originally, but 4 of the youngest children were subsequently dropped from the study because they failed to understand the task adequately. Of the remaining 26 children, 8 were female, mean age 4;10 (range 3;8–6;0). These children meet standard diagnostic criteria for Specific Language Impairment and were included on the following basis: (1) Clinical Evaluation of Language Fundamentals (CELF) Pre-School scores at least 1.5 SDs below normal: CELF Receptive, mean score = 75.4, SD = 9.20, range 64–100; CELF Expressive, mean score = 72.1, SD = 6.86, range 62–86; (2) non-verbal IQ within the normal range. British Abilities Scale (BAS) block-building, mean T-score = 44.96, SD = 11.51; and BAS picture similarities, mean T-score = 45.15, SD = 10.26; (3) normal articulation: all children had scores on the Goldman Fristoe Articulation test above the 10th percentile; and (4) no pragmatic difficulties, as assessed using the Children's Communication Checklist II.

*Group Two.* 116 children with Typical Language (TL) took part (mean age 5;9, range 3;1-7;9). These children were included on the following basis: (1) no reported language difficulties; (2) no identified special educational needs; and (3) BAS naming and verbal comprehension scores within the normal range (naming T-score: mean = 52.3, SD = 11.1; verbal comprehension T-score: mean = 47.0, SD = 9.6).

### *2.2 Grammaticality Judgement Task (GJT)*

Two versions of the task were designed (A and B) which constituted mirror images of each other. A sentence appearing as grammatical in Version A appeared as ungrammatical in Version B, and so on. Each version comprised ten practice items followed by thirty five sentences (see below). There were equal numbers of grammatical and ungrammatical sentences for each target structure.

*Test sentences.* Children were asked to judge 35 sentences, including 30 test sentences and five fillers. The test items comprised ten sentences for each target structure (five grammatical, five ungrammatical). The five filler sentences were based on present progressive –ING. Test sentences were rendered ungrammatical by dropping an obligatory target morpheme (e.g., *The seven dwarves looking at her*; auxiliary verb).

Sentences for each of the three targets were matched in a number of ways: mean length of utterance (MLU); number of syllables; phonological complexity; lexical frequency; and age of acquisition of lexical items. We also ensured that the target structures occurred in an approximately equal number of times in sentence-initial, mid-sentence and sentence-final positions.

*Mean Length of Utterance (MLU).* Sentences for each of the three target structures were matched for length in terms of three measures: MLU words; MLU morphemes; and MLU syllables. Grammatical and ungrammatical sentences for each target were also compared on these three measures, and no significant differences were found.

*Phonological complexity.* We ensured that test sentences were matched in terms of number of phonemes in both onset and offset for each word. Tests revealed no significant differences when the basis for comparison was either the different target structures or grammatical versus ungrammatical sentences for each target.

*Lexical frequency and Age of Acquisition (AoA):* Unfamiliar or complex lexical items might cause a child to reject a sentence, so we included only those words that we could reasonably expect to lie within our participants' vocabularies. Words were initially drawn from the MacArthur Communication Development Inventories (CDI), which is administered to children up to 30 months old. Other words were taken from Bird, Franklin & Howard (2001) and all others were either proper names familiar to the children (e.g., Snow White) or concrete referents depicted in the pictures used in the task. Once the words had been selected, we controlled for lexical frequency and AoA. Frequency data were taken from Burroughs (1957) since it is the only corpus available that is based on spoken child language. Only words figuring in the 500 most common words were included in our test sentences. With respect to AoA, we drew words, wherever possible, with an acquisition age of 6 years or less.

### *2.3 Procedure*

The task was presented to LD children on four occasions: pre-, mid-, post- and six-months-post the intervention. Versions A and B of the task were rotated through the four testing occasions for each child. A child who received Version A pre-intervention would thus be given Version A again immediately post-intervention. We did not consider that memory of the task would affect performance because the interval between these testing points was six weeks. Moreover, for this child, Version B would have been administered mid-intervention (three weeks after the start), reducing further the chances that memory for specific items would affect performance. Typical Language children completed the task on just one occasion (either Version A or version B on a rotation basis).

Children were first familiarized with the task by singing a well-known nursery rhyme with the experimenter (*Twinkle, twinkle little star*). On a second pass through the nursery rhyme, a puppet was used to sing the rhyme. At this point also red and green response cards were introduced. Based on traffic lights, which we assumed children would be familiar with, the red card was used for Stop, that is, for sentences that "sounded wrong", and the green card was used for Go, that is, for sentences that "sounded good". In these initial practice items, the word order of "red" sentences was grossly corrupted, to make errors easy to spot (e.g., *Up the high so above world*).

Children were then introduced to a computerized version of the task, prepared using DMDX software (Forster & Forster, 2003). Sentences were presented orally (recorded by an actor) and were contextualized by being paired with a picture taken from a cartoon version of a story familiar to the children (Snow White). Each was produced with normal sentence intonation (including ungrammatical versions). Children were centred on the screen prior to each item by presenting an image of a yellow star with a smiling face, accompanied by a spring-like sound. The experimenter retained control over the presentation of each item and so could be sure that the child was attending to the screen.

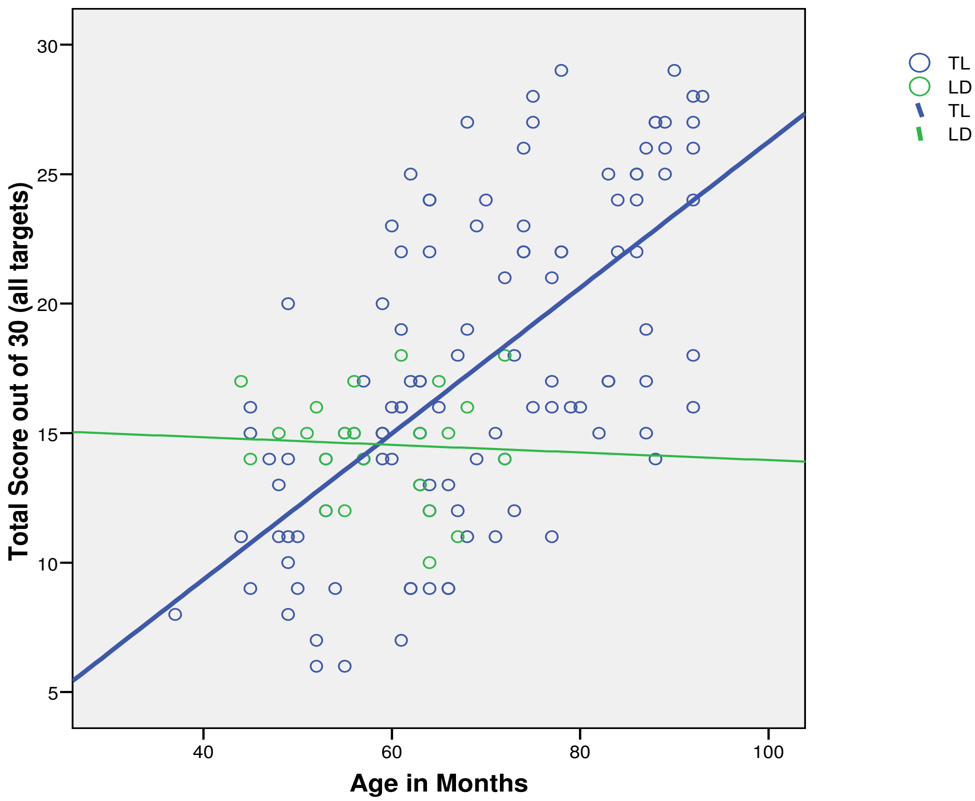
Children completed ten practice items on the computer, switching from the use of red and green cards to red and green response buttons connected to the laptop. The computer recorded child responses (red or green), as well as reaction times (RT) in milliseconds, timed from the offset of each stimulus sentence. The image depicting each sentence remained on screen until the child made a response in each case. The ten practice items included two with correct word order, two with wrong word order, plus six sentences using the target structures (two for each structure and an equal distribution of grammatical and ungrammatical sentences). Once children were familiar with the task, they proceeded to do either Version A or Version B of the task. The two versions were rotated through participants, and in either case, the presentation order of sentences was randomized for each child by the software.

## 3 Results

### *3.1 Trajectory Analyses: Accuracy*

With regard to the accuracy of child judgements, performance on each target structure yielded a total score out of 10 for each target structure, or a total out of 30 for the three structures combined. We first wanted to consider the developmental trajectories of both children with Typical Language (TL) and Language Delay (LD). Results for LD children pertain to their scores prior to the intervention. Following Thomas et al. (under review), we plotted child scores against age in months, with a view to comparing the two regression lines (see Figure 1). Curve estimation confirmed that both sets of data are best described by a linear function.

Figure 1. Developmental trajectories for grammaticality judgements on three target structures for children with typical language (TL) and children with language delay (LD) prior to an intervention




LD: $r^2$ = -.003    TL: $r^2$ = .418

Since analyses of scores on individual target structures mirror the findings when all three are combined, only the latter are reported here. Overall $r^2$ = .57 and the model explained a significant proportion of this variance: $F(3, 135) = 33.74$, $p<.001$, $\eta p^2 = .43$. Inspection of the results for each factor revealed an overall effect of Group (LD versus TL), indicating that the intercepts of the two groups are reliably different at the youngest age of measurement for LD children (44 months): $F(1, 135) = 4.22$, $p < .042$, $\eta p^2 = .03$. Hence, LD children experience a delayed onset in development with respect to grammaticality judgements on articles, auxiliaries and the copula. With the groups combined, age significantly predicted level of performance: $F(1, 135) = 5.01$, $p< .027$, $\eta p^2 = .04$. We also found a significant interaction between Group and Age: $F(1, 135) = 6.17$, $p < .014$, $\eta p^2 = .04$. Therefore, LD children exhibit a slower rate of development when compared with TL children.

*3.2 Comparisons of LD and TL Children: Accuracy*

Child responses were analysed according to the precepts of signal detection theory. The aim is to establish child sensitivity to the different categories of response (grammatical versus ungrammatical) and to control for any response bias (e.g., a predilection to press Green each time).

Previous studies have noted a “yes” response bias in children (e.g., Bishop, 1997; McDaniel & Cairns, 1996). Child responses are therefore allocated to one of four categories: (1) *Hit*: correct acceptance of a grammatical sentence; (2) *Miss*: incorrect rejection of a grammatical sentence; (2) *False Alarm:* incorrect acceptance of an ungrammatical sentence; and (4) *Correct Rejection*: correct rejection of an ungrammatical sentence. So called A' scores are then calculated based on the proportions of hits (y) and false alarms (x), according to the following formula from Linebarger, Schwartz & Saffran (1983): $A' = 0.5 + (y - x) / 4y (1 - x)$. Perfect discrimination (performance at ceiling) would result in an A' score of 1.0. A bias towards rejecting all items (Red response) produces an A' score less than .50. A tendency to accept as correct both grammatical and ungrammatical sentences (Green response) yields an A' score of .50. Mean A' scores for LD and TL children are shown in Table 1.

Table 1. Mean A' Scores for TL and LD children on grammaticality judgements for three target structures (SDs in parentheses)

| | TL Children | Intervention Phase for LD Children | | | |
|---|---|---|---|---|---|
| | | Pre- | Mid- | Post- | Six-Months-Post |
| Articles | .59 *(.20)* | .49 *(.15)* | .43 *(.11)* | .52 *(.16)* | .46 *(.21)* |
| Auxiliaries | .55 *(.20)* | .48 *(.14)* | .46 *(.20)* | .48 *(.15)* | .60 *(.13)* |
| Copula | .55 *(.20)* | .49 *(.13)* | .46 *(.15)* | .57 *(.15)* | .46 *(.12)* |

The use of A' scores provides a more sophisticated assessment of child sensitivity than simply reporting accuracy scores for grammatical and ungrammatical sentences conflated together (e.g., Smith & Tager-Flusberg, 1982). Accordingly, we conducted comparisons of TL and LD children based on A' scores. Comparisons revealed that LD children performed significantly worse than TL children prior to the intervention on all three target structures (see Table 2). This difference is still apparent mid-intervention for articles, but no longer obtains for auxiliaries and copula. Immediately after the intervention, there were no significant differences between TL and LD children for any of the target structures. And LD performance remained comparable to TL performance six months later for both auxiliaries and copula. The TL advantage on articles, meanwhile, re-emerged at this point. Overall, then, it is apparent that the intervention had a beneficial influence on LD children's ability to make grammaticality judgements. Moreover, these effects were still in evidence for two out of the three targets structures six months later.

Table 2. Comparisons of A' scores for TL and LD children on grammaticality judgements for three target structures

| | Intervention Phase for LD Children | | | |
|---|---|---|---|---|
| | Pre- | Mid- | Post- | Six-Months-Post |
| Articles | t(40.0) = 2.59, p = .015 | t(29.5) = 4.53, p < .001 | t(130) = 1.20, ns | t(125) = 1.99, p = .049 |
| Auxiliaries | t(41.1) = 2.13, p = .02 | t(130) = 1.79, ns | t(22.6) = 1.69, ns | t(12.5) = -.10, ns |
| Copula | t(47.2) = 1.83, p = .037 | t(130) = 1.66, ns | t(130) = -.37, ns | t(14.0) = 2.07, ns |

We next explored associations between LD children's language test scores and their performance on the task. Regression analyses were not conducted owing to the relatively small number of participants. Instead, for each structure, correlations were calculated between scores on sub-components of the CELF test and scores on the grammaticality judgement task (out of 10 for each structure). Aspects of CELF that we examined included: receptive language; expressive language; recall of sentences in context; sentence structure; formulating labels; linguistic concepts; basic concepts; and word structure. Before, during and six-months-post-intervention almost no significant relationships were found between these language measures and GJT performance. Two exceptions were found mid-intervention: for the copula, significant correlations were found with expressive language: $r(14) = .61$, $p < .01$; and formulating labels: $r(12) = .56$, $p < .025$). Immediately post-intervention, however, a number of significant positive correlations were found for two of the target structures (articles and copula). For articles, significant associations were found with receptive language: $r(15) = .64$, $p < .005$; sentence recall: $r(13) = .530$, p < 025; linguistic concepts: $r(13) = .48$, p < 05; and basic concepts: $r(13) = .63$, p < 01. For the copula, significant associations were found with sentence recall: $r(13) = .52$, p < 025; and sentence structure: $r(13) = .68$, $p < .005$. It appears, therefore, that LD children who score higher on various language measures at the outset benefit more from the intervention, at least for two of the three target structures. For LD children, age was not associated with performance at any of the four points of data collection.

With regard to TL children, we had two measures of language ability: BAS naming and BAS verbal comprehension. We entered these two variables, together with age, as predictors in separate regression analyses for each structure. For articles, we found that the overall model was significant: $F(3, 109) = 29.97$, $p < .001$, adjusted $r^2 = .44$. Age and BAS naming were both strong predictors of child performance. For age, standardized beta = .63, t = 8.77, p < .001, while for BAS naming, standardized beta = .20, t – 2.62, p < .01). BAS verbal comprehension was not a significant predictor. For copula, the same pattern was found, with a significant model overall: $F(3, 109) = 23.84$, $p < .001$, adjusted $r^2 = .38$. For BAS naming (copula), standardized beta = .19, t = 2.32, p < .02, while for age, standardized beta = .59, t = 7.79, p < .001. For auxiliaries, the overall model was again significant: $F(3, 112) = 14.32$, $p < .001$, adjusted $r^2 = .26$. However, only age functioned as a significant predictor of performance: standardized beta = .52, t = 6.31, p < .001.

*3.3 Reaction Time Data*

Table 3 shows mean reaction times (RTs) for LD and TL children. The pattern of findings was identical for each structure, so the data for all three are combined in the following analyses. Mean RT across all 30 stimulus sentences was thus taken as the dependent variable. We first conducted a trajectory analysis to consider if the developmental path of LD children differed from that of TL children. Overall $r^2$ = .82 and the model explained a significant proportion of this variance: $F(3,98) = 7.28$, $p < .001$, $\eta p^2 = .18$. However, the two groups (LD and TL) did not differ significantly: $F(1,98) = 1.12$, ns, $\eta^2 = .01$, indicating that LD children were not delayed with respect to reaction time. nor was there any interaction between age and group: $F(1,98) = .68$, ns, $\eta p^2 = .01$. Hence, LD children do not differ appreciably from their TL counterparts with respect to rate of development for RT. When both groups were combined, age approached significance; $F(1,98) = 2.93$, $p < .09$, partial $\eta^2 = .03$.

Table 3. Mean reaction times (seconds) for TL and LD children for three structures combined (articles, auxiliaries, copula) (SD in parentheses)

| | | TL Children | Phase of Intervention for LD Children | | | |
|---|---|---|---|---|---|---|
| | | | Pre- | Mid- | Post- | Six-Months-Post - |
| All Sentences Combined | | 4.93 *(.73)* | 4.99 *(.71)* | 4.82 *(.90)* | 4.69 *(.85)* | 4.75 *(.53)* |
| Grammatical | Correct Response | 4.73 *(1.04)* | 4.92 *(.93)* | 4.70 *(1.09)* | 4.71 *(.78)* | 4.69 *(.42)* |
| | Incorrect Response | 4.93 *(.94)* | 5.03 *(1.27)* | 4.92 *(1.09)* | 5.08 *(1.34)* | 5.06 *(.60)* |
| Ungrammatical | Correct Response | 4.85 *(1.28)* | 4.87 *(.80)* | 4.98 *(.78)* | 4.47 *(.97)* | 4.69 *(.91)* |
| | Incorrect Response | 4.86 *(1.05)* | 5.04 *(.92)* | 4.84 *(.91)* | 4.45 *(.96)* | 4.65 *(.64)* |

Separate analyses of the two groups suggest that an age effect *is* apparent for TL children, but not for LD children. For all 30 test sentences combined, mean RT is negatively correlated with age for TL children: $r(82) = -.46$, $p < .01$. As one might predict, then, TL children respond more quickly as they get older. This pattern is repeated, with one exception, when one considers different kinds of test sentence: Grammatical / Correct Response, $r(114) = -.40$, $p < .01$; Grammatical / Incorrect Response, $r(88) = -.48$, $p < .01$; Ungrammatical / Correct, $r(111) = -.33$, $p < .01$; and Ungrammatical / Incorrect, $r(108) = -.13$, ns. As can be seen, age does not confer any advantage when children erroneously accept an ungrammatical sentence (False Alarm). Of interest, RTs for LD children do not increase with age. For all 30 test sentences combined, r(16) = -.11, ns. This pattern is repeated for all kinds of target sentence, when considered either by target structure, grammaticality, or accuracy of response.

We also examined the relationship between speed and accuracy. For TL children, strong negative correlations between RT and GJT score were apparent, indicating that correct responses are produced more quickly than errors. Thus, the correlation of mean RT for all sentences with score out of 30 (all targets combined) was $r(82) = -.48$, $p < .01$. Breaking this down by target structure revealed a uniform pattern: articles, $r(82) = -.39$, $p < .01$; auxiliaries, $r(82) = -.37$, $p < .01$; and copula, $r(82) = -.49$, $p < .01$. For LD children (pre-intervention), on the other hand, no relationships between RT and accuracy could be discerned: articles, $r(16) = -.38$, ns; auxiliaries, $r(16) = .08$, ns; and copula, $r(16) = .07$, ns. Therefore, for TL children,. but not LD children, children become faster with age at producing correct responses.

## 4 Discussion

*4.1 Children with typical language*

Our starting point for this study was the desire to tap the intuitions of children with Language Delay with respect to three grammatical morphemes. In pursuing this aim it became apparent that baseline data from Typical language children were not available. This study thus provides the first such data on TL children and reveals, surprisingly, that development is not complete for articles, auxiliaries or the copula, by the time children reach school. The widespread assumption that development should be complete is therefore refuted. Even if assumptions on development are confined to speech production (rather than comprehension), it emerges that very little is known about TL performance in this arena. In one of the few studies to investigate spontaneous use of obligatory morphemes in TL children, Balason & Dollaghan (2004) report that morpheme production is not, in fact, perfect in children as old as four years. Our findings on child intuitions are in accord with Balason & Dollaghan (2004) and extend the age range of non-ceiling performance up to almost eight years. It appears that children take considerably longer to approximate to adult grammatical norms than is generally believed. Moreover, the linear growth curve we found suggests that development is gradual, with child intuitions converging slowly and only after many years on adult norms.

Our findings present puzzles for both nativist and non-nativist accounts of language development. From a nativist perspective, an explanation is needed for why language development is not, in fact, anywhere near as quick and effortless as conventional descriptions would suggest (e.g., Lightfoot, 2005). On the other hand, non-nativist approaches are also presented with a challenge, in particular those that rely strongly on frequency of exposure to input forms as the basis for development (e.g., Tomasello, 2003). The usage-based notion of *entrenchment* suggests that new structures become increasingly established (entrenched) in the child system with increasing exposure to particular input forms. The problem then is one of just how much input is needed to entrench the obligatory nature of grammatical morphemes? By the age of five, children will already have been exposed to hundreds of thousands, if not millions, of exemplars of obligatory morpheme use by adults. And yet, it would seem that this is not enough to furnish the child with a set of adult-like intuitions.

In the present study, we found strong correlations between Typical Language performance and both age and language level (specifically, naming ability). Smith & Tager-Flusberg (1982) also found a strong positive correlation between age and GJT performance for Typical Language children. However, they further found that the age effect washed out when language ability was taken into account. Unlike the present study, these authors assessed TL language via sentence comprehension and receptive vocabulary tests. The two studies are also distinguished by the particular grammatical structures investigated. Smith & Tager-Flusberg looked at word order and morphology, but in the latter case, they tested for sensitivity to overt (and arguably, implausible) errors like *walky* and *chairer*.

The further main finding with respect to TL children was the negative correlation between reaction time and accuracy. TL children produce correct responses more quickly than incorrect responses. Thus, processing time is increased when the child is faced with a problematic sentence, but that does not enhance the quality of decision making. Quite the reverse, in fact. A similar pattern for RTs on a GJT is reported for adult aphasic patients by Caplan, Waters, DeDe, Michaud & Reddy (2007). The fact that correct and incorrect response are distinguished by RT suggests that the child is sensitive (if not consciously aware) of the difference between the two. Items where the child is justifiably confident are dealt with relatively quickly. Items where the child is uncertain take longer to process and their uncertainty is confirmed by higher rates of failure.

*4.2 Children with Language Delay*

We found clear differences in the developmental pathways for LD versus TL children with respect to three grammatical morphemes: copula, auxiliaries and articles. Children with Language Delay exhibit a much flatter growth curve and their performance (pre-intervention) is at around chance. Arguably these LD children could be described as disordered, not simply delayed, on the grounds that their starting point (onset) is significantly lower than that for age-equivalent TL children and, moreover, their rate of subsequent growth is also significantly slower. Nevertheless, The LD group showed clear signs of susceptibility to our intervention. Performance improved for all three target structures as indexed by the non-significant differences between LD and TL children post-intervention. The improvements in grammatical sensitivity provide support for the view that direct contrasts (between child error and adult model) function as a form of negative evidence for the child.

One explanation for omissions in LD speech production is that certain factors (e.g., sentence length) increase the processing demands on the child, leading to errors. On this view, child competence might not be impaired at all, or would at least be in advance of child production abilities (e.g., Bishop, 1994). Certainly, our findings suggest that child competence is by no means perfect (neither LD nor TL). An alternative approach to child omissions suggests that production and comprehension abilities should run in parallel. For example, Rice, Wexler & Redmond (1999) provide evidence of this kind in support of their Extended Optional Infinitive account of omissions. With regard to our own findings on omissions of copula, auxiliaries and articles, we would need to make a detailed comparison of LD judgements with their productive abilities, in order to test processing limitation versus competence limitation accounts.

The standard nativist view of the input to language acquisition is that positive input only is available to the child. On this view, adults model grammatical forms for the child and the context in which they occur is irrelevant. In consequence, the only information supplied to the child is that certain, modelled forms are grammatical. However, our findings suggest that the context in which forms are modelled is, in fact, critical. When forms are modelled directly following a child error they fulfil two functions for the child: (1) they provide information on what forms are grammatical; and (2) they inform the child that their own form is ungrammatical. Error-contingent direct contrasts of this kind thus furnish a richer source of information for the child than non-contingent adult modelling. Our results bear out this interpretation of the input. The ability of children with Language Delay to make grammaticality judgements is enhanced by our intervention to the point that, statistically, it cannot be distinguished from TL ability. This suggests that their knowledge of grammar, with respect to our three target structures, is more closely allied with adult norms following an intervention programme in which negative evidence is supplied.

## Acknowledgements

We gratefully acknowledge the support of the ESRC (Award RES-000-23-1208). We are also very grateful to the children who agreed to participate and to parents and teachers who helped make the study possible. Thanks also to Gary Yershon for recording the test sentences used in the task.

## References

Balason, D. V., & Dollaghan, C. A. (2002). Grammatical morpheme production in 4-year-old children. *Journal of Speech Language and Hearing Research, 45(5),* 961-969.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33,* 73-79.

Bishop, D.V.M. (1994). Grammatical errors in specific language impairment: Competence or performance.

Bishop, D.V.M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children.* Hove: Psychology Press.

Bowey, J.A. (2005). Grammatical sensitivity: Its origins and potential contribution to early word reading skill. *Journal of Experimental Child Psychology, 90/4,* 318-343.

Caplan, D., Waters, G., DeDe, G., Michaud, J. & Reddy, A. (2007). A study of syntactic processing in aphasia I: Behavioral (psycholinguistic) aspects. *Brain and Language, 101,* 103–150.

Crain, S. & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63,* 522-543.

de Villiers, J.G. & de Villiers, P.A. (1974). Competence and performance in child language: Are children really competent to judge? *Journal of Child Language, 1,* 11-22.

Forster, K. I., & Forster, J. C. (2003). A Windows display program with millisecond accuracy. *Behavior Research Methods Instruments & Computers, 35(1),* 116-124.

Gleitman, L.R., Gleitman, H. & Shipley, E.F. (1972).The emergence of the child as grammarian. *Cognition, 1,* 137-164.

Gombert, J.E. (1992). *Metalinguistic development.* (Translated by T. Pownall). New York: Harvester Wheatsheaf.

Hodapp, R.M., Burack, J.A. & Zigler, E. (Eds.) (1990). *Issues in the developmental approach to mental retardation.* New York: C.U.P..

Johnson, J.S. & Newport, E.L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60-99.

Karmiloff-Smith, A., Thomas, M., Annaz, D., Humphreys, K., Ewing, S., Brace, N., et al. (2004). Exploring the Williams syndrome face-processing debate: The importance of building developmental trajectories. *Journal of Child Psychology and Psychiatry, 45(7),* 1258-1274.

Leonard, L.B., Camarata, S.M., Pawlowska, M., Brown, B. & Camarata, M.N. (2007). The acquisition of tense and agreement morphemes by children with Specific Language Impairment during intervention: Phase 3. *Journal of Speech, Language, and Hearing Research,* 51, 120-125.

Lightfoot, D. (2005). Plato's problem, UG, and the language organ. In J. McGilvray (Ed.*), The Cambridge companion to Chomsky.* (pp42-59). Cambridge: C.U.P..

Linebarger, M., Schwartz, M. & Saffran, E. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13,* 361-392.

McDaniel, D. & Cairns, H.S. 1996). Eliciting judgments of grammaticality and reference. In D. McDaniel;, C. McKee & H.S. Cairns (Eds.), *Methods for assessing children's syntax.* (pp233-254). Cambridge, MA: M.I.T. Press.

Nippold, M.A. (2007). *Later language development: School-age children, adolescents, and young adults (3rd edition).* Austin, TX: Pro-Ed.

Rice, M.L., Wexler, K. & Redmond, S.M. (1999). Grammaticality judgments of an Extended Optional Infinitive grammar: Evidence from English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 943-961.

Saxton, M. (1997). The Contrast theory of negative input. *Journal of Child Language 24,* 139-161.

Saxton, M., Backley, P. & Gallaway, C. (2005). Negative input for grammatical errors: effects after a lag of 12 weeks. *Journal of Child Language, 32,* 643-672.

Smith, C.L. & Tager-Flusberg, H. (1982). Metalinguistic awareness and language development. *Journal of Experimental Child Psychology, 34,* 449-468.

Strapp, C.M., Bleakney, D.M., Helmick, A.L. & Tonkovich, H.M. (2008). Developmental differences in the effects of negative and positive evidence. *First Language, 28/1,* 35-53.

Thomas, M.S.C., Annaz, D., Ansari, A., Scerif, G., Jarrold, C. & Karmiloff-Smith, A. (under review). Using developmental trajectories to understand genetic disorders.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Cambridge, Massachusetts: Harvard University Press.

van der Lely, H.K.J. & Ullman, M. (1996). The computation and representation of past-tense morphology in specifically language impaired and normally developing children. In A. Stringfellow, D. Cahana-Amitay, E. Hughes & A. Zukowski (Eds.) *Proceedings of the 20th annual Boston University conference on language development.* (pp804-815). Somerville, MA: Cascadilla Press.