Purpose: Sentence repetition is a reliable clinical marker of Specific Language Impairment (SLI). However, little is known about cognitive processes underpinning SR, or areas of breakdown in children with SLI (cwSLI). The study investigates causal mechanisms.

Method: 25 cwSLI (mean age 6;7) 19 age-matched and 21 language-matched children (mean age 4;8) repeated 180 sentences of varying length and complexity. Total words omitted, added or substituted were counted. Assessments of expressive and receptive language, Working Memory (WM), Phonological Short Term Memory (PSTM), vocabulary and sustained auditory attention were conducted. 20 sentences were presented in a delayed repetition condition.

Results: The cwSLI made more SR errors than controls, and found delayed repetition especially difficult. Their scores were strongly associated with expressive measures; MLU-in-words, and elicitation tests. Across all groups, comprehension and WM measures were good predictors. Attention and vocabulary were weak predictors. Specific errors made by cwSLI during repetition were also evident in other production tasks.

Conclusions: For cwSLI, performance on SR predominantly reflects expressive abilities and WM processes, with receptive abilities and PSTM playing a lesser role. Errors made during SR may reflect underlying linguistic knowledge, and may constitute appropriate intervention targets. SR likely involves reconstruction from representations in long-term memory.

Literature Review

About 7% of children have language difficulties of unknown aetiology (Tomblin, Records, Buckwalter, & Zhang, 1997), a condition referred to as “Specific Language Impairment” (SLI). Despite typical non-verbal abilities and hearing, performance on language assessments falls far behind age peers. One important clinical marker of SLI is “Sentence Repetition” (SR), otherwise called Elicited Imitation. This simple paradigm involves asking children to repeat sentences and counting their errors. SR scores are a better clinical marker of SLI than other widely used assessments such as non-word repetition (NWR), and a past-tense elicitation task, with greater sensitivity, specificity and accuracy (Archibald & Joanisse, 2009; Conti-Ramsden, Botting, & Faragher, 2001). Difficulties with SR also persist into adulthood (Poll, Betz, & C. A. Miller, 2010). However, despite its promise as a clinical marker, SR is poorly understood. Firstly, it is not certain why children with SLI perform so poorly. For example, does it assess verbal short-term memory (STM), or is it influenced by linguistic representations in long-term memory (LTM)? Secondly we do not fully understand how SR performance is related to a child’s language profile. For example, is it more closely linked to production or comprehension performance? In answering these questions, both researchers and clinicians will be in a better position interpret SR assessments.

Theoretically, an individual with a perfect memory for speech sounds could repeat a sentence by storing its acoustic form and converting this into a speech program. In other words, they could merely “parrot” the sentence. Phonological STM (PSTM), would facilitate this process, as a buffer which stores phonological information. The assumption is that to convert an acoustic form into an articulatory program, a more abstract, i.e. “phonological” level of representation is required. This account is supported by an association between SR performance and nonword repetition (NWR), a task designed to assess PSTM (e.g. Alloway, Gathercole, Willis, & Adams, 2004). Furthermore, sentence length-in-phonemes affects repetition accuracy in typical adults even when the number of words and syntactic structure remain constant (Panagos & Prelock, 1982), indicating that the sheer quantity of phonological information is an important factor. This finding implicates a limited capacity domain-specific buffer such as PSTM.

However, the “parroting” account cannot explain the recall of long sentences, which are beyond our PSTM cacpacity. For such stimuli, SR is likely to involve *reconstruction* from representations in LTM, e.g. syntactic and lexical representations (Potter & Lombardi, 1998). The role of syntactic representations is demonstrated by a priming study, which found that typical adults tended to alter a sentence during repetition to make its structure consistent with the most recently heard clause (Potter & Lombardi, 1998). According to the authors, syntactic priming is an important mechanism in the recall of sentences above a certain length as it allows the participant to replicate the syntactic structure of the stimulus without depending on limited information in STM. Further support for the role of syntax is provided by SR studies of relative clauses. In English these structures allow us to manipulate word order while keeping sentence length in words / phonemes constant, thereby controlling for PSTM load. Non-canonical object relatives (Object-Subject-Verb) are more difficult to repeat than subject relatives, an effect observed in typical adults (Hudgins & Cullinan, 1978), typical children (Kidd, Brandt, Lieven, & Tomasello, 2007), and adolescents with SLI and autism-plus-SLI (Riches, Loucas, Charman, Simonoff, & Baird, 2010). In view of its sensitivity to syntactic structure, SR has a long tradition in the assessment of syntactic development in typical children (Clay, 1971; Slobin & Welsh, 1968).

While primed syntactic representations in LTM allow an individual to recall sentence structure, activated lexical representations in LTM provide the content. This is demonstrated by a study which induced lexical substitutions by exposing a participant to "lure" items shortly before recall, i.e. words which overlap in meaning with words in the original sentence, e.g. *palace*→ *castle* (Lombardi & Potter, 1992). The disruptive effect of the lure item indicates that lexical representations are accessed during recall. It follows that individuals with good lexical representations may have a recall advantage. This was observed by Stokes, Wong, Fletcher and Leonard (2006) who found a strong relationship between SR performance and measures of vocabulary in Cantonese-speaking preschoolers.

The task of reconstructing a sentence is clearly complex. One must switch attention between syntactic and lexical representations in LTM, propositional representations, i.e. representations of sentence meaning, and phonological representations in STM. These different information sources must somehow be coordinated into a whole. Complex tasks involving attention-switching are generally thought to rely on executive functions, which in Baddeley and Hitch's original (1974) model reside in the Central Executive. The potential role of executive functions was demonstrated by Jefferies, Lambon-Ralph and Baddeley (2004) who found that a concurrent attention demanding task affected performance by typical adults on an SR task, but not a word span task. They argue that that the attention-demanding task interferes with executive functions in working memory (WM). SR being a more complex task than word span, involving the coordination of both syntactic and semantic representations, is hence is disrupted to a greater degree.

There is clearly controversy over cognitive mechanisms underlying SR. A parallel debate addresses the comprehension / production dichotomy. According to Vinther (2002) only partial comprehension is necessary for error-free repetition. By contrast, data gathered by McDade (1982) suggest a strong role for comprehension. 4-year-old children were more likely to pass a comprehension probe for sentences they had successfully repeated. In addition this effect was stronger in a delayed repetition condition, where the comprehension probe preceded repetition. One interpretation is that the children could not depend on STM, and were therefore more dependent on representations in LTM, which in turn were dependent on having successfully understood the sentence in the first place. Overall the data suggest a complex relationship between comprehension and repetition, the role of comprehension varying according to the child's ability to recruit information in STM.

Debates have likewise arisen over how closely SR correlates with measures of expressive language such as Mean Length of Utterance (MLU). Evidence to date suggests a close relationship between the two. Both Blake, Austin, Cannon, Lisus & Vaughan (1994) and Devescovi and Caselli (2007) identified a strong positive association between pre-school children's performance on an SR task, and their MLU in morphemes ($R = .56^{***}$ and $R = .70^{**}$ respectively). In addition to this quantitative data, there may be a qualitative relationship between a child's spontaneous language and their SR performance. Devescovi and Caselli found that those children who omitted articles in spontaneous speech also tended to do so during SR. This pattern suggests that both SR and spontaneous production engage similar underlying mechanisms. One possibility is that both SR and spontaneous speech are affected by a child's underlying syntactic competence. A child with imperfect knowledge of articles would therefore be likely to omit articles both during SR and spontaneous production. However, despite the evidence for a link between SR and expressive abilities, some authors have questioned this relationship. For example, Bates, Bretherton and Snyder (1988) have argued that imitativeness itself can be regarded as a unitary trait varying independently of linguistic abilities. Bloom, Lightbown and Hood (1975) report that a child aged 2;6 could not repeat sentence he had produced spontaneously the day before (see Devescovi & Caselli, 2007 for a summary).

The rich literature on the various processes supporting SR allows us to hypothesise potential areas of breakdown in children with SLI. However, this is a complex task given both the range of possible mechanisms involved in SR, and the range of difficulties proposed to account for SLI. Starting with LTM, many researchers have argued that children with SLI have impoverished syntactic and lexical representations. Therefore, assuming a *reconstruction* account, these children may not possess robust and detailed long-term representations which can effectively support recall. Moving on to the working memory (WM) system, children with SLI perform poorly on simple recall tasks, e.g. NWR, digit recall, and word span, thought to underlie various components of STM, and complex span tasks such as backwards digit recall, and complex listening span, which are used to assess WM and executive functions (see Montgomery, Magimairaj, & Finney, 2010 for a review). They therefore present with difficulties in all aspects of WM hypothesised to be involved in SR. In addition recent research has suggested difficulties with sustained attention using both visual and auditory paradigms, (Finneran, Francis, & Leonard, 2009; Spaulding, Plante, & Vance, 2008). Attention may be implicated in SR in two ways. Firstly, as argued by Jefferies et al. (2004), restricted attention may affect the functioning of WM, which is thought to underpin SR. Secondly, as SR is a complex and demanding task for children with SLI, limited attention may impact on their ability to stay focused on the task. Overall, this complex picture of multiple difficulties can explain why SR is so sensitive to language impairment, but it does little to aid clinicians or researchers seeking to interpret SR data, and identify the underlying causes of SR and language difficulties.

This study attempted to distinguish between competing hypotheses of poor SR performance, by identifying which underlying mechanisms are more closely associated with SR performance. It achieved this by conducting a range of cognitive assessments alongside SR, and partitioning the variance, in order to identify the relative contribution of particular mechanisms; the WM versus STM, and expressive versus receptive abilities. It also included a delayed repetition condition to investigate the role of PSTM. In particular, it addressed the following questions;

(i) Which memory mechanisms are implicated in SR?

(ii) To what extent does SR reflect production / comprehension abilities?

(iii) Do SR errors reflect those made during spontaneous production?

Methodology

Participants

25 children with SLI aged 6;0 to 7;3 were recruited from language units attached to mainstream schools in the South East of England. Recruitment letters were sent to Speech and Language Therapists, requesting that children meet criteria for SLI, with structural language difficulties, English as their main language, and no non-verbal learning difficulties, hearing difficulties, autism spectrum disorders, or other known syndrome. No child had been diagnosed with a disorder interfering with intelligibility, e.g. dyspraxia or oromotor difficulties, according to a screening questionnaire. Nonverbal abilities were assessed using the Wechsler Preschool and Primary Scale of Intelligence core subtests (WPPSI-3: Wechsler, 2002) with all children obtaining standard scores greater than 80. A variety of language assessments were used for screening; Word Structure (WS) and Recalling Sentences (RS) from the Clinical Evaluation of Language Fundamentals (Wiig, Secord, & Semel, 1992) the Renfrew Action Picture Task (Renfrew, 1997), the Test of Reception of Grammar-Electronic (TROG-E: Bishop, 2005), and the Children's Test of Nonword Repetition (CNRep: Gathercole & Baddeley, 1996). WS and RAPT assess expressive syntax, with both tests designed to elicit specific syntactic structures at both morpheme and sentence level. The

TROG-E was chosen to assess receptive syntax. Finally RS and CNRep were chosen as they are reliable diagnostic markers of language impairment. This particular version of the CELF was chosen as it is standardized across a wide age range, allowing the same assessment to be used by all children in the study. Children were diagnosed with SLI if they fell below -1.25 standard deviations on 2 or more of these assessments. All of the language-impaired children obtained a mean standard deviation across all tests of less than -1.2 standard deviations.

Age-matched (AM) and language-matched (LM) children (age 4;0 to 5;0) were recruited from mainstream schools via head teachers, with language matching accomplished via MLU-in-words (MLUw). Identical instruments were used, with every child scoring > 80 on the WPPSI, and no child scoring < -1.25 standard deviations on more than one language assessment. Table 1 shows psychometrics and significant group differences. According to norms from the Expressive, Reception and Recall of Narrative Instrument (ERRNI: Bishop, 2004), the mean MLU-w in the SLI group is at the 21st percentile (mean percentile from both ERRNI narratives based on the mean age in the SLI group). Percentiles for the other groups are 77 (AM) and 52 (LM).

Stimuli

Sentence stimuli were generated according to a 2 (complexity) x 2 (length) design. Complex sentences were object relatives, object questions and passives, all involving long-distance syntactic dependencies. Simple sentences were subject relatives, subject questions, and transitive sentences with either two or three arguments. The complex sentences were designed to investigate the effect of dependency length, but this is not focus of the current study. Sentence length ranged from 6 to 12 words (mean 8.2) and was manipulated using filler adjectives and adverbs. Examples are shown in Appendix 1. All nouns and verbs have a token frequency > 10 words per million on either the British National Corpus (2002) , or the CELEX database (spoken and written) (Burnage, 1990). All stimuli were spoken by a native female speaker of English with a local dialect, and recording was conducted in a sound-proof booth. Sentences were grouped into 8 blocks of 20 and pseudorandomised so that no two consecutive sentences had the same type, length and complexity characteristics.

Procedure

The SR task was demonstrated with a cuddly toy parrot and a story book called *the Gossipy Parrot* (Roddie & Terry, 2003). The experimenter read the story to the child, and at various stages the parrot commented on the story. This was achieved wirelessly via a Kensington conference pointer hidden inside the toy. The experimenter pretended not to understand the parrot, so the child had to help him out by repeating what the parrot had said. The parrot was also used for the SR task itself which was run on a laptop computer. The experimenter said "*Now the parrot is going to say some more sentences. I don't understand parrots so you have to tell me exactly what the parrot says"*. The child was then presented with a 5 x 4 grid, with a coloured band to show the half-way stage. As the child heard each sentence a number appeared in the grid. This technique was used to motivate the child by showing how many sentences remained. At the end of the task, a "reward" screen appeared with a picture of people clapping accompanied by applause. All sentences were heard via headphones (Sennheiser PC156), and the children's repetition attempts were recorded straight to the computer via the mouthpiece. The experiment was run using DMDX experimental software (K. I. Forster & J. C. Forster, 2003). It was found that the younger children acquired further motivation. Therefore, a clown peg board was used (personal communication, Shula Chiat and Penny Roy). After repeating each sentence, the child was allowed to put a peg on pole. This simple device greatly improved motivation.

20 of the stimulus sentences were also used in a delayed repetition condition. Equal numbers of simple and complex sentences were used with mean length 8.6 words. In this task a 3 second tone was played after each word. Children were requested to cover their mouth at the

onset of the tone, and take their hand away and repeat the word at the offset of the tone. This adaptation of the task was only administered to the AM and SLI children as the LM children were not capable of inhibiting an immediate response.

Assessments were conducted during 3 visits per child. Each visit consisted of two 30-40 minute sessions separated by a single break. Sentence repetition blocks were administered in one of 4 pseudorandomised orders, with orders evenly distributed within groups. The delayed condition was also presented in one of four orders.

Elicitation Task (ET)

The experimenter described one picture using the target structure, and the child was encouraged to describe a different picture, e.g. *EXPERIMENTER: This is the bread which the woman baked and this is the soup… CHILD: which the boy made, EXPERIMENTER: This is the picture was painted by the girl, and this photo… CHILD: was taken by the boy.* Importantly, all responses required a change of both verb and noun, so children could not merely repeat the structure, but had to go beyond the input. In this sense the task has good face validity as a measure of expressive language. The assessment contained 2 warmup items, 2 subject relatives, 2 object relatives and 2 passives, all complex structures also occurring in the SR task. A scoring protocol was devised to reflect the main syntactic components of each structure (see Appendix 2).

Narrative task

Narratives were elicited from the children in order to calculate their MLU-in-words (MLUw). They were also used to provide a measure of verb morphology. The two narratives were the *Bus Story* (Renfrew, 1991) and *Frog, Where Are You* (Mayer, 1969), often referred to as the *Frog Story*. While the Bus Story involves the experimenter telling the story first, the Frog Story involves the child building their own narrative from pictures. The children's speech was transcribed using the conventions proposed by Miller (1981). Average length of narratives was mean 69 utterances, s.d. 22 in the SLI group, mean 67, s.d. 14 in the AM group, and mean 55, s.d. 16 in the LM group.

Backwards Colour Span Task (BCS)

This is a measure proposed by Zoechsler, Seitz and Schumana-Hengsteler (2006) as an alternative to backwards digit recall not subject to variability in mathematical abilities. Nine colours were used; *black white red blue green pink grey brown yellow*. These are equivalent to digit recall stimuli, with 8 one-syllable stimuli, and 1 two-syllable stimulus (*yellow/seven)*. The task was demonstrated using coloured tennis balls, and testing using spoken stimuli commenced only once the child had understood the procedure. The scoring protocol was based on the Working Memory Test Battery for Children (Gathercole & Pickering, 2001). Items were administered in blocks of six stimuli, with 2 items per stimulus in block 1, 3 items per stimulus in block 2, and so on. Four correct responses were required to pass a block and proceed to the next. The test was discontinued after three failures in a block. Once a block was passed or failed the remaining responses in the block were automatically scored as correct or incorrect. This scoring system was adopted as it was sensitive to the children's abilities while avoiding administering numerous items which were either too easy or too difficult.

Modified Listening Span task (LS^{mod})

This is was an adapted version of reading and listening span tasks (Gaulin & Campbell, 1994; Just & Carpenter, 1992) which are designed to create competition between storage and processing. The most common version of listening span consists of remembering the last words in a set of sentences, while making a true/false judgment. The current task, based on Zoelch et al. (2006), manipulates the paradigm to make it better suited to young children. The children were shown a series of picture stimuli consisting of high-frequency food / non-food items and after each item were asked *Can you eat a [name of object]?* At the end of the set they were asked to recall the items. The task is more suited for young children as the verification task is semantically-specific and hence easier, and items are visually-presented to make the task more engaging. However, the linguistic complexity of the task is reduced, as the children are not required to parse a whole new sentence at each turn, and pictures could be visually encoded. Therefore the task may be regarded as an assessment of WM more broadly, as opposed to verbal WM. There were five blocks altogether and the first block starting at two items per stimulus. Like BCS, the scoring protocol was identical to the WMTB-C, except that testing was not discontinued until the second block.

Auditory Sustained Attention task (Asus)

Children listened to a computer audiofile containing a series of words, e.g. *...star, sock, fish…*, and were required to repeat only the animal words. There were 30 animal words separated by between 2 and 5 non-animal words. Words were 2 seconds apart and testing lasted 5 minutes. The children's score was the number of true positives (animals repeated) minus false positives (non-animals repeated). The test was devised by Atkinson, Braddick and Breckenridge (2010) as a task of auditory sustained attention suitable for young children.

Coding

Responses were transcribed from the audio recordings. Phonological errors were tolerated as long as the target word was recognisable. While this definition is subjective, it is validated by high levels of interrater agreement. Errors were counted using the Levenshtein Distance in words (LDw)(Levenshtein, 1966; Riches et al., 2010). This counts the minimum number of word substitutions, omissions and additions required to transform one sentence into another. It is identical to the CELF algorithm except there is no artificial ceiling on error counts, word swaps are not counted, and no semantic judgments are required, e.g. whether transpositions affect meaning. The algorithm has good concurrent and construct validity, demonstrating high correlations with WM / STM assessments, and sensitivity to syntactic structure (Riches et al., 2010). In addition, error counts based on the number of words added / omitted / substituted demonstrate better diagnostic accuracy than other protocols, e.g. the number of sentences containing one or more errors, or the percentage of words successfully recalled (Stokes, Wong & Fletcher, 2006). The decision was made to exclude morphological errors not affecting word counts; substitutions and omissions of bound closed-class morphemes and substitutions of freestanding morphemes (see Appendix 2 for a full list). This is because there is little theoretical support for a strong association between morphological difficulties in SLI, and the kind of cognitive mechanisms, e.g. WM and comprehension abilities, which may support SR, and hence are a focus of this paper. Therefore, by coding for all morphological errors we would be introducing additional variance which would diminish the sensitivity of analyses investigating these mechanisms.

Interrater reliability

2 raters trained in the coding scheme rated transcriptions for 20.3% of observations (5 AM participants, 4 LM, and 6 with SLI). Raters concurred on 95% of sentences. Where disagreements arose, error rates differed by mean 1.03 per item. Interrater agreement measures were also sought for the narratives. 2 raters listened to the narratives of 6 AM participants, 4 LM and 7 with SLI, covering 25% of total utterances. Raters concurred on 91% of sentences, with MLUs differing by mean 1.6 for each utterance where a disagreement arose.

Ethical approval

Ethical approval for the study was granted by both the University of Reading Research Ethics Committee, and the National Health Service Brighton East Research Ethics Committee.

Analysis

SR performance of the different groups was compared using one-way analyses. Then, Ordinary Least Squares (OLS) regressions were conducted to investigate the relationship between SR performance and additional assessments. Further one-way analyses investigated the effect of latency on SR performance and interactions between latency and Group. The final analysis, again using OLS regressions, investigated whether syntactic errors made by children with SLI during SR were also mirrored in their speech during elicitation and narrative tasks. All one-way analyses used non-parametric methods as data violated homogeneity of variance assumptions.

Analysis of error rates

Table 2 shows mean error rates by group. A Kruskall Wallis oneway ANOVA found a significant effect of Group ($\chi^2 (2) = 41.5, p = .001$), with a Shaich and Hamerle post hoc test finding the following significant differences; SLI > AM, SLI > LM ($p < .01$).

Linguistic and cognitive determinants of SR performance

OLS Regressions investigated the association between the linguistic and cognitive assessments and SR performance. Table 3 shows all significant regression models, with assessments grouped into 6 constructs; expressive language (MLUw, Elicitation, RAPT), receptive language (TROG-E), PSTM (CNRep), WM (LS^{mod}, BCS), attention (Asus), and vocabulary (BPVS raw). P-values were adjusted for multiple independent analyses using the Sidak method, with 6 independent analyses assumed, corresponding to the 6 constructs. Data were screened for outliers (cooks distance > 1). None of the significant regression models contained outliers according to this criterion, and non-significant models did not attain significance with outliers removed.

Three important patterns were evident. Firstly, in the SLI group there was a strong and significant association between LDw and expressive language (R^2 >=.40). By contrast, repetition performance in the control groups was not closely associated with expressive measures, with a significant pre-sidak effect observed only for RAPT scores in the LM group, which demonstrated a counter-intuitive positive association ($F(1,19) = 5.01$, $\beta = .06$, $p = .04$, $R^2 = .20$). In contrast to expressive measures, the TROG-E was a significant predictor across all three groups. However, R^2 values were smaller than those observed for expressive measures in the SLI group. Finally, significant effects were observed across groups for the WM assessments, but not the PSTM assessment (CNRep). The latter attained significance in the SLI group prior to the sidak adjustment ($F(1,23) = 6.2$, $\beta = -.07$, $p = .02$, $R^2 = .21$), but did not approach significance in the other groups (all p-values > .1). Both Asus and the BPVS were poor predictors of SR performance, with all unadjusted p-values > .1, with the exception of BPVS in the AM group ($p = .07$).

Further analyses investigated the relative contribution of expressive and receptive skills in the SLI group. MLUw was chosen as the expressive measure given that it accounts for the largest variance (R^2). The R^2 was calculated for a model containing MLUw and TROG-E blocks, and the decrease in R^2, i.e. the partial R^2, was calculated for models with either variable removed. Together, MLUw and TROG-E accounted for 60% of the variance ($R^2 = .60$), with partial R^2 values of .20 for MLUw and .10 for TROG-E. The remaining .30 represents the portion of shared variance. An identical method assessed the relative contribution of PSTM and the CE, with BCS chosen as the best measure of the CE. Together, the CNRep and BCS accounted for 47% of the variance ($R^2 = .47$), with partial R^2 values of .04 for the CNRep and .18 for BCS. The remaining .25 represents shared variance.

Analysis of latency

Descriptives are shown in Table 2. The children with SLI were more affected by the delay than the AM children with a mean difference of 2 versus 1.1. Before conducting any analyses, the data were visually inspected. Sample distribution varied by group, approximating a Poisson distribution in the AM group, and a Gaussian distribution in SLI group. Therefore non-parametric methods were used. A Wilcoxon test found a significant effect of Latency ($z = -5.5$, $p < .01$), and a Mann-Whitney test found a significant effect of Group ($z = -5.2$, $p < .01$). To investigate the interaction between Group and Latency, a difference score was calculated by subtracting the mean error rates for the immediate condition from mean error rates for the delayed condition on a child-by-child basis. Difference scores were subjected to a Mann-Whitney test, which found a trend towards a significant effect of Group ($z = -1.8$, $p =.08$).

In addition, the relationship between comprehension abilities (TROG-E blocks) and the difference score was investigated using an OLS regression. There was a significant positive association indicating that those individuals who passed more blocks on the TROG showed a greater effect of latency ($F(1,23) = 5.2$, $\beta = .38$, $p =.03$, $R^2 = .12$).

Morphosyntactic characteristics of repetition errors

To determine whether morphosyntactic errors made during SR were also found in more naturalistic contexts, performance on specific morphemes was coded for the SR, ET and narrative tasks. While the ET did not involve spontaneous production it is nonetheless a closer approximation of spontaneous production than SR as children were required to go beyond the input. The ET responses were coded for passive *by* omissions / substitutions, and complementiser omissions in the relative clauses. The narratives were used to calculate rates of verb morphology errors. Analyses are shown in Table 4, along with a more detailed description of the dependent variable. All morphosyntactic constructs were significantly associated across the different paradigms. Particularly noteworthy is the analysis of *by* omissions and substitutions, which accounted for a large proportion of the variance (R^2 = .49).

Discussion

SR performance was severely affected in the children with SLI, who made significantly more errors than even their LM peers. While expressive measures were better predictors of SR performance in SLI than the TROG-E, the latter was a more consistent predictor across all three groups. WM assessments were better predictors than CNRep, a measure of PSTM. Nonetheless, while CNRep was a relatively weak predictor, the performance of the children with SLI was more vulnerable to the effect of delay than the AM children, which may be due to PSTM limitations. Both vocabulary and attention (BPVS and Asus) were poor predictors. Morphosyntactic analyses found a close relationship between the morpheme errors on narrative and elicitation tasks, and errors during repetition tasks.

While numerous studies have observed that SR is a reliable clinical marker of SLI, to the author's knowledge only one other study has employed a language-matched control group (Stokes et al., 2006). While this latter study of Cantonese found no significant group differences, the current study identified an SR deficit, such that children with SLI performed worse than the LM group. If morphological errors affecting word counts had been included it is likely that the discrepancy would have been yet larger. This finding is in line with the theoretical claim that SLI may be characterised as a deficit, with poor performance compared to LM controls on certain measures, e.g. production of tense morphemes. One explanation for differing findings across the two studies is the marked typological difference between English and Cantonese, with the latter exhibiting extremely sparse morphology. If children with SLI have specific difficulties with morphology, then these will be more evident in morphologically richer languages, e.g. English versus Cantonese. While it is true that the current coding scheme discounted certain types of morphological errors, other types, e.g. errors affecting word counts (omission / addition of freestanding morphemes) *were* included, and therefore the coding scheme can detect morphological difficulties which may vary with language typology.

The study yielded conflicting findings with regard to the relative contributions of expressive and receptive abilities. While expressive measures (MLUw, RAPT and ET) yielded a larger effect size than the TROG-E in the SLI group, the TROG-E was nonetheless the only assessment which demonstrated a significant association with SR across all three groups. While the large effect sizes for expressive measures support the claim that SR accurately assesses expressive abilities (e.g. Blake et al., 1994; Devescovi & Caselli, 2007), the consistency of the TROG-E across groups indicates a substantial role for comprehension (McDade et al., 1982). A novel contribution of the current study is the analysis of the *relative* contribution of expressive and receptive skills. In the SLI group, the most highly correlated expressive measure, MLUw, uniquely accounted for twice as much variance as the TROG-E (R^2 = .20 versus .10). Discounting possible differences in the sensitivity and validity of these two assessments, the data support the contention that SR is more closely linked to expressive than receptive abilities, at least in children with SLI. This claim is further supported by the finding that children's morphosyntactic errors in the production tasks tended to be mirrored in their repetition performance.

However, in contrast with this finding, the consistency of the TROG-E across all three groups, as a predictor of LDw, was striking. There are two possible explanations of this relationship. Firstly, we could argue for a direct causal relationship between comprehension abilities and SR. We might assume that SR stimuli need to be successfully comprehended in order to be repeated, and therefore children with poor comprehension fail at an early stage in the processing chain. A second account proposes and indirect relationship. Comprehension scores derived from tests such as the TROG-E may be good measures of children's overall language abilities, which in turn determines repetition performance. The weak predictive power of TROG-E compared to MLUw suggests that this latter interpretation may be the correct one. However, a more fine-grained analysis of errors made in both comprehension and production is necessary to fully evaluate the extent to which comprehension determines SR performance.

Another striking pattern was the poor predictive power of the expressive assessments in the AM and LM groups, especially as these assessments were such good predictors in the SLI group. It is possible that both the expressive assessments and SR are sensitive to individual variation at the lower end of the ability scale, but less sensitive to individual variation within the normal range. The observation that standard deviations for LDw, MLUw, and RAPT scores were wider in the SLI group than the control groups attest to this possibility.

An analysis of the relationship between comprehension performance and latency was conducted to test the hypothesis that comprehension is more closely associated with delayed than immediate repetition, due to the diminished role of STM as a mechanism to support verbatim recall (McDade et al., 1982). A significant effect was found in the *opposite* direction to that predicted, with a weaker association for delayed recall. It is likely that this unusual finding is due to floor effect such that individuals in the SLI group with very poor language, and hence very poor TROG-E scores, performed close to floor in the immediate condition, and therefore showed a smaller discrepancy as a function of latency.

Consistent with Jefferies et al. (2004) the study identified a strong relationship between SR and WM tasks (BDR and LS^{mod}). These were better predictors of SR performance than the CNRep, were more consistent across groups, and uniquely accounted for a larger portion of variance (R^2 = .18 versus .04). While, in adapting LS for younger children, the linguistic demands of the task were reduced, it was nonetheless strongly associated with SR performance in the control groups, thereby demonstrating good concurrent validity as a measure of language processing. A simple explanation for the association between WM tasks and SR scores is that WM difficulties underlie both language difficulties and SR performance. However, this explanation is undermined by the fact that the WM difficulties in the SLI group were only moderate. For example, their performance on LS^{mod} was not significantly different to the LM controls, and it was only on the CNRep, that they performed significantly worse than the LM controls. This finding of moderate WM difficulties, combined with severe PSTM limitations has been observed in a number of studies of memory in SLI (e.g. Archibald & Gathercole, 2006). One is therefore faced with a paradox. SR is possibly the most reliable clinical marker of SLI (Conti-Ramsden et al., 2001), yet WM, the cognitive mechanism most closely associated with SR, is only moderately impaired.

A possible explanation is that WM difficulties do not cause SR difficulties, but both are a consequence of a third factor, underlying linguistic knowledge. SR is most adversely affected as it imposes greater linguistic demands than verbal WM tasks; for example, it involves the production of whole utterances as opposed to span tasks, which only involve the recall of word lists. WM performance can therefore be regarded as an epiphenomenon of one’s degree of expertise in the particular domain being assessed, in this case language (MacDonald & Christiansen, 2002). Likewise, it has also been proposed that lexical difficulties can result in poor performance on LS tasks (Mainela-Arnold & Evans, 2005). While the domain-specific account of MacDonald & Christiansen (2002) explains the current data well, it is at odds with a number of studies finding that, within individuals, WM scores are associated across the visual and verbal domains (e.g. Gathercole, Pickering, Ambridge, & Wearing, 2004), suggesting that WM abilities are indeed domain-general. Clearly, further research is needed to investigate the relationship between WM language.

PSTM was a relatively poor predictor of repetition performance in relation to the WM, though nonetheless significant prior to sidak adjustment ($p = .02$). This *comparatively* weak association suggests that SR depends only partially on verbatim recall using PSTM, and predominantly engages the wider language system, e.g. reconstruction from representations in LTM (Lombardi & Potter, 1992). However, one interesting finding in support of the role of PSTM is provided by the delayed recall task. Children with SLI were more affected by the delay, and while this effect did not reach significance, it is possible that the discrepancy measure was prone to a ceiling effect whereby those with poor performance in the immediate condition were incapable of performing that much worse in the delayed condition. Such a ceiling effect it is likely to mask the between-group differences observed. One interpretation of this effect is that children with SLI were less able to maintain information in PSTM during the delayed condition, possibly due to a limited capacity or poor rehearsal, and this impacted on their repetition performance.

Two constructs not closely associated with SR performance were vocabulary and attention. A discussion of vocabulary should bear in mind the limitations of static tests such as the BPVS, which can only ever test knowledge of a small number of items. Studies of lexical abilities in SLI have demonstrated a discrepancy between scores on static vocabulary assessments, which tend to be around -1 s.d., and performance on word-learning tasks, which may be severely impaired (e.g. Oetting, Rice, & Swank, 1995). This discrepancy might suggest that tests such as the BPVS are not very sensitive to lexical difficulties in SLI.

Therefore we should not over-interpret the lack of a strong association. It is possible that language typology may mediate the relationship between vocabulary scores and SR performance. For example, as Mandarin has virtually no grammatical morphology, factors leading to individual variation in the ability to use grammatical morphemes will not be operational. Therefore lexical knowledge will claim a correspondingly larger proportion of error variance than in morphologically richer languages such as English.

The attention test, like the vocabulary test, may also have been insufficiently discriminating. For example, a more complex divided attention test, as opposed to a sustained attention test may be more sensitive to attentional difficulties. Nonetheless, previous research has suggested that children with SLI have difficulties with sustained visual attention (Finneran et al., 2009), and also a sustained auditory attention task, albeit in a high-load condition using a degraded stimulus (Spaulding et al., 2008). In support of the current task, mean scores were well below ceiling in both groups, suggesting that the task was difficult enough to identify potential between-group differences. While the current data suggest that sustained auditory attention may not be an important factor in SR, it is clear that further research is necessary to probe attention SLI, and also to unpick the relationship between attention and WM difficulties as proposed by Jefferies et al. (2004).

Moving beyond analysis of raw error rates, there was a strong relationship between the kinds of errors the children with SLI made in the narrative and elicitation tasks, and the kind of errors they made during repetition. All of the morphosyntactic constructs; verb morphology, use of complementisers, and use of passive *by*, demonstrated a significant association across repetition, and narrative / elicitation contexts. This finding suggests that the kind of qualitative errors a child makes during repetition are also likely to be found in the child's speech in a more naturalistic context (Devescovi & Caselli, 2007). The reconstruction hypothesis, whereby SR exploits representations in LTM, again provides a convincing account of the data. Both SR and spontaneous speech depend on accessing syntactic representations in LTM, and therefore, where these representations are deficient, similar errors will occur in both paradigms.

It is interesting to note that the two measures of verb morphology were mutually exclusive, with tense errors excluded from the narrative data as they were impossible to verify, but only *–ed* omissions counted for repetition attempts. This suggests that in SLI, difficulties with tense-marking are also associated with a range of verb-related difficulties, e.g. difficulties with agreement and omissions of auxiliaries, as well as tense-marking.

The study has a number of implications for clinical practice. Firstly the close association with MLUw and other measures of expressive language suggests that error data from SR can provide a useful assessment of a child's expressive abilities. Furthermore, the observation that, in the SLI group, SR performance correlated with a wide range of language-related assessments, suggests that it is a good overall measure of language abilities. Finally, the close qualitative relationship between SR errors, and errors made during narrative and elicitation tasks, suggests that SR may be a valid tool for identifying the types of morphemes and constructions which the child has particular difficulties with, and hence determining items requiring intervention. Some SR assessments, e.g. the Early Repetition Battery (Seeff-Gabriel, Chiat, & Roy, 2008), encourage the coder to conduct a morphological analysis of errors, and the current study suggests that this kind of information may be clinically useful. It may even be appropriate to use SR to assess performance on intervention targets, given its close association with spontaneous speech, and its putative link with underlying linguistic knowledge.

While the current findings shed light on the cognitive mechanisms underpinning SR, it is clear that studies investigating correlations between different tasks can only go part way towards delineating causal relationships. Experimental manipulations of stimuli, or alterations to the basic SR paradigm, e.g. delayed repetition, can potentially help us make headway in identifying the contribution of specific mechanisms. Unfortunately the delayed repetition experiment in the current study proved difficult to analyse due to floor effects, and further studies may wish to adopt a simpler set of stimuli. Other potential manipulations include manipulating sentence structure, including nonwords in sentences, degrading the acoustic stimulus, accelerating delivery, or introducing concurrent tasks (e.g. Jefferies et al., 2004). Such manipulations need to be very gently introduced given the severe SR difficulties of children with SLI. It is likely that a two-pronged approach combining traditional cognitive tasks, e.g. listening span, and methodological innovations will further adumbrate the origins of repetition difficulties in SLI.

Acknowledgements

The author would like to thank the British Academy for funding this research project (award number PDF/2007/460), Susan Edwards for her mentorship and support for the research proposal, and Theo Marinis for additional assistance in writing the proposal. Many thanks also to various members of staff at Clinical Language Sciences at the University of Reading, both for their intellectual and moral support. The project could not have taken place without the valuable assistance of speech and language therapists in identifying children, staff at participating schools, and parents who gave their consent. Above all, many thanks to the children for their eager participation and sense of fun.

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*(2), 85–106.

Archibald, L., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899.

Archibald, L. M. D., & Gathercole, S. E. (2006). Short term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675-693.

Atkinson, J., Braddick, O., & Breckenridge, K. (2010). Components of attention in normal and atypical development. *Journal of Vision*, *10*(7), 492.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. Bower (Ed.), *The psychology of learning and motivation*. NY: Academic Press.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences in dissociable mechanisms*. New York: Cambridge UP.

Bishop, D. V. M. (2004). *Expression, Reception and Recall of Narrative Instrument (ERRNI)*. Pearson.

Bishop, D. V. M. (2005). Test of Reception of Grammar - Electronic. *Harcourt Assessment*.

Blake, J., Austin, W., Cannon, M., Lisus, A., & Vaughan, A. (1994). The relationship between memory span and measures of imitative and spontaneous language complexity in preschool children. *International Journal of Behavioral Development: special issue on the development of working memory*, *17*(1), 91-107.

Bloom, L., Lightbown, P., & Hood, L. (1975). Structure and Variation in Child Language. *Monographs of the Society for Research in Child Development*, *40*(2), 1-97.

BNC Consortium (2002). *British National Corpus, version 2.*

Burnage, G. (1990). *CELEX: A Guide for Users*. Nijmegen: CELEX.

Clay, M. M. (1971). Sentence Repetition: Elicited Imitation of a Controlled Set of Syntactic Structures by Four Language Groups. *Monographs of the Society for Research in Child Development*, *36*(3), 1-85.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(06), 741–748.

Devescovi, A., & Caselli, C. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language and Communication Disorders*, *42*(2), 187 - 208.

Finneran, D. A., Francis, A. L., & Leonard, L. B. (2009). Sustained Attention in Children With Specific Language Impairment (SLI). *Journal of Speech, Language, and Hearing Research*, *52*(4), 915.

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

Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of non-word repetition*. The Psychological Press.

Gathercole, S. E., & Pickering, S. (2001). *Working Memory Test Battery for Children (WMTB-C)*. Harcourt Assessment.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2).

Gaulin, C. A., & Campbell, T. F. (1994). Procedure for assessing the verbal working memory in normal school-age children: some preliminary data. *Perceptual and Motor Skills*, *79*, 55 - 64.

Hudgins, J. C., & Cullinan, W. L. (1978). The effect of sentence structure on sentence elicited imitation responses. *Journal of Speech and Hearing Research*, *21*, 809 - 819.

Jefferies, E., Lambon-Ralph, M., & Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: The role of long-term and working memory. *Journal of Memory and Language*, *51*(4), 623 - 642.

Just, A., & Carpenter, P. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*(1), 122 - 149.

Kidd, E., Brandt, S., Lieven, E., & Tomasello, M. (2007). Object relatives made easy: A cross-linguistic comparison of the constraints influencing young children's processing of relative clauses. *Language and Cognitive Processes*, *22*(6), 860 - 897.

Levenshtein, V. I. (1966). Binary codes capable of correcting delections, insertions, and reversals. *Soviet Physics - Doklady*, *10*, 707 - 710.

Lombardi, L., & Potter, M. C. (1992). The regeneration of syntax in short term memory. *Journal of memory and language*, *31*(6), 713-733.

MacDonald, M. C., & Christiansen, M. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, *109*(1), 35-54.

Mainela-Arnold, E., & Evans, J. L. (2005). Beyond Capacity Limitations: Determinants of Word Recall Performance on Verbal Working Memory Span Tasks in Children with SLI. *Journal of Speech Language and Hearing Research*, *48*(4), 897-908.

Mayer, M. (1969). *Frog, Where Are You?* Penguin.

McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: a question of validity. *Journal of Speech and Hearing Disorders*, *47*, 19 - 24.

Miller, J. F. (1981). *Assessing Language Production in Children: experimental procedures*. Assessing communicative behavior. 1.

Montgomery, J. W., Magimairaj, B. M., & Finney, M. C. (2010). Working Memory and

Specific Language Impairment: An Update on the Relation and Perspectives on Assessment and Treatment. *American Journal of Speech-Language Pathology*, *19*(1), 78.

Oetting, J., Rice, M., & Swank, L. (1995). Quick Incidental Learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech and Hearing Research*, *38*, 434 – 445.

Panagos, J. M., & Prelock, P. (1982). Phonological constraints on the sentence productions of language-disordered children. *Journal of Speech and Hearing Research*, *25*, 171-177.

Poll, G. H., Betz, S. K., & Miller, C. A. (2010). Identification of Clinical Markers of Specific Language Impairment in Adults. *Journal of Speech, Language and Hearing Research*, *53*(2), 414.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*, 265 - 282.

Renfrew, C. (1991). *The Bus Story: a test of continuous speech* (2nd ed.). Oxford: Published by author.

Renfrew, C. (1997). *Action Picture Test*. Speechmark.

Riches, N. G., Loucas, T., Charman, T., Simonoff, E., & Baird, G. (2010). Sentence Repetition in Adolescents with Specific Language Impairments and Autism: an Investigation of Complex Syntax. *International Journal of Language and Communication Disorders*, *45*(1), 47 - 60.

Roddie, S., & Terry, M. (2003). *The Gossipy Parrot*. Bloomsbury.

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

Seitz, K., & Schumana-Hengsteler, R. (2006). From Rag (Bag) s to Riches: Measuring the Developing Central Executive. In W. Schneider, R. Schuman-Hengsteler, & B.

Sodian (Eds.), *Young Children's Cognitive Development: Interrelationships Among Executive Functioning, Working Memory, Verbal Ability, and Theory of Mind*. Routledge.

Slobin, D. I., & Welsh, C. A. (1968). Elicited imitation as a research tool in developmental psycholinguistics. *Working papers of the Language Behavior Research Laboratory, University of California, Berkeley*, *10*.

Spaulding, T. J., Plante, E., & Vance, R. (2008). Sustained Selective Attention Skills of Preschool Children With Specific Language Impairment: Evidence for Separate Attentional Capacities. *J Speech Lang Hear Res*, *51*(1), 16-34.

Stokes, S. F., Wong, A. M., Fletcher, P., & Leonard, L. B. (2006). Nonword Repetition and Sentence Repetition as Clinical Markers of Specific Language Impairment: The Case of Cantonese. *J Speech Lang Hear Res*, *49*(2), 219-236.

Tomblin, B., Records, N., Buckwalter, P., & Zhang, X. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language and Hearing Research*, *40*(6), 1245-1260.

Vinther, T. (2002). Elicited imitation: a brief overview. *International Journal of Applied Linguistics*, *12*(1), 54 - 73.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence - 3rd Edition*. Pearson.

Wiig, E., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals - Preschool*. The Psychological Corporation.

Table 1

Psychometrics

	SLI n = 25	AM n = 19	LM n = 21	sig. differences
Age	79.2 (=6;7)	77.9 (=6;5)	56 (=4;8)	LM<(SLI=AM)
	4.3	3.7	1.5	
WISC PIQ	101	109	112	SLI<LM
	14.8	14.1	11.5	
MLUw	6.5	7.9	6.6	SLI=LM<AM
	1.4	0.8	1.2	SLI<AM
CELF RS z	-1.9	-0.3	0.9	SLI<AM<LM
	0.6	0.7	0.6	
CELF WS z	-2.1	-0.1	-0.1	SLI<(AM=LM)
	0.5	0.7	1	
RAPT raw (z)[a]	19.6 (-2.1)	26.8 (-0.1)	23.4 (0.7)	SLI<LM<AM
	5	2.3	3.3	
TROG blocks passed (ss)	5.0 (67.6)	9.8 (91.4)	8.0 (109)	SLI<(LM=AM)
	2.3	3.2	2.8	
BPVS raw (ss)[b]	58.4 (93.9)	72.6 (106)	62.9 (116)	(SLI=LM)<AM
	11.6	9.9	10.3	
CNRep raw (z)[c]	13.0 (-1.3)	25.0 (0.6)	17.8 (-0.3)	SLI<LM<AM
	7.3	4.9	4.1	
LS^{mod} raw	7	8.9	4.8	LM<AM
	3.6	2.3	2.7	
BCS raw	5.8	10.6	no data	SLI<AM
	3.5	3.5		
Asus raw	20.6	22.9	no data	
	8.2	6.8		

raw = raw score, z = z-score, ss = standard score
Statistics in brackets in the first column are shown in brackets in the remaining columns

(a) Data missing for 1 child with SLI due to non-cooperation and 1 AM child due to experimenter error
(b) Data missing for 5 LM children due to time constraints
(c) Data missing for 1 LM child due to non-cooperation

Table 2

Mean error rates per stimulus by Group

Means and *standard deviations*

Brackets show mean increase in error rates as a function of latency

		SLI	AM	LM
Immediate	all stimuli n = 160	2.9 *3.1*	0.9 *1.9*	1.1 *2.2*
Immediate	subset of 20 stimuli	3.5 2.7	.9 1.7	*not applicable*
Delayed	subset of 20 stimuli	5.5 (+2) *3.3*	2.0 (+1.1) *2.7*	*no data*

Table 3

Regression table showing the relationship between psychometric assessments and repetition performance (LD^{lex})

P-values adjusted for 6 independent analyses using the Sidak method ($\alpha^{sidak} = 1 - (1- \alpha^{original})^6$)

Construct and measures	SLI	AM	LM
1. Expressive skills			
MLUw	$F(1,23) = 23.2$, $\beta = -.60$, $p < .01$, $R^2 = .50$	*n-s*	*n-s*
ET raw	$F(1,23) = 16.88$, $\beta = -2.53$, $p < .01$, $R^2 = .42$	*n-s*	*n-s*
RAPT raw	$F(1,22)$[a] $= 14.32$, $\beta = -.17$, $p < .01$, $R^2 = .40$	*n-s*	*n-s*
2. Receptive skills			
TROG-E blocks	$F(1,23) = 9.68$, $\beta = -.27$, $p = .03$, $R^2 = .30$	$F(1,17) = 11.3$, $\beta = -.10$, $p = .02$, $R^2 = .39$	$F(1,19) = 13.6$, $\beta = -.09$, $p = .01$, $R^2 = .39$
3. PSTM			
CNRep	*non-significant*	*n-s*	*n-s*
4. WM			
LS^{mod} raw	*n-s*	$F(1,17) = 8.9$, $\beta = -.13$, $p = .049$, $R^2 = .34$	$F(1,18)$[b] $= 10.4$, $\beta = -.09$, $p = .03$, $R^2 = .37$
BCS raw	$F(1,23) = 17.3$, $\beta = -.22$, $p < .01$, $R^2 = .43$	*n-s*	*no data*
5. Vocabulary			
BPVS raw	*n-s*	*n-s*	*n-s*
6. Attention			
Asus	*n-s*	*n-s*	*no data*

(a) 1 child did not complete the task due to non-cooperation
(b) 1 observation missing due to experimental error

Table 4

Morphosyntactic error patterns across SR and narrative / elicitation tasks

Descriptives show means, standard deviations and minimum – maximum values

Ind. Var.	Desc.	Dep. Var.	Desc.	Regression output
Rate of *by* omission or substitution in passive ET stimuli *(2 stimuli)*	.9 .8 0 - 2	Rate of *by* omission or substitution in passive SR stimuli *(5 stimuli)*	2.7 2.1 0 – 5	$F(1,23) = 22.6$ $\beta = -1.7$ $p <.01$, $R^2 = .49$
Rate of complementiser omissions in relative clause ET stimuli *(4 stimuli)*	1.7 1.7 0 - 4	Rate of complementiser omissions in relative clause SR stimuli *(60 stimuli)*	9.8 12.9 0 – 34	$F(1,23) = 5.3$ $\beta = -3.2$ $p =.03$, $R^2 = .18$
Verb form errors[a] per utterance in narrative task	.3 .3 .06 - .73	Rate of past tense –ed omissions in SR stimuli *(48 stimuli)*	10.9 9.7 0 - 31	$F(1,23) = 4.4$, $\beta = 22.8$, $p =.047$, $R^2 = .16$

(a) 1 error counted for each sentence containing any of the following; agreement errors, e.g. *they is*, incorrect auxiliary + main verb combinations, e.g. *he has doing*, missing auxiliaries, e.g. *he doing*. Additionally, 1 error was counted for non-exclamative sentences without a main verb. Tense errors were not counted as they are contextually-dependent, and therefore impossible to code reliably.

Appendix 1

Examples of stimuli (filler items shown in brackets)

Simple	Complex
There's **the horse** that pulled *him*	There's the *mother* that the **child** hugged
The **child** that woke the *mother* was very friendly	The *dog* that the (nice old) **cat** woke was very grumpy
Which **child** is splashing the *mother*?	Which *elephant* is the (big friendly) **hippo** following?
The (small) **monkey** chased the *pig* up the path	The hare was carried over the puddle by the tortoise.
The man kicked a ball	
The man poured a drink for his friend	

Appendix 2

Scoring Protocols

(A) Scoring protocol for elicitation task

1 point scored for each essential morphemes	
Relative clauses (1) Any complementiser, allowing for dialect (2) A new verb, not in the prompt (3) An appropriate NP argument	*Passive* (1) Auxiliary *be* (2) A new verb, i.e. not in the prompt (3) Preposition *by* (4) Appropriate NP argument
1 point subtracted for each structural error	
(1) Word order errors, e.g. VS (-1 point)	(1) Word order errors, e.g. SV (-1 point) (2) Incorrect morphology on participle, e.g. *taked*

(B) Scoring protocol for the transcription of repetition attempts

Ignore substitutions of the following freestanding morphemes: articles, e.g. a → the, demonstratives, e.g. that → there, question words, e.g. what → which, complementisers, e.g. that → which, pronouns, e.g. she → her, and verbs where only tense and/or aspect is affected, e.g. do → done.

Ignore substitution / omission / addition of bound grammatical morphemes of tense, agreement, aspect or number, e.g. kicked → kick / kicks, dogs → dog.