Probed Serial Recall in Williams Syndrome: Lexical Influences on Phonological Short-Term Memory

Jon Brock
Teresa McCormack
Jill Boucher
University of Warwick, Coventry, United Kingdom

Williams syndrome is a genetic disorder that, it has been claimed, results in an unusual pattern of linguistic strengths and weaknesses. The current study investigated the hypothesis that there is a reduced influence of lexical knowledge on phonological short-term memory in Williams syndrome. Fourteen children with Williams syndrome and 2 vocabulary-matched control groups, 20 typically developing children and 13 children with learning difficulties, were tested on 2 probed serial-recall tasks. On the basis of previous findings, it was predicted that children with Williams syndrome would demonstrate (a) a reduced effect of lexicality on the recall of list items, (b) relatively poorer recall of list items compared with recall of serial order, and (c) a reduced tendency to produce lexicalization errors in the recall of nonwords. In fact, none of these predictions were supported. Alternative explanations for previous findings and implications for accounts of language development in Williams syndrome are discussed.

KEY WORDS: phonological short-term memory, vocabulary knowledge, Williams syndrome

Williams syndrome (WS) is a rare genetic disorder that is linked to a microdeletion in the 7q11.23 region of chromosome 7 (Ewart et al., 1993). Individuals with WS typically have IQs in the mid 50s to low 60s (e.g., Howlin, Davies, & Udwin, 1998; Mervis, Morris, Bertrand, & Robinson, 1999) and face particular difficulties in aspects of visuospatial construction (see Farran & Jarrold, 2003). They have been credited, however, with “an unusual command of language” (Von Arnim & Engel, 1964, p. 367), and it has been suggested that “linguistic functioning is selectively preserved” (Bellugi, Bihrlle, Neville, Doherty, & Jernigan, 1992, p. 201). In fact, it is becoming increasingly apparent that such claims are overstated (Bates, 2004; Brock, 2005; Karmiloff-Smith, Brown, Grice, & Paterson, 2003). Although older children and adults with WS often achieve receptive vocabulary scores that are better than expected given overall or nonverbal mental age (e.g., Bellugi, Bihrlle, Jernigan, Trauner, & Doherty, 1990; Jarrold, Baddeley, Hewes, & Phillips, 2001; Robinson, Mervis, & Robinson, 2003), their performance is rarely at age-appropriate levels (see Bishop, 1999). Moreover, performance on tests of morphological and syntactical abilities is typically in line with overall or nonverbal mental age (e.g., Grant et al., 1997; Robinson et al., 2003; Thomas et al., 2001).

A further issue is whether language development in WS is simply delayed or whether it follows an atypical course. Thomas and Karmiloff-Smith (2003) have suggested that WS is characterized by an imbalance
between phonological and semantic processing. In keeping with this view, individuals with WS appear to be better at learning the phonological forms of words and phrases than they are at learning their meaning. Children with WS often produce words that they do not understand (Singer-Harris, Bellugi, Bates, Jones, & Rossen, 1997; see also Paterson, 2000), and older individuals with WS have a tendency to use low-frequency words, clichés, and idioms in contexts that are not entirely appropriate (Udwin & Yule, 1990; see also Bellugi et al., 1990; Rossen, Klima, Bellugi, Bihrlé, & Jones, 1996).

The precise nature of this phonology–semantics imbalance is currently unclear. One suggestion is that access to lexical-semantic representations is atypical in WS (Rossen et al., 1996; Temple, Almazan, & Sherwood, 2002). This hypothesis has been used to account for a wide range of findings, including evidence for specific deficits in irregular morphology (e.g., Clahsen & Almazan, 1998; but see Thomas et al., 2001), difficulties in naming objects (Temple et al., 2002; but see Thomas et al., in press), and a tendency to produce unusual exemplars in category fluency tasks (Bellugi et al., 1990; but see Jarrold, Hartley, Phillips, & Baddeley, 2000). A different claim is that language acquisition in WS is excessively reliant on phonological short-term memory (Grant et al., 1997; Mervis et al., 1999; Vicari, Brizzolara, Carlesimo, Pezzini, & Volterra, 1996; Vicari, Carlesimo, Brizzolara, & Pezzini, 1996; cf. Baddeley, Gathercole, & Papagno, 1998). Evidence used to support this hypothesis comes from the relatively good performance of individuals with WS on serial-recall tasks (e.g., Mervis et al., 1999; but see Jarrold, Cowan, Hewes, & Gunn, 2004), the failure of individuals with WS to show a primacy effect in verbal free recall (Vicari, Brizzolara et al., 1996; but see Brock, Brown, & Boucher, in press), and an unusual pattern of correlations between receptive vocabulary knowledge and nonword repetition ability (Grant et al., 1997; but see Brock, 2002).

The current study investigated the relationship between lexical and phonological processing in WS within the context of immediate serial-recall tasks. Such tasks are considered to be measures of phonological short-term memory; however, performance is also influenced by long-term knowledge of the phonological structure of words and by knowledge of word meaning (Hulme et al., 1997; Walker & Hulme, 1999). This is illustrated by the lexicality effect, whereby words are recalled better than nonwords or unfamiliar foreign words (e.g., Brener, 1940; Hulme, Maughan, & Brown, 1991). Similarly, high-frequency words are recalled better than low-frequency words (the word-frequency effect; e.g., Hulme et al., 1997; Watkins & Watkins, 1977), and nonwords that contain high-frequency phoneme combinations are recalled better than those that contain low-frequency combinations (the phonotactic frequency effect; e.g., Gathercole, Frankish, Pickering, & Peaker, 1999).

Two studies, however, have reported that the effects of lexical knowledge on serial recall are reduced in WS. Vicari, Carlesimo, et al. (1996) reported that children with WS demonstrated a reduced word-frequency effect when compared with a typically developing (TD) control group matched on nonverbal mental age. Similarly, in a multiple-case study, Majerus, Barisnikov, Vuilleme, Poncelet, and Van der Linden (2003) reported that the effects of lexicality, word frequency, and phonotactic frequency were numerically smaller (and in some cases statistically smaller) in children with WS than in TD controls matched on chronological age or vocabulary mental age. Both sets of authors interpreted their findings in terms of a reduced contribution of lexical–semantic knowledge to phonological short-term memory.

Majerus et al. (2003) also noted that children with WS performed better on serial-recall tasks when stimuli were drawn from a limited pool rather than being sampled without replacement. In the former condition, they argued, participants can anticipate the items in the list but cannot anticipate the order of the items. Consequently, individual variation in performance primarily reflects differences in order memory (cf. Bjork & Healy, 1974), and lexical influences are reduced (cf. Roodenrys & Quinlan, 2000). This is because lexical knowledge influences item memory but appears to have relatively little effect on order memory (cf. Brock & Jarrold, 2004; Gathercole, Pickering, Hall, & Peaker, 2001; Saint-Aubin & Poirier, 1999). The implication here is that individuals with WS have good phonological short-term memory, but they perform poorly relative to controls on tasks that tap item memory because the beneficial effects of lexical knowledge are reduced.

Further evidence for a reduced influence of lexical–semantic knowledge on phonological short-term memory comes from a study by Karmiloff-Smith et al. (1997). These authors noted that when required to repeat nonwords, TD children made numerous lexicalization errors (i.e., they misrepeated nonwords as similar-sounding real words). In contrast, participants with WS almost always repeated the nonwords correctly. Karmiloff-Smith et al. suggested that presentation of nonwords normally activates the lexical-level representations of similar-sounding real words and that top-down feedback to phonological representations (cf. McClelland & Elman, 1986) leads to lexicalization errors. Therefore, they argued, the apparent immunity of participants with WS to this effect could be interpreted in terms of a reduced interaction between lexical and phonological representations.
Such findings have assumed an important role in theoretical debates concerning language development in WS, particularly regarding the notion of an imbalance between phonological and semantic processing (cf. Thomas & Karmiloff-Smith, 2003). For example, both Vicari, Carlesimo, et al. (1996) and Majerus et al. (2003) suggested that their findings provided evidence for impaired lexical–semantic processing in WS (see Bello, Capirci, & Volterra, 2004; Laing et al., 2002; Pleh, Lukács, & Racsmány, 2003, for similar interpretations). Vicari, Carlesimo et al. (1996) further suggested that individuals with WS can be considered to be hyperphonological—relying excessively on phonological as opposed to semantic recoding of words to achieve a comparable level of performance to controls (see also Karmiloff-Smith et al., 1997). Because phonological short-term memory is thought to play an important role in language acquisition, particularly in learning the phonological forms of new words (e.g., Baddeley et al., 1998), Vicari, Carlesimo et al.’s (1996) findings have also been taken as evidence that language acquisition in WS is excessively reliant on phonological short-term memory (see Grant et al., 1997; Laing et al., 2002).

Given the theoretical importance attached to these findings, the aim of the current study was to further investigate the hypothesis that there is a reduced influence of lexical knowledge on phonological short-term memory in WS. To this end, children with WS were tested on two probed recall tasks adapted from a study by Turner, Henry, and Smith (2000) investigating similar issues in typical development. In the probed position-recall task, participants were presented with a list of words or nonwords and were then given an item and asked to indicate its position in the list. This task, therefore, is primarily a test of order memory. Turner et al. reported that TD children did not show a significant lexicality effect on this task, indicating that it provides a test of phonological memory that is independent of any contribution from lexical–semantic knowledge. In the probed item-recall task, after list presentation, participants were required to recall a single item that occurred at a given position in the list. Thus, a correct response required that item as well as order information was retained, and consequently, the performance of TD children was subject to a lexicality effect (Turner et al., 2000).

The hypothesis of a reduced influence of lexical knowledge on short-term memory in WS leads to three predictions. First, individuals with WS will show a smaller lexicality effect than controls on the probed item-recall task (i.e., they will have a reduced advantage for recall of words over nonwords). Second, relative to controls, individuals with WS will perform more poorly on the probed item-recall task than on the probed position-recall task, because only the former task is affected by lexical knowledge. Third, analysis of erroneous responses in the probed item recall of nonwords will reveal a reduced tendency to produce lexicalization errors among individuals with WS.

Given that the aim of the current study was to investigate the interaction between lexical knowledge and short-term memory, it was critical that groups were equated for the extent of their lexical knowledge. Two control groups were therefore used—both matched to the WS group on vocabulary mental age, which was assessed using the British Picture Vocabulary Scale, Second Edition (BPVS–II; Dunn, Dunn, Whetton, & Burley, 1997). First, as in previous studies, children with WS were compared with a group of younger TD children. Turner et al. (2000) reported, however, that the size of the lexicality effect in TD children increased with age, so the differences in chronological age between the WS and TD groups could be critical. Consequently, a second control group consisting of children with learning difficulty (LD) was also tested. These children were matched to the WS group on chronological age as well as vocabulary mental age and could therefore be considered to have had comparable language exposure.

**Method**

**Participants**

The WS group was composed of 9 boys and 5 girls with WS, ranging in age from 10;5 (years;months) to 17;4, who were recruited through the Williams Syndrome Foundation of the United Kingdom. Four of the WS group had received genetic confirmation of their diagnosis via a positive fluorescence in situ hybridization (FISH) test for deletion of the ELN gene which lies within the 7q11.23 region of chromosome 7 (Lowery et al., 1995). The remaining participants had not undertaken such a test but had all received a formal clinical diagnosis of WS.

The TD control group was composed of 10 boys and 10 girls, ranging in age from 5;10 to 8;8, who were recruited from mainstream primary schools. Class teachers were consulted to select children who represented the middle ability range of their class and exclude children with any documented or suspected learning disability. Table 1 shows that the TD group was closely matched to the WS group on vocabulary mental age, $t(32) = -0.01, p = .996$, 95% confidence interval (CI) = $-0.87$ to $0.87$, although the TD group was much younger (nonoverlapping distributions).

The LD control group comprised 8 boys and 5 girls, ranging in age from 9;9 to 16;3, who were recruited from
a school for children with special educational needs. One child in this group had mild cerebral palsy, and another later received a diagnosis of Asperger syndrome. The remaining 11 children in this group had no specific diagnosis. The more open term learning difficulty is therefore preferred to learning disability to describe this group. Table 1 shows that the WS and LD groups had similar chronological ages, $t(25) = 1.06, p = .300$, 95% CI = $-0.84 \leq \mu_{WS} - \mu_{LD} \leq 2.61$, and were closely matched on vocabulary mental age, $t(25) = 0.09, p = .930$, 95% CI = $-0.91 \leq \mu_{WS} - \mu_{LD} \leq 1.00$.

Table 1 also shows performance on two background measures of verbal short-term memory: serial recall of digits (raw scores on the Forward Digit Recall subtest of the Wechsler Intelligence Scale for Children—Revised; Wechsler, 1974) and nonword repetition (Children’s Test of Nonword Repetition; Gathercole & Baddeley, 1996). One-way analysis of variance (ANOVA) revealed a significant effect of group on serial recall of digits, $F(2, 44) = 4.21, p = .021$, $\eta_p^2 = .161$, which reflected a significant difference between the WS and TD groups (as measured with the Newman–Keuls test); however, there was no significant effect of group on nonword repetition, $F(2, 44) = 1.00, p = .376$, $\eta_p^2 = .043$.

### Stimuli

Forty one-syllable CVC words and 40 one-syllable CVC nonwords were recorded by a native English-speaking female. The words had high familiarity ratings (>500; Coltheart, 1981) and a mean age of acquisition rating (Bird, Franklin, & Howard, 2001) of 278 ($N = 30, SD = 41$), indicating that they are typically acquired between the ages of 3 and 5 years. Therefore, given the range of vocabulary mental ages in this study, all participants should have been familiar with the words. In addition, following Turner et al. (2000), the words all had low concreteness and imageability ratings (<500; Coltheart, 1981) to discourage participants from engaging in mnemonic strategies such as imagining a visual representation of each item in the corresponding spatial position on the screen. The nonwords were constructed by changing the initial-consonant sounds in each of the words.

The stimuli were divided into two sets of 20 words and two corresponding sets of 20 nonwords (see Appendix). For each of the four stimulus sets, 2 two-item practice lists, 9 three-item test lists, and 12 four-item test lists were constructed by sampling items an approximately equal number of times from the stimulus set, subject to the constraint that items in each list were phonologically distinctive.

The stimuli were presented on a Macintosh PowerBook G3 laptop computer using the PsyScope application (Cohen, MacWhinney, Flatt, & Provost, 1993). Several of the children with WS disliked wearing head-phones. Therefore, to ensure that these children were not at a disadvantage, items were presented via the computer’s internal speakers for all participants.1

### Procedure

Informed consent was obtained from the caregivers of all participants before the commencement of the study. Children with WS were tested in a quiet room in their home, whereas the TD and LD control children were tested individually in a quiet room in their respective schools. Where possible, all testing was conducted within a single session with short breaks if necessary. The order of testing was counter-balanced, as was the allocation of stimulus sets to the different tasks.

**Probed position recall.** Participants were first familiarized with the paradigm by playing a card game.

---

1A naive adult was able to correctly identify all the stimuli when presented individually through the computer’s internal speakers, confirming that the speakers provided a clear signal.
The experimenter placed either two or three picture cards in a row, facedown, in front of the child. He then named the cards from the child’s left to right, pointing at each card as he named it, before repeating one of those words and asking the child to point to the corresponding card. The practice stage continued until the child had successfully completed five trials of the card game, whereupon they were told that they would now play the same game on the computer.

In the words condition, participants played a card game with a cartoon lady who appeared at the top of the screen. On each trial, the cards appeared facedown, side-by-side in the center of the screen, a hand moved across the bottom of the screen pointing to each card in turn, and the participant heard each card named as it was pointed to. Items were presented at a rate of one every 1.5 s (interonset times). After a delay of 2.5 s from the onset of the last word, one of the words was presented again, and the child was required to point to the corresponding card. Initially, there were two practice trials, each with two cards. Most children completed both practice trials correctly and proceeded to the test phase; however, if either of the practice trials was incorrectly answered, then both practice trials were repeated. In the test phase, there were 9 trials with three cards followed by 12 trials with four cards. Participants therefore received a score out of 21 for each condition. In the nonwords condition, the procedure was identical, but the game was played with a cartoon martian lady who said “funny martian words.”

**Probed item recall.** The procedure for the probed item-recall task was similar to that in the probed position-recall task; however, in the initial card game, after the experimenter had named the cards, he pointed to one of the cards and asked the child to name it. Correspondingly, in the computerized item-recall task, 2.5 s after the onset of the last item in the list, the hand pointed to one of the cards, and the child was required to produce the corresponding word. The experimenter noted all responses and scored them as correct only if they exactly matched the probe.

**Results**

**Overall Performance**

Table 1 shows measures of overall performance on the two probed recall tasks. Results were subjected to a two-way ANOVA with group as a between-subjects factor and task as a repeated measure. The effect of group was significant, \( F(2, 44) = 5.96, p = .005, \eta^2 = .213, \) reflecting a significant advantage for the TD group over the other two groups (as measured with the Newman–Keuls test); however, there was no significant interaction between group and task, \( F(2, 44) = 0.17, p = .845, \eta^2 = .008, \) indicating that the WS group did not perform relatively poorly on the item-recall task.

**Lexicality Effects**

Figure 1 shows the effect of lexicality on performance on the probed position-recall task (A) and the probed item-recall task (B). Results were analyzed using a two-way ANOVA with group as a between-subjects factor and lexicality as a repeated measure. For position recall, the effect of group was significant, \( F(2, 44) = 6.37, p = .004, \eta^2 = .225, \) reflecting a significant advantage for the TD group over the other two groups (as measured with the Newman–Keuls test); however, there was no significant effect of lexicality, \( F(1, 44) = 2.60, p = .114, \eta^2 = .056, \) and no significant interaction between group and lexicality, \( F(2, 44) = 0.83, p = .444, \eta^2 = .036. \) For item recall, there was again a significant effect of group, \( F(2, 44) = 4.11, p = .023, \eta^2 = .157 \), reflecting superior performance in the TD group compared with the WS and LD groups (as measured with the Newman–Keuls test), and words were recalled significantly better than nonwords, \( F(1, 44) = 200.62, p < .001, \eta^2 = .820. \) Crucially, however, there was no significant interaction between group and lexicality, \( F(2, 44) = 0.08, p = .923, \eta^2 = .004. \) Thus, the three groups showed similar effects of lexicality.

**Item and Order Errors in Probed-Item Recall**

The contributions of item and order memory to performance on the probed item-recall task were investigated by means of error analysis. Each erroneous response was classified as an order error (another item from the same list) or an item error (including extra-list intrusions and omissions). Figure 2A shows the corrected proportion of order errors. This was calculated by dividing the number of order errors by the total number of responses that corresponded to one of the items in the list (i.e., correct responses + order errors), because the probability of an order error is contingent on the probability of correctly recalling an item from the list (cf. Saint-Aubin & Poirier, 1999). Two-way ANOVA revealed a significant main effect of group, \( F(2, 44) = 3.80, p = .030, \eta^2 = .147, \) although there were no significant pairwise differences (as measured with the Newman–Keuls test). There was no significant effect of lexicality, \( F(1, 44) = 0.54, p = .467, \eta^2 = .012, \) and the interaction between lexicality and group was not
significant, $F(2, 44) = 0.98, p = .384, \eta_p^2 = .043$. Figure 2B shows item errors as a proportion of total responses. There was a significant effect of lexicality, $F(1, 44) = 165.41, p < .001, \eta_p^2 = .790$, but the effect of group, $F(2, 44) = 0.62, p = .543, \eta_p^2 = .027$, and the interaction between lexicality and group, $F(2, 44) = 0.18, p = .834, \eta_p^2 = .008$, were both nonsignificant.

**Lexicalization Errors**

Lexicalization errors in the probed item recall of nonwords were investigated by calculating the proportion of item errors that were real words. The proportion of lexicalization errors was slightly smaller in the WS group ($M = .473, SD = .225$) than in the TD group.

**Figure 2.** Proportion (±1 SEM) of order errors (A) and item errors (B) on the probed item-recall task.
group \((M = .531, SD = .255)\) or the LD group \((M = .531, SD = .253)\); however, a one-way ANOVA showed that the effect of group was not significant, \(F(2, 44) = 0.27, p = .767, \eta^2_p = .012\). Lexicalization errors were further investigated by looking at the mean frequency (Kucˇera & Francis, 1967) of the errors produced by each participant. Participants were excluded from this analysis if they made less than three lexicalization errors for which frequency norms were available. The log(10)-transformed frequencies of errors in the WS group \((n = 9, M = 1.48; SD = 0.34)\) were lower than those in the LD group \((n = 9, M = 1.70; SD = 0.30)\) and were slightly higher than those in the TD group \((n = 12, M = 1.46, SD = 0.32)\), but the effect of group on frequency was not significant, \(F(2, 27) = 1.72, p = .198, \eta^2_p = .113\).

**Individual Variation Within the WS Group**

Figure 3 shows \(z\) scores (calculated on the basis of data from the combined control group) for each child with WS on three measures of interest. By convention, \(z\) scores that are greater than 1.64 or less than \(-1.64\) are considered to be outside the normal range. In a sample of 14, however, the probability of finding individuals outside this range is relatively high (note also that this normality criterion is less conservative than the modified \(t\) tests adopted by Majerus et al., 2003; cf. Crawford & Garthwaite, 2002). First, looking at the magnitude of the lexicality effect in probed item recall (i.e., recall of words – recall of nonwords), the hypothesis predicted that children with WS would show a reduced lexicality effect (low \(z\) scores) on this measure; however, all of the children with WS were within the normal range. Second, regarding the advantage for probed position recall over probed item recall, the hypothesis predicted high \(z\) scores as a consequence of relatively poor item memory. In fact, only 1 individual was outside the normal range in the predicted direction. Third, it was predicted that individuals with WS would show a reduced proportion of lexicalization errors in recall of nonwords, but again none of them were outside the normal range in the predicted direction. These observations suggest that the discrepancy between our results and those of previous studies cannot be explained in terms of heterogeneity within the WS population. Moreover, Figure 3 shows that none of the 4 children with a positive FISH test were outside the normal range in the predicted direction. Thus, even if we restrict the analyses to participants with a genetically confirmed diagnosis of WS, there is no evidence to support the hypothesis.

**Discussion**

This study investigated phonological short-term memory abilities in children with WS using two probed recall tasks. It has been claimed that the influence of lexical knowledge on phonological short-term memory is reduced in WS (Karmiloff-Smith et al., 1997; Majerus et al., 2003; Vicari, Carlesimo, et al., 1996), and this hypothesis led to three predictions. First, children with WS would demonstrate relatively small effects of lexicality on the probed item-recall task. Second, they would perform relatively poorly on the probed item-recall task compared with the probed position-recall task. Third, they would show a reduced tendency to produce lexicalization errors in the probed item recall of nonwords. In fact, none of these predictions were supported.

**Prediction 1: Reduced Lexicality Effect**

The overall pattern of results was consistent with that reported by Turner et al. (2000), who used similar tasks in a study of TD children. Performance on the probed item-recall task was significantly better for words than for nonwords, but there was no significant lexicality effect on the probed position-recall task. If one assumes that the probed item-recall task taps both item and order memory, whereas the probed position-recall task is a relatively pure measure of order memory, then these findings are consistent with the view that lexical
knowledge primarily affects item rather than order memory (cf. Saint-Aubin & Poirier, 1999). This interpretation was further supported by error analysis of responses in the probed item-recall task, which showed that lexicality influenced the proportion of item errors but had no significant effect on the correct proportion of order errors. Crucially, however, children with WS and controls demonstrated comparable effects of lexicality on the probed item-recall task, indicating a normal effect of lexical knowledge on serial recall in WS.

This finding contrasts with the reduced word-frequency effect reported by Vicari, Carlesimo et al. (1996). One possible explanation for this discrepancy is that the design of the current study was simply less sensitive; however, this seems unlikely. First, the current study ($n = 14$) included more participants with WS than the Vicari, Carlesimo et al. (1996) study ($n = 12$). Second, in the current study, participants were scored on the number of trials correct, which is more sensitive to individual differences than the span measure (i.e., longest list correct) adopted by Vicari, Carlesimo et al. (1996; cf. Oberauer & Süß, 2000). Third, the lexicality effect is essentially an extreme word-frequency effect (cf. Hulme et al., 1991) so should be more sensitive than the word-frequency effect to differences in the top-down influence of lexical knowledge. Instead, we suggest that Vicari, Carlesimo et al. ’s (1996) findings may be a consequence of matching their two groups on nonverbal mental age. Given that receptive vocabulary is a relative strength in WS, it is likely that the TD children had poorer vocabulary knowledge than the children with WS. Consequently, many of the low-frequency words may have been unfamiliar to the TD children (i.e., they were effectively nonwords), leading to exaggerated word-frequency effects in these individuals. Indeed, closer inspection of reported means and standard deviations shows that many of Vicari, Carlesimo et al. ’s (1996) control children had a span of zero for low-frequency words. This potential difficulty was avoided in the current study because vocabulary knowledge was controlled for. In addition, the words should have been highly familiar to all participants, and the nonwords were necessarily unfamiliar.

The current findings also contrast with the reduced effects of word frequency, lexicality, and phonotactic frequency reported by Majerus et al. (2003). It seems unlikely that their findings can be explained in terms of group differences in familiarity with the stimuli because their study, like ours, included a control group of vocabulary-matched TD children. However, Majerus et al. ’s results should be treated with caution because they only tested 4 children with WS and were thus unable to conduct groupwise statistical comparisons of performance. Furthermore, it has been shown that effect sizes in serial-recall tasks often increase in magnitude with overall performance levels (Logie, Della Salla, Laiacona, Chambers, & Wynn, 1996). The children with WS tested by Majerus et al. performed more poorly than controls, so this may have contributed to the relatively small effect sizes that were observed. A similar criticism cannot be leveled at the current study because the three groups produced comparable numbers of item errors and comparable effects of lexicality on item errors. Nevertheless, in the current study, we only looked at the lexicality effect, and it is unclear at present whether the same processes are involved in lexicality and phonotactic frequency effects (cf. Gathercole et al., 1999; Roodenrys & Hinton, 2002). Therefore, it remains possible that future studies will identify a typical phonotactic frequency effects in WS.

**Prediction 2: Impaired Item Memory**

The second prediction was that individuals with WS would have relatively poor item memory compared with their order memory, but there was again no evidence to support this prediction. When performance on the probed item-recall and probed position-recall tasks were compared, there was no Group × Task interaction, indicating that the additional item memory demands of the probed item-recall task had similar effects in all three groups. This view was supported by error analysis of performance in the probed item-recall task, which showed comparable numbers of item errors in the three groups and, if anything, fewer order errors in the TD group.

These findings contrast with those of Majerus et al. (2003), who reported that children with WS were more likely to be within the normal range of performance for tasks where order memory was relatively important than for tasks where item memory (and therefore lexical influences) were more important. Note, however, that these authors did not directly compare item and order memory, as we have done. Furthermore, their results can potentially be explained by the differential sensitivity of the tasks. More specifically, Majerus et al. measured performance on tasks with ostensibly high item memory demands in terms of the number of items correctly recalled, whereas performance on tasks with high order memory demands was measured using a less sensitive span measure (cf. Oberauer & Süß, 2000). Consequently, the order memory tasks were much less likely to find significant differences between children with WS and controls.

Of course, it remains to be explained why the children with WS in the current study demonstrated poor order memory compared with children in the TD group. One potential concern here is that both tasks involved the use of spatial position to represent the serial position of items in the list, so some participants
may have remembered serial-position information by associating each item in the list with its spatial position. Poor order memory in the WS group could then be explained in terms of a deficit in visuospatial memory (cf. Jarrold, Baddeley, & Hewes, 1999) or mental imagery (cf. Farran, Jarrold, & Gathercole, 2001); however, this seems unlikely for a number of reasons. First, the use of low-imageability words and nonwords should have discouraged all participants from using such a strategy. Second, if controls were using visual imagery as a mnemonic, then their order memory would be better for words than for nonwords, and this was not the case. Third, a similar pattern of results was found for performance on a conventional serial recall of digits task that had no visuospatial component (see also Jarrold et al., 2004; Laing, Hulme, Grant, & Karmiloff-Smith, 2001; for evidence of poor serial recall in WS relative to vocabulary-matched TD children). Instead, the relatively poor order memory of children in both the WS and the LD groups may simply reflect the matching procedures used. More specifically, receptive vocabulary tests are likely to overestimate the mental age of these children because they benefit from their extra experience compared with the TD children. Consequently, matching groups on this measure means that children in the WS and LD groups are at a disadvantage on tests of fluid (i.e., experience independent) intelligence such as recall of the arbitrary order of items in a list.

**Prediction 3: Reduced Lexicalization Errors**

The third prediction concerned lexicalization errors. Analysis of errors in the nonwords condition of the probed item-recall task showed that the three groups had similar tendencies to produce lexicalization errors, again indicating similar interfering effects of vocabulary knowledge on short-term memory. This result contrasts with the reduction in lexicalization errors in WS reported by Karmiloff-Smith et al. (1997); however, these authors did not report any statistical analyses of error data. Moreover, in their study, the WS group made far fewer errors as a whole, so had fewer opportunities to make lexicalization errors. This concern was avoided in the current study because the groups made comparable numbers of item errors overall, and individuals’ lexicalization errors were normalized against their overall item errors.

**Conclusion**

To summarize, the current study failed to find any evidence to support the claim that the influence of lexical knowledge on phonological short-term memory is reduced in WS. Our conclusions, therefore, differ from those of several previous studies of WS, but in each case, we have been able to provide alternative explanations for earlier findings. Of course, with all null results, it is difficult to say with absolute certainty that there are no differences between individuals with and without WS in terms of the interaction between short-term memory and lexical knowledge, but there is currently little sustainable evidence to suggest otherwise.

As noted in the introduction, findings concerning the influence of lexical knowledge on phonological short-term memory have played an important role in theoretical accounts of language development in WS. The current findings, therefore, have important implications. At a minimum, they serve to narrow the scope of the proposed dissociation between phonological and semantic processing. More generally, however, they form part of a growing body of research suggesting that language and related abilities in WS may not be as unusual as has been claimed (e.g., Brock, 2002; Brock et al., in press; Jarrold, Cowan et al., 2004; Jarrold, Hartley et al., 2000; Thomas, Dockrell, et al., in press; Thomas, Grant et al., 2001; see Brock, 2005, for a review). Indeed, there currently appears to be few reliable data on language in WS that support the notion of an imbalance between phonology and semantics.

In discussing various possible manifestations of a phonology–semantics imbalance in WS, Thomas and Karmiloff-Smith (2003) introduced what they termed the conservative hypothesis—essentially a null hypothesis against which claims of specific abnormalities in WS language could be compared. According to the conservative hypothesis, deficits in vocabulary, syntax, and pragmatics in WS are what one might expect for the level of learning disability in these individuals, and anomalies in the WS language system are a consequence of other features of the disorder. For example, recent evidence for deficits in the use of spatial language (Landau & Zukowski, 2003; Lukács, Pléh, & Raemány, 2004; Phillips, Jarrold, Baddeley, Grant, & Karmiloff-Smith, 2004) may be explained in terms of more general impairments in visuospatial processing. Obviously, future research may prove otherwise, but in our opinion, there is little evidence at present to warrant the rejection of this null hypothesis. Arguably, therefore, the challenge for future research is to determine why language appears to develop relatively normally in WS whereas other developmental disorders such as Down syndrome are associated with specific impairments in language (Brock, 2005; cf. Chapman, 1997; Laws & Bishop, 2003). Indeed, the investigation of the similarities and differences between such syndromes may prove to be a powerful tool in determining those factors that play a critical role in determining language development and its disorders.
Acknowledgments

This research was supported by a doctoral studentship awarded to Jon Brock by the Williams Syndrome Foundation of the United Kingdom. We thank the families from the Williams Syndrome Foundation and the staff and pupils of Round Oak School, Leamington Spa, and Woodloes Infant and Junior Schools, Warwick, United Kingdom, for their cooperation in this work. We are also grateful to Gordon Brown, Chris Jarrold, Alan Baddeley, and Steve Majerus for comments on the manuscript and to Judy Turner for helpful suggestions concerning the probed recall tasks.

References


Received March 29, 2004
Accepted July 13, 2004
DOI: 10.1044/1092-4388(2005/025)

Contact author: Jon Brock, who is now at the Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, United Kingdom.
E-mail: jon.brock@psy.ox.ac.uk

**Appendix.** Stimuli for the probed recall tasks.

<table>
<thead>
<tr>
<th>Set A</th>
<th></th>
<th>Set B</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cheat</td>
<td>reet</td>
<td>cool</td>
<td>sool</td>
</tr>
<tr>
<td>fine</td>
<td>hine</td>
<td>cut</td>
<td>lut</td>
</tr>
<tr>
<td>get</td>
<td>pull</td>
<td>down</td>
<td>chown</td>
</tr>
<tr>
<td>gone</td>
<td>rude</td>
<td>feel</td>
<td>cheal</td>
</tr>
<tr>
<td>have</td>
<td>same</td>
<td>got</td>
<td>fat</td>
</tr>
<tr>
<td>keep</td>
<td>meep</td>
<td>hard</td>
<td>sard</td>
</tr>
<tr>
<td>large</td>
<td>targe</td>
<td>less</td>
<td>wess</td>
</tr>
<tr>
<td>live</td>
<td>tiv</td>
<td>look</td>
<td>wook</td>
</tr>
<tr>
<td>loud</td>
<td>goud</td>
<td>miss</td>
<td>niss</td>
</tr>
<tr>
<td>make</td>
<td>pake</td>
<td>move</td>
<td>koove</td>
</tr>
</tbody>
</table>

Note. Each nonword rhymes with the corresponding word to its left.