

*The role of non-initial clusters in the
Children's test of Nonword Repetition:
evidence from children with language
impairment and typically developing
children*

Article

Accepted Version

Cilibrasi, L., Stojanovik, V. ORCID: <https://orcid.org/0000-0001-6791-9968>, Loucas, T. ORCID: <https://orcid.org/0000-0001-8130-6690> and Riddell, P. ORCID: <https://orcid.org/0000-0002-4916-2057> (2018) The role of non-initial clusters in the Children's test of Nonword Repetition: evidence from children with language impairment and typically developing children. *Dyslexia*, 24 (4). pp. 322-335. ISSN 1099-0909 doi: 10.1002/dys.1599 Available at <https://centaur.reading.ac.uk/79282/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/dys.1599>

Publisher: Wiley

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

The role of non-initial clusters in the Children's test of Nonword Repetition (CNRep): Evidence from children with language impairment and typically developing children

Luca Cilibrasi^{*+}, Vesna Stojanovik[°], Tom Loucas[°], Patricia Riddell[°]

^{*} Charles University, Department of English-ELT and Department of Linguistics

⁺ University of Cambridge, Department of Theoretical and Applied Linguistics

[°] University of Reading, School of Psychology and Clinical Language Sciences

Abstract (max 200 words)

The CNRep (Gathercole et al., 1994) is one of the most popular tests of nonword repetition. The test is composed of nonwords of different length, and normative data suggest that children experience more difficulties in repeating long nonwords. An analysis of the distribution of phonological clusters in the test shows that non-initial clusters are unequally distributed in the test: they only appear in long nonwords (4-5 syllable). For this reason, we hypothesised that the difficulties children encounter with long nonwords may be influenced by the phonological complexity of the clusters, and not just by the challenge for working memory associated with length. To test the hypothesis, we compared repetition performance in long nonwords with and without a non-initial cluster in 18 children with language impairment and 18 typically developing children. The analysis shows that long nonwords with non-initial clusters are repeated less accurately by both groups. In addition, there was an interaction between cluster and age: the effect of cluster is absent in younger children and gradually increases with age. These findings suggest that phonological complexity may be having an impact on the length effect normally observed in the CNRep, and this impact may be particularly evident in older children.

Introduction

Nonword repetition tests play an important role in the assessment of developmental language disorders and a link has been made between the ability to repeat nonwords and language impairment (Norbury, Bishop & Briscoe, 2001). Poor performance in nonword repetition tasks is regarded as a clinical marker of developmental language disorder (Conti-Ramsden & Hesketh, 2003).

Dollaghan and Campbell (1998) showed that purposely developed nonword repetition tasks can be more accurate than traditional language tests in assessing the presence of language impairment. In a large sample including 581 children, Weismer et al. (2000) confirmed the reliability of nonword repetition tasks as a measure of phonological working memory in the identification of language impairment. Botting and Conti-Ramsden (2001) replicated these findings by comparing low and high performing children with

Specific Language Impairment (SLI) and concluded that nonword repetition is a good predictor of language performance.

Nonword repetition tasks can also be reliably used in younger, pre-school children, and these tasks have good sensitivity and specificity to discriminate between children with and without SLI (Gray, 2003). A meta-analysis including 23 studies, showed that nonword repetition is a very reliable measure of language impairment (Estes et al. (2007), with children with SLI performing on average more than 1 standard deviation (SD) below typically developing (TD) children. A consistent length effect was observed with long nonwords being more problematic in clinical groups. Even when controlling for differences in short-term memory span, children with SLI show poorer performance in a nonword repetition task (Archibald & Gathercole, 2007).

Nonword repetition abilities have also been linked to reading. Gallon and Marshall (2009) showed that reading performance can be predicted by performance in nonword repetition. The authors claim that nonword repetition tasks are extremely important in the assessment of reading disorders, particularly in the assessment of dyslexia. In a twin study, Bishop et al. (2004) showed that nonword repetition tasks can be used to distinguish genetic and environmental factors in the development of early reading skills. In a longitudinal study, Gathercole (1995) investigated the relation between lexical knowledge, phonology and nonword repetition. The study showed that lexical knowledge and working memory are differentially correlated with nonword repetition performance, depending on the nature of the nonwords. Performance on nonwords that were previously evaluated as highly word-like correlated with performance on a lexical knowledge task whereas performance on nonwords with low word-likeness correlated with phonological working memory measures. Further, performance on the nonword repetition task was a significant predictor of reading performance in young children, suggesting a crucial role of working memory in learning to read.

One of the most commonly used tests of nonword repetition in the UK is the Children's Test of Nonword Repetition (CNRep), developed by Gathercole et al. (1994). The CNRep assesses working memory (which has been shown to correlate with both language and reading abilities) and phonological processing, and is often used as part of a battery in the assessment of developmental language and reading disorders (Gallon & Marshall, 2009, Gathercole & Baddeley, 1996). In this test, there are 4 types of nonwords, divided according to number of syllables: 10 two syllable nonwords, 10 three syllable nonwords, 10 four syllable nonwords and 10 five syllable nonwords. Normative data suggest that longer nonwords are repeated less accurately by all age groups (Gathercole & Baddeley, 1996).

According to the model of working memory proposed by Baddeley (2003) that we adopt in this study, nonword repetition tasks rely on what is known as the phonological loop, a subsystem of working memory. Working memory can be defined as "a brain system that provides temporary storage and manipulation of the information necessary for complex cognitive tasks such as language comprehension, learning, and reasoning", (Baddeley, 1992, p. 556). The term "working memory" is preferred to the term "short-term memory" since it is usually used when referring to a multi-component account of memory (Baddeley, 2000). According to Baddeley (2003) working memory can be divided in three distinct sub-components. There is a priority component, defined as the central executive component, which inhibits

stimuli that are not relevant in a given context and focuses the system on the stimuli that need to be processed in that specific moment. In other words, the central executive component focuses the attention on what is relevant to the current task. There are two further subcomponents, defined as slave components: one has a visual specialisation (the visual scratchpad), and it manipulates visual images for a short time, allowing us to perform actions in the world we move in; the other one has a phonological specialisation (the phonological loop), and it manipulates speech sounds and crucially rehearses speech sounds when needed, for instance when we need to store information until a person has finished uttering a sentence, or during first and second language acquisition, when new stimuli need to be learned and gradually moved to long-term memory.

In relation to working memory, it is well established that long nonwords require more resources than short nonwords (Gathercole et al., 1991). It is therefore possible to interpret the finding that longer nonwords are repeated less accurately by children with language impairment assessed with the CNRep (Gathercole & Baddeley, 1996) as a consequence of a decrease in efficiency in working memory. The CNRep is thus a successful test aimed at measuring working memory. Despite its wide use, the test has been criticized.

An early critique to a previous version of the test (Snowling, Chiat & Hulme, 1991) stresses that it is simplistic to regard the CNRep as simply a measure of working memory, even if referring only to phonological memory. The authors suggest that it is necessary also to refer to phonological processing, which is representationally and computationally independent from memory. Children may find the CNRep difficult for perceptual reasons, for segmentation reasons, or because they struggle in creating phonological representations of the nonwords. This may be independent from phonological memory. Archibald and Gathercole (2006) explored this possibility, and showed that children with SLI experience greater difficulties in repeating the nonwords in the CNRep that contain phonological clusters, suggesting that phonological complexity has a role in predicting the difficulties of children with SLI with nonwords. More specifically, the authors point toward the impact of output speech motor processes, since they found the effect of cluster to apply only for children with SLI, and not in the control group.

Another critique of the CNRep can be found in the analysis of Dollaghan and Campbell (1998), who notice that the nonwords chosen for a well-balanced nonword repetition task should not contain syllables that appear in real words, nor should they form minimal pairs with real words. This is because wordlike stimuli might be more affected by lexical factors and might therefore influence children's performance, since children with larger lexica would be expected to perform better in tasks that contain nonwords with high neighbourhood density, even when their phonological working memory is limited. The CNRep does not, however, control for this confound, and some of the words in the test form minimal pairs with real words, such as *diller* vs *dealer* and *differ*, or *prindle* vs *Pringles*.

Another factor that requires to be better controlled in nonword repetition tasks is prosody (Roy & Chiat, 2004). It is well established that prosody plays a crucial role in early speech segmentation (Echols, 1996). Prosody is partially controlled for in the CNRep task, in the sense that stress appears in different positions across nonwords, and the position was chosen after checking the distribution of stress in real words of similar lengths. Nonetheless, it is worth noting that the majority of words in English are monosyllabic, and as such present stress in their only vowel. In addition, given a certain length, stress appears in different

positions depending on the word class (Roach, 2000), but this is a rather complex issue to control when using nonwords.

A more recent phonological analysis of the test reveals that the consistently reported length effect might be influenced by phonological complexity and not just be a reflection of increased working memory requirements. An inspection of the distribution of consonant clusters in the CNRep task shows that non-initial clusters are not balanced across syllable length (Cilibrasi, 2016). Specifically, these clusters appear only in long nonwords (4 and 5 syllable nonwords).

The present study builds on this analysis and focuses on the effect on repetition of non-initial clusters. Non-initial clusters are of particular interest because, according to several studies, these positions are psycholinguistically non-salient and thus lack confounds that may be present in initial positions (Beckman, 1998, Cilibrasi, 2016, Marshall & van der Lely, 2009, Smith, 2004). Clusters appearing at the beginning of a word may be repeated accurately, despite their inherent phonological complexity, because they are presented in a psycholinguistically salient position.

For example, Marshall & van der Lely (2009) found that both TD children and children with SLI and/or dyslexia repeated nonwords with clusters in medial position less accurately compared to nonwords with clusters in initial position, showing ceiling effects in stressed initial clusters. For this reason, we have chosen to focus on non-initial clusters rather than on clusters in general.

Non-initial clusters are only found in the four and five syllable nonwords in the CNRep test, and never in the two and three syllable nonwords. A Fischer's exact-test (chi-square could not be used because the count of cells is smaller than 5 in 4 cases) shows that the distribution of clusters in non-initial position is significantly unbalanced: $\chi(3) = 11.3$, $p = .004$ two tailed (as previously reported by Cilibrasi et al., 2015, with a similar analysis). Considering the unbalanced distribution of clusters presented above, it becomes difficult to assess whether the difficulties found in children with long nonwords in the CNRep are the result only of workingmemory limitations, only to the phonological complexity related to the presence of clusters or to a combination of these factors.

In order to investigate the potential impact of the unbalanced distribution reported in the CNRep, we analysed the assessment data of 18 children with language impairment and the data of 18 age-matched controls (based on clinical reports). The aim of this study is to test whether the presence of non-initial clusters has an negative impact on the repetition of 4 and 5-syllable nonwords in the CNRep, and whether the effect is mediated by the presence of a language impairment. Our hypothesis is that the presence of non-initial clusters makes repetition of nonwords more difficult, particularly for children with language impairment.

Methods

Design: We examined the contribution of non-initial clusters to the length effect observed in the CNRep by comparing children's performance in long nonwords containing non-initial clusters vs long nonwords without non-initial clusters. The sample consisted of two groups of children: typically developing

children and children with language impairment. The design of the study is thus 2 x 2 x 2 (cluster: yes/no x group: Control group/Clinical Group x length: 4/5 syllables). Age was additionally included in the model.

Participants: Data from 18 children with language impairment and 18 controls were used. The clinical group consisted of 10 boys and 8 girls, mean age 130 months, SD 33.16; the control group consisted of 10 boys and 8 girls, mean age 130 months, SD 32.23. Control children were chronological age-matched at the individual level to the children in the clinical group. All children were monolingual speakers of English.

The data of the children in the clinical group was available under the agreement of the Linguistic Assessment Clinic, stating that anonymised data collected in the University of Reading Clinic can be used for research purposes by academic staff. The original sample contained data from 22 children, but 4 children were excluded because their language difficulties had a non-linguistic origin (hearing loss, childhood stroke). During the revision of this paper 3 more children were excluded (due to ambiguous profiles) and were substituted with 3 children with language impairment from the sample of Loucas et al. (2010). The children seen as part of the Linguistic Assessment Clinic sessions were all currently on the case load of speech and language therapists. They were referred to the Linguistic Assessment Clinic because they presented with complex linguistic profiles and the speech and language therapists required further, more detailed assessments of the children's linguistic skills. All children had a history of speech and language difficulties and were attending language units and receiving speech and language therapy support. These children are representative of the heterogeneity and variability of developmental language disorders (Wells & Peppé, 2003, Bishop & Snowling, 2004). They were assessed with different standardised and non-standardised language tests, depending on their presenting symptoms. Despite this variability, all 18 children whose data are used for the purposes of this study met 3 crucial criteria:

1. They had language difficulties that were not caused by factors external to language (such as hearing loss or neurological factors)
2. They were all administered the CNRep test (Gathercole & Baddeley, 1996)
3. They were assessed with at least one other standardised language test, such as the Test for the Reception of Grammar-2 (Bishop, 2000), Clinical Evaluation of Language Fundamentals-4 (Semel, Wiig & Secord, 2004), Test of Word Finding (German, 1989), and performed at least 1 SD below the mean in at least one test.

All children had unimpaired non-verbal abilities. Detailed information about each participant is presented in Appendix.

The data of the 18 typically developing controls were provided by Lisa Archibald and Susan Gathercole (2006) and by Loucas et al. (2010). These data have already been described, and participants were categorised as typically developing children using standardised tests (specifically: Raven's Coloured Progressive Matrices, Goldman Fristoe Test of Articulation, British Picture Vocabulary Scale, Test for Reception of Grammar; Expressive Vocabulary Test, in Archibald and Gathercole, 2006; British Picture Vocabulary Scale, Clinical Evaluation of Language Fundamentals and the Wechsler Intelligence Scale for

Children in Loucas et al., 2010). Control children performed within 1SD from the mean for their age. The sample of Archibald and Gathercole (2006) contained 24 TD children, divided in two groups according to their age. We initially selected only the group that had a mean age closer to that of our clinical sample. We then added three children (during the revision of the manuscript) from the other group, who were age-matched individually to 3 more children from our clinical sample (total of 15 children). During the second revision, 2 children were excluded as a result of exclusion of their age-matched counterparts from the clinical group. The data of the remaining five children in our control group come from Loucas et al. (2010), and were also chosen as individual age matches to children in the clinical group. The final group tested therefore included 18 children with language impairment individually age matched to 18 typically developing children.

Stimuli: We noted above that non-initial clusters are not systematically distributed across item-lengths in CNRep, and that this makes results difficult to interpret since both clinical and non-clinical groups tend to find it difficult to process consonant clusters in a non-initial position. To test whether a word position effect was present (and overlooked) in the clinical assessment, we analysed the distribution of errors in the CNRep reports of the 18 children with language impairment, and compared their performance to that of the 18 age-matched controls. The stimuli used in our analysis are the 4 and 5-syllable nonwords used in the CNRep test (Gathercole & Baddeley, 1996). As discussed above, non-initial clusters are not present at all in the CNRep in words composed of 2 and 3 syllables. Fivesyllable words offer a clear ground for analysis, since half of the tokens contain a non-initial cluster, and half do not.

In 4 syllable words, the distribution of non-initial clusters is unbalanced: 4 words out of 10 contain non-initial clusters, and 6 words do not. For this reason, we coded the responses of the children as the proportion of correct answers. Other phonological properties are well controlled in the task. Specifically, nonwords are matched for syllabic length (Gathercole & Baddeley), and the task was created so that the stress patterns of each nonword would be the most appropriate for a nonword of that length (Gathercole et al., 1991). Wordlikeness was evaluated during the creation of the task (Gathercole et al., 1991 and was found to range from an average of 1.1 for /loddernapish/ to an average of 3.95 for /defermication/, independently of word length. In addition, wordlikeness measures correlated with neighborhood density. The fact that the test is well controlled on many phonological properties makes the analysis of non-initial clusters particularly reliable.

It should be noted that the 5 syllable nonword, “pristoractional”, has been categorized as belonging to the group without clusters for syllabification reasons. This is because most speakers would consider the phonemes creating the cluster as belonging to two different syllables, as in the real word “reaction”, syllabified “re-ac-tion” (Fallows, 1981, Treiman & Danis, 1988, Treiman & Zukowski, 1990). The main reason for this syllabification lies in the phonotactic constraints of English: the sequence of sounds /kj/ is not allowed either as a syllable coda or as a syllable onset in English. This is demonstrated by the fact that there are no real words starting or ending with the sequence /kj/. This claim can also be tested using a phonotactic probability calculator, such as Vitevitch and Luce (2004), which shows that the phonotactic

probability of /kj/ in word initial or in word final position equals “0”. In comparison, the clusters in the other nonwords belong to categories that speakers consistently process as units. Specifically, the other clusters are mostly composed of a plosive + liquid consonant. The consonants in these types of clusters are processed as belonging to the same syllable onset by English native speakers (Roach, 2000), i.e. they are not split between two syllables.

Results

Descriptive statistics for accuracy in nonwords with and without clusters for the two groups is presented in Table 1.

Table 1. Descriptive statistics for the two groups in nonwords with and without clusters

	TD - Mean (SE)	SD	LI - Mean (SE)	SD
with cluster	0.41 (0.04)	0.26	0.59 (0.05)	0.34
without cluster	0.36 (0.04)	0.29	0.46 (0.05)	0.32

Note: TD-typically developing, LI-Language Impaired

The assumptions of normality and homogeneity of variance were checked. The repetition data were normally distributed in the clinical group, $KS(36)=.109$, $p > .05$, and skewed to the left in the control group, $KS(36)=.178$, $p = .01$. According to several authors (Glass et al. 1972, Harwell et al. 1992, Lix et al. 1996, Feir-Walsh & Toothaker, 1974; Havlicek & Peterson, 1977), these kinds of violations are not problematic for ANOVA. The homogeneity of variance assumption was tested using Levene’s tests and the results were not significant ($p > .05$ for no-cluster, $p > .05$ for cluster), so homogeneity of variance can be assumed.

A four-way ANOVA was performed, including group, length, age and cluster as factors. The ANOVA revealed a main effect of group, $F(140,1) = 10.13$, $p = .001$, a main effect of age $F(140, 1)$, $F = 22.84$, $p < .001$, and an interaction between length and cluster, $F(140, 1)= 4.28$, $p = .040$. No other main effect or interaction reached significance. Post-hoc tests showed that children with language impairment were overall less accurate in repetition than TD children, $t(142) = 2.84$, $p = .005$ (see Figure 1) and age had an overall positive effect on performance with decreasing errors in repetition with increasing age (see Figure 2), as shown by the negative estimate (estimate = -0.004). The effect of cluster only reached significance in 5 syllable nonwords: $t(35) = -2.64$, $p = .01$ for 5 syllable nonwords, $t(35) = -1.03$, $p > .05$ for 4 syllable nonwords (see Figure 3).

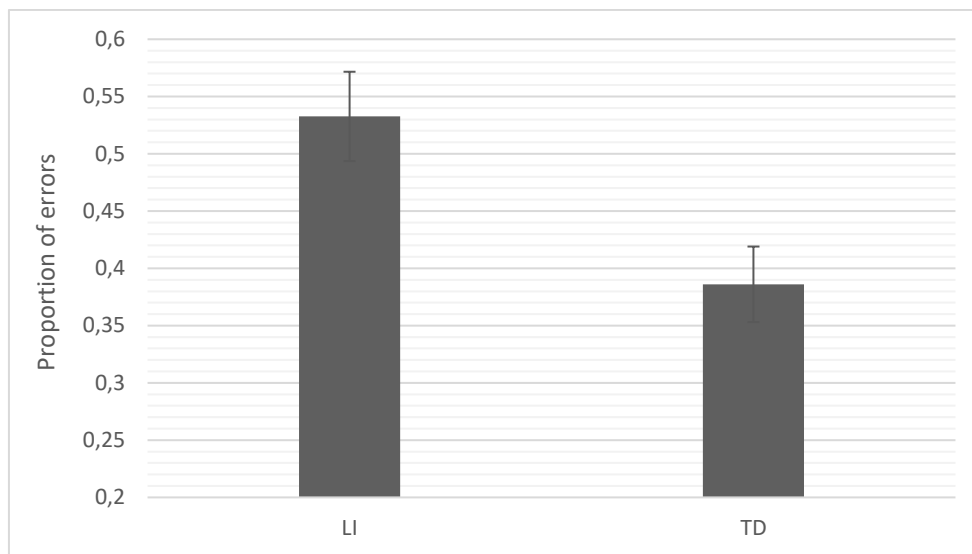


Figure 1. Proportion of errors across different groups. This figure shows the presence of a main effect of group. Children with language impairment made overall a larger number of errors than typically developing children. The histogram plots proportion of errors, rather than raw number of errors. The proportion of errors was obtained dividing the raw number of errors by the number of items repeated.

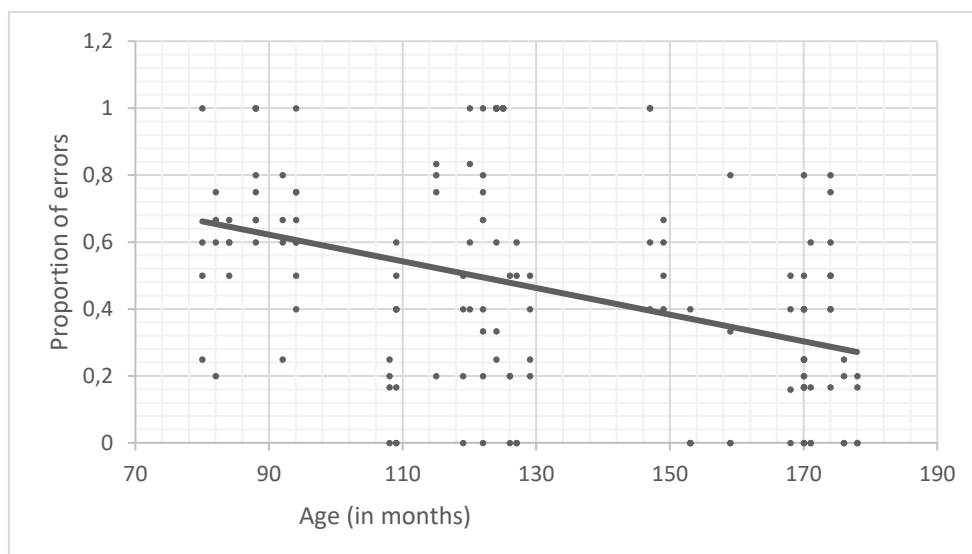


Figure 2. Proportion of errors across different ages. This figure represents the main effect of age. Overall, as age increases, the number of errors gradually decreases. The graphs plots proportion of errors, rather than raw number of errors. The proportion of errors was obtained dividing the raw number of errors by the number of items repeated.

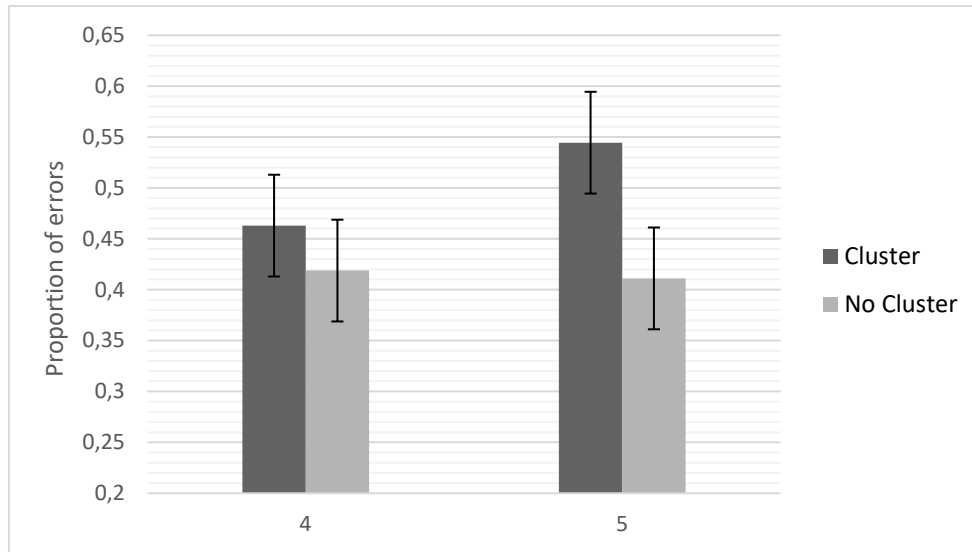


Figure 3. Proportion of errors in nonwords with and without clusters across different word lengths. This figure represents the interaction between length and cluster. While the effect of cluster is significant in 5 syllable nonwords, it does not reach significance in 4 syllable nonwords. The histogram plots proportion of errors, rather than raw number of errors. The proportion of errors was obtained dividing the raw number of errors by the number of items repeated (for each condition).

In order to capture individual variation in the model, data were further analysed with a linear mixed model (LME) that included random effects of participant (Baayen et al., 2008). This analysis revealed a main effect of cluster, $F(140, 1) = 5.139$, $p = .026$, a main effect of age, $F(140, 1) = 7.51$, $p = .009$, and an interaction between cluster and age, $F(140, 1) = 6.53$, $p = .012$. No other main effects or interactions reached significance. Post-hoc tests showed that nonwords containing clusters were repeated overall less accurately than nonwords without clusters, $t(71) = -2.67$, $p = .009$. As in the previous analysis, increasing age results in a significant improvement in accuracy. The sign of the estimate of the interaction (+0.001) shows that the effect of cluster is mediated by age: it is absent in young children and it becomes gradually larger as age increases.

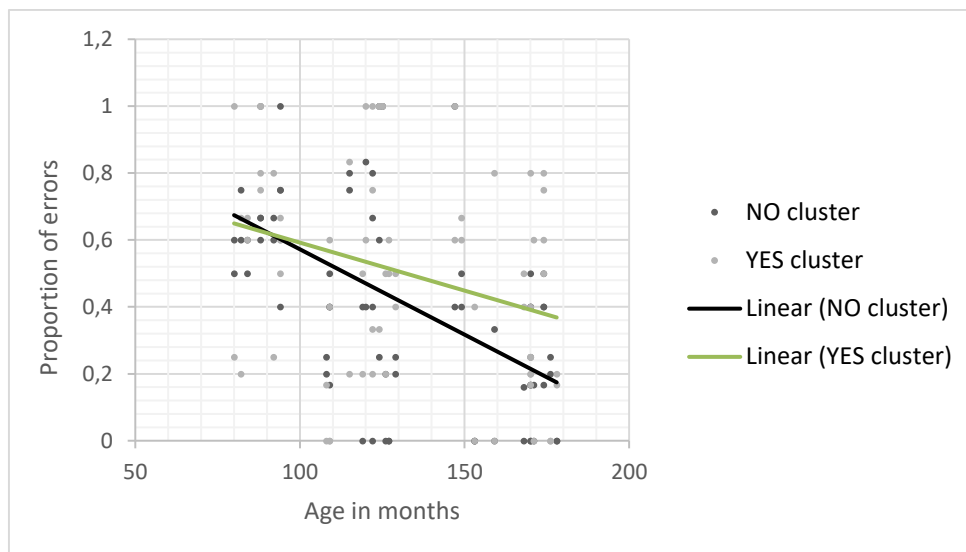


Figure 4. Proportion of errors in nonwords with and without clusters across different ages. This figure represents the main effect of cluster and the interaction between age and cluster. Nonwords without clusters were repeated overall more accurately than nonwords with clusters, as shown by the position of the grey line, above the black line. The interaction is represented by the different slope of the two lines: Performance improved with age more steadily in nonwords without clusters than in nonwords with clusters: While the black line shows a steep descent with the increase in age, the descent of the grey line is moderate.

The main effect of cluster and the interaction of age and cluster are shown in Figure 4. The fact that the line reporting accuracy for nonwords with clusters is nearly exclusively above the other line (which reports accuracy for nonwords without clusters) indicates that nonwords with clusters were repeated overall less accurately than nonwords without clusters. In the meantime, the fact that the slope of the line representing nonwords without clusters is steeper in its descent than the other line indicates that the improvement in repeating with nonwords without clusters found with increasing age is significantly larger than the improvement found in repeating nonwords with clusters. While children of 100 months of age (8 yrs 4 months) make a similar number of errors in words with and without clusters, as age increases the number of errors made reduces more for words without clusters than for words with clusters. Thus, older children show a significant difference in errors made, with more errors made in words with clusters.

Discussion

The aim of the study was to better understand whether the presence of non-initial clusters plays a role in the length effect observed in the CNRep test. It was hypothesised that long nonwords containing non-initial clusters would be more difficult to repeat than long nonwords that do not contain non-initial clusters, particularly for children in the clinical group. Our data analysis shows that all children experienced more difficulties with repeating long nonwords that contained a non-initial cluster, whether they had a language impairment or not.

The results presented in this paper, when combined with the analysis of the distribution of clusters in nonwords within the CNRep test presented in Cilibrasi et al. (2015) may help us to interpret what is measured by the CNRep test. The results presented in this study suggest that the presence of non-initial clusters contributes to the number of errors made by children in long nonwords. Since the non-initial clusters are only found in nonwords of 4 and 5 syllables, children might make more errors in longer words either because they are long or because they contain non-initial clusters. While the possible contribution of clusters in the patterns of performance observed in the CNRep task was previously reported by the authors of the test in a reanalysis of the test (Archibald & Gathercole, 2006), our analysis is novel because we also make an association between the pattern of performance and the distribution of clusters across different word lengths.

It should be noted that we found that the effect of cluster interacted with length: it was larger in 5 syllable nonwords, and smaller in 4 syllable nonwords. Post-hoc analysis shows that, when taking the two subsets in isolation, the comparison between presence and absence of cluster is significant only in 5 syllable nonwords. This difference across number of syllables may be due to lack of power, or it might be a genuine effect of length: the longer the nonwords, the larger the importance of non-initial clusters. Further studies with purposely developed tasks may answer this question.

The validity of the assessment as an important contributor to a battery of tests designed to identify children with language impairment is not questioned: the test was previously evaluated on a large number of participants and revealed to be a powerful diagnostic tool (Conti-Ramsden & Hesketh, 2003). Our study is in line with this work: children with language impairment performed worse than controls in our analysis (the effect of group was significant in the ANOVA analysis). However, the unbalanced distribution of clusters in 4-syllable words, as well as the absence of a natural way of controlling for this, requires attention. Thus, our results suggest that part of the word length effect often observed might be a consequence of phonological complexity rather than just working memory.

Other relevant patterns emerged once we included the random effect of participant in the model: this analysis showed that the effect of cluster interacts with age. Specifically, the effect of cluster appears to be absent in very young children and it gradually increases with age. This interaction reflects impact of age in the two different types of nonwords: in nonwords with clusters, an increase in age corresponds only to a modest increase in performance; in nonwords without clusters, however, an increase in age corresponds to a large increase in performance. In other words, while nonwords without clusters gradually become easier as children grow older, nonwords with clusters remain difficult. This finding is consistent with previous

literature on the processing of clusters. Ebbels et al. (2012), for example, showed that adolescents with language impairment experience long-standing difficulties with phonological clusters, and these difficulties remain evident up until the age of 15. Considering our findings, it appears important to further investigate whether the word length effects observed in CNRep result only from individual differences in working memory effects, whether they are a consequence of phonological complexity, or if both of these effects are affecting performance. It also is important to determine how these phenomena interact with age in order to properly diagnose language impairment in children of different ages.

Our study complements the work of Archibald and Gathercole (2006) and Cilibrasi et al. (2015) and it further explores the issues raised by Snowling et al. (1991). As Snowling et al. (1991) pointed out, it is clear that any analysis of the CNRep task must take into account phonological processing and not uniquely phonological working memory. Our result is in line with the work of Archibald and Gathercole (2006) who showed that nonwords containing clusters in the CNRep are more difficult for children with SLI, and has the additional advantage of avoiding the potential confound of initial clusters (as suggested by Cilibrasi, 2016). Contrary to Archibald and Gathercole (2006), we found the effect of cluster to hold for both children with and without language impairment. This may be a consequence of the fact that we focused on different kinds of clusters, or it may be a consequence of the fact that the samples in the two studies differed in age.

Our pattern of results, as well as the analysis of the distribution of the clusters, should be taken into account in future updates of the CNRep, or indeed when developing new nonword repetition tasks. Our results show that when creating nonword repetition tasks, it is important to balance the distribution of non-initial clusters. Particularly, it is important to include non-initial clusters also in nonwords which are 2 and 3 syllables in length. By balancing the distribution of clusters, one can be more confident about the measure of length effects. In the absence of a balanced distribution of non-initial clusters the risk of overestimating length effects is increased.

Conclusion

Our analysis shows that, when comparing the repetition of long nonwords (4-5 syllable) with and without non-initial clusters, both children with language impairment and TDs are more likely to make an error if the nonword contains a non-initial cluster. In addition, our analysis shows that, when random effects of participants are included, age interacts significantly with the cluster effect: more specifically, the effect of cluster appears absent in young children and it gradually increases with age. Performance on repeating nonwords without clusters increases steadily with age, while performance on repeating nonwords that contain clusters undergoes a more moderate improvement. These findings do not question the reliability of the CNRep as a contributor in the assessment of language disorders, but they suggest further ways of interpreting its outputs: Length effects in the CNRep are normally interpreted as working memory effects, but the fact that non-initial clusters appear only in longer (i.e. 4- and 5- syllable) nonwords, and the fact that nonwords with these clusters are repeated less accurately, suggest that phonological complexity may be interacting with the length effect, together with working memory.

Acknowledgements

This project was funded by the University of Reading, Faculty of Social Sciences studentship 21012649, and the revision was completed under the funding of the De Vincenzi Foundation postdoctoral fellowship. We are very thankful to both these funding bodies.

We are also very thankful to Lisa Archibald and Susan Gathercole for sharing their TD data with us.

References

- Archibald, L. M., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, 49(5), 970-983.
- Archibald, L. M., & Gathercole, S. E. (2007). Nonword repetition in specific language impairment: More than a phonological short-term memory deficit. *Psychonomic bulletin & review*, 14(5), 919-924.
- Baddeley, A. (1992). Working memory. *Science*, 255(5044), 556-559.
- Baddeley, A. D. (2000). Short-term and working memory. *The Oxford handbook of memory*, 77-92.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience*, 4(10), 829-839.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4), 390-412.
- Beckman, J. N. (2013). Positional faithfulness: an Optimality Theoretic treatment of phonological asymmetries. Routledge.
- Bishop, D. V. (1989). Test for the Reception of Grammar:(TROG). Medical Research Council.
- Bishop, D. V., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *Journal of child Psychology and Psychiatry*, 37(4), 391-403.
- Bishop, D. V., Adams, C. V., & Norbury, C. F. (2004). Using nonword repetition to distinguish genetic and environmental influences on early literacy development: a study of 6-year-old twins. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 129(1), 94-96.
- Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin*, 130(6), 858-886.
- Botting, N., & Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment (SLI). *International Journal of Language & Communication Disorders*, 36(4), 421-432.
- Cilibrasi, L., Stojanovik, V., & Riddell, P. (2015). Word Position and Stress Effects in Consonant Cluster Perception and Production. *Dyslexia*, 21(1), 50-59.
- Cilibrasi L. (2016). Word position effects in speech perception. PhD dissertation. University of Reading.
- Conti-Ramsden, G., & Hesketh, A. (2003). Risk markers for language difficulties: a study of young language-learning children. *International Journal of Language & Communication Disorders*, 38(3), 251-263.
- Di Simoni, F. (1978). The Token Test for Children (TTC). DLM Teaching Resources.

- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41(5), 1136-1146.
- Dunn, L. M., & Dunn, D. M. (2009). *The British Picture Vocabulary Scale (BPVS)*. GL Assessment Limited
- Ebbels, S. H., Dockrell, J. E., & van der Lely, H. K. (2012). Non-word repetition in adolescents with specific language impairment (SLI). *International journal of language & communication disorders*, 47(3), 257-273.
- Echols, C. H. (1996). A role for stress in early speech segmentation. In Morgan J. L. & Demuth K. (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 151–170). Mahwah, NJ: Erlbaum.
- Estes, K. G., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177-195.
- Fallows, D. (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics*, 17(2), 309-317.
- Feir-Walsh, B. J., & Toothaker, L. E. (1974). An empirical comparison of the ANOVA F-test, normal scores test and Kruskal-Wallis test under violation of assumptions. *Educational and Psychological Measurement*, 34(4), 789-799.
- Frederickson, N., Frith, U., & Reason, R. (1997). *Phonological Assessment Battery: (PHAB): Manual and Test Materials*. NFER Nelson..
- Gallon Stuart, N., & Marshall, C. (2009) Using Non-word Repetition Tasks with Children with Dyslexia, *Dyslexia Review*, 21, 1, 4-8.
- Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, 23(1), 83-94.
- Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of nonword repetition*. Psychological Corporation.
- Gathercole, S. E., Willis, C., Emslie, H., & Baddeley, A. D. (1991). The influences of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12(03), 349-367.
- Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. *Memory*, 2(2), 103-127.
- German, D. J. (1989). *Test of Word Finding (TWF)*. Pro-Ed.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of educational research*, 42(3), 237-288.

Gray, S. (2003). Diagnostic accuracy and test–retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of communication disorders*, 36(2), 129-151.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one-and two-factor fixed effects ANOVA cases. *Journal of educational statistics*, 17(4), 315-339.

Havlicek, L. L., & Peterson, N. L. (1977). Effect of the violation of assumptions upon significance levels of the Pearson *r*. *Psychological Bulletin*, 84(2), 373.

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in nursing & health*, 31(2), 180-191.

Lix, L. M., Keselman, J. C., & Keselman, H. J. (1996). Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance *F* test. *Review of educational research*, 66(4), 579-619.

Marshall, C. R., & van der Lely, H. K. (2009). Effects of word position and stress on onset cluster production: Evidence from typical development, specific language impairment, and dyslexia. *Language*, 85(1), 39-57.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., & Hodges, J. R. (2000). Addenbrokes Cognitive Examination (ACE). *Neurology*, 55(11), 1613-1620.

Norbury, C. F., Bishop, D. V., & Briscoe, J. (2001). Production of English Finite Verb Morphology. A Comparison of language difficulties and Mild-Moderate Hearing Impairment. *Journal of Speech, Language, and Hearing Research*, 44(1), 165-178.

Reynell, J. K., & Gruber, C. P. (1985). *Reynell Developmental Language Scales*. Western Psychological Services.

Roach, P. (2010). *English Phonetics and Phonology Fourth Edition: A Practical Course*. Ernst Klett Sprachen.

Roy, P., & Chiat, S. (2004). A prosodically controlled word and nonword repetition task for 2-to 4-year-olds: Evidence from typically developing children. *Journal of Speech, Language, and Hearing Research*, 47(1), 223-234.

Schmider, E. Z., Danay, M., Beyer, E., & Bühner, L. M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 6, 147.

Semel, E. M., Wiig, E. H., & Secord, W. (2004). *CELF 4: Clinical Evaluation of Language Fundamentals 4 Screening Test*. Pearson, PsychCorp..

Smith, J. L. (2004). *Phonological augmentation in prominent positions*. Routledge.

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, 12(03), 369-373.

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. *Journal of Speech, Language, and Hearing Research*, 49(2), 219-236.

Loucas, T., Riches, N. G., Charman, T., Pickles, A., Simonoff, E., Chandler, S., & Baird, G. (2010). Speech perception and phonological short-term memory capacity in language impairment: preliminary evidence from adolescents with specific language impairment (SLI) and autism spectrum disorders (ASD). *International journal of language & communication disorders*, 45(3), 275-286.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of memory and language*, 27(1), 87-104.

Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language*, 29(1), 66-85.

Vitevitch, M. S., & Luce, P. A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods*, 36(3), 481-487.

Weismer, S. E., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 43(4), 865-878.

Wells, B., & Peppé, S. (2003). Intonation abilities of children with speech and language impairments. *Journal of Speech, Language, and Hearing Research*, 46(1), 5-20.

Wiig, E. H., & Secord, W. (1992). *Test Of Word Knowledge (TOWK)*. New York: Psychological Corporation.

Appendix: Participants with language impairment

SUBJ	ID	CA (sex)	CNRep raw Δ	TROG-2 SS	TTC raw	RDLS raw (SS)	CELF SS	ACE raw (SS)	TWF raw (SS)	TOWK raw (SS)	BPVS SS	PhAB
1	Ef1lac	10;4 (F)	15	55*				7 (3)* Naming 24 (8) Synt comp				
2	Mm2lac	7;4 (M)	12 Δ	55*			7 WC		16 (60)*		58*	
3	Af18lac	9;1(M)	30	76*			11 UP 3* WS					
4	Tm4lac	12;9(M)	38	92							<70*	
5	Wm5lac	14;2(M)	34	81*			7 UP					
6	Ef6lac	14;6(F)	29							36 (10) rec 21 (5)* exp		
8	Lf8lac	7;4 (F)	18 Δ	55*				5 (4)*Naming			85	
9	Gm9lac	10;7(M)	33		45 (IV-V*)			9 (8) Inf comp				138 (72)* Pic. naming
10	Cm10lac	6;10 (M)	25 Δ			55 (78)*	6* RS			68 (101)		
11	Sf11lac	10;0 (F)	21	65*				19 (6)*Syntactic formulation				
12	Om12lac	7;8 (M)	17 Δ	88			79*○	12 (10) Naming			100	
13	Cm13lac	10;2(M)	16				5* C&FD		41 (68)*		88	
14	Sf14lac	10;5 (F)	2	78*			7 WC1 8 WC2				71*	
15	Lf15lac	14;3 (F)	34	81*			7 RS				74*	

18	Jm3lac	9;1(F)	33	81*	8 RS	8 (81)* Rhyme
1R	1354mdlou	14;02(M)	25		6* CD 6* RS	81*
2R	7005tdlou	13;03(M)	31		4* CD 3* RS	73*
3R	7008khlou	14;06(F)	32		4* CD 3* RS	87

Note: *-more than 1SD below the mean; CA-chronological age; CNRep-Children's Test of Nonword Repetition; TROG- Test for the Reception of Grammar; TTC – Token Test for Children; RDLS – Reynell Developmental Language Scales; CELF-Clinical Evaluation of Language Fundamentals (WC-word classes; UP-understanding paragraphs; RS-recalling sentences; C&FD- Concepts and Following Directions); ACE-Assessment of Comprehension and Expression (Synt comp-Syntactic Comprehension; Inf comp-inferential comprehension); TWF-Test of Word Finding; TOWK-Test of Word Knowledge; BPVS-British Picture Vocabulary Scale (short or long form); PhAB-Phonological Assessment Battery

○ this is a standard score obtained by summing up the standard scores of the core subtests and converting the value on a 100 point scale. The value obtained by Om12lac is 1.55 SD below the mean for his age (8th percentile)

△ Although the test can be used to assess older children with language related learning difficulties (www.pearsonclinical.co.uk), standard scores are available only for children aged 4-8 (Gathercole & Baddeley, 1996). For this reason, only children within this age range were flagged if they performed 1SD below the mean for their age.

