



Original Articles

The influence of children's exposure to language from two to six years: The case of nonword repetition



Gary Jones*

Nottingham Trent University, UK

ARTICLE INFO

Article history:

Received 28 November 2014
 Revised 26 April 2016
 Accepted 26 April 2016
 Available online 4 May 2016

Keywords:

Language exposure
 Nonword repetition
 Statistical learning
 Phonological sensitivity
 CLASSIC
 Phonological working memory

ABSTRACT

Nonword repetition (NWR) is highly predictive of vocabulary size, has strong links to language and reading ability, and is a clinical marker of language impairment. However, it is unclear what processes provide major contributions to NWR performance. This paper presents a computational model of NWR based on Chunking Lexical and Sub-lexical Sequences in Children (CLASSIC) that focuses on the child's exposure to language when learning lexical phonological knowledge. Based on language input aimed at 2–6 year old children, CLASSIC shows a substantial fit to children's NWR performance for 6 different types of NWR test across 6 different NWR studies that use children of various ages from 2;1 to 6;1. Furthermore, CLASSIC's repetitions of individual nonwords correlate significantly with children's repetitions of the same nonwords, NWR performance shows strong correlations to vocabulary size, and interaction effects seen in the model are consistent with those found in children. Such a fit to the data is achieved without any need for developmental parameters, suggesting that between the ages of two and six years, NWR performance measures the child's current level of linguistic knowledge that arises from their exposure to language over time and their ability to extract lexical phonological knowledge from that exposure.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Vocabulary acquisition is an essential part of language learning, enabling the child to build a lexicon that can be used by other processes such as sentence production. Vocabulary size can be indexed by performance on nonword repetition (NWR), a simple task whereby children repeat aloud nonwords that are spoken to them. Although children's NWR performance has very strong links with vocabulary learning in particular (e.g., [Baddeley, Gathercole, & Papagno, 1998](#); [Gathercole, 2006](#); [Hoff, Core, & Bridges, 2008](#)), it is also predictive of general language ability (e.g., [Marton & Schwartz, 2003](#); [Thal, Miller, Carlson, & Vega, 2005](#)), reading success (e.g., [Hansen & Bowey, 1994](#); [Kamhi & Catts, 1986](#)) and difficulties with language or reading (e.g., [Bishop, North, & Donlan, 1996](#); [Montgomery, 1995](#); [Snowling, Goulandris, Bowlby, & Howell, 1986](#)). Performance on NWR tests therefore capture key mechanisms that are involved in the child's vocabulary learning that ultimately influence language acquisition more generally. However, the underlying processes involved in repeating nonwords are quite broad ([Bowey, 2001](#); [Coady & Aslin, 2004](#);

[Snowling, Chiat, & Hulme, 1991](#)), leading to differing accounts of what NWR actually measures. Resolving this issue is the focus of the current paper.

Phonological working memory is seen as playing a pivotal role in vocabulary learning because in order to repeat a sequence of sounds one must first be able to store the sequence in temporary memory ([Baddeley et al., 1998](#)). The dominant view of NWR (see [Melby-Lervåg et al., 2012](#)) is that it is a pure measure of phonological working memory. Under this explanation, differences in NWR performance that are seen within and across ages is largely due to differences in phonological working memory (e.g., [Baddeley et al., 1998](#); [Gathercole & Baddeley, 1989](#)). The phonological working memory account also explains robust length effects that are found in NWR whereby long nonwords are consistently repeated less accurately than short nonwords (e.g., [Gathercole & Baddeley, 1989](#); [Jones, Tamburelli, Watson, Gobet, & Pine, 2010](#)). Nevertheless, this explanation is somewhat confounded by NWR also showing strong links to long-term lexical phonological knowledge, defined here as knowledge of the individual sounds, sound sequences and lexical items of the native language. For example, children repeat nonwords that are judged as wordlike more accurately than nonwords that are not judged as wordlike (e.g., [Gathercole, 1995](#); [Munson, Kurtz, & Windsor, 2005](#)) and similarly nonwords constructed from phoneme sequences that occur

* Address: Division of Psychology, Nottingham Trent University, Chaucer Building, Nottingham NG1 4BU, UK.

E-mail address: gary.jones@ntu.ac.uk

frequently in the native language are repeated more accurately than nonwords constructed from relatively infrequent phoneme sequences (e.g., Jones et al., 2010; Munson, 2001).

An alternative view of NWR is that it measures phonological sensitivity because children's performance on phonological awareness tasks has been shown to be more predictive of vocabulary size than NWR (e.g., Metsala, 1999). Children's vocabulary learning is seen by many to begin with holistic forms (e.g., Fowler, 1991a, 1991b; Storkel, 2002; Treiman & Breaux, 1982; Vihman & Velleman, 1989; Walley, 1993). Later restructuring of lexical items to include segmental detail is driven by a need to have a more fine-grained account of similar words in order for them to be differentiated (Charles-Luce & Luce, 1990; Metsala & Walley, 1998; Walley, 1993). Consistent with this account, Metsala found superior phonological awareness for familiar over unfamiliar words and for words from dense over sparse neighborhoods. The segmental detail of words from dense neighborhoods, because their characteristics overlap with other words, is likely to be learned more quickly than words from sparse neighborhoods (see also Edwards, Beckman, & Munson, 2004). Since familiar words tend to be from dense neighborhoods (Vitevitch & Luce, 1998), segmental detail will also be learned more quickly for familiar over unfamiliar words. This contrasts with a phonological working memory account because it suggest that the driving force in NWR performance stems from elaborating long-term linguistic knowledge rather than constraints on temporary storage of information. However, it is unclear how the phonological sensitivity account explains length effects that are routinely seen in NWR performance (but see Metsala & Chisholm, 2010, for discussion on this point).

Both of these accounts either implicitly or explicitly recognize the role of the child's exposure to language. For phonological working memory, findings such as greater repetition accuracy for word-like nonwords over non-wordlike nonwords suggest that exposure to language must influence the NWR process. Combined with theoretical positions that suggest the effective size of phonological working memory is influenced by long-term knowledge (e.g., Miller, 1956; Gobet et al., 2001; Cowan, 2001), one could argue whether NWR truly measures phonological working memory or whether it is a reflection of the child's current level of linguistic exposure (see also Gupta, Lipinski, Abbs, & Lin, 2005; Snowling & Hulme, 1994). For the phonological sensitivity account, holistic representations of words are elaborated based on their similarity to other words, a process that is driven by increased exposure to language. Empirical investigations of vocabulary learning have also found a major role for language exposure. For example, Fernald and colleagues have shown that children who receive extensive child-directed speech or diversity in their language input have larger vocabularies than children who do not (e.g., Hurtado, Marchman, & Fernald, 2008; Weisleder & Fernald, 2013); Hoff and Naigles (2002) highlight the quantity and richness of the input, suggesting language exposure may play a greater role than social factors in children's language learning; Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991) show that individual differences in children's vocabulary growth are linked to the amount that is spoken to them by their mother; and numerous computational accounts have used language input to simulate a range of language phenomena (e.g., Brown, Preece, & Hulme, 2000; Goldwater, Griffiths, & Johnson, 2009; Hartley & Houghton, 1996; Monaghan & Ellis, 2010; Perruchet & Vinter, 1998).

Although the child's exposure to language could be seen as supporting the phonological sensitivity account, this explanation suggests that NWR performance is influenced by phonological information that emerges from restructuring of the lexical item. There is now sufficient evidence to challenge this view, suggesting that children's segmental knowledge is present from a very early age (e.g., Basirat, Dehaene, & Dehaene-Lambertz, 2014; Coady &

Aslin, 2003, 2004; Gervain, Macagno, Cogoi, Pen, & Mehler, 2008; Yoshida, Fennell, Swingley, & Werker, 2009).

There are two key computational accounts of NWR that learn from their exposure to language and capitalize on research that supports a bottom-up approach.¹ Gupta and Tisdale (2009) adapted a neural network model of serial order by Botvinick and Plaut (2006) using as input 125,000 syllabified words. Long-term knowledge (the patterns of weights across the units of the network) represented gradually more detailed phonological representations of the individual syllabified words. This interacted with phonological working memory (the temporary activations across units) such that over time, longer words and nonwords were able to be recalled. The model showed differences in NWR performance for nonwords of different lengths, for high and low phonotactic probability nonwords, and for different levels of exposure to input – effects that are also seen in children's NWR performance.

Jones and colleagues have used an alternative modeling environment originally labelled EPAM-VOC (Jones, Gobet, & Pine, 2007) but later given the more meaningful acronym CLASSIC (Chunking Lexical and Sublexical Sequences in Children, Jones, Gobet, Freudenthal, Watson, & Pine, 2014).² As the name suggests, this account is very much embedded in chunking (e.g., Cowan, 2001; Gobet et al., 2001; Miller, 1956) and chunk-based modeling environments (e.g., French, Addyman, & Mareschal, 2011; Servan-Schreiber & Anderson, 1990) whereby larger units of information are learned over time. CLASSIC uses phonemically-coded large-scale naturalistic input aimed at young children (e.g., mother's utterances) and learns increasingly larger phoneme sequences from an input that is constrained by phonological working memory. The model again captures many of the NWR effects seen in children, such as nonword length, wordlikeness, and age differences.

The Gupta and Tisdale (2009) model explicitly targets the phonological working memory and phonological sensitivity accounts, showing how both may apply in the NWR process. The Jones et al. (2007) model on the other hand targets the link between long-term knowledge and phonological working memory in the NWR process. Nevertheless, both accounts illustrate how exposure to language is potentially a critical factor in NWR performance. However, because the simulations in both models are largely qualitative, the extent to which exposure to language can explain NWR performance at the empirical level is left unanswered.

The purpose of this paper is twofold. First, to focus on exposure to language by using large-scale naturalistic language input aimed at children between the ages of 2 and 6 years within a model that does not utilize any developmental parameters. By keeping all parameters constant, the only "developmental" change is the linguistic knowledge that the model learns, which increases with greater exposure to language. Any differences in NWR performance over time are therefore caused by the learning that takes place on the language input rather than from developmental changes per se. Second, to provide an extensive examination of the fit between model and child by using 6 different NWR studies involving children between the ages of 2 and 6 years. If the model is able to provide both qualitative and quantitative fits to the majority of the child data, it would provide strong evidence that NWR performance is a measure of the child's current level of linguistic knowledge that is accrued from exposure to language and is not a reflection of any mechanistic developmental change.

¹ The computational accounts also address one of the problems inherent in verbal explanations of NWR: phonological working memory and long-term linguistic knowledge interact with one another (see also Chen & Cowan, 2005).

² This modeling environment is based on the same principles as MOSAIC (e.g., Freudenthal, Pine, Aguado-Orea, & Gobet, 2007; Freudenthal, Pine, Jones, & Gobet, 2015) and uses a similar input set. However, the Jones et al. model is based on phonological input and focuses on NWR performance whereas MOSAIC is based on lexical input and focuses on syntactic processing.

The remainder of this paper is as follows. First, we detail the modeling environment together with the parameter settings used. Second, we introduce a set of NWR studies to compare the model against, detail what baseline performance might be for these studies, and illustrate how the model improves on baseline performance by capturing the vast majority of children's repetition performance together with associated effects such as increases in vocabulary size and interactions that are present in the child data. Finally, we discuss how the model provides such a good fit to the children's data and why the knowledge learned in the model influences many of the processes involved in NWR.

2. Model

CLASSIC is based on previous work (e.g., Jones et al., 2007, 2014) but the phonological working memory component has been simplified for parsimony and to emphasize the role of exposure to language on NWR performance. Phonological working memory is therefore a means by which input to the model can be constrained and should not be viewed as a proposal for how phonological working memory should be construed in the developing child. The simulations presented below illustrate that a simplified view of phonological working memory does not affect model performance.

2.1. Learning long-term lexical phonological knowledge

The learning mechanism within the model is a simple sequence learner that learns new knowledge by joining adjacent items in phonological working memory (sequence learning of this nature has a wealth of support in developmental literature, e.g., Aslin, Saffran, & Newport, 1998; Newport & Aslin, 2004; Saffran, Aslin, & Newport, 1996; see also Altmann, 2002). Take as example the input utterance 'what's she doing?', the phonemic representation of which is /w ɒ t s/ /ʃ i:/ /d u: ɪ ŋ/, with the utterance being word-delimited. The model begins with the initial phonemes of the native language. The utterance is therefore represented in phonological working memory as a word-delimited sequence of ten individual phonemes (/w, ɒ, t, s/ /ʃ, i:/ /d, u:, ɪ, ŋ/) where order is preserved. The learning mechanism acquires new information by joining adjacent items without crossing word boundaries, unless the adjacent items are themselves words. New knowledge would therefore be learnt consisting of the sequences /wɒ/, /ɒt/, /ts/, /ʃ i:/, /du:/, /u:ɪ/, and /ɪŋ/. Note that as 'she' is only two phonemes in length, it has been learnt as a new word. If the same utterance appears a second time in the input, the model recognizes that its existing knowledge is able to represent the utterance using fewer items (/wɒ/, /ts/, /ʃ i:/, /du:/, and /ɪŋ/). The learning mechanism is then applied, adjoining adjacent items – again preserving word boundaries – such that 'what's' and 'doing' become new words. The only time that learning can occur across word boundaries is when the adjacent items are themselves words, which occurs if the same utterance appears a third time in the input. The model can now represent the utterance using only three items (/wɒts/, /ʃ i:/, and /du:ɪŋ/), and the phrases 'what's she' and 'she doing' are learnt.

The above example of learning only involves the same utterance occurring three times in the input. In reality, the language input is vast and varied, resulting in a broad range of linguistic knowledge being learned. Note that the model presented here will learn at every opportunity as per the example, but this is because of constraints on the amount of language input that is available (e.g., one-year-old children may be exposed to as many as half a million words in just a 3 week period, Swingley, 2007). One would naturally expect children to require several exposures before learning a linguistic input.

2.2. How phonological working memory constrains long-term lexical phonological knowledge

The number of items (or 'chunks') that can be represented in phonological working memory always averages 4.5 items. However, since each item can vary in the amount of information it holds, the *information* in phonological working memory subsequently varies (analogous to chunking, Miller, 1956). Averaging 4.5 items is achieved using a simple probabilistic sigmoidal function whereby each item has an associated probability of being accessed. The probability function is detailed in Fig. 1 and the effect this has on recall of items in phonological working memory is detailed in Fig. 2. As can be seen in the figures, phonological working memory capacity will average to be 4.5 items (Fig. 2 illustrates that for 9 items, 4.5 on average will be accessed). The capacity setting is merely a method by which the input is constrained in a fixed way and is not intended as a suggestion that a capacity of 4.5 items may be realized in young children. This is particularly important because phonological working memory interacts with learned knowledge, which for the current model is based on linguistic exposure that falls far below that of the developing child (i.e., a higher capacity setting may be required simply because learned knowledge will fall short of that of children).

Note that the sigmoidal probabilities give preference to accessing just-heard items (i.e., recency effects, see for example Baddeley, 1987). Although memory strategies such as rehearsal enable additional items to be recalled, young children fail to use such strategies (e.g., Gathercole & Adams, 1994; Gathercole, Adams, & Hitch, 1994).

Let us now repeat the example above but with the constraint of a phonological working memory capacity of 4.5 items. Table 1 illustrates this situation, showing that phonological working memory limits the amount of learning that can take place because learning a new item is dependent upon adjacent items in phonological working memory both being accessed on the basis of the probabilistic sigmoidal function.

Key to the model's performance is diversity in the linguistic input such that a range of long-term knowledge can be learnt. When only presented with 'what's she doing', the model can quickly learn a lot about this utterance but without diversity in the input it will fail to broaden its knowledge to apply to other linguistic inputs. Similarly, without chunked knowledge, the model's fixed phonological working memory will not be able to capitalize on its learning (e.g., initially, 'what's she doing' is represented using 10 chunks but this is dramatically reduced with chunking).

2.3. Input to the model and training regime

The model is trained on linguistic input that is directed at children between the ages of 2 and 6 years. The input is word-delimited, reflecting evidence that by the age of two, children are able to closely monitor phonetic information to identify word boundaries in continuous speech (e.g., Jusczyk & Aslin, 1995; Jusczyk, Houston, & Newsome, 1999; Saffran, 2001; Swingley, Pinto, & Fernald, 1999). Word-delimited input focuses on learning lexical phonological knowledge relating to the phoneme sequences that constitute novel words. While this ignores information that is relevant for identifying word boundaries (e.g., low co-occurrence of word-final and word-initial phonemes, see for example Hockema, 2006), since nonwords are always constructed using phoneme sequences that are attested in the native language, this should not be an issue for model performance. Nevertheless, we will compare word-delimited versus non-delimited input in the results section.

Spoken mother's utterances are taken from the 12 children of the Manchester corpus (Theakston, Lieven, Pine, & Rowland,

$$p = \frac{1}{1 + \exp(\text{distance of item from utterance-final item} - \text{phonological working memory capacity})}$$

Fig. 1. Probability of correctly representing a chunk in temporary memory.

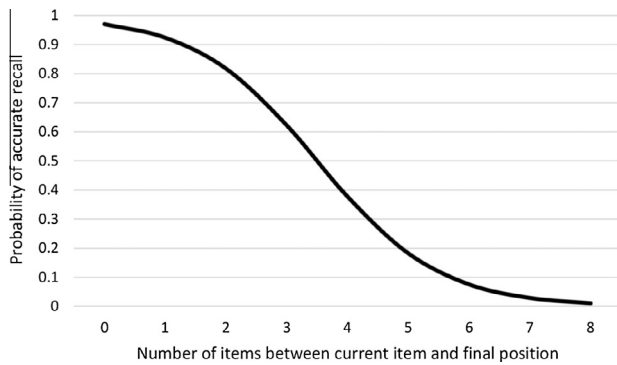


Fig. 2. Sigmoidal function indicating the probability of correctly representing an item in temporary memory. The figure shows the probability of accurate recall for each item position based on phonological working memory capacity being 4.5. A zero on the x-axis represents the item that is in the utterance-final position, 1 represents the item that immediately precedes the utterance-final item, and so on. Note an item can be an individual phoneme, phoneme sequence, word or phrase, depending on the model's learning.

Table 1

Hypothetical learning when the model is repeatedly exposed to 'what's she doing' with learning being constrained by phonological working memory. Items underlined are those that are not accessed based on the sigmoidal function from Fig. 2.

Input	Items placed in phonological working memory	New knowledge learned
/w ɒ t s/ ʃ i: /d u: ɪ ŋ/	<u>w</u> , <u>d</u> , <u>t</u> , <u>s</u> / ʃ i: /d, <u>u</u> , ɪ, ŋ [10 items, 5 accessed]	ʃ i:, ɪŋ
/w ɒ t s/ ʃ i: /d u: ɪ ŋ/	<u>w</u> , <u>d</u> , <u>t</u> , <u>s</u> / ʃ i: /d, u:, ɪŋ [8 items, 4 accessed]	du:, u:ɪŋ
/w ɒ t s/ ʃ i: /d u: ɪ ŋ/	<u>w</u> , <u>d</u> , t, s/ <u>ʃ</u> i: /du:, ɪŋ [7 items, 5 accessed]	ɒt, ts, du:ɪŋ

2001) on CHILDES (MacWhinney, 2000), which involve a total of approximately 34 h of mother-child recordings for each child taken over the period of one year when the children were 2–3 years of age. All maternal utterances (the 'younger input') are converted to a phonemic representation where possible (range 16,417–33,591 maternal phonemic utterances). Since the corpus involves children between the ages of 2 and 3, the input also includes the phonemic form of older children's spoken language (from 3;0 upwards for Forrester, Lara, Smith, and Thomas from the English UK transcripts on CHILDES, the 'older input'). The age of the oldest child in these transcripts is 4;11 hence the utterances were supplemented with a variety of children's literature aimed at children aged 4–6 years (e.g., Snow White, Alice in Wonderland), with each phonemically-coded sentence being the equivalent of an utterance. The ratio of spoken versus story book input was approximately 2:1. Information on the characteristics of the different inputs is given in Table 2, showing that MLU and vocabulary size are larger for the older input than the younger input even when matched for number of utterances.

Sets of input are created individually for each mother by randomly sampling utterances from the relevant mother and the older input while maintaining the absolute number of utterances produced by the mother. The model is presented with the input from beginning to end, one utterance at a time. Sampling is therefore biased such that early in the input, utterances are predominantly

Table 2

Mean number of utterances, unique words, total words, and MLU for the inputs used. Standard deviations are in parentheses.

Input	Utterances	Unique words	Total words	MLU
Younger input ^a (N = 12)	25,568 (6507)	2972 (593)	101,632 (29,722)	3.97 (0.38)
Older input	75,981	9966	336,170	4.40
Older input sampled (N = 10) to match younger input	25,568	5832 (70)	107,354 (667)	4.20 (0.03)
Input used (N = 10 for each mother)	25,568 (6507)	3303 (698)	107,264 (28,506)	4.18 (0.20)

^a These figures are based on an automatic phonemic conversion of the maternal utterances whereby utterances that could not be automatically converted to phonemes were discarded. The figures are therefore likely to be lower than those based on the original transcripts.

from the mother and late in the input it is predominantly from the older set, reflecting the gradual change in child-directed input with development. Ten input versions are created for each mother to ensure that results are not an artifact of a fortuitous input. For example, Anne's mother produces 33,390 utterances and therefore ten input files are created for Anne's mother, all containing 33,390 utterances but each different from the other. The characteristics of the input files is given in the lowest row of Table 2.

The model is run separately for each of the ten mother input files and for each of the 12 mothers. However, as shown above, although phonological working memory always averages 4.5 items it is probabilistic and therefore it is unlikely that the model will learn the same information when it is presented with the same input file. Each input file is therefore presented to the model 10 times, meaning that for each mother, there are 100 runs of the model (10 input files each presented 10 times).

2.4. Performing NWR

Performing NWR in the model is straightforward. The nonword is presented to the model as per any of the input utterances exemplified above (i.e., maintaining the serial order of items) and is converted to as few items as possible based on the lexical phonological knowledge that the model holds at that particular time. The probabilistic sigmoidal function is then used to identify whether or not all of the items in phonological working memory can be accessed correctly. If this is the case, the nonword is repeated correctly; if it is not the case a repetition error has occurred. Note that the NWR process itself involves accurate perception of the novel sequence of sounds, encoding of the sequence, temporal marking of the novel sequence, temporary storage of the sequence, speech/motor planning, and finally production of the sequence via articulatory mechanisms (e.g., Coady & Aslin, 2004). The model focuses on only a subset of these processes and as such should not be viewed as an instantiation of NWR in its entirety. This is a point to which we will return in the discussion because processes that lie outside of CLASSIC are largely influenced by the existing lexical phonological knowledge that is the focus of the model.

Whether or not a nonword is repeated correctly is probabilistic, because it relies on the sigmoidal function in phonological working memory. Each NWR test is therefore performed 10 times for every run of the model. In effect, this means that for each mother, a NWR test is completed 1000 times (10 input files * 10 presentations of each input file * 10 repetitions of every nonword).

Although there are 12 mothers, performance on any one nonword is averaged across the full set of repetitions for that nonword – i.e., across 12,000 repetitions (12 mothers * 1000 repetitions for each nonword). Child comparisons will be made to the ‘averaged model’ because there is likely to be a great deal of overlap in the utterances of each of the mother input files due to sharing the same older input.

3. NWR performance in the model compared to typically-developing children

The input to the model is based on UK input to children aged 2–6 years therefore we compare its performance to typically-developing children on six NWR studies that have been carried out in the UK within these age ranges and that include a list of stimuli used. These are detailed in Table 3. Of particular note is that: (1) the age of the children used is well dispersed across the age range of interest; (2) there are 38 datapoints to compare against; and (3) the nonword sets differ on various features such as length, structure, wordlikeness, and phonotactic probability. The characteristics of these studies are therefore suitably challenging for any explanation of NWR performance.

The sampling bias means that the input begins with language directed at children of 2;0 and ends with language directed at children of 6;0. This enables CLASSIC’s performance to be examined at varying levels of input. For example, NWR performance of 3-year-old children can be compared after 25% of the input has been

processed (an ‘input age’ of 3 years). Before doing so, however, the characteristics of the input will first be used to estimate a baseline performance that reflects the repetition accuracy expected of children if they simply ‘tuned in’ to the frequency characteristics that are present in the input.

3.1. Baseline NWR performance

How often an item appears in the input is reflected in subsequent performance on a test containing that item (e.g., Jones et al., 2010; Munson, 2001). Given that CLASSIC learns sequence information from phonological input, the most appropriate ‘items’ to examine are phonemes, biphones, and triphones. Baseline performance will therefore be established that accounts for the frequency by which phonemes, biphones, and triphones appear in the input at each of the relevant stages of input that correspond to a relevant child age (e.g. after 25% of the input for comparison to 3 year old children). This will be accomplished by use of a Monte Carlo style simulation where every instance of an item results in it being placed in a pool such that the pool is dominated by frequent items relative to infrequent ones. Accounting for frequency in this way should favor baseline performance because children repeat nonwords with high phonotactic probability or wordlikeness more accurately than nonwords low in phonotactic probability or wordlikeness.

For example, in the case of phonemes and in a comparison to 3 year old children, how often each phoneme is encountered in the first 25% of the input determines how many instances of that phoneme are placed in the pool. Accurate repetition of a nonword is reflected in the probability of drawing each of the constituent items from the pool (where the probability of drawing one correct item is the number of instances of the item divided by the total number of instances of all items in the pool). In the case of

Table 3

Pertinent details of the studies to which CLASSIC’s NWR performance is compared. For child details, TD indicates typically-developing (e.g., where children were ruled out on the basis of poor performance on hearing or language tests). For nonword types, nonword structure is denoted by either ‘CV’ (each consonant in all nonwords is always flanked by either a vowel or silence), ‘CC’ (all nonwords include at least two consonants in succession, or a ‘consonant cluster’), or ‘CCV’ (some nonwords are CV and some CC); and nonword characteristics are denoted by ‘WL’ if one or more nonwords are wordlike (i.e., contain lexical items or morphemes), ‘NWL’ if not, and ‘UNK’ if this information is not known. Unless stated otherwise, accuracy in the study was averaged across all nonwords. Standard deviations (when given for the study) are in parentheses.

Study	Age range for comparison to model	Nonword types	NWR performance (% in same order as nonword types)	NWR-vocabulary correlation
SK09 (Stokes & Klee, 2009)	2;1–2;5 (N = 172)	12 CV NWL (4 × 1, 2, and 3-syllables) ^a	2;1 59; 2;2 64; 2;3 68; 2;4 73; 2;5 75	Only given for late-talkers
GA93 (Gathercole & Adams, 1993) ^b	3;0 (N = 54)	5 CCV UNK 2-syllable, 5 CCV WL 3-syllable	74.80 (27.60); 52.60 (29.60)	$r = 0.34, p < 0.05$
RC04 (Roy & Chiat, 2004)	2;5 TD (N = 27); 3;5 TD (N = 39)	12 CV WL ^c (6 × 1-syllable, 6 × 2-syllable), 6 CCV UNK 3-syllable	2;5: 78.33 (18.33), 61.67 (28.33), 41.67 (30.00); 3;5: 88.33 (15.00), 75.00 (18.33), 56.67 (25.00)	$r = 0.45, p < 0.001$ (N = 66)
GB89 (Gathercole & Baddeley, 1989)	4;7 (N = 134); 5;7 (N = 134)	15 CV WL ^d (5 × 2, 3, and 4-syllable); 15 CC WL (5 × 2, 3, and 4-syllable)	4;7: 77, 57, 33 and 63, 53, 28; 5;7: 90, 75, 52 and 84, 74, 58	4;7 $r = 0.53$, 5;7 $r = 0.49$
J14 (Jones et al., 2014) ^e	6;1 (N = 25)	16 CV NWL 3-syllable	43.00 (17.44)	Not given
J10 (Jones et al., 2010)	6;1 TD (N = 18)	As per GB89 plus 18 CV NWL high phonotactic probability (6 × 2, 3, and 4-syllable), 18 CV NWL low phonotactic probability (6 × 2, 3, and 4-syllable)	GB89: 66.67 (15.34), 63.33 (23.01), 45.56 (17.90), 68.89 (18.44), 44.17 (21.85), 22.22 (20.45); 18 CV NWL: 70.92 (19.80), 57.59 (21.20), 23.15 (19.08), 59.26 (25.71), 44.08 (22.54), 19.07 (20.06)	Not given

^a Scores in this test were given for children on the 16th centile (1 standard deviation below the mean) and for phoneme rather than nonword accuracy. Subsequent comparisons are therefore made using the same criteria.

^b The monosyllabic nonwords are omitted due to recognized problems with them (Gathercole & Baddeley, 1989).

^c Stokes and Klee (2009) suggest that these nonwords should be considered wordlike.

^d Various researchers have suggested that these nonwords are wordlike (e.g., Graf Estes, Evans, & Else-Quest, 2007; Jones et al., 2010).

^e Three nonword sets were used in this study. Since the performance of each was captured by a previous version of the model illustrated in the current paper, we use the nonword set with the least number of datapoints to minimize the influence of previous fits to data on the RMSE values reported here.

phonemes, for example, accurate repetition occurs if each phoneme within a nonword is correctly drawn from the pool. A five phoneme nonword where each phoneme has a probability of 0.9 of being selected will therefore have an accuracy of $0.9 \times 0.9 \times 0.9 \times 0.9 \times 0.9 = 0.59$ (59%; this will increase to 66% if the nonword comprised 4 phonemes and decrease to 53% if the nonword comprised 6 phonemes). Computing repetition accuracy in this way gives baseline performance a good opportunity to fit the child data because as noted earlier, NWR performance is strongly influenced by length, with long nonwords consistently having lower repetition accuracy than short nonwords (see also Table 3). Emphasis on nonword length reduces when biphones and triphones are examined because the effective length of the nonword is reduced (e.g., a 5 phoneme nonword need only be represented using 3 biphones and 2 triphones). Note that for biphones and triphones, calculations are based on all possible combinations while maintaining the minimum number of items required. For example, a five phoneme sequence ABCDE requires only 3 biphones (the possible combinations of which are AB/BC/DE, AB/CD/DE) and 2 triphones (where only ABC/CDE is possible).

However, NWR involves hearing the sequence of sounds in a nonword stimulus, therefore the constituent items in the nonword need to be salient. To account for this, each phoneme/biphone/triphone from the nonword is taken in turn and the number of instances of the item in the pool is multiplied by a weighting to increase the likelihood of its selection (e.g., if /t/ occurs 100 times in the pool and the weighting is 10, the number of instances of /t/ in the pool is increased by 900).

Let us assume we are computing phoneme baseline performance to compare to 3 year old children (i.e., after 25% of the input) and at this point there are 1090 phonemes in the pool. If the first phoneme of a nonword is /t/, there are 90 instances of /t/ in the pool, and if the weighting is set to 100, then instances of /t/ in the pool would increase by $99 \times 90 = 8910$ resulting in the total number of phonemes in the pool increasing to $1090 + 8910 = 10,000$. The probability of selecting /t/ from the pool is therefore $(100 \times 90)/10,000$, or 0.9. Baseline performance is therefore derived probabilistically based on the weighted frequency by which each item (phoneme, biphone, or triphone) appears in the appropriate input (e.g., 25% of input for comparing to 3 year olds) relative to all other items. A low weighting will tend towards NWR accuracy of 0% and a high weighting will tend towards 100%; the question is whether a weighting can be found where the child data can be suitably approximated.

Simulations are run for 10,000 iterations and the value of the weighting is varied from 0 to 200,000 in increments of 50. Fig. 3 shows Root Mean Square Error (RMSE) across the 38 datapoints in Table 3 for different weight settings and items. RMSE represents the average amount of error or difference between predicted and actual performance (i.e., low RMSE indicates a good fit to the child data). The best fit to the child data is a RMSE of 12.9% for phonemes at a weighting of 350, indicating that on average, each datapoint is only within 13% accuracy of the child. The goodness of fit decreases with increasing item detail. As one moves from phonemes to biphones to triphones, the number of items in the pool exponentially increase and the relative influence of nonword length decreases; while this may be counteracted for a particular datapoint by selecting an appropriate weighting, it is impossible to do so in a way that maintains a low RMSE across all datapoints. Of interest now is how CLASSIC performs against the child data because the model has a fixed phonological working memory capacity that acts against the robust length effect seen in NWR; and since the model does not monitor how often items occur in the input, the only way it can account for frequency effects is from the chunking aspect of its learning that is derived from exposure to language.

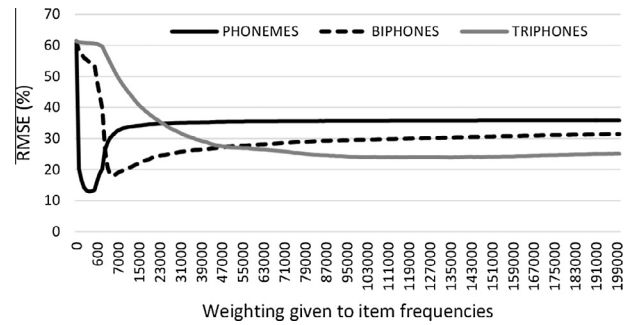


Fig. 3. RMSE (%) for baseline performance of phonemes, biphones, and triphones across different weight values.

3.2. CLASSIC NWR performance

Prior to detailing model performance, we first consider the contribution of different processes and methods that are used by CLASSIC. The model involves learning chunked lexical and sublexical sequences from an input that is constrained by a fixed sigmoidal probability function; in addition, the input is word-delimited. Table 4 shows RMSE when one or more of these factors is excluded and also RMSE for the full model. The Table shows three things of note: (1) RMSE is only substantially reduced when learning interacts with phonological working memory (rightmost two columns of Table 4); (2) word-delimited input only marginally influences RMSE; and (3) the full model averages to be within 7%³ of children's repetition scores across all 38 datapoints even though phonological working memory is fixed and performance is not mediated by the raw frequency of particular items in the input. Given the simplicity of the model it is not too surprising that without any learning (i.e., individual phoneme knowledge only) or without any internal constraint for representing learned knowledge (i.e., no phonological working memory) the model is unable to provide good RMSE estimates.

Fig. 4 shows a more in-depth examination of the model's performance for each of the 38 datapoints of Table 3. Over 50% of datapoints are within 5% of children's performance and over 85% (33 of 38 datapoints) are within 10% of children's performance. This compares favorably to optimal baseline performance, where only 37% of datapoints are within 5% and only 50% of datapoints are within 10% of the child data. One reason for the superior performance of the model relative to baseline performance is that the model can use all of its knowledge of phonemes, biphones, triphones etcetera to reconstruct nonwords in phonological working memory. However, this is not a complete explanation because one must remember that the model does not monitor frequency information.

CLASSIC also shows a good approximation to children's performance on an item-by-item basis. Accuracy data for individual items was available for the J10 study, where nonwords vary in terms of four different characteristics and three different lengths. Nevertheless, the repetition accuracy of each nonword in the model shows a highly significant correlation with those of the children ($r = 0.72$, $N = 66$, $p < 0.001$).

Item-level correlations for different nonword types and lengths show that for the nonwords also used by GB89, model and child

³ Table 3 and Fig. 4 both show that achieving a RMSE lower than 7% is challenging because NWR performance for the GB89 five year olds is superior to the six year olds of J10 for the same nonwords (i.e., fitting one set of nonwords will likely mean failing to fit the other). One possible reason for the discrepancy in the children's data is that the GB89 nonwords were scored correct/incorrect whereas the J10 nonwords were phonetically transcribed before reaching a correct/incorrect judgement. RMSE is 5.59% when excluding the GB89 five year old children.

Table 4

Average RMSE across 38 datapoints for different model settings. L = Learning; PWM = Phonological Working Memory; WD = Word-delimited input. A '+' indicates the presence of the model setting, a '-' the absence of it.

Model	L-, PWM-, WD-	L-, PWM-, WD+	L-, PWM+, WD-	L-, PWM+, WD+	L+, PWM-, WD-	L+, PWM-, WD+	L+, PWM+, WD-	L+, PWM+, WD+
RMSE	40.82	40.82	45.04	44.96	40.82	40.82	8.25	7.13

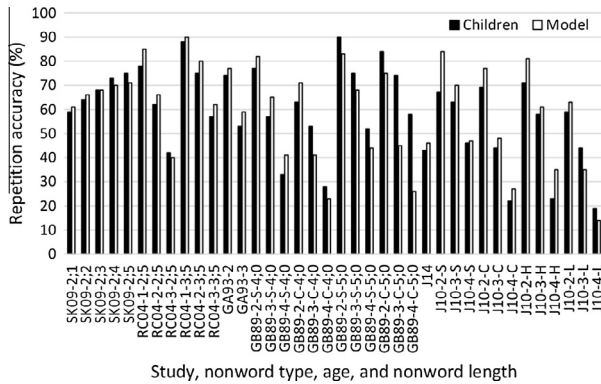


Fig. 4. Model-child comparisons for each of the 38 datapoints. Each study is abbreviated as per Table 3. Further abbreviations where applicable indicate number of syllables (e.g., -2-), nonword type (-S- = single consonant (CV); -C- = clustered consonant (CC); -H- = high phonotactic probability; -L- = low phonotactic probability) and age (e.g., 2;3).

repetitions correlate more readily for CC nonwords ($r = 0.76$, $N = 15$, $p = 0.001$) than CV nonwords ($r = 0.45$, $N = 15$, $p = 0.095$). For low lexicality nonwords, model and child repetitions correlated well irrespective of nonword type ($r = 0.76$, $N = 18$, $p < 0.001$ for high phonotactic probability; $r = 0.79$, $N = 18$, $p < 0.001$ for low phonotactic probability). The relationship between model and child NWR performance increases in strength as nonword length increases ($r = 0.28$, $N = 22$, $p = 0.208$; $r = 0.47$, $N = 22$, $p = 0.027$; and $r = 0.64$, $N = 22$, $p = 0.001$ for 2, 3, and 4 syllable nonwords respectively).

One key element of NWR performance is that it shows a strong relationship to vocabulary size, with several of the studies in Table 3 finding significant correlations between the two. Table 5 shows the correlation between mean repetition accuracy and vocabulary size across the different sets of maternal input ($N = 12$ for each correlation) for those nonword tests that were correlated with vocabulary size. For each test, vocabulary size is the number of chunked phoneme sequences that correspond to lexical items in the input. As the table shows, strong correlations exist between NWR performance and vocabulary size on all NWR tests.

Two of the nonword sets included for analysis show interesting interactions across a number of variables (the GB89 and J10 sets). Until now, model-child comparisons have been made based on an 'averaged model' where there is little variation in performance because of overlap across the various input sets and a fixed capacity of 4.5 items. In order to create variation in the model's performance so that interactions could be investigated, the same input regime as above was presented when capacity was varied from 3 to 6 in 0.25 increments, maintaining an average capacity of 4.5 items as per the original model. Model runs proceeded on the same basis as above but for each of $N = 13$ capacity settings. The resulting fit was almost identical to the original data when averaged across the different capacity settings: mean scores were broadly similar to those shown in Fig. 4 and RMSE was 7.01%.⁴

⁴ For comparison to the original model (where capacity = 4.5), the individual RMSE values for capacity = 3, 4, 5, and 6 respectively were: 30.06, 12.79, 12.53, and 25.76.

Table 5

Average repetition accuracy and vocabulary size in the model at stages of learning that correspond to the age of the children tested, together with correlation coefficients between repetition accuracy and vocabulary size across the 12 mother input sets.

NWR study	Model repetition accuracy (%)	Model vocabulary size (# lexical items)	Model correlation co-efficient
Roy and Chiat (2004): age 2;5 and 3;5	2;5: 72.83; 3;5: 77.92	2;5: 677.42; 3;5: 1390.71	2;5: $r = 0.88$, $p < 0.001$; 3;5: $r = 0.89$, $p < 0.001$
Gathercole and Adams (1993)	68.22	1115.50	$r = 0.78$, $p = 0.003$
Gathercole and Baddeley (1989): age 4;7 and 5;7	4;7: 53.80; 5;7: 56.82	4;7: 1716.96; 5;7: 2272.29	4;7: $r = 0.86$, $p < 0.001$; 5;7: $r = 0.87$, $p < 0.001$

Fig. 5 shows means and SDs for the capacity-varied model for the datapoints of concern in examining interaction effects. For GB89, a 2 (age: 4 or 5 years) \times 2 (nonword type: CV or CC) \times 3 (nonword length: 2, 3, or 4-syllables) repeated measures ANOVA examined CLASSIC's NWR data. The interaction between age and nonword type ($F(1,12) = 16.74$, $p = 0.001$, $\eta_p^2 = 0.58$) indicated that the difference in performance between CV and CC nonwords diminished with age; the interaction between nonword type and nonword length ($F(1,12) = 6.63$, $p = 0.005$, $\eta_p^2 = 0.36$) indicated a greater difference between CV and CC performance as length increased; and the interaction between age and nonword length ($F(1,12) = 5.54$, $p = 0.011$, $\eta_p^2 = 0.32$) indicated that increases in nonword length hampered younger children more than older children. All three interaction effects are consistent with patterns of data from other studies using the same nonwords but excluding the monosyllabic items (e.g., Briscoe, Bishop, & Norbury, 2001; Jones et al., 2010), suggesting that all effects can plausibly be derived from the child's exposure to language even when capacity is fixed and frequency information is not monitored.

For J10, one nonword test used the same nonwords as GB89 so we focus on the nonword set that varied in phonotactic probability, plus the comparison across the two nonword tests. A 2 (nonword type: high or low phonotactic probability) \times 2 (nonword length: 2 or 3 syllables⁵) repeated measures ANOVA was performed on CLASSIC's NWR data. There was a significant interaction between nonword type and nonword length ($F(1,12) = 5.53$, $p = 0.037$, $\eta_p^2 = 0.32$) where differences in repetition accuracy between high and low phonotactic probability words increased with nonword length. A further 2 (nonword test: WL or NWL) \times 2 (nonword length: 2 or 3 syllables) repeated measures ANOVA⁶ showed an interaction between nonword test and nonword length ($F(1,12) = 103.11$, $p < 0.001$, $\eta_p^2 = 0.90$) whereby superior repetition accuracy for WL over NWL nonwords increased with nonword length. Both interactions are consistent with the J10 data and further illustrate how interaction effects can be captured primarily from exposure to language.

⁵ The original study omitted four syllable nonwords because the analysis included children with SLI who were at floor for these items.

⁶ The original analysis used 2- and 3-syllable CV nonwords and high phonotactic probability nonwords to match nonwords as much as possible for structure (CV) and frequency of constituent items.

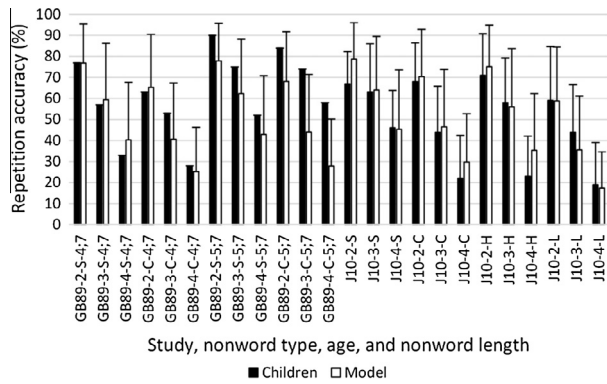


Fig. 5. Model and child repetition accuracy for the varied capacity simulations and for those studies that illustrate interaction effects. Error bars (where available) indicate standard deviation.

4. Discussion

NWR has broadly been construed as a measure of phonological working memory or phonological sensitivity. Computational modeling work has shown that these two approaches are potentially compatible by illustrating qualitatively similar NWR effects when taking into account language exposure. Using large-scale naturalistic input aimed at young children, we now provide a significant step in explaining NWR performance by showing how a computational model that has no developmental parameters is able to quantitatively fit child data across 6 NWR studies involving children between the ages of two and six years where nonword characteristics vary widely within and across studies. Furthermore, the model shows various qualitative fits to the children: a strong relationship between NWR performance and vocabulary size; repetition accuracy for individual nonwords correlates well with children's repetitions of the same nonwords; and the model shows the same interaction effects that are seen in the child studies. All of this is accomplished in a model where the only developmental change is the knowledge gained from language exposure. Under this view, the reason that NWR is so predictive of a range of language abilities is because it is very sensitive to the child's current level of linguistic knowledge.

These results show the power of lexical phonological knowledge in the repetition of nonwords because (for example) the robust length effects seen in CLASSIC's NWR performance can only be captured from the way in which accrued knowledge interacts with a fixed phonological working memory capacity. Furthermore, because CLASSIC does not explicitly monitor how often a sequence occurs in the input, wordlikeness and phonotactic probability effects in NWR are explained by frequently occurring sequences being more likely to form large chunks than those that occur relatively infrequently. That is, effects of length and frequency – the central hallmarks of NWR performance – are captured from a system that does not vary in its capacity and where the only influence of frequency is with regard to learning increasingly larger chunks of information. Chunked phonological sequences are the driving force behind the model's performance because they determine the amount of information that can be represented in temporary memory. Performance of the model clearly illustrates that viewing NWR in this way – as an aggregation of the relevant known phonological sequences that comprise the nonword – captures enough of the repetition process to enable the model to simulate a large amount of child data.

In line with the work of Gupta and Tisdale (2009), CLASSIC's account is consistent with phonological working memory and phonological sensitivity accounts of language learning in many

regards. Although of constant capacity, phonological working memory has a role in the model's vocabulary acquisition because it serves to constrain the amount of language that can be processed both as input and output. However, one should note that while some form of capacity limitation is necessary to fit the child data (see Table 4), the role of working memory, at least as operationalized by CLASSIC, is very much overshadowed by the role of accrued linguistic knowledge because the latter is pivotal in providing a substantial fit to children's performance across ages.

With regard to the phonological sensitivity account, while the model's learning mechanism begins at the level of the individual segment rather than holistically, both the model and the phonological sensitivity account highlight the role of sublexical information in vocabulary learning and NWR performance. Moreover, the model is consistent with the familiarity and neighborhood findings outlined by Metsala (1999). Familiar words and dense neighborhood words occur more often than unfamiliar words and sparse neighborhood words (for dense neighborhoods, the words themselves may not appear frequently but the sublexical parts do; for example, the rime of *make*, *take*, and *rake* is encountered whenever any of these words appear). During learning, the size of chunked phoneme sequences that represent familiar words and dense neighborhood words are therefore likely to be larger than for unfamiliar words and sparse neighborhood words. This enables performance on onset-rime blending and similar phonological awareness tasks to be performed with greater accuracy because the largest component part (the rime) is highly likely to already exist as a chunked unit, somewhat consistent with dense neighborhood rimes being represented in greater detail than those from sparse neighborhoods (Storkel, 2002). This view contrasts slightly with the phonological sensitivity account because it explicitly targets the size of sublexical phoneme sequences as important for onset-rime blending tasks rather than a more general ability to extract segmental detail from lexical items (see also De Cara & Goswami, 2003). Note that the shared information that is inherent in dense neighborhood words should also lead them to be learned more quickly than sparse neighborhood words, particularly while the information is still being learned as sublexical phonological knowledge. This is consistent with young children's lexicons containing a greater proportion of dense neighborhood words than the adult lexicon (Coady & Aslin, 2003) and also with hypotheses that word learning will be facilitated when new words contain familiar sounds and sound sequences (e.g., Lindblom, 1992; Menn, 1978).

Given that the NWR process is far more complex than that which is implemented in the model, it is perhaps surprising that developmental change in NWR is primarily captured by changes to lexical phonological knowledge. When some of these processes are examined, however, it becomes clear that their efficacy is greatly influenced by what one already knows about language. For example, resolution of ambiguities in speech perception are heavily influenced by linguistic knowledge (e.g., Cristia, Seidl, Junge, Soderstrom, & Hagoort, 2014; Mitterer & McQueen, 2009) and knowledge of a sound sequence is of greater benefit to speech production than knowledge of the individual sounds in different phonological contexts, since the former presents the listener with greater information on which to base any articulation of the phoneme sequence (e.g., Catts & Kamhi, 1984; Jakobson, 1941; Johnson & Wilson, 2002; Macken & Barton, 1980). The role that long-term lexical phonological knowledge plays in its interaction with phonological working memory is also crucial: maintaining a stored sequence of sounds will be easier when those sounds are encoded into few items rather than many items; and similarly the effort involved in maintaining the temporal order of a sequence will decrease as the number of items that are required to encode the sequence decrease. Lexical phonological knowledge is therefore pervasive in many of the processes involved in NWR and this

contributes to CLASSIC readily simulating children's NWR performance.

Contrary to the model, children's NWR performance is quite variable and determining the source of this variation is important to understanding the NWR process. While lexical phonological knowledge may influence speech perception and production, it can do so in ways that may either increase or decrease NWR performance. In this case, one must remember that the model represents the 'average child' and speech perception and speech production processes, in addition to the lexical phonological knowledge the child acquires and their individual phonological working memory capacity, all contribute to individual differences across children. Three parameters affect how quickly lexical phonological knowledge is learned in the model: the amount of linguistic information to which the child is exposed, the size of phonological working memory, and the learning rate. It may therefore be possible to capture individual differences by varying these parameters rather than keeping them constant as per the current model description.⁷ Fernald and Marchman (2012) have already shown that differences in processing efficiency (which could plausibly arise from individual differences in language exposure, phonological working memory, or learning rate) predict later vocabulary learning. Further study therefore needs to examine how changes to these parameters affect the NWR performance, and whether or not this reflects that seen in individuals.

Finally, the model suggests that NWR performance between the ages of two and six years can be captured without any need for increases in phonological working memory (see also French & O'Brien, 2008; Ottem, Lian, & Karlsen, 2007, though see Cowan, Ricker, Clark, Hinrichs, & Glass, 2015, for capacity increases in older children). Although some constraint on processing appears to be important in providing quantitative fits to children's NWR performance (see Table 4), the model simulates the data across ages without any change to phonological working memory. Rather, increases in repetition accuracy with age are achieved solely from increases in lexical phonological knowledge that enable a greater amount of information to be represented within a fixed 'chunk' capacity. Interestingly, these increases in linguistic knowledge give the perception of an increase in phonological working memory because the model is more able to repeat long nonwords when it holds a greater amount of lexical phonological knowledge. This may explain why NWR is viewed by many as a measure of phonological working memory because behaviorally, the child's performance is consistent with a capacity explanation. It also gives a possible explanation for span differences across stimuli in children's serial recall (e.g., Bachevalier & Denny, 1977; Crannell & Parrish, 1957; Dempster, 1981). Span size for random sequences of digits may be consistently greater than span size for random sequences of other stimuli such as words because children have been exposed to pseudo-random sequences of digits (e.g., dates, times, numerical calculations) whereas grammatical rules generally prevent exposure to random sequences of words. CLASSIC has been extended to adult input to illustrate such an effect (Jones & Macken, 2015).

In summary, we have shown how a model that focuses on linguistic exposure is able to provide substantial quantitative and qualitative fits to children's NWR performance across the ages of two to six years and for six different NWR studies that vary in nonword characteristics. Lexical phonological knowledge is the driving force behind the model's developmental change in NWR performance, suggesting that children's NWR performance is primarily a measure of the child's current level of linguistic knowledge that

is derived from their exposure to language and their ability to extract lexical phonological knowledge from that exposure.

References

- Altmann, G. T. M. (2002). Statistical learning in infants. *Proceedings of the National Academy of Sciences*, 9, 15250–15251.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321–324.
- Bachevalier, B. L., & Denny, M. R. (1977). A theory of intelligence: I. Span and the complexity of stimulus control. *Intelligence*, 1, 127–150.
- Baddeley, A. D. (1987). *Working memory*. Oxford, UK: Clarendon Press.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105, 158–173.
- Basirat, A., Dehaene, S., & Dehaene-Lambertz, G. (2014). A hierarchy of cortical responses to sequence violations in three-month-old infants. *Cognition*, 132, 137–150.
- 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, 391–403.
- Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, 113, 201–233.
- Bowey, J. A. (2001). Nonword repetition and young children's receptive vocabulary: A longitudinal study. *Applied Psycholinguistics*, 22, 441–469.
- Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 42, 329–340.
- Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107, 127–181.
- Catts, H. W., & Kamhi, A. G. (1984). Simplification of /s/+stop consonant clusters: A developmental perspective. *Journal of Speech and Hearing Research*, 27, 556–561.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205–215.
- Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1235–1249.
- Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, 30, 441–469.
- Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, 89, 183–213.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 97–185.
- Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*, 18, 132–145.
- Crannell, C. W., & Parrish, J. M. (1957). A comparison of immediate memory span for digits, letters, and words. *The Journal of Psychology*, 44, 319–327.
- Cristia, A., Seidl, A., Junge, C., Soderstrom, M., & Hagoort, P. (2014). Predicting individual variation in language from infant speech perception measures. *Child Development*, 85, 1330–1345.
- De Cara, B., & Goswami, U. (2003). Phonological neighbourhood density: Effects in a rhyme awareness task in five-year-old children. *Journal of Child Language*, 30, 695–710.
- Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin*, 89, 63–100.
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47, 421–436.
- Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, 83, 203–222.
- Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. In *Haskins laboratories status report on speech research* (pp. 53–64).
- Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. In S. Brady & D. Schankweiler (Eds.), *Phonological processing in literacy: A tribute to Isabelle Y. Liberman* (pp. 91–117). Hillsdale, NJ: LEA.
- French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, 118, 614–636.
- French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics*, 29, 463–487.
- Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the developmental pattern of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science*, 31, 311–341.
- Freudenthal, D., Pine, J. M., Jones, G., & Gobet, F. (2015). Simulating the cross-linguistic pattern of Optional Infinitive errors in children's declaratives and Wh-questions. *Cognition*, 143, 61–76.
- 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, 83–94.

⁷ The simulations that alter phonological working memory capacity illustrate one way in which variance in performance can be captured.

- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513–543.
- Gathercole, S. E., & Adams, A. (1993). Phonological working memory in very young children. *Developmental Psychology*, 29, 770–778.
- Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, 33, 672–688.
- Gathercole, S. E., Adams, A. M., & Hitch, G. J. (1994). Do young children rehearse? An individual-differences analysis. *Memory & Cognition*, 22, 201–207.
- Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language*, 28, 200–213.
- Gervain, J., Macagno, F., Cogoi, S., Pen, M., & Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences*, 105, 14222–14227.
- Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, 5, 236–243.
- Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, 112, 21–54.
- Graf Estes, K., 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, 177–195.
- Gupta, P., Lipinski, J., Abbs, B., & Lin, P.-H. (2005). Serial position effects in nonword repetition. *Journal of Memory and Language*, 53, 141–162.
- Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of Memory and Language*, 61, 481–502.
- Hansen, J., & Bowey, J. A. (1994). Phonological analysis skills, verbal working memory, and reading ability in second-grade children. *Child Development*, 65, 938–950.
- Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language*, 35, 1–31.
- Hockema, S. A. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development*, 2, 119–146.
- Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20- to 24-month-olds. *Journal of Child Language*, 35, 903–916.
- Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, 73, 418–433.
- Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, 11, F31–F39.
- Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, 27, 236–248.
- Jakobson, R. (1941/1968). *Child language, aphasia, and phonological universals* (A.R. Keiler, Trans.). The Hague: Mouton
- Johnson, C. E., & Wilson, I. L. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, 6, 271–289.
- Jones, G., Gobet, F., Freudenthal, D., Watson, S. E., & Pine, J. M. (2014). Why computational models are better than verbal theories: The case of nonword repetition. *Developmental Science*, 17, 298–310.
- Jones, G., Gobet, F., & Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science*, 10, 853–873.
- Jones, G., & Macken, W. J. (2015). Questioning short-term memory and its measurement: Why digit span measures long-term associative learning. *Cognition*, 144, 1–13.
- Jones, G., Tamburelli, M., Watson, S. E., Gobet, F., & Pine, J. M. (2010). Lexicality and frequency in Specific Language Impairment: Accuracy and error data from two nonword repetition tests. *Journal of Speech, Language, and Hearing Research*, 53, 1642–1655.
- Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, 29, 1–23.
- Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, 39, 159–207.
- Kamhi, A. G., & Catts, H. W. (1986). Toward an understanding of developmental language and reading disorders. *Journal of Speech and Hearing Disorders*, 51, 337–347.
- Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 131–163). Mahwah, NJ: LEA.
- Macken, M. A., & Barton, D. (1980). The acquisition of the voicing contrast in English: A study of voice onset time in word-initial stop consonants. *Journal of Child Language*, 7, 41–74.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 46, 1138–1153.
- Melby-Lervåg, M., Lervåg, A., Lyster, S. A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science*, 23, 1092–1098.
- Menn, L. (1978). Phonological units in beginning speech. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments*. Amsterdam, Netherlands: North-Holland.
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91, 3–19.
- Metsala, J. L., & Chisholm, G. M. (2010). The influence of lexical status and neighborhood density on children's nonword repetition. *Applied Psycholinguistics*, 31, 489–506.
- Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Hillsdale, NJ: LEA.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–97.
- Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 244–263.
- Monaghan, P., & Ellis, A. W. (2010). Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, 63, 506–525.
- Montgomery, J. W. (1995). Sentence comprehension in children with Specific Language Impairment: The role of phonological working memory. *Journal of Speech, Language, and Hearing Research*, 38, 187–199.
- Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research*, 44, 778–792.
- Munson, B., Kurtz, B. A., & Windsor, J. (2005). The influence of vocabulary size, phonotactic probability, and wordlikeness on nonword repetitions of children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research*, 48, 1033–1048.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.
- Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology*, 19, 233–270.
- Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39, 246–263.
- 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, 223–234.
- Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, 81, 149–169.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-Month-old infants. *Science*, 274, 1926–1928.
- Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 592–608.
- Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, 12, 369–373.
- Snowling, M., Goulandris, N., Bowlby, M., & Howell, P. (1986). Segmentation and speech perception in relation to reading skill: A developmental analysis. *Journal of Experimental Child Psychology*, 41, 489–507.
- Snowling, M., & Hulme, C. (1994). The development of phonological skills. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 346, 21–27.
- Stokes, S. F., & Klee, T. (2009). The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *Journal of Speech, Language, and Hearing Research*, 52, 872–882.
- Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, 29, 251–274.
- Swingle, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, 43, 454–464.
- Swingle, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, 71, 73–108.
- Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword repetition and language development in 4-year-old children with and without a history of early language delay. *Journal of Speech, Language, and Hearing Research*, 48, 1481–1495.
- Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of Verb-Argument structure: An alternative account. *Journal of Child Language*, 28, 127–152.
- Treiman, R., & Breaux, A. M. (1982). Common phoneme and overall similarity relations among spoken syllables: Their use by children and adults. *Journal of Psycholinguistic Research*, 11, 569–598.
- Vihman, M. M., & Velleman, S. L. (1989). Phonological reorganization: A case study. *Language and Speech*, 32, 149–170.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9, 325–329.
- Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, 13, 286–350.
- Weisleder, A., & Fernald, A. (2013). Talking to children matters early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24, 2143–2152.
- Yoshida, K. A., Fennell, C. T., Swingle, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, 12, 412–418.