Cognitive Science 32 (2008) 1148–1176

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1080/03640210802073689

Computer Simulations of Developmental Change: The Contributions of Working Memory Capacity and Long-Term Knowledge

Gary Jones[a], Fernand Gobet[b], Julian M. Pine[c]

[a]*Department of Psychology, Nottingham Trent University*
[b]*School of Social Sciences, Brunel University*
[c]*School of Psychology, University of Liverpool*

Received 16 April 2007; accepted 28 February 2008; received in revised form 10 January 2008

Abstract

Increasing working memory (WM) capacity is often cited as a major influence on children's development and yet WM capacity is difficult to examine independently of long-term knowledge. A computational model of children's nonword repetition (NWR) performance is presented that independently manipulates long-term knowledge and WM capacity to determine the relative contributions of each in explaining the developmental data. The simulations show that (a) both mechanisms independently cause the same overall developmental changes in NWR performance, (b) increase in long-term knowledge provides the better fit to the child data, and (c) varying both long-term knowledge and WM capacity adds no significant gains over varying long-term knowledge alone. Given that increases in long-term knowledge must occur during development, the results indicate that increases in WM capacity may not be required to explain developmental differences. An increase in WM capacity should only be cited as a mechanism of developmental change when there are clear empirical reasons for doing so.

Keywords: Computational modeling; Developmental change; Nonword repetition; Child development

1. Introduction

Although it is clear that cognitive changes occur during the course of a child's development, it is less clear precisely what develops. Indeed, this issue is central to developmental psychology, and has generated a considerable amount of empirical data and theoretical debate. Theories range from changes in knowledge structures (e.g., Piaget, 1950, 1952; Vygotsky, 1978), changes in a combination of knowledge and working memory (WM) capacity (e.g.,

Correspondence should be sent to Gary Jones, Department of Psychology, Nottingham Trent University, Burton Street, Nottingham NG1 4BU, UK. E-mail: gary.jones@ntu.ac.uk

Case, 1985; Halford, 1993), changes in adaptive strategy choice (Siegler, 1995), and changes in processing speed (e.g., Kail, 1988), to name but a few. This article concentrates on developmental change occurring via increases in knowledge and increases in WM capacity.[1] As we will see, these two mechanisms are inextricably linked such that it is difficult to examine the effects of one in the absence of the other.

No serious researcher would argue against the idea that increases in children's knowledge play a central role in development. Piaget (1950, 1952) first put forward the hypothesis that knowledge structures are continually updated by the child, with the vast majority of subsequent research supporting this view in one form or another (e.g., Klahr & Wallace, 1976; Siegler, 1995). For example, although Siegler (1995) suggested development via adaptive strategy choice, this encompasses general knowledge that develops through task experience within a domain.

However, further mechanisms of development have also been proposed. In particular, increases in WM capacity have consistently been cited as a separate mechanism of development in a wide range of domains, such as reasoning (e.g., Halford, Maybery, & Bain, 1986), vocabulary learning (e.g., Gathercole & Baddeley, 1989), arithmetic (Passolunghi & Siegel, 2001), and spelling (Ormrod & Cochran, 1998). Furthermore, Cowan (2000) argued, based on a wealth of previous literature, for both individual and developmental differences in WM capacity; and Cantor and Engle (1993) argued that individual differences in capacity arise from variations in the amount of activation that is available to distribute among long-term memory traces.

One problem in examining mechanisms of development other than knowledge is the extent to which knowledge pervades these other hypothesized developmental mechanisms. WM capacity, in particular, is sensitive to knowledge changes—the chunking hypothesis, for example, suggests that our capacity to hold meaningful chunks for recoding material is based on our long-term knowledge (e.g., Miller, 1956; Simon, 1974). Developmental theorists also acknowledge the interplay between knowledge and WM capacity. For example, Case (1985) argued that WM capacity remains fixed across childhood, but the amount of information that can be stored in WM increases as knowledge increases.

If WM capacity is strongly influenced by long-term knowledge, then developmental increases in WM capacity will be hard to differentiate from developmental increases in knowledge. Any empirical assessment of WM capacity must, therefore, account for the child's existing knowledge, because failure to do so may lead to tests of WM capacity that inadvertently capture differences in knowledge rather than capacity. However, although it is possible to lay out tasks that estimate a child's knowledge in a specific domain, it is almost impossible to be certain what knowledge a child may bring to bear when completing a task within that domain. For example, in the balance scale domain (e.g., Siegler, 1976), knowledge of weight and distance are seen as critical in completing the task successfully, but there is a variety of other types of knowledge that also help performance, such as knowledge of numbers and the concepts of greater-than and less-than.

There are therefore two related problems in providing an accurate measure of WM capacity. First, there is a strong interplay between WM capacity and long-term knowledge. Second, measuring WM capacity independently of long-term knowledge is difficult because it is almost impossible to derive all of the pieces of knowledge a child may use when completing a WM capacity task. Taken together, these two issues raise questions about whether tests of WM capacity, in part or whole, are tests of long-term knowledge.

This article examines the relative contributions of long-term knowledge and WM capacity in explaining developmental change. Given that increases in the child's long-term knowledge must take place during the course of development, we ask whether additional assumptions need to be made regarding developmental increases in WM capacity or whether increases in long-term knowledge are sufficient to account for the developmental data. To address this issue, a computational model of development will be presented that independently examines the roles of long-term knowledge and WM capacity and compares the results of each with developmental data.

As stated previously, empirically assessing WM capacity independently of long-term knowledge is difficult. Computational modeling can help because a model requires all necessary task knowledge to be specified in order to complete a task, enabling a clear-cut analysis of how long-term knowledge influences performance. In addition, plausible assumptions regarding WM capacity can be included within a model. A computational model that includes both long-term knowledge and WM capacity can, therefore, independently manipulate each to see how increases in long-term knowledge and increases in WM capacity are able to match the developmental differences in the child data. In particular, we can ask which phenomena in the child data can be explained by changes in long-term knowledge and which phenomena can be explained by changes in WM capacity. Furthermore, we can vary both long-term knowledge and WM capacity simultaneously to see whether the interplay between the two is able to provide a better explanation of the child data than either increasing long-term knowledge or increasing WM capacity alone.

The domain we use to examine long-term knowledge and WM capacity is one where both mechanisms are cited as being the dominant explanation for age-related changes: nonword repetition (NWR; Gathercole, Willis, Baddeley, & Emslie, 1994). NWR studies involve a nonsense word being read aloud to the child, who is asked to repeat it back accurately. Across a range of studies, NWR performance has consistently been shown to improve with age and to be inversely related to nonword length (e.g., Gathercole & Adams, 1993; Gathercole & Baddeley, 1989; Roy & Chiat, 2004). These results appeared to support the view that phonological WM capacity increased with age (e.g., Gathercole & Adams, 1993; Gathercole & Baddeley, 1989). However, it quickly became clear that there were long-term knowledge influences on NWR ability, because performance was significantly better for nonwords rated as being wordlike and nonwords containing high-frequency phonemes (e.g., Frisch, Large, & Pisoni, 2000; Gathercole, 1995; Gathercole, Willis, Emslie, & Baddeley, 1991). The results of studies of NWR in children thus appear to support both the idea that increasing WM capacity is the dominant factor (e.g., Baddeley, 2002; Baddeley, Papagno, & Vallar, 1988; Gathercole & Adams, 1993; Gathercole & Baddeley, 1989; Gathercole & Pickering, 1999; Gathercole, Willis, et al., 1994) and the idea that increasing long-term knowledge is the dominant factor (e.g., Bowey, 1996; Metsala, 1999; Munson, Edwards, & Beckman, 2005; Munson, Kurtz, & Windsor, 2005). NWR performance therefore provides an ideal domain to examine the relative contributions of long-term knowledge and WM capacity.

The remainder of this article is organized as follows. First, the computational model of NWR performance is outlined. Second, we report three simulations of NWR performance (varying long-term knowledge, WM capacity, and both) together with comparisons across simulations. Third, we discuss the results of the simulations and their implications for theory, highlighting the respective roles of long-term knowledge and WM capacity.

2. The model: EPAM-VOC

EPAM-VOC (Elementary Perceiver and Memorizer-Vocabulary; Jones, Gobet, & Pine, 2007) is a phoneme sequence learner that takes speech in phonemic form as input and builds a hierarchical network of phoneme sequences (or "chunks") that represents long-term knowledge of the linguistic input. The model has previously been used to simulate NWR performance in 2- to 5-year-old children (Jones et al., 2007). EPAM-VOC is based on the EPAM modeling architecture (Feigenbaum & Simon, 1984), which, together with related discrimination-net models such as CHREST (Gobet & Simon, 2000) and MOSAIC (Freudenthal, Pine, Aguado-Orea, & Gobet, 2007; Freudenthal, Pine, & Gobet, 2006),[2] has been used to simulate psychological phenomena in a variety of domains such as learning, memory, and perception in chess, verbal learning behavior, the digit-span task, the context effect in letter perception, and the acquisition of syntactic categories (for overviews, see Gobet & Lane, 2005; or Gobet et al., 2001). We first provide an overview of EPAM before describing EPAM-VOC in order to highlight areas where EPAM-VOC has been simplified from the original EPAM architecture.

2.1. The EPAM architecture

EPAM (e.g., Feigenbaum & Simon, 1984) is a modeling architecture consisting of a short-term memory and a discrimination network giving access to long-term memory; it also postulates attention mechanisms that will impact on the construction of the discrimination network. The discrimination network is built based on the features of a given input; the links contains tests on these features, and the nodes (or "chunks") contain the internal description of the item. For example, a large red triangle might have the three features *large, red,* and *triangle.* After learning, these features will be represented in the network as a sequence of tests, each related to a feature of the input item. The sequence of tests can be used to determine whether or not a given input is familiar (i.e., is similar to an input that has been seen before). The features of the input item would be sorted through the sequence of tests and the resulting information, if it matched the sequence of features of the input, would determine that the input was familiar. However, if the resulting information mismatched the features in the input, then this gives EPAM an opportunity for learning something about the input. There are two methods of learning: If the information held at the resulting node under-represents the sequence of input features, then a process of familiarization adds more information to the node; if the resulting information over-represents the input, in the sense that it contains features not shared by the input, then a process of discrimination creates a new test containing the mismatched part of the input, and a new node below that test.

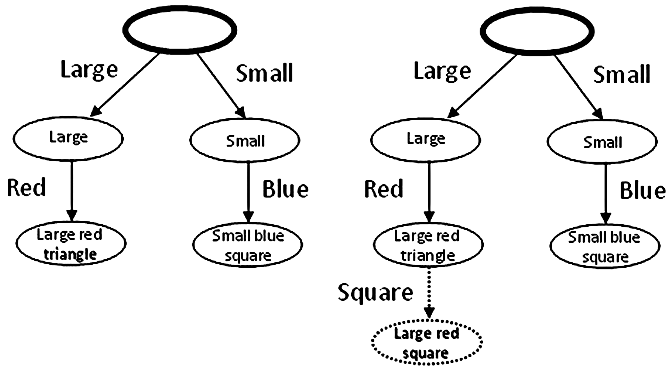

Fig. 1. Learning in EPAM. *Note:* For both graphs, the starting node for traversing the network is indicated by the topmost node (indicated in bold) that contains no information. Arrows indicate tests and ellipses indicate nodes, which contain the resulting information after a test. New information added via learning is represented either by bold text or by arrows and ellipses using dashed lines.

Fig. 1 shows how the familiarization and discrimination processes work, and how the sequence of input features would be tested in the discrimination network. In this figure, nodes are represented by ellipses. If the network was as shown in the left graph of Fig. 1 and the input was "large red triangle," EPAM would first look for a test that satisfies the first feature of the input ("large") below all tests emanating from the topmost node. As such a test exists, EPAM traverses to the "large" node and processes the next feature of the input ("red"). Again, the "red" test can be satisfied, and EPAM traverses to the "large red" node. The next feature is now processed ("triangle"), but no tests emanate from the "large red" node, so EPAM cannot traverse any further. However, as the information in the final node ("large red") mismatches the sequence of features in the input ("large red triangle"), EPAM familiarizes by adding the feature "triangle" to the "large red" information in the node. If the network was as shown in the right graph of Fig. 1, and the input was "large red square," EPAM would satisfy the "large" and "red" tests, but would then find that the resulting information "large red triangle" mismatched the features in the input ("large red square"). At this point, EPAM would discriminate the two by adding a test ("square") and a node with the new input sequence ("large red square").

EPAM provides a simple means of determining whether a given input is recognized by the network (i.e., has been seen before) by traversing the network. For example, in the resulting network on the right side of Fig. 1, and the input "large red square," EPAM would apply the first feature of the input ("large") to all tests below the null top node. Such a test exists, and the "large" node now becomes the current set of information, and EPAM moves on to the next feature ("red"). Such a test exists below the current node, and so the "large red" node now becomes the current node. The input moves on to the final feature ("square"), which exists as a test and so the input can be said to have been recognized by the model.

EPAM, therefore, provides a method by which a set of input features can be learned while preserving the pattern within that set of input features. Furthermore, any given input can be applied to the model to determine whether the knowledge gained by the model makes it possible to recognize the input pattern. For EPAM-VOC, it is vocabulary that is being learned, and so the input features will be phonemes. The patterns that the model will learn will therefore be sequences of phonemes, and we will see that these can be used effectively to help in vocabulary acquisition.

2.2. EPAM-VOC and EPAM

EPAM-VOC is a simplified version of EPAM that dispenses with the familiarization process. This means that the information returned after fulfilling a test is the accumulation of all the preceding tests (i.e., the network can no longer under- or over-represent the features of the input). Given that EPAM-VOC is applied to vocabulary learning and that learning new words involves the short-term storage of sound patterns, more attention will be given to short-term memory mechanisms than in the standard EPAM. We now detail how EPAM-VOC learns sequences of phonemes and how short-term memory is implemented.

2.3. Learning phoneme sequences in EPAM-VOC

The simulations we present compare the model's performance against 2- to 5-year-old children, so we assume that at the beginning of the simulations, EPAM-VOC has knowledge of the phonemes used in English (an assumption that has support in the vocabulary acquisition literature; e.g., Bailey & Plunkett, 2002). Before any learning takes place, the network therefore consists of a null top node plus all the constituent phonemes in English as tests and nodes below the null top node.

In keeping with EPAM, EPAM-VOC examines each feature (for vocabulary learning, each phoneme) of the input sequence in turn, until it can learn something from that sequence. When a sequence of phonemes is presented to the model, EPAM-VOC traverses as far as possible down its existing hierarchy of nodes by examining each input phoneme in turn, until it cannot traverse any more. At this point, something is learned regarding the current phoneme in the sequence, and the remainder of the sequence now becomes a new input that is processed by the top node.

As an example, consider the utterance "Where?," which has a phonemic equivalent of "W EH1 R" (speech is converted to a phonemic equivalent using the CMU Lexicon database, available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict). Traversal in EPAM-VOC involves selecting a test below the current node that is equal to the next phoneme in the sequence. When "W EH1 R" is presented, EPAM-VOC attempts to find a test below the null top node equal to "W." Because a "W" test exists, the node "W" now becomes the current top node in the network. The input now becomes "EH1 R" and the "EH1" phoneme is considered for traversal. However, there are no tests below the "W" node (remember that the network contains only the top node and nodes for the constituent phonemes in English) and therefore traversal ends. EPAM-VOC now learns "W EH1" by adding an "EH1" test and a node with the sequence "W EH1" below the "W" node. Some learning has occurred, so processing reverts to the null top node, and the input proceeds to the last phoneme, "R," but as this already exists below the top node, learning ends.

Presenting the input a second time results in the actual sequence "W EH1 R" being learned. The first phoneme, "W," is examined, and the "W" test is taken from the null top node to the

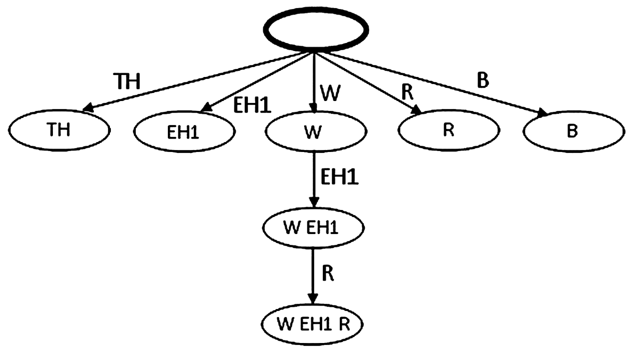

Fig. 2. The resulting EPAM-VOC network after receiving the input "W EH1 R" twice. *Note:* Although only five individual phonemes are illustrated below the root node, the model knows all phoneme primitives.

"W" node. This now becomes the top node, and the input moves on to the "EH1" phoneme. An "EH1" test can be taken below the "W" node, and so "W EH1" now becomes the top node, with the input moving on to the "R" phoneme. No further tests exist below this node, and so "R" is added as a test below "W EH1," and a new node "W EH1 R" is added at the end of the test. The resulting network after two presentations of "W EH1 R" is shown in Fig. 2.

Learning in EPAM-VOC therefore involves the creation of tests and nodes. Tests specify phonemes to be matched in the input in order to traverse the network. Nodes represent phonemes and phoneme sequences that are known in the network. Traversal of the network begins when EPAM-VOC is presented with an input (e.g., a mother's utterance). This input is then used to traverse the network until no further traversal is possible, at which point a new test and node will be created below the furthest traversed node. Once learning has occurred, processing reverts to the null top node, and the traversal and learning process begins again using the remainder of the input.

Because of the way EPAM-VOC learns, the contents of any one node are the concatenation of all the tests that lead to that node (e.g., the "W EH1 R" node in Fig. 2 comprises all of the phonemes contained in the tests that lead to the node). There is therefore only ever one test that leads to any one node. The learning mechanism within EPAM-VOC means that a word containing seven phonemes would require six learning passes (the initial phoneme in the word would already be known below the null top node). Although it may seem that EPAM-VOC learns very quickly, it is possible to reduce the rate of learning (e.g., by altering the probability of learning a new node), and this has been successful for other variants of EPAM/CHREST models (e.g., Croker, Pine, & Gobet, 2003; Freudenthal, Pine, & Gobet, 2002). Slowing down the rate of learning yields similar networks, but over a longer period of time. The input sets used in the simulations contain a very small subset of the input that a child hears, so it is reasonable to have learning take place in the way that has been illustrated.

The learning mechanism within EPAM-VOC is sensitive to the input it receives. For example, words or phrases that occur often in the input are likely to be represented at a single node, whereas words or phrases that occur rarely in the input are unlikely to be represented at a single node (unless they consist of very few phonemes). Sensitivity to the frequency characteristics of the input will be important when we consider how EPAM-VOC simulates WM capacity limitations.

2.4. Providing WM capacity limitations within EPAM-VOC

The model uses a fixed duration WM capacity based on the phonological store component of the WM model (Baddeley & Hitch, 1974). The phonological store is implemented rather than the phonological loop in line with findings that children of 5 years or younger show no reliable rehearsal strategy (e.g., Gathercole & Adams, 1994; Gathercole, Adams, & Hitch, 1994). The phonological store has a temporal duration of 2,000 msec (Baddeley, Thompson, & Buchanan, 1975) that is implemented within EPAM-VOC as a time to match the input using the nodes in long-term knowledge (the hierarchical network). To match a node takes 400 msec, and to match a phoneme within that node takes an additional 30 msec, excluding the first phoneme (these timing estimates are based on those of Zhang & Simon, 1985). For example, matching the "W EH1 R" node in the network shown in Fig. 2 would take 460 msec. Because it takes 400 msec to match any node in the network, the "W EH1 R" node is allocated a time of 400 msec to match the node itself, but added to this time is the time to match each constituent phoneme bar first (i.e., 30 msec for "EH1" and 30 msec for "R")—resulting in a time allocation of 460 msec.

Consider the input "Where's baby?" (phonemic equivalent "W EH1 R Z B EY1 B IY0?") and the network as shown in Fig. 2. The "W EH1 R" part of the input can be matched using the contents of a single node, and is allocated a time of 460 msec. The remainder of the input contains phonemes that exist only as single item nodes in the network, which are therefore allocated a time limit of 400 msec each. The input presented to EPAM-VOC for learning is therefore "W EH1 R Z B EY1" and has a temporal duration of 1,660 msec. The phonemes at the end of the utterance, "B" and "IY0," are not included, as these would exceed the 2,000 msec limit—that is, once the time limit of the phonological store is exceeded, no further input is able to be processed.

By using long-term knowledge to mediate the amount of information that can be represented within a fixed capacity limit, EPAM-VOC is able to concretely specify how WM capacity and long-term knowledge interact. The absence of a detailed specification of the link between WM and long-term memory has been acknowledged as a problem with current accounts of NWR performance (e.g., Gathercole, Willis, et al., 1994) and, although there have been attempts to provide verbal descriptions (e.g., Gathercole, 2006; Metsala, 1999), EPAM-VOC offers a precise specification of the interaction.

At the early stages, after EPAM-VOC has been presented with a small amount of mother's speech, its hierarchy of nodes is not very large and therefore long-term memory is of minimal aid to offset WM capacity limitations. The nodes at this point will only contain small sequences of phonemes, and so any given input to the model is likely to require many nodes to represent it, resulting in only some of those nodes being captured within the 2,000 msec limit of the phonological store. However, after the model has been presented with a large amount of speech, the hierarchy becomes more extensive such that nodes can contain long sequences of phonemes—if part of the input can be represented using these nodes, this will reduce the amount of time allocated to the input such that more of it can now be captured within the phonological store. Furthermore, EPAM-VOC's sensitivity to the variation in the input means that more will be learned from speech containing a large rather than a small set of vocabulary even when, for example, the number of utterances and mean length of utterances are matched. This is because any diversity within the input results in more opportunity for the model to learn nodes containing different phoneme sequences. It is worth noting that the time to match a node and the time to match constituent phonemes in a node do not vary with vocabulary size. Rather, vocabulary size itself drives how much information can fit into WM capacity.

2.5. How EPAM-VOC performs the NWR test

NWR is achieved by presenting the model with the phonemic representation of each individual nonword in the same way that normal speech input is presented to the model. EPAM-VOC therefore attempts to capture as much of the nonword as possible using existing nodes by traversing the network in exactly the same way as with standard speech input—including the same time-limited capacity. If the whole nonword can be captured in the phonological store within the given time-limited capacity, it is assumed to have been repeated correctly, otherwise the nonword is assumed to have been repeated incorrectly. Nonwords that are repeated correctly obtain a score of 1, and nonwords repeated incorrectly obtain a score of 0. This is the same method of scoring as per the children. Each group of nonwords contains five stimuli, so scores are out of 5. Multiplying these scores by 20 results in a percentage of repetition accuracy for the model and for the children.

For young children, errors are made on the NWR test even for the simplest stimuli (single-syllable, wordlike nonwords). Errors are believed to occur either from inaccurate encoding/storage (Gathercole & Baddeley, 1990b) or inaccurate articulation of the nonword (particularly for nonwords containing consonant clusters; Gathercole et al., 1991). In fact, NWR studies often make allowances for articulation difficulties (e.g., Roy & Chiat, 2004). Encoding/storage/articulation difficulties have been incorporated within EPAM-VOC by adding a probability of error when making traversals in the network. This means that when trying to represent a nonword in as few nodes as possible, an incorrect test may be taken, resulting in an incorrect response. Error probabilities are the same as those used by Jones et al. (2007).

Children's NWR errors can be categorized in terms of phoneme substitutions, phoneme deletions, and the combination of the two (phoneme addition rarely occurs in NWR; Gathercole, Willis, et al., 1994). EPAM-VOC is also able to produce these categories of error. By having the possibility of selecting an incorrect node when traversing the network, the model is able to produce phoneme substitutions. Phoneme deletions occur when the nonword is unable to fit in the time-limited phonological store.

2.6. Alternative models of NWR and psycholinguistic phenomena

There are other models that examine NWR and also a variety of models that are concerned with phenomena from psycholinguistics and memory research, such as serial order effects. We consider both varieties of model here in order to give a perspective as to how EPAM-VOC fits in with these models.

There exist at least three models of NWR. First, Hartley and Houghton (1996) described a connectionist network that incorporates a decay element. Nonwords are presented to the model in the training phase, and recall of the nonwords is determined in a later test phase. Decay in the model means that long nonwords have a lower probability of correct recall than short nonwords, consistent with the NWR literature (e.g., Gathercole & Baddeley, 1989). The model also includes competition at the phoneme level such that, for example, phonological substitutions can take place. Based on data from Treiman and Danis (1988), the model makes similar types of error to those made by children and adults.

Second, Brown and Hulme's (1995, 1996) trace decay model represents a given nonword (or other item) as a sequence of time slices based on the time taken to articulate the nonword. Each time slice begins with a high activation strength that declines as time progresses, meaning the beginning segment of a nonword decays more rapidly than the middle and end segments. However, activation strength can be increased based on relationships to long-term memory traces. For nonwords that share similarities to real words (i.e., wordlike nonwords), activation would therefore be higher than that of nonwords sharing little similarity to real words (i.e., non-wordlike nonwords). The resulting effect, as seen in children's NWR, is that wordlike nonwords have a higher repetition accuracy than non-wordlike nonwords (e.g., Gathercole, 1995).

Third, Gupta and colleagues (Abbs, Gupta, Tomblin, & Lipinski, 2007) detailed a recurrent connectionist network that combines long-term phonological knowledge (the weights in the network) and phonological short-term memory (the recurrence in the network). The training set comprised 4,386 English words varying in length from two to four syllables. By including units in this network that in some sense represent phoneme features, it was possible to examine phonological discrimination effects as well as NWR effects. A significant relationship between phonological discrimination and NWR was found, independent of any involvement of vocabulary learning. This finding in the model mirrors that of human participants. Although no examination of specific NWR effects was carried out, it should be noted that this model is in its infancy, and further work is due to come out.

If we consider the first two models (given that the third does not yet examine NWR phenomena), both models are able to capture some of the central phenomena that are seen in the NWR literature, such as differences in performance depending on nonword length and wordlikeness. EPAM-VOC is also able to capture these effects (Jones et al., 2007). For example, better performance is found for short nonwords because these are more likely to fit in the model's time-limited phonological store. Wordlike nonwords show an advantage over non-wordlike nonwords because they are captured in fewer nodes (thereby receiving a lower time allocation in the phonological store). The main advantage of EPAM-VOC over the models listed above is that EPAM-VOC captures all of the necessary NWR effects while at the same time explaining how phonological knowledge is actually acquired through exposure to naturalistic stimuli.

Further models exist that attempt to simulate short-term memory phenomena other than NWR. For example, OSCAR (Brown, Preece, & Hulme, 2000) is able to simulate a wide range of serial order phenomena, such as item similarity and grouping effects. The primacy model (Page & Norris, 1998) simulates word length, list length, and phonological similarity effects in serial recall. Burgess and Hitch's (1999) network model simulates the same phenomena as the primacy model but also includes effects of articulatory suppression. All of these models specifically address phenomena seen primarily in the serial recall literature rather than phenomena in the NWR literature, and so this is a clear difference compared to EPAM-VOC. The lack of a mechanism by which EPAM-VOC can simulate serial recall is a limitation of the model that we will return to in our general discussion. However, it should be noted that the models described earlier, although they provide an explanation of serial recall phenomena, do not explain how the material relevant to this phenomena is learned—that is, how phonological knowledge is acquired and how new words are learned. This is a major advantage of EPAM-VOC over all of the models covered in this section.

3. Simulations of the NWR data

Before presenting the analyses, we first describe the child data that were used to compare the simulations, how the simulations were performed, and how the analyses that compare the simulations to the child data were carried out.

3.1. Selecting appropriate comparison data

EPAM-VOC is a computational model that emphasizes the role of the input in the child's development. To provide as close an approximation to the input as possible, NWR performance will be compared to young children, because their input is easier to estimate (older children receive input from a variety of sources such as books, television, etc.). The children's NWR data from Jones et al. (2007) are used because this study uses children between 2 and 5 years of age, and it is the only study we are aware of that uses the same NWR test and methodology across these ages. The data compares 2- to 3-year-old children and 4- to 5-year-old children on nonwords that are either wordlike or non-wordlike and that vary from one to three syllables in length. Older children show better NWR performance and there are effects of wordlikeness and nonword length for both ages, with better performance for wordlike nonwords and shorter nonwords.

3.2. Method of simulation

The simulations attempt to match 2- to 3-year-olds NWR performance at an early stage in the model's development (when WM capacity is small or the model is at an early stage in its learning) and to match 4- to 5-year-olds NWR performance at a later stage in the model's development (when WM capacity is large or the model is at a late stage in its learning). An indicative estimate of the input that a 2- to 3-year-old child receives is the speech from the primary caregiver, so an input solely based on mother's utterances is used in the early stages of the model's learning. However, to approximate the input that a 4- to 5-year-old receives, words from a vocabulary frequency database for 8-year-old children (available at http://www.essex.ac.uk/psychology/cpwd/) are used in conjunction with the mother's utterances. The simulations begin by using only a mother's utterances, but gradually introduce words from the vocabulary database at later stages in the model's learning. We assume that,

Table 1
Stage of learning, amount of input seen by the model, the ratio of mother's utterances to vocabulary items used in the input, and the probability of making a traversal error at each stage of the EPAM-VOC's learning

Stage of learning	Amount of input seen by the model (%)	Percentage of pairs of vocabulary items included in the input	Probability of selecting an incorrect link when traversing
1	0–12.5	0	0.10
2	12.5–25	0	0.10
3	25–37.5	10	0.09
4	37.5–50	20	0.08
5	50–62.5	30	0.07
6	62.5–75	40	0.06
7	75–87.5	50	0.05
8	87.5–100	60	0.04

with age, children become better at encoding and articulating words, and so the probability of making an error when traversing the discrimination network is reduced at later stages of the model's learning.

Table 1 shows the stage of learning, the amount of input seen by the model, the ratio of mother's utterances to vocabulary items used in the input, and the probability of making a traversal error at each stage of the model's learning. Note that the probability of making a traversal error is not based on an arbitrary figure but reflects children's error rates for single-syllable nonwords. For example, the 2- to 3-year-olds have a 28% average error rate for single-syllable nonwords (Gathercole & Adams, 1993; Jones et al., 2007). Single-syllable nonwords average 3.1 phonemes, and assuming one traversal per phoneme, with each phoneme having a probability of error of .10, then the probability of making a correct traversal is $.90 \times .90 \times .90 = .73$, or a 27% error rate. Although the error rates for single-syllable nonwords can be said to have been "fit," the actual comparisons are made on nonwords of one to three syllables and nonwords that are both wordlike and non-wordlike, so error rates mainly arise from the dynamics of the model.

EPAM-VOC was trained individually on each of 12 sets of mother's utterances taken from mother–child interactions with 2- to 3-year-old children across the period of one year (Theakston, Lieven, Pine, & Rowland, 2001). The number of utterances varied for each mother–child (range 17,474–33,452; $M = 25{,}519$). When introducing vocabulary items into the input, pairs of vocabulary items were used so that the number of phonemes in the input would be roughly equal to the number of phonemes in a mother's utterance. The average number of phonemes in a mother utterance is 12.03; the average number of phonemes in a pair of vocabulary items is 10.46. The vocabulary items selected for use as input were scaled based on the frequency of occurrence of the item. For example, *lake* has a frequency of 181 and is, therefore, over three times as likely to be selected for use as input than *laid,* which has a frequency of 59.

Consider as an example the mother–child interactions for *Anne,* which contain 33,390 mother's utterances. The model is presented with all of the first 25% of these utterances. For the next 12.5% of the mother utterances, one in every 10 utterances is replaced by a pair of vocabulary items (in accordance with the figures in Table 1). Similarly, the subsequent 12.5% of the mother utterances have two in every 10 utterances replaced with pairs of vocabulary items. If a NWR test were to be carried out at this stage (i.e., after 50% of the input has been seen), the probability of selecting an incorrect test when traversing the network would be .08.

The model was run 10 times for each of the 12 sets of mother–child data. Ten simulations give a representative estimate of NWR performance for each set of mother–child data given that there are two random elements in the model: the vocabulary items selected for use as input and the probability of making a traversal error. The results of the 10 simulations from each set of mother–child data were then averaged in order to arrive at a mean NWR performance score for each dataset.

The simulations need to vary both long-term knowledge (by manipulating the amount of input seen by the model) and WM capacity (by manipulating the time limit of the phonological store), and so simulations were run at each of the following phonological store time durations: 1,500 msec, 1,600 msec, 1,700 msec, 1,800 msec, 1,900 msec, and 2,000 msec. We also allowed for the possibility that higher values of WM capacity might allow better matches to the data, and therefore also included durations of 2,100 msec and 2,200 msec. Altogether, there were 960 simulations (8 time durations $\times$ 12 children $\times$ 10 runs per child), or 96 when the NWR tests are averaged for each child. For each simulation, a NWR test was carried out for every 12.5% of the input seen by the model so that performance could be analyzed at different levels of knowledge. For each simulation, this resulted in eight NWR tests, one for each "stage of learning" (12.5% of input, 25%, 37.5%, 50%, 62.5%, 75%, 87.5%, and 100%). The NWR test used the nonwords from Jones et al. (2007), as NWR comparisons are being made to the children from this study.

3.3. Method of analysis

There exist a number of methods for measuring the goodness of fit between the simulations of a model and the observed data. Here, to assess the success of different parameter assignments of EPAM-VOC in replicating the child NWR data, we use four methods that appear natural.

First, computing the root mean squared error (RMSE) between the child data and the simulations gives an indication of how well the simulations map onto the child data in terms of raw NWR performance. RMSE estimates the overall error between two sets of data. For each condition (e.g., one-syllable wordlike nonwords) the RMSE value represents the difference in repetition accuracy between the simulations and the child data. RMSE values therefore give an estimate of how closely the simulations match the child data, with low RMSE values indicating that the model matches the child data closely. For analysis purposes, the 10 simulation runs for each set of mother–child data are averaged, as are the RMSE values for each condition. This results in one overall RMSE value for each set of mother–child data that represents the difference in repetition accuracy between the 10 simulations and the child data across all types and lengths of nonwords.

Second, computing correlations for each set of NWR results gives an indication of how well the simulations map onto the trends shown in the child data. If the model perfectly predicts the pattern of variation in the observed data, the correlation should be equal to 1.

Third, subjecting the simulation data to the same analyses of variance (ANOVAs) as those used with the original child data will confirm (or not) that the NWR phenomena that are seen in the child data are also seen in the simulations. The target phenomena are improvement in performance with age, decrease in performance as nonword length increases, and decrease in performance as wordlikeness decreases.

Fourth, examining the types of errors produced by the model and comparing them to the types of error children make provides a finer indication of how well the model simulates the children's data. We compare error data with the kinds of error that 5-year-old children produce (Gathercole, Willis, et al., 1994). Although this makes comparison to 2- to 3-year-old children difficult, the only 2- to 3-year-old error data we know of (Roy & Chiat, 2004) examines syllable errors rather than errors at the phonemic level.

In the first part of the analyses, we examine three different variations of simulation. First, we vary long-term knowledge while keeping WM capacity constant to see what NWR phenomena are explained by increases in long-term knowledge alone. Second, we vary WM capacity while keeping long-term knowledge constant to see what NWR phenomena are explained by increases in WM capacity alone. Third, we allow both variables to vary—that is, we are interested in the interaction of these variables to see if the combination of knowledge and WM capacity provides a better explanation of the data than either increases in long-term knowledge or increases in WM capacity alone.

In all three cases, we are interested in finding the levels of long-term knowledge and WM capacity that minimize RMSE both for 2- to 3-year-olds and 4- to 5-year-olds. Once these levels are found, we use a correlation analysis to help select the simulation that best approximates the child data. The 'winning' simulation is then subjected to an ANOVA in order to test whether the main NWR phenomena seen in the child data are also found in the simulation data. Note that the ANOVA analyses concentrate on the main effects of age, nonword length, and the wordlikeness of the nonwords because these are the central phenomena of interest. This is partly to show whether or not the simulations capture these main effects, and partly to keep the analyses concise. Finally, we examine the error pattern for the 'winning' simulation and compare it to the error patterns in the children.

In the second part of the analyses, we compare the three 'winning' models in more detail by examining RMSE for each type of nonword and each length of nonword. Analyzing the data in finer detail will help to establish (a) which model provides the best fit to the child data, (b) which aspects of the data the models fit best, and (c) where the most important differences between the models lie. The analyses in this section will report interactions because we are now interested in the dynamics of how each model fits the child data.

3.4. Analyses 1: Varying long-term knowledge

This section examines the extent to which long-term knowledge alone can account for the developmental changes in NWR performance. By varying long-term knowledge while WM capacity is held constant, we hope to find a simulation at one level of knowledge that approximates 2- to 3-year-olds' NWR performance and find a simulation at a higher level of knowledge that approximates 4- to 5-year-olds' NWR performance. However, this also involves finding an appropriate WM capacity across the two simulations. As can be seen in

Table 2
RMSE averaged across simulations at varying stages of long-term knowledge (stages of learning within EPAM-VOC) and at varying durations of WM capacity. WM capacities 2,100 and 2,200 are given to examine whether optimal simulations are found for variations that lie outside of normal ranges.

Stage of learning	WM capacity 1,500 ms	1,600 ms	1,700 ms	1,800 ms	1,900 ms	2,000 ms	2,100 ms	2,200 ms
	RMSE comparisons, EPAM-VOC vs. 2–3-year-olds							
1	25.41	23.82	12.24	11.12	10.31	10.33	11.09	11.29
2	24.34	22.18	10.82	10.39	9.79	9.59	10.94	11.72
3	22.88	22.04	9.71	10.04	10.89	11.17	14.40	14.38
4	20.23	20.32	11.82	13.04	13.83	11.98	18.64	18.59
5	17.14	18.22	13.37	14.24	13.52	16.93	22.10	22.22
6	14.17	17.18	16.60	17.29	16.54	19.35	26.31	26.74
7	13.74	16.81	18.77	20.60	22.25	23.09	30.25	29.45
8	12.59	16.24	20.79	24.33	25.54	25.34	33.96	34.02
	RMSE comparisons, EPAM-VOC vs. 4–5-year-olds							
1	47.59	45.68	32.93	29.24	30.41	28.91	21.84	22.16
2	46.04	44.27	29.66	25.91	26.93	27.24	22.07	21.10
3	44.18	42.99	26.93	24.32	22.21	21.71	16.66	17.20
4	41.38	41.12	23.30	20.88	20.52	19.48	13.14	13.35
5	38.02	38.88	21.16	18.54	18.93	15.69	10.65	9.84
6	34.99	36.70	19.02	14.29	15.59	14.15	8.86	9.42
7	31.57	34.73	18.15	12.66	11.90	10.26	8.89	9.53
8	29.27	32.54	16.70	11.19	10.62	10.08	8.92	9.51

Table 2, the simulations that give the lowest combined RMSE values are at WM capacity durations of 2,000 msec and 2,100 msec. For example, keeping WM capacity constant at a duration of 2,000 msec and varying long-term knowledge results in an average RMSE of 9.59 for simulations at Stage 2 of the model's learning (compared to 2- to 3-year-olds) and 10.08 for simulations at Stage 8 (compared to 4- to 5-year-olds). Similarly, keeping WM capacity constant at a duration of 2,100 msec also results in low RMSE values for simulations at Stages 2 and 8 of the model's learning (10.94 and 8.86, respectively).

We use a correlation analysis to further establish the quality of the best assignments of model values. The simulations show good correlations to the 2- to 3-year-old and 4- to 5-year-old data. The 2,000 msec WM capacity duration simulations compare well at Stage 2 to 2- to 3-year-old children, $r(4) = .89$, $p < .02$; and at Stage 8 to 4- to 5-year-olds, $r(4) = .71$, $p > .05$.[3] The 2,100 msec WM capacity duration simulations also account for the data well—$r(4) = .88$, $p < .03$ and $r(4) = .63$, $p > .05$, respectively. Although there is little difference between the RMSE and correlation data, the 2,000 msec WM capacity simulations are slightly better in both cases, and so are analyzed further. The left graph of Fig. 3 shows a comparison between the 2- to 3-year-olds and the Stage 2 simulations, and the right graph shows a comparison between the 4- to 5-year-olds and the Stage 8 simulations.

A 2 (stage of learning: 2 or 8) × 2 (nonword type: wordlike or non-wordlike) × 3 (nonword length: 1, 2, or 3 syllables) mixed ANOVA was performed on the 2,000 msec simulation data.

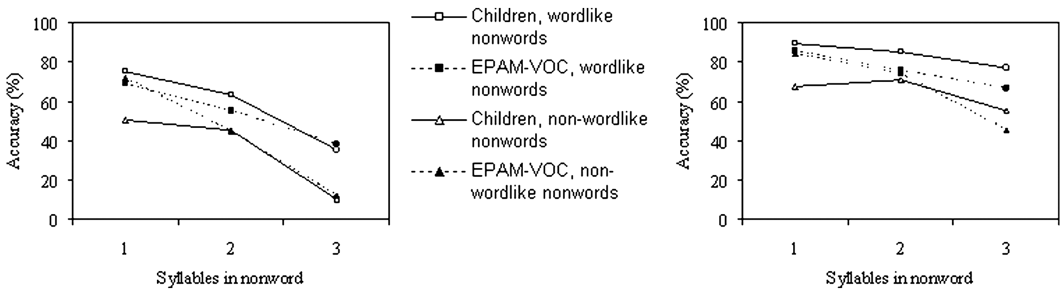

Fig. 3. Two thousand msec capacity duration NWR performance at Stages 2 and 8, compared to 2- to 3-year-old (left graph) and 4- to 5-year-old children (right graph), respectively.

There was a significant effect of stage of learning—$F(1, 22) = 313.17, p < .001$—with better performance at Stage 8; and a significant effect of nonword type—$F(1, 22) = 75.83$, $p < .001$—with better performance for wordlike nonwords. There was also a significant effect of nonword length—$F(2, 44) = 348.76, p < .001$—with post hoc Bonferroni tests showing better performance for one-syllable nonwords over both two- and three-syllable nonwords (both ps $< .001$) and better performance for two-syllable nonwords over three-syllable nonwords ($p < .001$). Importantly, the simulations show the same pattern of performance as the children: There is better performance with age (with age in these simulations corresponding to the amount of knowledge), there is better performance for wordlike nonwords over non-wordlike nonwords, and there is better performance for short nonwords over long nonwords. Variations in long-term knowledge are sufficient to capture the main developmental phenomena in the child data.

The error data of Gathercole, Willis, et al. (1994) indicate that the majority of errors involve phoneme substitution (38%), followed by phoneme deletion (28%) and phoneme deletion and substitution (22%). All other error types occur relatively infrequently (7% or lower). The error types for the simulations also follow this pattern. At Stage 2, substitutions were the most common error (60%), followed by phoneme deletion and phoneme deletion and substitution (both 17%). A similar pattern was found at Stage 8, where substitutions were again the most common error (62%), followed by phoneme deletion (14%) and phoneme deletion and substitution (12%).

3.5. *Analyses 2: Varying WM capacity*

In this section, we examine whether variations in WM capacity alone can account for the developmental NWR data. In a similar manner to the first analyses, we first need to find suitable levels of long-term knowledge in order to establish an appropriate level of knowledge that results in low RMSE values. Taking into consideration RMSE values across both the 2- to 3-year-old and 4- to 5-year-old data (see Table 2), the simulations that give the lowest combined RMSE are at Stages 7 and 8. When the model has seen 87.5% of the input (i.e., Stage 7), a comparison to the 2- to 3-year-old data shows a RMSE of 13.74 at a WM capacity duration of 1,500 msec; and a comparison to the 4- to 5-year-old data shows a RMSE of 8.89 at a WM capacity duration of 2,100 msec. After the model has seen 100% of the input (i.e.,

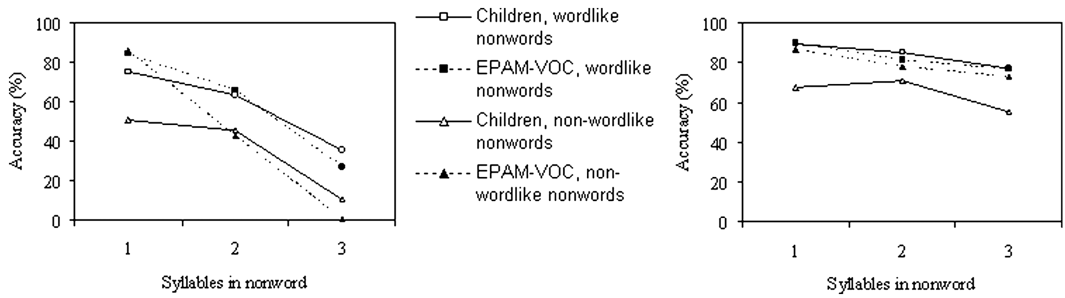

Fig. 4. Stage 8 NWR performance at 1,500 msec and 2,100 msec capacity durations, compared to 2- to 3-year-old (left graph) and 4- to 5-year-old children (right graph), respectively.

Stage 8), the 2- to 3-year-old comparison has a RMSE of 12.59 at a WM capacity duration of 1,500 msec; and the 4- to 5-year-old comparison has a RMSE of 8.92 at a WM capacity duration of 2,100 msec.[4]

The simulations where only WM capacity is varied also show good correlations to the 2- to 3-year-old and 4- to 5-year-old data. The Stage 7 simulations at WM capacity 1,500 msec compare well to 2- to 3-year-olds, $r(4) = .82$, $p < .05$; and reasonably well to 4- to 5-year-olds at a 2,100 msec WM capacity duration, $r(4) = .52$, $p > .05$). The Stage 8 simulations also compare well-$r(4) = .90$, $p < .02$ and $r(4) = .63$, $p > .05$, respectively. Both sets of simulations compare favorably with the child data in terms of correlations, and so analysis is carried out on the Stage 8 simulations, which have slightly lower RMSE values overall and also slightly better correlations. The left graph of Fig. 4 shows NWR performance for the 2- to 3-year-olds compared to the simulations at a WM capacity duration of 1,500 msec; and the right graph shows NWR performance for the 4- to 5-year-olds compared to the simulations at a WM capacity of 2,100 msec.

Finally, we carry out an ANOVA to examine whether the model with the selected value assignment reproduces the phenomena observed with children. A 2 (WM capacity duration: 1,500 msec or 2,100 msec) $\times$ 2 (nonword type: wordlike or non-wordlike) $\times$ 3 (nonword length: 1, 2, or 3 syllables) mixed ANOVA was performed on the Stage 8 simulation data. There was a significant effect of WM capacity duration—$F(1, 22) = 941.94$, $p < .001$—with better performance at 2,100 msec; and a significant effect of nonword type—$F(1, 22) = 156.94$, $p < .001$—with better performance for wordlike nonwords. There was also a significant effect of nonword length—$F(2, 44) = 879.93$, $p < .001$—with post hoc Bonferroni tests showing better performance for one-syllable nonwords over both two- and three-syllable nonwords (both ps $< .001$) and better performance for two-syllable nonwords over three-syllable nonwords ($p < .001$). As with the simulations where long-term knowledge was varied, the data show the same pattern of NWR performance as for the children: There is better performance with age (with age in these simulations corresponding to the phonological store duration), there is better NWR performance for wordlike nonwords over non-wordlike nonwords, and there is better performance for short nonwords over long nonwords. Variations in WM capacity alone are also able to capture the developmental NWR phenomena that are seen in children.

The error data show a different pattern to that of Gathercole, Willis, et al. (1994), who found that the main order of error frequency is phoneme substitution followed by phoneme deletion and then phoneme deletion and substitution. Although the three central error types are again predominant, the order of frequency is different. At 1,500 msec duration, the primary form of error is phoneme deletion (57%), followed by substitution (23%) and deletion and substitution (19%). At 2,100 msec, there are only two main forms of error: substitution (87%) and addition and substitution (11%).

3.6. Analyses 3: Varying both WM capacity and long-term knowledge

If developmental change involves both long-term knowledge and WM capacity, then the best simulation of the children's data would be expected to arise from the interaction between WM capacity and knowledge. This is what we investigate in this analysis by allowing both variables to change as the model learns as a function of time. It is interesting that Table 2 shows that when comparing to the 2- to 3-year-olds, the lowest possible RMSE is obtained with a 2,000 msec WM capacity duration at Stage 2 of the model's learning—the exact same parameter settings as for the long-term knowledge analysis. When comparing to the 4- to 5-year-olds, the lowest possible RMSE is obtained with a 2,100 msec WM capacity duration at Stage 6 of the model's learning. We have already ascertained that the model at Stage 2 with a WM capacity duration of 2,000 msec matches the main effects seen in the children, so we only analyze the model at Stage 6 with WM capacity duration 2,100 msec here.

A 2 (nonword type: wordlike or non-wordlike) $\times$ 3 (nonword length: 1, 2, or 3 syllables) mixed ANOVA was performed on the Stage 6 WM capacity duration 2,100 msec simulation data. There was a significant effect of nonword type—$F(1, 11) = 4.96$, $p < .05$—with better performance for word-like nonwords; and a significant effect of nonword length—$F(2, 22) = 93.05$, $p < .001$. Post hoc Bonferroni tests showed better performance for one-syllable nonwords over both two- and three-syllable nonwords (both *p*s $< .001$) and better performance for two-syllable nonwords over three-syllable nonwords ($p < .003$). The results fit the same pattern of results for the 4- to 5-year-old children. Variations in both long-term knowledge and WM capacity are able to capture the developmental NWR phenomena that are seen in children.

For our analysis of errors, we concentrate on the sixth stage 2,100 msec simulation data because we already know that the second stage 2,000 msec simulation data compare well to the children for types of errors. The sixth stage 2,100 msec simulation data show a different pattern of errors to the children. Only two types of error are predominant: phoneme substitution (86%) and phoneme addition and substitution (10%).

3.7. Examining the match between simulation and child data

We have now identified the simulations that best approximate the children's performance in three cases: when varying levels of long-term knowledge, when varying levels of WM capacity, and when varying levels of long-term knowledge and WM capacity simultaneously. Not surprisingly, given the strong links between knowledge and WM capacity outlined earlier, independently varying long-term knowledge and WM capacity allowed us to match

Table 3
Average RMSE values for each 'winning' simulation type, computed by nonword type and nonword length

	Syllables in wordlike nonwords			Syllables in non-wordlike nonwords		
Simulation	1	2	3	1	2	3
RMSE for simulations of 2–3-year-old NWR performance						
Long-term knowledge	.35	.44	.38	1.06	.26	.39
WM capacity	.46	.30	.51	1.75	.26	.50
Interaction	.35	.44	.38	1.06	.26	.39
RMSE for simulations of 4–5-year-old NWR performance						
Long-term knowledge	.28	.49	.53	.81	.28	.64
WM capacity	.14	.19	.18	.95	.35	.88
Interaction	.30	.59	.53	.80	.19	.25

the developmental NWR phenomena in both cases. When both were allowed to vary simultaneously, we determined that the best simulations to 2- to 3-year-old children were the same as those seen when only long-term knowledge was varied, but the best simulation to 4- to 5-year-old children was for a new pair of parameter settings. We now further investigate the pattern of results by examining the knowledge, WM capacity, and interaction simulations in more detail to see how well each is able to match the intricacies of the child data.

To provide a more fine-grained analysis of the closeness of fit for the simulations, the analyses in this section focus on performance for each nonword type and for each nonword length. Rather than use raw NWR performance scores, RMSE values are used because these will indicate how well each simulation matches the child data for each type and length of nonword—that is, we already know that all the 'winning' simulations match the basic findings seen in the children. The goal here, therefore, is to examine the pattern of error across each of the simulations by examining the RMSE error rates across each nonword type and length. Analyzing the data in this much detail will not only indicate which model provides the best fit to the child data, but also which aspects of the child data the models are most—and least—successful in accounting for. Table 3 shows RMSE figures for each type of nonword at each of the three syllable lengths, for the 'winning' knowledge, WM capacity, and interaction simulations.

3.7.1. Two- to three-year-old data

We only analyze the long-term knowledge and WM capacity simulations here because the 'winning' interaction simulations were the same as those for long-term knowledge. A 2 (simulation type: long-term knowledge or WM capacity) $\times$ 2 (nonword type: wordlike or non-wordlike) $\times$ 3 (nonword length: 1, 2, or 3 syllables) ANOVA was computed on the RMSE data for the 2- to 3-year-old simulations. There was a main effect of simulation type—$F(1, 22) = 13.72$, $p < .002$—with the long-term knowledge simulation having lower RMSE rates than the WM capacity simulation. There was also a main effect of nonword type—$F(1, 22) = 64.40$, $p < .001$—with RMSEs being lower for wordlike nonwords. Finally, there was a main effect of nonword length, $F(2, 44) = 103.82$, $p < .001$. Post hoc Bonferroni tests showed that

RMSE rates for two-syllable nonwords were lower than those for one-syllable ($p < .001$) and three-syllable nonwords ($p < .007$), and RMSE rates for three-syllable nonwords were lower than those for one-syllable nonwords ($p < .001$).

There were also significant interactions involving the simulation variable, indicating the areas where the long-term knowledge simulations provided a better fit to the data than the WM capacity simulations. Specifically, whereas the mean difference in RMSE error rates between the long-term knowledge simulations and the WM capacity simulations was only –0.18 for the wordlike nonwords, it was –0.42 with the non-wordlike nonwords, $F(1, 22) = 9.80$, $p < .006$. The interaction between type of simulation and nonword length—$F(2, 44) = 15.20$, $p < .001$—comes from the fact that the difference between the two simulations is larger with one-syllable (–0.40) than with two-syllable (0.07) and three-syllable nonwords (–0.12). There was also an interaction between nonword type and nonword length—$F(2, 44) = 65.83$, $p < .001$—indicating that for both of the simulations, RMSE error rates were particularly high for non-wordlike one-syllable nonwords.

3.7.2. Four- to 5-year-old data

The 4- to 5-year-old data were subjected to the same analysis as above, but this time the 'winning' interaction simulation is included because the three 'winning' simulations in the 4- to 5-year-old comparisons are all different from each other. A 3 (simulation type: long-term knowledge, WM capacity, or interaction) × 2 (nonword type: wordlike or non-wordlike) × 3 (nonword length: 1, 2, or 3 syllables) ANOVA was computed on the RMSE data for the 4- to 5-year-old simulations. As per the previous analyses, there were main effects of nonword type, $F(1, 33) = 35.68$, $p < .001$; and nonword length, $F(2, 44) = 103.82$, $p < .001$. However, the most important finding was the fact that there was no effect of simulation type—$F(2, 33) = 1.39$, $p > .05$—illustrating that RMSE rates were similar across all three types of simulation.

There were also significant interactions involving the simulation variable. The interaction involving nonword type—$F(2, 33) = 25.27$, $p < .001$—illustrated that the WM capacity simulations provided the lowest error rates for wordlike nonwords and yet the highest error rates for non-wordlike nonwords. The interaction involving nonword length—$F(4, 66) = 3.51$, $p < .02$—illustrated that there were no differences across simulations for one- and two-syllable nonwords but the interaction simulation had lower error rates for three-syllable nonwords. Similar to the 2- to 3-year-old analysis, there was also an interaction between nonword type and nonword length—$F(2, 66) = 41.89$, $p < .001$—indicating particularly high error rates for non-wordlike one-syllable nonwords.

3.8. Summary

In summary, the analyses in this section showed that the knowledge and interaction simulations provided a closer match to the 2- to 3-year-old child data compared to the WM capacity simulations. There were no major differences across simulations in comparisons to the 4- to 5-year-old data. This indicates that, overall, the long-term knowledge simulations provided a closer match to the child data than the WM capacity simulations, with little benefit arising from allowing both knowledge and WM capacity to vary. The analyses also revealed the general success of all three types of simulation in matching the child data. First, the simulations matched the child data best for word-like nonwords and for nonwords of two and three syllables. The latter is particularly important because it illustrates that the probability of making a traversal error in EPAM-VOC, which was based on children's error rates for one-syllable nonwords, is influenced by the dynamics of the model. Second, the simulations are poor for one-syllable non-wordlike nonwords, indicating a specific area where the model needs further development.

4. Discussion

The goal of this article was to investigate the respective roles of increasing long-term knowledge and increasing WM capacity in explaining developmental change. A computational model of vocabulary learning was presented that was able to simulate children's NWR performance. Long-term knowledge and WM capacity were each systematically varied independently of one another, showing that both were able to capture the central findings in NWR performance: improved performance with age; improved performance for shorter nonwords; and improved performance for wordlike nonwords. Allowing both knowledge and WM capacity to vary (i.e., allowing the two to interact) revealed that the best simulations of the 2- to 3-year-old children were the same as those where only knowledge was varied, although the best simulations of the 4- to 5-year-old data arrived at a new set of parameter assignments. However, an analysis of the patterns of error made by the children and the simulations showed that variations in task knowledge provided the best fit to the types of error made by children. Taken as a whole, these findings suggest that long-term knowledge alone may be sufficient to match the developmental data.

A further, more fine-grained analysis was performed on the three 'winning' simulations to examine where each simulation matched the child data for type and length of nonword. These analyses indicated that for the 2- to 3-year-old data, both the long-term knowledge simulations and the interaction simulations provided a significantly better fit to the intricacies of the child data than the WM capacity simulations. By contrast, no differences between any of the simulations were found for the 4- to 5-year-old children. The results suggest that increases in WM capacity may not be necessary to explain developmental change given that increases in long-term knowledge must occur during development. The results are important not only for NWR and vocabulary learning, but also for developmental psychology in general. We now discuss implications for each of these areas.

4.1. Implications for developmental psychology

The clear finding from the results presented is that independent changes in long-term knowledge and WM capacity are both sufficient to simulate developmental data. This is important because it illustrates that long-term knowledge and WM capacity share strong links with each other, suggesting that it may well be very difficult to measure each of these factors independently of the other.

Upon closer inspection, the results showed that increases in long-term knowledge provided a significantly better match to the child data. Given that no serious developmental theory would argue against changes in long-term knowledge, we can assume that increases in the child's knowledge base must constitute a significant part of the child's development. If this is the case, then one can legitimately ask whether changes in long-term knowledge cause perceived changes in other mechanisms of development. The findings here illustrate that, at least for simulations of NWR performance, changes in long-term knowledge can account for apparent changes in WM capacity. It is, therefore, possible that any changes in performance on developmental tasks that are hypothesized to arise from increases in WM capacity may simply arise from increases in long-term knowledge.[5]

The results show support for both the chunking hypothesis and Case's (1985) idea that WM capacity remains fixed throughout development. As knowledge develops, the units used for measuring WM capacity change, where previously independent units of knowledge become grouped into a new memory structure that can now be used as a single unit (Lane, Gobet, & Cheng, 2001). WM capacity therefore remains constant, but through long-term chunking the amount of information that can be held in WM increases over time. Increases in long-term domain knowledge therefore give rise to the perception that there are associated increases in WM capacity because expansions in long-term knowledge result in an ability to hold more information in WM.

Knowledge effects have been seen in a variety of domains, particularly with regard to expertise. For example, children who have expertise in chess are able to hold more information in WM than their non-chess-playing peers in chess-related memory tests, whereas in domains where both sets of children are non-experts, no differences are seen in tests of WM capacity (Chi, 1978; Schneider, Gruber, Gold, & Opwis, 1993).

The possibility that changes in WM capacity are an artifact of changes in long-term memory is also consistent with previous results in the memory literature. For example, Swanson (1999) found clear relationships between both verbal and visuospatial WM capacities and reading and mathematics ability—high scores on the WM capacity tasks were therefore related to high scores on the ability tasks that tap into long-term knowledge. Although it could be argued that reading and mathematics ability do not directly relate to the WM tasks that Swanson carried out, they may be indicative of a larger knowledge base and hence the possibility that long-term knowledge provided a significant contribution to the apparent age-related differences in WM capacity.

Bayliss, Jarrold, Gunn, and Baddeley (2003) tested complex span (memory span involving both WM capacity and a processing component) together with traditional tests of WM capacity. They found that domain-specific WM capacity tasks made significant contributions to both children's and adults' performance of complex span tasks that involved the same form of storage. For example, performance on verbal WM capacity tasks made significant contributions to performance on complex span tasks involving a verbal storage component. This is consistent with the hypothesis that long-term knowledge within a domain (the domains here being rather general—verbal or visuospatial) may influence the amount of information that can be held in WM capacity for that domain—that is, more verbal long-term knowledge results in a larger WM capacity for verbal information and thus, better performance on complex tasks involving verbal storage of information.

The idea that WM capacity tasks may contain a long-term component has also been put forward for traditional WM capacity tasks. For example, digit-span tasks have been criticized for involving long-term knowledge, such as familiarity with the digits used (e.g., Case, Kurland, & Goldberg, 1982). However, there is evidence to suggest that there may be WM capacity differences over and above any long-term knowledge influences. Although young children have been shown to have more knowledge for lower numbers than higher ones (Dehaene & Mehler, 1992), Cowan, Nugent, Elliott, Ponomarev, and Saults (1999) found age-related differences for a version of the digit-span task, but found no evidence of digit preference in children of younger ages. This would suggest that, although long-term knowledge plays a significant role in the child's development, there may also be developmental increases in WM capacity.

The evidence thus far suggests that, although long-term knowledge may explain age-related differences in WM capacity, there is still the possibility of developmental differences in WM capacity itself. As previously mentioned, studies involving the measurement of WM capacity are difficult to interpret because of the influence of long-term knowledge. In this respect, studies that appear to show age-related differences in WM capacity should be treated with caution unless there are clear empirical reasons for preferring an explanation in terms of increases in WM capacity over an explanation in terms of increasing knowledge. On the basis of the results that have been presented here, we would argue that "clear empirical reasons" are not only phenomena that give the appearance of a WM capacity explanation but also those where computer simulations have shown that the target phenomena cannot be simulated using an explanation involving only long-term knowledge. Only when both of these stipulations are met can one legitimately conclude that age-related WM capacity differences are required to explain developmental change.

One final finding within the memory literature that may appear not to fit easily with a long-term knowledge explanation is the decline in memory performance for older adults (e.g., Salthouse, 1990; Swanson, 1999). However, this can be explained from a pure knowledge view of WM capacity if one assumes the knowledge itself is difficult to access in older populations because of interference. Hasher and Zacks (1988) suggested that older adults have difficulty removing items from WM and, as such, these items interfere with others. Similar views are also held by Dempster (1993) and Bjorklund and Harnishfeger (1990).

4.2. Implications for NWR and vocabulary learning

Although NWR involves both WM capacity and long-term knowledge, WM capacity is seen by many as being the most important factor (e.g., Baddeley, Gathercole, & Papagno, 1998; Gathercole & Baddeley, 1990a, 1990b; Gathercole, Willis, et al., 1994). The results here suggest that the relationship between WM capacity and long-term knowledge is actually a complex one that changes over time. As more phonological knowledge is acquired, more information can be captured in a fixed WM capacity and, thus, shifts in performance are seen. These shifts do not require any alteration in WM capacity—they only require increases in long-term phonological knowledge.

Although EPAM-VOC supports views of vocabulary learning that highlight phonological knowledge as the key mediator (e.g., Bowey, 1996; Metsala, 1999), more recent theoretical explanations have attempted to clarify the respective roles of long-term knowledge and WM capacity. In particular, Gathercole (2006) suggested that auditory processing and phonological analysis are used to construct a phonological representation of the nonword; and, on the basis of this, redintegration may occur based on the amount of overlap between the phonological form and stored lexical entries (i.e., words). This suggests that the relative role of WM capacity depends upon the type of nonword—those nonwords that share few features with lexical items will place more reliance on WM capacity. This explanation is somewhat true of all of the simulations presented here—when nonwords had strong links to long-term knowledge (i.e., word-like nonwords), there was a closer match to the child data in terms of lower RMSE rates. However, the results also suggest WM capacity may need further investigation. If non-wordlike nonwords emphasize the role of WM capacity, then the simulations where only WM capacity was varied should have shown a better fit to non-wordlike nonwords than wordlike nonwords. In fact, the better fit to the data was seen for word-like nonwords.

One line of research that may help in identifying the roles of long-term knowledge and WM capacity in NWR performance involves specific language impairment (SLI). For example, Archibald and Gathercole (2006) found that children with SLI have a WM capacity deficit that is restricted to the verbal domain (implicating phonological WM capacity deficits), and Marton and Schwartz (2003) also implicated WM in suggesting that children with SLI have problems of simultaneous processing. Further research also suggests WM capacity problems for language-impaired learners (e.g., De Beni, Palladino, Pazzaglia, & Cornoldi, 1998; De Jong, 1998).

At first blush, these results speak against the role of long-term knowledge. However, the interpretation of these studies suffers from the same problems as those highlighted in the introduction of this article—namely, that as phonological long-term learning occurs, the units used to measure WM capacity change. We believe that computational modeling is a tool that can be used to help in examining language impairments. We have supported the view that WM capacity is closely linked to long-term knowledge, and it now needs to be ascertained whether language impairments lie in WM capacity limitations (as suggested earlier) or alternatively general language learning limitations (such as slow learning; Gray, 2006), degraded phonological long-term representations (as suggested by Service, 2006), a combination of these, or some other form of deficit. Computer models such as EPAM-VOC can be used to examine the effects that each has upon subsequent NWR performance—based on the fit of the model to the data, specific hypotheses can be generated to help pinpoint potential areas of impairment.

4.3. Limitations of the model

The results presented provide an indication that changes in long-term knowledge may be sufficient to account for developmental changes in the NWR task among 2- to 5-year-old children. However, there are some limitations of the model that one needs to consider before accepting this conclusion.

First, WM is represented as a simple time-limited store that allocates a temporal duration to each part of the input. Once the time allocation for the input exceeds the duration of the phonological store (2,000 msec), the remainder of the input is not processed. This does not harmonize with recall effects in the adult literature, where primacy and recency effects have been found for nonwords (Gupta, 2005). Although there is some contention concerning primacy effects and rehearsal in young children (e.g., Siegel, Allick, & Herman, 1976), recency effects have been found (e.g., Hagen & Kingsley, 1968). Future versions of EPAM-VOC therefore need to incorporate a recency mechanism whereby the most recent part of the input is available for processing.

Second, the NWR test carried out by the model involves it being able to encode the nonword within the time-limited capacity of the phonological store. It could be argued, therefore, that rather than the model performing NWR, EPAM-VOC is performing nonword recognition. In fact, a fuller account of the NWR process should not only include an encoding process but also an articulation process. This is a goal to be achieved in future versions of the model.

Third, the model does not account for memory effects, such as serial order effects (e.g., Thorn & Frankish, 2005). The instantiation of WM in EPAM-VOC is most aligned to chaining accounts—items in WM are recalled based on the context of preceding items (this is most applicable when several items exist as the contents of a node). However, chaining accounts have been criticized in terms of their adequacy in explaining serial recall effects. For example, Henson, Norris, Page, and Baddeley (1996) found that confusable items in a list (e.g., phonologically similar items) have no obvious influence on the likelihood of correctly recalling non-confusable items. As such, when a non-confusable to-be-recalled item exists in a list containing confusable and non-confusable items, the preceding–succeeding items do not predict the recall likelihood of the to-be-recalled item. As Henson et al. noted, these findings present difficulties for EPAM-like models that are predominantly of the chaining variety. Future versions of EPAM-VOC need to consider how its account of WM can deal with the type of serial recall findings presented earlier.

4.4. Overall summary

A computational model of NWR performance has shown that developmental changes in vocabulary learning are likely to be mediated by long-term phonological knowledge rather than WM capacity. It is, therefore, possible that WM capacity explanations of developmental change actually arise from differences in long-term knowledge. These results suggest the need for caution when evaluating WM capacity explanations of developmental change, with researchers only invoking developmental changes in WM capacity when there are clear empirical reasons for doing so.

The use of computational models can help in examining the relative contributions of long-term knowledge and WM capacity within developmental tasks because they allow the two to be independently manipulated so that the relative influence of each can be examined. Using the domain of vocabulary learning, we compared variations in long-term knowledge and variations in WM capacity, showing that it is likely that the key mediator in age-related differences is long-term vocabulary knowledge.

Specific language impairment is a key area where further examination of vocabulary learning is necessary because there is a wealth of research that points toward WM capacity impairments, whereas alternative explanations could exist relating to long-term knowledge. Computational modeling techniques could be of particular value in this domain because they provide a key resource in helping to identify where the deficits lie.

Notes

1. Working memory capacity in the context of this article refers to the storage component of working memory.
2. EPAM stands for Elementary Perceiver and Memorizer, CHREST for Chunk Hierarchy and REtrieval STructures, and MOSAIC for Model Of Syntax Acquisition In Children.
3. Given such small sample sizes (6 data points), only high correlation coefficients (.81 or above) are significant.
4. Note that the error probabilities decrease as the stages of learning increase. The error probabilities reflect improvement in the long-term processes of encoding and articulation and, therefore, reflect increases in long-term knowledge. A correlation between the error probabilities at each stage of learning and the number of nodes in the model at each stage confirm this relation, $r(6) = -.99$, $p < .001$.
5. Note that we are referring here to developmental differences in working memory capacity rather than individual differences.

References

Abbs, B., Gupta, P., Tomblin, J. B., & Lipinski, J. (2007). A behavioral and computational integration of phonological, short-term memory, and vocabulary acquisition processes in nonword repetition. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual conference of the Cognitive Science Society* (pp. 59–64). New York: Lawrence Erlbaum Associates, Inc.

Archibald, L. M. D., & Gathercole, S. E. (2006). Visuospatial immediate memory in specific language impairment. *Journal of Speech, Language, and Hearing Research, 49*, 265–277.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85–97.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 47–90). New York: Academic Press.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586–595.

Baddeley, A. D., Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behaviour, 14*, 575–589.

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*, 1265–1282.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*, 71–92.

Bjorklund, D. F., & Harnishfeger, K. K. (1990). The resources construct in cognitive development: Diverse sources of evidence and a theory of inefficient inhibition. *Developmental Review, 10*, 48–71.

Bowey, J. A. (1996). On the association between phonological memory and receptive vocabulary in five-year-olds. *Journal of Experimental Child Psychology, 63*, 44–78.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34*, 594–621.

Brown, G. D. A., & Hulme, C. (1996). Nonword repetition, STM, and word age-of-acquisition: A computational model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 129–148). Hove, UK: Psychology Press.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127–181.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551–581.

Cantor, J., & Engle, R. W. (1993). Working memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1101–1114.

Case, R. (1985). *Intellectual development: Birth to adulthood.* Orlando, FL: Academic Press.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Cowan, N., Nugent, L. D., Elliott, E. M., Ponomarev I., & Saults, J. S. (1999). The role of attention in the development of short-term memory: Age differences in the verbal span of apprehension. *Child Development, 70*, 1082–1097.

Croker, S., Pine, J. M., & Gobet, F. (2003). Modelling children's negation errors using probabilistic learning in MOSAIC. In F. Detje, D. Dörner & H. Schaub (Eds.), *Proceedings of the 5th International Conference on Cognitive Modeling* (pp. 69–74). Bamberg, Germany: Universitäts-Verlag.

De Beni, R., Palladino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 51*, 305–320.

De Jong, P. (1998). Working memory deficits of reading disabled children. *Journal of Experimental Child Psychology, 70*, 75–95.

Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition, 43*, 1–29.

Dempster, F. N. (1993). Resistance to interference: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*, 45–75.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like models of recognition and learning. *Cognitive Science, 8*, 305–336.

Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modelling the developmental patterning of finiteness marking in English, Dutch, German and Spanish using MOSAIC. *Cognitive Science, 31*, 311–341.

Freudenthal, D., Pine, J. M., & Gobet, F. (2002). Modelling the development of Dutch optional infinitives in MOSAIC. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th annual meeting of the Cognitive Science Society* (pp. 328–333). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Freudenthal, D., Pine, J. M., & Gobet , F. (2006). Modelling the development of children's use of optional infinitives in English and Dutch using MOSAIC. *Cognitive Science, 30*, 277–310.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language, 42*, 481–496.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory and Cognition, 23*, 83–94.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics, 27*, 513–543.

Gathercole, S. E., & Adams, A.-M. (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 and 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.

Gathercole, S. E., & Baddeley, A. D. (1990a). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336–360.

Gathercole, S. E., & Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*, 439–454.

Gathercole, S. E., & Pickering, S. J. (1999). Estimating the capacity of phonological short-term memory. *International Journal of Psychology, 34*, 378–382.

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*, 103–127.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1991). The influence of number of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics, 12*, 349–367.

Gobet, F., & Lane, P. C. R. (2005). The CHREST architecture of cognition: Listening to empirical data. In D. Davis (Ed.), *Visions of mind: Architectures for cognition and affect* (pp. 204–224). Hershey, PA: IPS.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C. H., Jones, G., Oliver, I., et al (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences, 5*, 236–243.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24*, 651–682.

Gray, S. (2006). Commentary on keynote. *Applied Psycholinguistics, 27*, 562–564.

Gupta, P. (2005). Primacy and recency in nonword repetition. *Memory, 13*, 318–324.

Hagen, J. W., & Kingsley, P. R. (1968). Labeling effects in short-term memory. *Child Development, 39*, 113–121.

Halford, G. S. (1993). *Children's understanding: The development of mental models*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Halford, G. S., Maybery, M. T., & Bain, J. D. (1986). Capacity limitations in children's reasoning: A dual-task approach. *Child Development, 57*, 616–627.

Hartley, T., & Houghton, G. (1996). A linguistically constrained model of short-term memory for nonwords. *Journal of Memory and Language, 35*, 1–31.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Henson, R. N. A., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Quarterly Journal of Experimental Psychology, 49A*, 80–115.

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.

Kail, R. (1988). Developmental functions for speeds of cognitive processes. *Journal of Experimental Child Psychology, 45*, 339–364.

Klahr, D., & Wallace, J. G. (1976). *Cognitive development: An information processing view*. Hillsdate, NJ: Lawrence Erlbaum Associates, Inc.

Lane, P. C. R., Gobet, F., & Cheng, P. C. H. (2001). What forms the chunks in a subject's performance? Lessons from the CHREST computational model of learning. *Behavioral and Brain Sciences, 24*, 128–129.

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.

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.

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.

Munson, B., Edwards, J., & Beckman, M. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61–78.

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*, 1033–1047.

Ormrod, J. E., & Cochran, K. F. (1998). Relationship of verbal ability and working memory to spelling achievement and learning to spell. *Reading Research and Instruction, 28*, 33–43.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory and inhibitory control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology, 80*, 44–57.

Piaget, J. (1950). *The psychology of intelligence*. London: Routledge & Kegan Paul.

Piaget, J. (1952). *The origins of intelligence in children*. New York: International Universities Press.

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.

Salthouse, T. A. (1990). Working-memory as a processing resource in cognitive aging. *Developmental Review, 10*, 101–124.

Schneider, W., Gruber, H., Gold, A., & Opwis, K. (1993). Chess expertise and memory for chess positions in children and adults. *Journal of Experimental Child Psychology, 56*, 328–349.

Service, E. (2006). Phonological networks and new word learning. *Applied Psycholinguistics, 27*, 581–584.

Siegel, A. W., Allik, J. P., & Herman, J. F. (1976). The primacy effect in young children: Verbal fact or spatial artifact? *Child Development, 47*, 242–247.

Siegler, R. S. (1976). Three aspects of cognitive development. *Cognitive Psychology, 8*, 481–520.

Siegler, R. S. (1995). Children's thinking: How does change occur? In W. Schneider & F. Weinert (Eds.), *Memory, performance and competencies: Issues in growth and development* (pp. 405–430). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Simon, H. A. (1974). How big is a chunk? *Science, 183*, 482–488.

Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental Psychology, 35*, 986–1000.

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.

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 729–735.

Treiman, R., & Danis, C. (1988). Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 145–152.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, NY: Cambridge University Press.

Zhang, G., & Simon, H. A. (1985). STM capacity for Chinese words and idioms: Chunking and acoustical loop hypothesis. *Memory and Cognition, 13*, 193–201.