



Published in final edited form as:

Q J Exp Psychol (Hove). 2023 January ; 76(1): 196–219. doi:10.1177/17470218221090483.

The development of lexical competition in written- and spoken-word recognition

Keith S Apfelbaum¹, Claire Goodwin¹, Christina Blomquist², Bob McMurray^{1,3}

¹Department of Psychological and Brain Sciences, University of Iowa, Iowa City, IA, USA

²Department of Communication Sciences and Disorders, University of Maryland, College Park, MD, USA

³Department of Communication Sciences and Disorders, Department of Linguistics, Department of Otolaryngology, University of Iowa, Iowa City, IA, USA

Abstract

Efficient word recognition depends on the ability to overcome competition from overlapping words. The nature of the overlap depends on the input modality: spoken words have temporal overlap from other words that share phonemes in the same positions, whereas written words have spatial overlap from other words with letters in the same places. It is unclear how these differences in input format affect the ability to recognise a word and the types of competitors that become active while doing so. This study investigates word recognition in both modalities in children between 7 and 15. Children complete a visual-world paradigm eye-tracking task that measures competition from words with several types of overlap, using identical word lists between modalities. Results showed correlated developmental changes in the speed of target recognition in both modalities. In addition, developmental changes were seen in the efficiency of competitor suppression for some competitor types in the spoken modality. These data reveal some developmental continuity in the process of word recognition independent of modality but also some instances of independence in how competitors are activated. Stimuli, data, and analyses from this project are available at: <https://osf.io/eav72>.

Keywords

Development; word recognition; visual world paradigm; visual word recognition; spoken word recognition

Corresponding author: Keith S Apfelbaum, Department of Psychological and Brain Sciences, University of Iowa, G60 PBSB, Iowa City, IA 52242, USA. keith-apfelbaum@uiowa.edu.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Supplementary material

The supplementary material is available at qjep.sagepub.com.

Introduction

Effective language use relies on rapid and accurate word recognition. This is true for both spoken and written words. For spoken words, the rapid arrival of the input leads listeners to activate an array of words as soon as any input arrives and to incrementally update this array as additional input arrives. This represents a form of competition in which multiple candidates are briefly active and compete until one remains. For written words, the fact that most words share letters with other words requires a reader to suppress partially overlapping items as they read.

Accurate and efficient word recognition is fundamental to higher level language. Thus, the ability to manage competition is a foundational skill that should develop rapidly. However, recent work suggests that spoken-word recognition—and even speech perception—shows a surprisingly protracted development that continues to develop through adolescence (McMurray et al., 2018; Rigler et al., 2015), despite years of language input. And of course, written-word recognition depends on reading abilities that develop across many years of childhood.

The slow development of these processes raises important questions about the nature of lexical competition across development. Previous studies have shown that late teenagers and adults better navigate lexical competition for spoken words than elementary school children (McMurray et al., 2018; Rigler et al., 2015; Sekerina & Brooks, 2007). These studies confirm that word recognition is changing at some point during childhood, supporting the idea that the development of spoken-word recognition is far from complete when children enter school. However, these studies included groups with large differences in age and so are unable to pinpoint when these changes occur.

Moreover, little work directly assesses the development of competition for written-word recognition. Written language relies on a substantially different input format: words are typically static, without the temporal structure that seems to govern aspects of spoken-word recognition. Nonetheless, written-word recognition has long been thought to rely on similar competition mechanisms as spoken-word recognition (e.g., in interactive activation models for spoken words: McClelland & Elman, 1986; and written words: Harm & Seidenberg, 2004; McClelland & Rumelhart, 1981), although competition in written words is governed by spatial overlap of letters rather than temporal overlap of phonemes. This raises important developmental questions. In particular, it is unclear how letter position affects which classes of words compete: early readers may process letters more serially (left to right) while older readers could process them in parallel. Moreover, it is unclear if spoken and written words rely on similar competition resolution mechanisms. As we argue, this lack of evidence is in part because there are few good methods for assessing competition in written-word recognition.

The current study attempts to fill these gaps by measuring the real-time dynamics of word recognition in both the spoken and the written modalities over development. This offers novel insight into the nature and development of word recognition in the written modality and a means of directly comparing its development to concurrent spoken-language

recognition. This study also includes smaller age gaps to help identify how word recognition develops during childhood.

Spoken-word recognition

The dynamics of spoken-word recognition are well understood, especially for adults (Alloppenna et al., 1998; Marslen-Wilson & Warren, 1994; Marslen-Wilson & Zwitserlood, 1989; Weber & Scharenborg, 2012). Listeners activate words as soon as any acoustic information arrives and modulate this activation in response to later-arriving information. Critically, because processing proceeds incrementally, at early points in recognition the system works on the basis of only partial information (e.g., the /wi-/ in *wizard*); consequently, multiple words are activated in parallel (*wizard*, *whistle*, *window*). Activation levels of candidate words are graded, such that words are activated to different degrees (Marslen-Wilson, 1987), and these activations are updated dynamically throughout processing (Dahan & Gaskell, 2007; Frauenfelder et al., 2001).

A prominent feature of spoken-word recognition is the dominance of information early in words. Words that share onsets (cohorts) exhibit the strongest competition (Alloppenna et al., 1998; Marslen-Wilson, 1987), with other types of words (e.g., rhymes and anadromes) showing less (Alloppenna et al., 1998; Hendrickson et al., 2021; Marslen-Wilson et al., 1996; Simmons & Magnuson, 2018; Toscano et al., 2013).¹ This can also be observed in interference patterns: the number of onset competitors is a better predictor of recognition speed than the number of traditional neighbours (Magnuson et al., 2007). Although onset competitors are eventually suppressed as later mismatching acoustic information arrives, they exert considerable competition during the earliest stages of word recognition and may limit consideration of words that don't overlap at onset (Simmons & Magnuson, 2018).

Online spoken-word recognition in toddlers shows profiles of competition that are qualitatively similar to those of adults (Fernald et al., 1998, 2006). Given this, it was expected that by school age, children should be able to apply this in a mostly adult-like manner. However, recent investigations suggest a more protracted developmental time course. Rigler and colleagues (2015) investigated lexical competition in 9 and 16 year olds using eye-tracking in the Visual World Paradigm (VWP). In their study, children selected a picture corresponding to a spoken word (e.g., *bees*) from a set of pictures including the target, a cohort competitor (*bean*), a rhyme (*peas*), and an unrelated object (*cap*). Children's word recognition showed the same general pattern as adults: words were processed incrementally, and cohort competitors had a stronger effect than rhymes. However, there were consistent age-related differences in the subtle dynamics of this process. Older children showed more rapid fixations to spoken targets, and younger children showed increased looks to competitors and unrelated objects. The younger group also persisted in looking at non-target items for substantially longer than the older group. The faster target looks are perhaps unsurprising—children get faster at many processes as they age. However, the

¹The provenance of rhyme activation is also debated: Desroches and colleagues (2009) argue that rhymes are not activated by the bottom-up phonological input, but instead via spreading activation from other words; Huettig and colleagues (2011) suggest that rhymes are not activated at all in typical situations and that evidence of such may reflect task-specific strategies. At best, rhymes are activated weakly, as the incremental processing of onset information swamps their ability to reach high levels of activation; at worst, rhymes might not even be activated by bottom-up lexical activation processes, and may not directly compete for recognition.

changes in competitor fixations suggest that their *lexical recognition processes* also become more efficient with age—older children show lower consideration of competitors initially, and they suppress these competitors more quickly. Contrary to the prevailing ideas that the process of word recognition develops in early childhood, these data indicate improvements in the efficiency of word recognition even between 9 and 16 years.

Written-word recognition

The emphasis on word-initial information makes sense for spoken-word recognition, where the temporal structure of the input forces word recognition to operate quickly and start activating candidate words before the end of the word arrives. In contrast, written words are present in their entirety throughout processing; a reader does not need to preferentially attend to word onsets as the words are static, and onsets and offsets are co-present in the input. Skilled readers seem to read words in parallel: they tend to make a single fixation to most words, and this fixation tends to be near the centre of the word (McConkie et al., 1988)—they do not scan the word letter-by-letter or left-to-right, but instead visually parse it with a single fixation.

Nonetheless, even under assumptions of completely parallel processing, one should still expect competition although arising from a different source. The C and A in CAT, for example, will lead to partial activation of CAP and CAN (which share those letters), although CAT will ultimately receive more activation. But lexical activation of written words might not show the same emphasis on word initial overlap that is seen in spoken words; HAT and COT share the same number of letters with CAT, and the lack of temporal structure of the input means they can be activated simultaneously. Consequently, profiles of lexical competition might show greater balance among words that overlap in different places (i.e., different letter positions within a word).

There is mixed evidence for this. Some studies suggest a broader emphasis on word onsets, akin to the effects in spoken-word recognition. Rastle and Coltheart (1999) found greater disruption from irregularities in spelling early in words than those late in words; they attribute this to serial processing of letters. In a masked priming paradigm, Carreiras and Perea (2002) found evidence of inhibition for prime words that shared the first syllable with targets, and facilitation from nonwords with similar overlap. However, this same study failed to find such inhibitory effects for single-syllable primes that share the first several letters, suggesting that the effects may be related to syllable structure more than letter order. In contrast, another study using this paradigm found inhibitory effects only for words that differed in medial letters (e.g., frost-FRONT; Perea, 1998), which runs counter to evidence for prominence of word-initial information.

Yet other studies suggest heightened sensitivity to information at the beginnings *and* ends of words relative to word medial information, which suggests prominence of these letter positions during recognition (Estes et al., 1976; Perea & Lupker, 2003). The supposed “Cambridge University effect” reports a (likely apocryphal) study showing that people can read text with relative ease when the letters in the middle of the words are rearranged (e.g., *porbelm* for *problem*) as long as the first and last letters are maintained. Despite debate about the veracity of this effect, several studies do show evidence of increased attention to

initial and final letters, with greater flexibility when interior letters vary. Perea and Lupker (2003) found that letter transpositions in the interior of a word (*uhser* for *usher*) were less disruptive than transpositions at the ends of words (*ushre*), and similar asymmetries arise between word medial transpositions compared to those at both the beginnings and ends of words (White et al., 2008). In general, these findings suggest that at some places in the word, letter position may be coded fairly coarsely (and see Toscano et al., 2013 and Dufour & Grainger, 2019, for analogues in spoken words). However, they also suggest that letters at word onset and offset may be more precisely encoded.

On the whole, this seems to suggest that competition in written-word recognition reflects the position of overlap but not in the same way as for spoken words. There are also hints suggesting some positional effects in recognising written words, with evidence of emphasis on word beginnings and endings during reading. However, results are mixed, and other potentially relevant competitor types have not been considered. For example, competitors that overlap at onset *and* offset, but not medially (e.g., CAP and CUP), might be particularly active in the written modality, where letters are read in parallel, but the edges get more attention. This may not be the case for spoken words, where phonemes arrive over time.

One possible reason for the conflicting results is that the tasks used in these studies only provide indirect insight into competition processes. For example, the masked inhibitory priming work (e.g., Carreiras & Perea, 2002; Perea & Lupker, 2003) relies on the fact that words interfere with each other. That is, the prime is expected to activate one word (e.g., CAT) that then slows recognition of the target (CAP). So, activation for CAT (the competitor) is only indirectly measured via changes in the recognition of the target. However, it is possible that words can be co-activated (all that is necessary for competition) without inhibition—in fact many models of word recognition do not include inhibition (e.g., Marslen-Wilson, 1987). In contrast, in spoken-word recognition, tasks like cross-modal priming or the VWP can directly assess how active *cap* is while hearing *cat*, without assuming that they inhibit each other.

What is needed is a task that can directly assess the consideration of competitors as they arise during word recognition so that we can more directly observe what types of words are activated and elicit competition during word recognition. Such a more direct measure could better trace out which competitors are active—and hence the degree to which position matters—over development as children are learning to read. The VWP offers a promising avenue, as it allows concurrent measurement of recognition of a target word and competitors during spoken-word recognition without assuming inhibition or other second-order effects. However, until recently, it there has been no way to use it to study written-word recognition.

Using the VWP to index spoken- and written-word recognition

Hendrickson et al. (2021) recently developed an analogue of the VWP for written-word recognition. In this task, the participant sees four pictures corresponding to potential competitors of the target word. For example, for the target word, *cat*, the images might represent *cat* (the target), and an onset competitor like *cap*, along with two unrelated items (*mug* and *bed*). Across trials different types of competitors can be studied (e.g., *hat*, a rhyme). The participant then sees a printed word at screen centre. After a short time (~100

msec), it is covered by a mask (###). This is designed to force participants to rapidly recode the word to working memory and to release the eyes to fixate the intended responses. Much like the traditional VWP for spoken words, they then click on the intended picture, and their eye movements to each competitor are used as an index of how strongly that word is considered during lexical access.

Hendrickson et al. (2021) used this method to show strikingly similar patterns of competition across spoken and written word recognition. In both modes, participants showed heightened consideration of cohort competitors. Anadromes, in which the order of letters or phonemes is swapped (e.g., *pat* and *tap*; Toscano et al., 2013), also exhibited competition in both modalities but somewhat less than cohorts. Rhyme competitors showed little consideration in either modality. These findings revealed remarkable continuity in lexical competition between the two modalities. In each, onsets appear privileged, offsets less so, and temporal order shows some flexibility, as indexed by the anadromes. The spoken modality showed a slightly larger cohort effect, suggesting that the temporal nature of the spoken input may play some role, but otherwise processing seemed to proceed quite similarly.

These data suggest that when using comparable measures, the types of words that are coactivated and compete for recognition are fairly consistent between modalities in adults. A remaining question is why these modalities are so similar. Perhaps the similarities arise because word recognition relies on modality-independent representations. Alternatively, modality differences might be apparent for less skilled readers, but then become more aligned with development, as mappings between modalities become better learned. Developmental changes in how people read words could alter the nature of competition throughout learning.

Compared to older readers, early readers have smaller “perceptual spans” (the number of letters they can process to the right of a fixation; Rayner, 1986). This could lead to more widespread competitor activation—a more skilled reader can activate a broader range of competitors from a single fixation. Or it could reduce competition as more letters offer more information to disambiguate targets from competitors. The number of fixations within a word also decreases throughout development (Huestegge et al., 2009; Vorstius et al., 2014). Consequently, poorer readers (or readers earlier in their development) who make multiple fixations might show different patterns of lexical activation; their more piecemeal intake of information across multiple fixations may alter which competitors become active and when. Specifically, novice readers might show competition letter-by-letter, as they are processed. Reading strategies also change developmentally. More skilled readers recognise printed words with minimal effort, without letter-by-letter decoding, and extremely quickly (Roembke et al., 2019, 2021). Developing readers might thus move from a more left-to-right activation of words to a more parallel form of recognition, where all letters are accessed in parallel. Thus, despite the commonalities found between the modalities by Hendrickson and colleagues (2021) in adults, we might see more divergences in lexical recognition between modalities developmentally.

The present study

The present study uses the VWP to assess the development of word recognition in spoken and written modalities during childhood. The study builds on the work of Hendrickson et al. (2021) to more directly assess competition dynamics in developing children (including early readers) and determine whether patterns of competition are qualitatively similar throughout development. Spoken-word recognition has only been studied using larger age differences, so this offers the potential to refine our understanding of the slow development observed by Rigler et al. (2015). Word recognition for written stimuli has not previously been measured in this way for children at all. Thus, there may be methodological issues that need to be addressed. However, this new approach to reading offers unique insight into the development of competition in the written domain, by allowing us to measure what types of words compete for recognition in each modality across development.

This study addresses several novel aspects of the development of word recognition. First, it establishes whether lexical competition in the written modality is apparent for children in this paradigm. The VWP has not been applied with written stimuli to this age group nor has the consideration of competitors during written-word recognition been directly measured in developing readers. We thus can determine whether developmental changes in written-word recognition can be indexed with this paradigm.

Second, this study assesses several types of lexical competitors, including one that has not been assessed in other studies: words that overlap at onset and offset but differ in the vowel (hereafter *frame competitors*; e.g., *cub* and *cab*). In written-word recognition for adults and spoken-word recognition for both children and adults, observed lexical competition has proven strongest for cohorts, whereas rhyme competition is weaker and inconsistent. However, both cohorts and rhymes contain the same vowel as the target (in single-syllable words). It remains to be seen how critical the vowel (or central phoneme/letter) is for recognition in both modalities. Vowel recognition develops slowly (Havy et al., 2011), so young children may be less reliant on vowels during word recognition. This suggests that words that share everything *except* the vowel might be particularly strong competitors. In addition, given the evidence that readers are particularly sensitive to letters at the beginning and ends of words (Chambers, 1979; Perea & Lupker, 2003), frame competitors may be particularly strong competitors in the written domain; when reading the word *cub*, the reader might rely on the *c* and *b* more than the *u*, and thus more strongly activate words that share these, like *cab*. This study measures consideration of this new competitor type, as well as cohorts and rhymes, giving a more comprehensive assessment of what's activated during spoken- and written-word recognition in children.

Third, this study traces changes in lexical competition across a tight age band, with a focus on ages particularly relevant to reading development. Previous research on the development of spoken-word recognition showed differences between children across a wide age gap (9 year olds vs. 16 year olds; Rigler et al., 2015). This study focuses on a finer granularity of age (7–9 vs. 10–12 vs. 13–15) to better trace how competition changes through childhood. This age range also covers very early readers to fairly skilled readers, allowing us to track whether competition reflects a child's level of reading attainment (as well as whether such links occur in the spoken modality).

Fourth, this study allows more direct comparison of competition in spoken- and written-word recognition within individual participants. While past studies of children's lexical processing have shown patterns of activation and competition within modalities and age groups, the present study includes both modalities, in a directly comparable task and with identical item sets (similar to Hendrickson et al., 2021, with adults). We assess how competition plays out within modalities across age but also whether individual differences in word recognition are consistent across modalities or if they vary independently.

Finally, this study includes several measures of language, reading, and general cognitive ability. These measures allow exploratory analyses to begin to investigate individual differences are related to changes in lexical processing in the two modalities.

Method

Participants

Participants were 92 children between 7 and 15 years old ($M = 11.28$, $SD = 2.64$). Children were divided into young, middle, and older groups of approximately equal size. The young group ($n = 33$, 13 females) included 7–9 year olds ($M = 8.44$, $SD = 0.92$); the middle group ($n = 33$, 15 females) 10–12 year olds ($M = 11.48$, $SD = 0.89$); and the older group ($n = 26$, 19 females) 13–15 year olds ($M = 14.62$, $SD = 0.97$). This study was begun prior to the widespread embrace of power analyses in response to the replication crises. Thus, sample sizes were determined on the basis of prior studies, and we sought per-age-group samples that were slightly larger than two prior studies of spoken word recognition development (McMurray et al., 2018; Rigler et al., 2015).

All participants were native monolingual speakers of English with no history of speech, language, or hearing deficits. All children completed a hearing screening before testing to ensure that their hearing was within typical bounds. Participants were recruited via mass email and word of mouth. Before data collection, participants' parents provided informed consent, and participants agreed to participate. Participants received \$30 for completing the study.

Five participants were excluded from the analysis due to technical error (three in the middle age group, two in the young group). In addition, participants whose accuracy in a modality was too low (below 85% correct) were excluded from the analyses of that modality and correlational analyses comparing modalities. Interpretation of fixations in the VWP assumes accurate word recognition to determine what was considered in the course of this recognition. Participants with large numbers of errors have increased noise in their data, as many of their correct responses may have arisen because of guessing. On this basis, 12 children were excluded from the written condition (11 young and 1 middle); no children were excluded from the spoken condition. The children excluded from the written condition tended to be substantially younger than those not excluded (in the younger group: $M_{excluded} = 7.71$, $M_{included} = 8.85$). It is likely that many of these children had not yet achieved sufficient reading ability to perform this task.

Design

Items consisted of 24 pairs of picturable monosyllabic words (see Appendix A for complete item lists). Words were selected that were likely to be known by young children. Most words had a log frequency of at least 1.6 in the Child Mental Lexicon corpus (Storkel & Hoover, 2010), corresponding to around 40 occurrences or more in the corpus. Four words (*toad*, *bait*, *soil*, and *beak*) had lower frequencies or were not present in the database. These words were retained to allow appropriate counterbalancing. To ensure including these words did not confuse participants, we assessed accuracy for these items. Participants were slightly *more* accurate on trials with these words as targets (low frequency words: $M = 99.3\%$ correct in spoken, 97.2% in written; high frequency words: $M = 97.9\%$ spoken, 95.9% written), suggesting that they were sufficiently familiar to children for use in the study.

Half of the item pairs were three-letter words with a CVC format in both written and spoken form; the other half were four-letter words, with a written CVVC format, and a spoken format of CVC (in some cases, the V was a diphthong). In each pair, the words differed by a single phoneme in the onset (Rhymes: e.g., *dog/fog*), vowel (Frames: e.g., *hoop/heap*), or offset (Cohorts: e.g., *leaf/leak*) position. There were eight pairs of each overlap type, evenly split between three- and four-letter pairs.

An item set consisted of two pairs, with minimal phonological or semantic overlap between the pairs. Consequently, on every trial, the target word (e.g., *dog*) always had exactly one other displayed item with orthographic/phonological overlap (*fog*), and two items with minimal overlap (*pan* and *pen*). As a result, each trial had a target, a competitor, and two unrelated items. These unrelated items later served as experimental targets on other trials. This eliminated the need for filler trials—the unrelated items on one trial are meaningful targets with competitors on other trials—and it allows straightforward comparison between conditions. In analyses, unrelated fixations are the mean of fixations to the two unrelated objects. The two pairs comprising a set always had different competitor types; four sets included a Cohort and Rhyme pairs; four included Cohort and Frame pairs; and four included Rhyme and Frame pairs. Within a set, all words had the same number of letters. Item sets were consistent between modalities for each child. Two different assortments of pairings were made to counteract effects of idiosyncratic relationships between pairs; these were randomly assigned by participant.

Each item in a set served as a target on six trials: three in the auditory modality and three in the visual modality. This led to $8 \text{ pairs/type} \times 2 \text{ words/pair} \times 3 \text{ competitor types} \times 3 \text{ reps} = 144$ trials in each modality, and 288 trials total. The trials in each modality were randomly ordered and then divided into four blocks of 36 trials. Blocks were interleaved between modalities; not, which modality was delivered first was counterbalanced between participants. The item locations in the display were randomly chosen on each trial.

Procedure

Testing was completed in two sessions conducted on different days, approximately a week apart. Each session took approximately 1 hr. In the first session, participants provided informed consent and we conducted the hearing screening. They then completed the primary

eye-tracking study (examining both spoken- and written-word recognition). At the beginning of the word-recognition VWP study, participants were given instructions for both spoken and written trials and then completed the full complement of blocks of these trials. After eye-tracking, participants completed the oral language standardised assessments. During their second session, participants completed the visual-only control VWP study, and then completed the remaining standardised assessments, as time permitted.

Spoken trials.—For spoken trials, participants saw a display with four images in the four corners of the screen and a red dot in the centre. After 500 ms, the dot turned blue, and participants clicked on this blue dot to begin the trial. After they clicked the dot, the dot disappeared, and the auditory stimulus was played. Participants then clicked on one of the response options. After a response was registered, a blank screen was shown for 115 ms before the next trial began.

Written trials.—Written trials were identical to spoken trials, except that 50 ms after the dot was clicked, the target word was printed in the centre of the screen in lower-case 40-pt Times New Roman text. The text remained on the screen for 100 ms, and then a visual masker consisting of seven # symbols replaced it in the centre of the screen. The masker remained on the screen for another 100 ms, and then was also removed. The inclusion of the backward mask ensures that participants cannot continue to fixate the printed word and thus provides better time-locking of fixations to lexical processing. Previous research shows that children are able to complete reading tasks with similarly timed backward masking (Roembke et al., 2019, 2021).

Stimuli

Stimuli are available at <https://osf.io/eav72>. Auditory stimuli were recordings of the target words, isolated from context. Words were initially recorded in a neutral sentence context (“he said [target]”) by a female native speaker of American English. Each sentence was repeated five times. The recordings of the sentences were noise reduced in Audacity to eliminate background noise. For each word, an experimenter selected a repetition that was clear, with relatively neutral affect, and with no apparent artefacts, based on listening and visually inspecting the waveform. The selected repetitions were excised from their sentence context and amplitude normalised to have equivalent peak amplitudes in Praat. Finally, 50 ms of silence was added to the beginning and end of each file.

Visual stimuli were colour drawings of the words. These drawings were developed based on lab protocol for developing visual stimuli for VWP studies (McMurray et al., 2010). For each word, several images were identified and downloaded from a commercial clipart database. A group of four to six people examined the images and selected the one that best represented the intended sense of the target word. This group also suggested any edits to the image, including changing brightness, removing unnecessary detail, resizing, and ensuring that the image stylistically matched other images used in the study. After these edits were made, a senior lab member who was not involved in developing the stimulus approved the image.

Visual analogue of the VWP

The standard VWP requires general skill in decision-making, speed of processing, attention, and oculomotor control. Farris-Trimble and McMurray (2013) found that some aspects of looking behaviour in the VWP show consistent individual differences even absent linguistic input—there are general parameters governing things like speed of target fixation. As these parameters might vary developmentally, it is important to determine whether changes seen in the present study arise because of changes in linguistic processing, or if instead, they reflect more general developmental patterns in how eye movements are planned and launched.

We thus adapted their visual-only VWP to be more accessible for children. On each trial, three coloured objects were shown in a triangle configuration around the centre of the screen (one just above centre and two below). This differed slightly from the Farris-Trimble and McMurray (2013) version of the task, which included four objects in the display; the smaller number used here was expected to simplify task demands for the young participants. All objects in the display were different shapes, but two shared a colour. A red dot in the centre of the screen turned blue after 500 ms; participants clicked this dot to begin the trial. A coloured object matching one of those displayed appeared in the centre, remained on the screen for 100 ms, and then disappeared. Eye movements were monitored as participants selected the colour-matching object from the three displayed options. On critical trials, the participants had to select a target when another displayed object shared the same colour. This provides a measure of target identification in the face of competition as well as of the degree of competitor consideration and suppression.

Materials consisted of 16 shapes that could appear in 12 different colours. These were allocated to create 48 sets of three items, with two shapes in one colour and the third shape in a different colour. Participants completed trials with each item as the target for each item set, leading to 144 trials (96 with a colour-matching competitor of the target). Trials were randomly ordered for each participant.

Assessments

Children's language, reading and cognitive abilities were assessed using a series of standardised assessments conducted by a trained experimenter. Vocabulary was measured using the Peabody Picture Vocabulary Test (PPVT), Version 4. Oral language abilities were measured using the CELF Recalling Sentences subtest. Reading ability was measured using the Woodcock-Johnson Reading Mastery Test (WRMT) Word Identification and Word Attack subtests. Nonverbal cognitive measures consisted of the Wechsler Abbreviated Scale of Intelligence II Block Design and Matrix Reasoning subtests. For some children, reading fluency was measured with the WRMT Oral Reading Fluency subtest; however, this test was conducted last, so a large proportion of participants did not complete it. These data were thus not analysed.

Eye movement recording and processing

Eye movements were recorded with an SR Research Eyelink 1000 table-mounted eye-tracker, at a 500 Hz sampling rate. Before both the primary experiment and the visual-only control study, the standard nine-point calibration procedure was conducted. Drift correction

was performed every 48 trials of the primary experiment and every 32 trials during the control study. Fixations were automatically parsed into saccades, fixations and blinks using the tracker's default cognitive parameter set. Saccades and fixations were then combined into "looks," which started at the beginning of the saccade and lasted until the end of the subsequent fixation. Looks initiated before the onset of the target stimulus were ignored, as these could not be driven by lexical activation processes, and eye movement data were curtailed when a response was given. If the trial ended prior to the 2000 ms analysis window, the final look was extended to the end of the window. Looks to objects were defined by fixations that fell within 100 pixels of the outer bounds of the displayed objects. This extension of object boundaries did not result in any overlap between regions. Only trials where the correct target was clicked were included in the analyses.

Results

Data and analysis scripts are available at <https://osf.io/eav72>. We first examine the accuracy of the overt response to establish that children could perform the task well enough to examine the fixations. Next, we examine the degree of competitor fixations for the various competitor types in the two modalities separately. Because these analyses are concerned with what effects emerge within a modality, we include all data for all children who surpassed the accuracy threshold for the modality of interest. These analyses ask what types of competitors are considered during word recognition in each modality at each age.

Then, we examine the degree to which lexical activation is correlated between the spoken and written modalities. We also examine correlations between competitor effects for individual items between the modalities. Finally, we present exploratory analyses investigating the degree to which language, reading, and general cognitive abilities relate to lexical competition.

Accuracy

Eye movement data in this paradigm require accurate responses; we want to measure what listeners consider along the way to correctly recognising a word. As such, we only consider participants with high accuracy rates in analyses. Twelve participants fell below 85% accuracy on the written condition and so were not included in analyses of the written condition or those comparing the written and spoken conditions. All participants exceeded the 85% threshold in the spoken condition, so all were retained for analyses of the spoken condition.

For the remaining participants, accuracy was overall quite high ($M_{spoken} = 98.4\%$; M_{spoken} only those with written above 85% = 99.2%; $M_{written} = 96.0\%$). The data for children surpassing the accuracy threshold in the written condition were analysed in a 3 (age group) \times 2 (modality) analysis of variance (ANOVA) (Figure 1). This revealed a significant effect of modality, $F(1,72) = 100.58$, $p < .0005$, with significantly higher performance on spoken trials. There was also a significant effect of age group, $F(2,72) = 10.54$, $p < .0005$, as accuracy was greater for the older age groups. The interaction of modality by age was also significant, $F(2,72) = 8.86$, $p < .0005$, as the effect of age was more apparent for written trials than for spoken trials.

Analysis of fixations

Figure 2 shows the time course of fixations across all ages as a function of time for spoken (Panel A) and written (Panel B) words separately. Panel A suggests that spoken-word recognition showed the expected profile. By about 350 ms into the trial (100 ms after word onset if you account for the 200 ms oculomotor delay and the fact that auditory stimuli began 50 ms into the trial), children were fixating the target and cohort above other competitors. The cohort began to be suppressed at around 600 ms, and the target reached asymptote around 1200 ms. Fixations to rhymes and frames stayed near the unrelated. In written-word recognition (Panel B), we also see convergence on the target word by around 1200 ms. Competitor fixations (especially for cohorts) were lower and persisted throughout the trial.

The distinct morphology of the target and competitor fixations required distinct statistical approaches. Thus, we start with the target, before turning to the competitor.

Target fixations

Target fixations show how quickly and effectively participants begin to consider the correct word, and how stably they settle there by the end. Previous research showed faster target fixations between 9 and 16 years of age in spoken-word recognition (Rigler et al., 2015), but both groups ultimately reached the same asymptote (a pattern we previously termed *activation rate*: (McMurray et al., in press)). Target fixations from the present study are displayed in Figure 3 and closely match this pattern. There were clear age-related changes in the speed of fixating the targets, but all groups reached similar maxima.

For analysis, we fit a four-parameter logistic function to the time course of target looks for each participant in each modality (Farris-Trimble & McMurray, 2013). This function provides meaningful parameters defining the shape of the target curve. The minimum specifies the lower asymptote; as this value is necessarily near zero because of the task, it is of little theoretical interest, and so we do not consider differences here. The maximum specifies the upper asymptote, showing the peak amount of target fixations the participant reaches. This parameter varies systematically as a function of language abilities—people with developmental language disorder (DLD) show lower maxima than people with typical language (McMurray et al., 2010; McMurray, Klein-Packard, & Tomblin, 2019). The slope specifies the speed of rise of target fixations, and the crossover specifies the midpoint of this rise. Both of these parameters index aspects of the speed of target recognition. These speed parameters vary as a function of age, with younger children showing slower target fixation curves than older children and adults (Rigler et al., 2015). We combined them into a single “timing” parameter (McMurray, Ellis, & Apfelbaum, 2019). To do so, we first log normalised slope. Next, we computed z-scores for the log slope and the crossover (using all participants within a modality). Z-scores for crossover were multiplied by -1 , as slower target identification leads to higher crossover values but lower slopes. The average of these z-scored values was then used as the timing index, with higher values signifying faster responding.

Prior to fitting, target curves were averaged across all competitor conditions (within a child and modality). Fits were performed via constrained gradient descent, minimising the least squared error between the estimated and the actual data (McMurray, 2020; <https://osf.io/4atgv/>). Fits were visually inspected, and refitted if needed. Final fits were quite good (Spoken: mean $r = .997$, $SD = .0017$, minimum = .992; Young $r = .997$; Middle $r = .997$; Old $r = .998$; Written: mean $r = .997$, $SD = .0016$, minimum = .991; Young $r = .997$; Middle $r = .997$; Old $r = .997$). No data was dropped for poor fit.

We analysed the parameter estimates in a series of one-way ANOVAs, considering each of the two theoretically motivated indices (target timing and maximum/resolution) separately in each modality. We did not attempt to combine spoken and written data in a single ANOVA as we were concerned that the different visual displays in each would drive some overall differences in looking (which were not of interest; c.f., Hendrickson et al., 2021). For each ANOVA, age group was the independent variable (IV) and the curvefit parameter was the dependent variable (DV).

These analyses revealed no significant differences in target maxima for either modality, spoken: $F(2,84) = .92$, $p = .40$, $\eta_{\text{ges}}^2 = .021$; written: $F(2,72) = .003$, $p = .997$, $\eta_{\text{ges}}^2 < .00001$. However, there were significant differences of timing in both modalities, spoken: $F(2,84) = 23.04$, $p < .00001$, $\eta_{\text{ges}}^2 = .35$; written: $F(2,72) = 14.12$, $p < .00001$, $\eta_{\text{ges}}^2 = .28$, as children's speed to fixate targets increased with age. We conducted follow-up simple-effects comparisons of adjacent age groups to identify the ages when significant changes occurred. These analyses were Bonferroni corrected for multiple comparisons, producing a significance threshold of $p < .025$. Both modalities revealed a significant difference between young and middle age groups, Spoken: $t(59) = 5.68$, $p < .00001$; Written: $t(47) = 4.41$, $p = .000060$, but not between middle and old age groups, Spoken: $t(54) = -1.49$, $p = .14$; Written: $t(53) = -1.08$, $p = .28$. It appears that much of the speed up in target looks is occurring between the young age range (7–9) and the middle age range (10–12), with more stability thereafter.

This replicates Rigler et al. (2015) showing faster target recognition with age for spoken words, and it extends it by suggesting that these developmental changes may peak around 11–12. It also extends this finding to the written domain, showing that children also become faster to recognise written words across development, despite little change in asymptotic looking.

Competitor effects

Figure 4 shows the overall pattern of competitor fixations for each age group in each modality. It reveals several things that guided our analysis.

First, fixations to cohorts clearly exceeded those to other competitor types, especially in the spoken trials. In fact, fixations to frames and rhymes only exceeded unrelated fixations in a few cases and not by much. Thus, competitor effects were examined by measuring the likelihood of fixating a particular competitor type (e.g., cohort), relative to that of fixing either unrelated item. This difference score approach (competitor—unrelated) is motivated by prior work by Rigler and colleagues (2015) which found that younger children exhibited

increased fixations to phonologically related *and* unrelated objects, suggesting that younger participants are more prone to fixate anything in the display. Subtracting out baseline looks thus provides a more appropriate measure of how much a competitor is considered over and above the tendency to fixate anything in the display. This approach ignores potential main effects of the amount of looking to all non-target objects, as the primary questions of interest concern what type of competitors are considered and to what degree.

Second, Figure 4 also reveals potential differences in the time course of competitor consideration by age and modality. Younger participants (Figure 4(a) and (d)) show ongoing consideration of competitors late in the time course—for instance, they have elevated cohort looks up to around 1750 ms. Similarly, even older participants may have extended competitor fixations in the written modality (Figure 4(f)). For this reason, analyses of the competitors used a wide time window from 250 ms after stimulus onset (to account for the 200 ms oculomotor delay plus 50 msec of silence after trial onset) until 1750 ms after stimulus onset.

The competitor effect (competitor—unrelated) across time for each competitor type in each modality is shown in Figure 5(a) and (b). A few patterns are immediately apparent. First, the largest effect was for cohorts in spoken-word recognition. This finding is consistent with incremental activation of competitors in the spoken modality. Rhymes also showed a small amount of competition but frames did not. Meanwhile, competition in the written modality was quite low across the board. This pattern differs from that found by Hendrickson and colleagues (2021) who showed similar patterns of competition between the two modalities in adults.

Our analysis of competitor fixations used a two-step approach. First, we computed the degree to which competitor looks exceeded unrelated looks across the entire analysis window for different competitor types. This asked whether competitors are considered at all, but not when they are considered (for a given competitor type at a given age). We then assessed the extent and timing of competitor consideration using the approach used by Rigler and colleagues (2015), for only those conditions that have significant competitor consideration across the analysis window.

In the first step, we analysed the mean of the competition effect across the time window 250–1750 ms. We computed these mean proportions for each included participant in each modality and each competitor type (Figure 5(c) and (d), for spoken and written). These data were entered into separate 3 (age group) \times 3 (competitor type) ANOVAs for each modality. This revealed a significant effect of competitor type in each modality, spoken: $F(2,168) = 49.93, p < .00001, \eta_{\text{ges}}^2 = .31$; written: $F(2,144) = 6.59, p = .002, \eta_{\text{ges}}^2 = .059$. However, neither modality showed an effect of age, spoken: $F(2,84) = 2.50, p = .088, \eta_{\text{ges}}^2 = .015$; written: $F(2,72) = .005, p = .99, \eta_{\text{ges}}^2 = .000043$, nor an interaction of competitor type and age, spoken: $F(4,168) = .41, p = .80, \eta_{\text{ges}}^2 = .007$; written: $F(4,144) = .36, p = .84, \eta_{\text{ges}}^2 = .0070$.

We further examined the effect of competitor type by conducting simple-effects analyses contrasting the competitor types in each modality (ignoring age). These analyses took the

form of pairwise comparisons of each pair of competitors; Bonferroni corrections were used to account for multiple comparisons, setting a significance threshold of .017. For spoken words, this revealed significantly greater competitor effects for cohorts than for frames, $t(86) = 10.00$, $p < .00001$, and rhymes, $t(86) = 6.25$, $p < .00001$, as well as significantly greater competitor effects for rhymes than for frames, $t(86) = 3.56$, $p = .00062$. In the written modality, cohorts had larger competitor effects than frames, $t(74) = 3.34$, $p = .0010$, as did rhymes, $t(74) = 2.81$, $p = .0060$, cohorts and rhymes did not significantly differ, $t(74) = -.83$, $p = .41$.

We then examined the overall competitor effect for each competitor type across age groups to assess which conditions showed significant competitor effects in each modality. This was done using one-sample t tests comparing each condition to 0. Spoken words showed significant competitor effects for cohorts, $t(87) = 11.50$, $p < .00001$, and rhymes, $t(87) = 3.41$, $p = .00098$, but not for frames, which were marginally significantly *below* 0: $t(87) = -1.86$, $p = .067$. Written words showed a similar pattern: significant competitor effects for cohorts, $t(75) = 3.75$, $p = .00034$, and rhymes, $t(75) = 2.98$, $p = .0039$, but not for frames, $t(75) = -.94$, $p = .35$.

Time course of competitor effects.—These analyses suggest that only cohorts and rhymes are potential avenues of development of competitor effects, as frames do not elicit overall greater looks than unrelated. We thus investigated patterns of looks to these competitor types using the same approach as Rigler et al. (2015). They conducted two detailed assessments of the time course of processing to investigate how competitors are active across time. First, to assess the *peak* amount of consideration, they assessed the maximum of difference between the competitor and unrelated looking. Second, to assess the *duration* of time that a competitor was considered, they considered the amount of time that the difference between the competitor and unrelated exceeded a fixed threshold (0.03^2).

We thus conducted parallel analyses of the present data. Figure 6 shows the competitor effect by age for cohort and rhyme competitors in each modality. Our two key markers of competition (peak and extent) were computed using the following procedures.

First, we computed the average competitor effects for the different competitor types across time for each participant, in each modality. We then applied a 48-ms smoothing window to these data to avoid idiosyncratic effects of short-lived differences in looking behaviour altering outcomes. From these data, we identified the maximum difference between the competitor and unrelated for each competitor type for each participant in the two modalities as the *peak* competition. We next assessed the duration of competitor effects (or *extent*). We did this the same way as Rigler et al. (2015): We used the smoothed time courses and measured the number of 4-ms time windows between 250 and 1750 ms when this difference was above .03 for each cell. We then subjected these data to a set of ANOVAs with age and competitor type as IVs, and the maximum competitor effect or extent as the DV.

²Rigler and colleagues tested several thresholds and had similar results with each. We use just the 0.03 value for consistency with their reported data.

We started by examining peak competition. In the spoken modality, this revealed significant main effects of competitor type, $F(1,84) = 45.23, p < .00001, \eta_{\text{ges}}^2 = .23$, with greater peak competition for cohorts than rhymes. There was also a main effect of age, $F(2,84) = 3.63, p = .031, \eta_{\text{ges}}^2 = .036$, with decreasing competition across ages. There was no interaction, $F(2,84) = 1.15, p = .32, \eta_{\text{ges}}^2 = .015$, suggesting similar effects for cohorts and rhymes (as was observed by Rigler et al., 2015). Follow-up pairwise comparisons between age groups used a Bonferroni-adjusted significance of .017. This did not reveal significant differences between adjacent age groups, young versus middle: $t(120) = 1.34, p = .18$; middle vs. old: $t(110) = .83, p = .41$, although the distinction between young and old approached significance, $t(112) = 2.28, p = .025$.

In the written modality, no effects were significant, age: $F(2,72) = 2.24, p = .11, \eta_{\text{ges}}^2 = .024$; competitor type: $F(1,72) = .31, p = .58, \eta_{\text{ges}}^2 = .003$; age \times competitor type: $F(2,72) = 1.74, p = .18, \eta_{\text{ges}}^2 = .029$.

We next examined competitor duration or extent. In the spoken domain, this revealed significant main effects of competitor type, $F(1,84) = 35.89, p < .00001, \eta_{\text{ges}}^2 = .18$, with greater extent for cohorts than rhymes. There was also a main effect of age, $F(2,84) = 5.32, p = .0070, \eta_{\text{ges}}^2 = .060$, driven by decreasing extent across ages. There was no interaction, $F(2,84) = .00067, p = .99, \eta_{\text{ges}}^2 < .00001$. Again, we examined the effect of age with simple effects comparisons (with a Bonferroni-adjusted alpha of .017). The older age group showed a smaller competitor extent than the young age group, $t(112) = -2.71, p = .008$. However, the old and middle age groups did not differ, $t(110) = -.69, p = .49$, nor did the middle and young age groups, $t(120) = -2.19, p = .030$, suggesting a gradual decline throughout this period.

The change in extent (at least between the largest age differences) is apparent in Figure 6(b) and aligns with the findings of Rigler et al. (2015). The fact that this did not interact with the type of competitor suggests development in managing competition as a whole, not the improvement in dealing with a specific class of competitors.

In the written modality, no effects were significant, although the effect of age approached significance, age: $F(2,72) = 3.01, p = .056, \eta_{\text{ges}}^2 = .035$; competitor type: $F(1,72) = 1.47, p = .23, \eta_{\text{ges}}^2 = .011$; age \times competitor type: $F(2,72) = .074, p = .93, \eta_{\text{ges}}^2 = .001$.

Summary.—The analyses of the detailed time course of the competition revealed notable effects of age and competitor type in the spoken modality, and intriguing differences from other studies. First, in the spoken modality, cohort competition predominated. This replicates previous studies and is in line with theories of incremental processing. Surprisingly, frame competitors received minimal consideration, despite considerable phonological overlap with targets and the relatively poor vowel encoding of young children. Even rhyme competitors, which do not overlap with the target in the first phoneme, received stronger consideration than frame competitors. It seems that mismatches in the central vowel for these stimuli reduced or eliminated competition. We return to this finding in the General Discussion.

In the written modality, cohort effects were substantially less pronounced, with statistically comparable competitor effects for cohorts and rhymes. This stands in contrast to the results of Hendrickson and colleagues (2021), who found that for adults, cohorts were much stronger competitors than rhymes in the written modality. This may indicate that early readers engage in different processes for written-word recognition than later readers. Specifically, early readers seem to have less positional bias in their processing, as they show similar consideration of items that overlap at onset or at offset. However, frame competitors again received little consideration in this modality. This finding was particularly unexpected given the substantial previous evidence for the precedence of onset and offset letters during word reading. Despite these findings, it seems that early readers rely heavily on word-medial vowels, such that words that mismatch only there are not strongly considered.

Age effects were limited to a few specific areas. First, target fixations became faster across age in both modalities, and there was no effect of age on the ultimate asymptote. This accords with previous developmental work showing increases in the efficiency of word recognition throughout childhood and extends this to the written modality. Second, spoken-word competition showed developmental changes in the peak of competitor consideration and the extent of competitor activation over time. Importantly this was seen for both cohorts and rhymes (as in Rigler et al., 2015), suggesting that the relevant development is in the overall ability to manage competition, not in changes to how particular classes of words engage in competition.

Interestingly, competition in the written modality did not exhibit substantial age-related changes, despite the considerable growth in reading skills that are characteristic of these age groups.

Contributions of general visuo-cognitive development

One potential issue with our analyses above is that changes with age might arise for reasons other than similarities in word recognition. For example, development in the dynamics of visual search, decision-making, or speed of processing could account for some of the effects of age. Thus, we start with a short discussion of the visual-only task to understand the relevant development in non-linguistic domains before asking if the effects of age can be seen over and above this development. Critically, it is important to point out that some of the domain-general processes tapped by this visuo-cognitive task are likely relevant to language or reading: decision-making, inhibition, speed of processing, and so forth. We return to this point later.

Visual-only VWP.—First, we analysed the data from the visual-only VWP to determine how performance in this task varies with age. Data from this task by age group are shown in Figure 7. One participant (in the young age group) did not complete the visual-only experiment; the other 74 participants are included here. We followed a highly similar analysis strategy to that used in the word-recognition versions: we fit four-parameter logistic curves to the targets, and assessed the overall extent of consideration of competitor objects. Target curves were fit using the same process described for the word-recognition data above. These fits were good (mean $r = .997$), and no fits were excluded from the analysis. The slope

and crossover parameters were combined into a single timing parameter using the same approach taken with the word recognition data. These data were analysed in ANOVAs with age group as the IV and the parameter values as the DV.

The effect of age group was not significant for the maximum, $F(2,71) = 1.96, p = .15, \eta_{\text{ges}}^2 = .052$, but it was significant for timing, $F(2,71) = 20.24, p < .00001, \eta_{\text{ges}}^2 = .36$. Older children were faster to fixate the target in this task, much as they were in both modalities for the word recognition tasks. This suggests that at least part of the developmental effect seen for those tasks may have been driven by changes in non-linguistic visuo-cognitive processes.

Competitors were measured as the difference between looks to the colour-matching competitor object and looks to the unrelated object (Figure 7(d) and (e)). There were clear differences in competitor looks by age: Younger children had more looks to competitors, and these looks persisted for a longer period of time. We followed the same approach to analysing these data as in the word recognition tasks: we computed the mean value of the competitor effect (using a time window of 200–1000 ms after the stimulus, as competitor looks dropped to near 0 thereafter for all groups). We then subjected these data to an ANOVA with age group as the IV. This revealed a significant effect of age, $F(2,71) = 10.87, p = .000076, \eta_{\text{ges}}^2 = .23$, as the size of the competitor effect decreased with age.

Does word recognition change with age?—The developmental patterns in the non-linguistic VWP task are similar to those seen for both modalities in the linguistic tasks. This raises a question of whether the developmental changes in target timing reflect changes in word recognition or more general changes in eye movement dynamics.³ To assess this, we used hierarchical regression to ask whether age predicts unique variance in target timing after accounting for variance predicted by the non-linguistic VWP task. This variance could arise from multiple sources. First, the shared variance between the linguistic and non-linguistic VWP tasks might reflect basic eye movement dynamics, like the speed of fixations. Second, it might reflect overlap in cognitive processes between the tasks—processes like target identification, visual search, and competitor suppression that are needed for both linguistic and non-linguistic variants of the VWP might depend on similar substrates.

We attempted to partition out these possible sources of shared variance. First, we estimated the dynamics of fixations using fixations made during the non-linguistic VWP prior to target onset. Recall that in this task, there is a self-paced preview period when response options are displayed but before the participant clicks the centre dot to initialise the trial. During this pre-trial period, participants could freely look around the display. The duration of these fixations gives a coarse indicator of how individual participants move their eyes before they have the information needed to complete the task. We thus computed the mean duration of all fixations made by each participant prior to clicking the centre dot in the non-linguistic VWP task and used this measure of fixation duration as a measure of eye movement dynamics in the first step of the hierarchical models. In a subsequent step, we added the relevant derived measures of target or competitor looks from the non-linguistic

³. We did not conduct similar analyses for target max, as max did not vary with age.

VWP. For these models, we input continuous age rather than using the age group. We conducted separate models for each modality.

In the spoken modality, the first model, predicting target timing from the mean fixation duration during preview of the non-linguistic task was not significant, $R^2 = .016$, $F(1,72) = 1.16$, $p = .29$. Adding the timing parameter from the non-linguistic task revealed a significant increase in model fit, $R^2 = .48$, $R^2 = .47$, $F(1,71) = 64.04$, $B = .00074$, $p < .00001$. Adding age at the next step did not improve model fit, $R^2 = .48$, $R^2 = .0014$; $F(1,70) = .19$, $B = .016$, $p = .66$.

The written modality revealed a different pattern: The first model predicting target timing from the fixation durations during preview of the non-linguistic task *was* significant, $R^2 = .057$, $F(1,72) = 4.37$, $B = -.0037$, $p = .040$,—participants with longer fixation durations had slower target timing. Adding the timing parameter from the non-linguistic task improved model fit, $R^2 = .58$, $R^2 = .52$, $F(1,71) = 48.94$, $B = .75$, $p < .00001$. Further adding age did not improve model fit, $R^2 = .58$, $R^2 = .0002$; $F(1,70) = .044$, $B = .0080$, $p = .83$. It appears that much of the increased speed for target timing in word recognition reflects more general changes visuo-cognitive performance, rather than age-specific changes to word recognition. Interestingly, basic eye movement dynamics account for variance in the written modality but not in the spoken modality. But even here, eye-movement control accounted for a small proportion of the variance, whereas the more abstract aspects of visuo-cognitive processing accounted for a relatively large amount. We return to this pattern in the discussion.

Correlations across modalities

We next asked whether profiles of word recognition and competition are consistent between the two modalities for individual children. One possibility is that these domains are largely independent: Spoken-word recognition depends largely on auditory processes, which develop independently of reading, which reflects exposure to text and the reading teaching environment. Alternatively, if both forms of input access some form of shared lexical system, it is possible that they would relate. We thus correlated performance in the modalities for each child. For these analyses, we thus only included children who surpassed the accuracy threshold in both modalities ($n = 75$).

Again, some of these correlations may be due to general visuo-cognitive processes relevant to both tasks. This might be particularly apparent across age; the developmental sampling of the current study could thus lead to spurious correlations between modalities that are driven by age-related improvements in eye movement control. To account for this, we used the visual-only VWP study as a control measure of fixation dynamics and sought correlations between modalities over-and-above what is predicted by non-linguistic measures of eye movements. This led to an approach that used hierarchical regression to predict the fixation patterns in one modality from those of another modality, after accounting for visuo-cognitive processing.

Target fixations.—These analyses focused on the maximum and timing indices from the preceding analyses. In each analysis, we first constructed a regression of one of the two

indices for the written modality from the fixation duration and the corresponding index in the visual-only task, as well as age. This accounts for age-related changes in eye movement dynamics and other non-linguistic factors that might impact fixation patterns. Next, we added the corresponding parameter from the spoken modality to the model to determine whether significant additional variance was accounted for by the addition of the spoken predictor.

For the maximum parameter, the first model (age + visuo-cognitive performance) accounted for significant variance in the written maximum, $R^2 = .12$, $F(2,71) = 4.74$, $p = .012$. The visual-only task was a significant predictor ($B = .25$, $p = .0022$), but fixation duration was not ($B = .000050$, $p = .72$), nor was age ($B = .0029$, $p = .43$). This is consistent with the earlier finding here and in prior work that age is not significantly related to ultimate asymptote. However, the addition of the spoken maxima accounted for substantial additional variance, $R^2 = .39$, $R^2 = .27$; $F(1,69) = 30.27$, $B = .69$, $p < .00001$. Thus, for maximum, there is a strong correspondence between the two modalities, independent of effects of age and non-linguistic eye movement control.

Timing showed similar patterns. The initial model including the visual-only task and age as predictors of written-word timing accounted for significant variance in the timing parameter for the written trials, as reported above, $R^2 = .58$, $F(2,70) = 32.20$, $p < .00001$. In this model, the timing parameter from the non-linguistic task was a significant predictor ($B = .73$, $p < .00001$), but fixation duration was not ($B = -.00072$, $p = .56$) nor was age ($B = .0080$, $p = .83$). Note that this model accounted for a much larger portion of the variance than the target maximum model; it appears that timing is more consistent across even non-linguistic tasks than maximum. However, the addition of the spoken parameters accounted for additional variance ($R^2 = .77$, $R^2 = .19$) and significantly improved model fit, $F(1,69) = 55.21$, $B = .70$, $p < .00001$. Although some of the developmental effects on target timing appear to reflect the development of general (non-linguistic) cognitive processes, there is substantial correspondence between the speed of target looks between modalities, even when accounting for effects of age and non-linguistic cognitive processes and eye movement dynamics. This suggests that individual differences in target timing are consistent across modalities in a way that cannot be fully explained by purely visuo-cognitive processes.

Correlations for competitor effects.—We assessed correlations in competitor effects in a similar manner, using the magnitude of the competitor effect as the primary variable of interest. Our primary research question is whether the degree of competitor effect is correlated between modalities. However, our design included the three different types of competitors, raising the possibility that each could be specifically correlated across modalities. To attempt to reduce the number of necessary comparisons, we first asked whether competitor effects were correlated across competitor types *within* modalities; this could allow us to collapse across competitor type to more directly assess correlations between modalities. However, these correlations were quite small (and often negative; see Table 1).

We thus assessed correlations for each competitor type separately. In each case, we first predicted the orthographic competitor effect from the fixation duration and competitor effect

in the visual-only task as well as age; then we added the auditory competitor effect to the model. The results are presented in Table 2. The only significant predictor was the visual-only VWP task predicting written rhyme effects ($B = -.26, p = .028$); however, this effect was in the opposite direction such that participants with larger competitor effects in the visual-only VWP tended to have *smaller* written rhyme effects. It is unclear why these would be anticorrelated. All other predictors, including those of age and all relationships between modalities, were not significantly correlated with the orthographic competitor effects (although there was a trend towards a negative correlation of age and rhyme effects). One possible explanation is that competitor effects were small and unreliable: Correlating two unreliable effects may just be too difficult. However, alternatively, if we take this null effect seriously, this suggests that competitor consideration (for speech, written words, and even visual decisions) is not amodal, but is specific to a modality, despite the commonalities in target activation.

Correlations within words.—A final way that the modalities might show correlations is within items. The design included the same words with the same competitors in the two modalities. We can thus ask whether items that engender high competition in one modality do so in the other modality as well. To investigate this, we computed the competitor effect for each item in each modality over the same 250–1750 msec time window used in the participants' analyses. Correlations might emerge because cohorts typically receive more consideration than other competitors; to avoid these driving effects, we assessed the correlation between modalities separately for each competitor type (Figure 8). This revealed strong correlations between the modalities for each competitor type (Table 3). The strength of these correlations is rather surprising, particularly given the lack of correlation between modalities by participants. Although individuals vary in the degree of competition they exhibit for spoken and written trials, the mean competition an item engenders is quite consistent. This could arise because of some shared underlying representational space in which competition takes place. However, this effect should be interpreted with some caution; the visual images used for the target and competitor were identical between the modalities, and facets of these images are likely to affect eye movement dynamics. It is likely that at least a portion of the correlation between modalities for items resulted from image characteristics rather than lexical processing.

Individual differences

Finally, we asked if individual differences in word recognition were related to oral language, reading, and non-verbal intelligence. These were exploratory analyses to determine potential avenues of future research. We thus investigated the correlation between measures of different abilities and word recognition independently. We formed composite scores for measures that tapped related constructs. These included a composite score for oral language, (the PPVT and CELF-Recalling sentences); a composite for reading (the WRMT Word ID and Word Attack subtests); and a composite for non-verbal intelligence (WASI Block Design and Matrix Reasoning subtests). Separate analyses were conducted for the data from the two modalities, for each of the word-recognition variables computed previously. In each case, we included all participants that exceeded the accuracy threshold for that modality and that completed the relevant assessments. For these analyses, we first conducted a linear

regression predicting the word recognition measure from age. Then, we added the standard scores from the individual difference measure to assess whether that measure accounted for significant variance after age was accounted for. As our sample size was modest, we were unable to include multiple predictors simultaneously in the second step, so separate models examined each predictor.

Overall, our sample proved to be rather high performing on all measures and had limited variance ($M_{language} = 112.1$, $SD = 11.5$; $M_{reading} = 104.77$, $SD = 13.0$; $M_{non-verbal} = 53.5$, $SD = 7.4$).⁴ This bias likely reflects the convenience sampling approach, which included many children of university employees, who are likely to be high performing. Due to this reduced range and the modest sample, these correlational analyses should be interpreted with caution.

Results are presented in Table 4. Overall, the individual difference measures were not highly correlated with the measures of word recognition. However, the oral language composite score significantly predicted auditory target timing. This suggests that general oral language ability is related to the way that spoken words are recognised in real time. This is somewhat surprising; prior work has linked high-level language abilities like the oral language composite here to target asymptote, but not to target speed (McMurray et al., 2010; McMurray, Klein-Packard, & Tomblin, 2019). Those in the DLD range of performance often show decreased target resolution. This may reflect differences in how language ability maps to word recognition in younger children—in earlier development, language ability might more directly link to the ability to initially activate relevant word forms, and only come to govern resolution after this process has improved. Alternatively, this may reflect decreased sensitivity to detect relationships with the maximum, given the high-performance levels of our sample. Further research is needed to investigate this relationship.

Discussion

These data demonstrate a complex pattern of development of word recognition throughout childhood. The developmental patterns in spoken-word recognition found by Rigler and colleagues (2015) were replicated here: target recognition speed increased with age, there were no effects of age on the asymptotic performance, and competitors showed more rapid resolution. The tendency for spoken cohorts to show outsized competition compared to other competitor types also replicated here, in line with various other studies of adults (Alloppenna et al., 1998) and children (Rigler et al., 2015). However, several results from this study extended this in novel ways and revealed new insights into lexical development.

What competes?

First, this study extends our understanding of the classes of words that compete in spoken and written word recognition. In this case, we observed clear evidence for the competition from cohort and rhyme competitors in both modalities but not frame competitors in which the vowel is not overlapping.

⁴The WASI subtests are standard-scored with a mean of 50 and SD of 10.

The cohort effects are not surprising. These are extremely robust in spoken words at all ages (Fernald et al., 1998; Rigler et al., 2015; Sekerina & Brooks, 2007). However, rhyme effects—though small—are perhaps more surprising. Using a highly similar paradigm with a comparable design and similar words, Hendrickson et al. (2021), showed no evidence for rhyme activation in either spoken or written word recognition. While rhyme effects have been demonstrated with adults (e.g., Allopenna et al., 1998; Connine et al., 1993) and older children (McMurray et al., 2010; Rigler et al., 2015), rhyme effects are often not seen with shorter (one-syllable) words (Simmons & Magnuson, 2018). This may explain the lack of effects in Hendrickson et al. (2021), but if that were the case, we should have also not observed them here. However, recall that the general developmental pattern—observed both here and in Rigler et al. (2015) and Sekerina and Brooks (2007) is for greater competitor fixations of all types earlier in development. Thus, one possibility is that these effects are more robust in children and gradually wane by adulthood.

In addition, this study introduced a novel competitor type, frames. These competitors were motivated by work in written-word recognition (Perea & Lupker, 2003) and were predicted to receive quite strong consideration, especially in that domain. This prediction was not supported. Frame competitors received minimal consideration in both modalities across all age groups (and were numerically *less* likely to be fixated than unrelated objects).

Why did frames receive such little consideration? The total amount of overlap between frames and other competitor types was generally comparable (although some 4-letter frames differed in both medial letters—*heap* vs. *hoop*—rather than a single letter—*sail* vs. *soil*). One possibility is that vowels play an outsized role in word recognition for both spoken and written words. In the spoken domain, vowels have been conceptualised as “islands of reliability” (Diehl et al., 1987)—vowels carry consistent acoustic information that can be used to help interpret the surrounding consonants (Galle et al., 2019; McMurray & Jongman, 2011; Schreiber & McMurray, 2019). A similar pattern emerged for a study investigating competitors with different phoneme orders (Gregg et al., 2019). This study used the VWP to assess consideration of anadromes when the vowel was in a different position (e.g., *leaf* vs. *flea*). They found no evidence of consideration of these types of competitors but did find evidence for activation when the vowel was in a constant position (*leaf* vs. *feel*). It seems that listeners may rely heavily on the vowel for word recognition.

The lack of frame effects for written words is perhaps more surprising. The static nature of written words allows readers to see the onset and offset letters simultaneously (particularly for the short three- and four-letter words used here). In principle, this could elicit just as much competition as cohort competitors with similar numbers of shared letters; the nature of reading allows a skilled reader to sample letters in parallel. Furthermore, many prior studies, such as those reviewed in the introduction to this article, suggested that onset and offset information is prioritised during word reading (e.g., Estes et al., 1976; Perea & Lupker, 2003). These studies imply that word medial information should play a smaller role than onset and offset consonants. This is the exact opposite of the pattern found here, in which both cohort and rhyme competitors were considered more than unrelated, whereas frames were not. This may indicate continuity with the spoken domain, in which vowels help drive word recognition; and in principle, it may derive from developmental processes that build

direct orthographic mappings on the basis of phonology (e.g., Seidenberg & McClelland, 1989).

It may also suggest that word medial information is more critical than previously thought. Many prior studies relied on transposition effects to show that word-medial transpositions were less disruptive than word-onset or word-offset transpositions. Perhaps letter *order* is less crucial word medially, but letter *identity* is more crucial. This might suggest that transposed-medial-letter competitors would be much more highly considered for written words (e.g., *liar* vs. *lair*). However, such words are quite rare in English CVVC words. Future research is needed to understand the roles of letter order and vowels in written-word recognition.

One important caveat to these claims is that competition effects were overall fairly small, especially in the written modality. Even the oldest age group showed much weaker cohort consideration than adults in the Hendrickson et al. (2021) study. This might reflect aspects of stimulus selection. In the present study, half of the items were three-letter words, and half were four-letter words. This was chosen to allow careful control of overlap between competitor types—we did not have to consider the effects of consonant clusters or consonant digraphs. However, the use of very short words might affect the nature of word recognition and competition (c.f., Pitt & Samuel, 2006; Simmons & Magnuson, 2018). These words might elicit extremely automatic word recognition and thus bypass some competition processes that might emerge for words whose recognition is more protracted. Consequently, it is possible that longer words may yet reveal activation of frames.

Nonetheless, even with these short words, cohorts and rhymes obeyed the typical patterns (cohort > rhyme, and both increasing over development), so it is not likely that the use of longer words would change the qualitative pattern of results.

Development

This study extended our understanding of development in several ways. First, in spoken words, in line with previous work, children become faster to activate targets. This speed-up was identified to take place especially between the young and middle age groups in our sample, roughly between 8 and 11. Thereafter, we saw much less change in target timing between the middle and older age groups (10–12 vs. 13–15 years). We would not argue that development is complete by this age. Indeed, McMurray and colleagues (2018) show changes between 11–12 and 17–18 year olds in aspects of speech perception and spoken-word recognition for highly similar minimal pairs (*beach/peach*, *sip/ship*) with some phonetic ambiguity. Nonetheless, the similarity of the two older groups here suggests more gradual changes in these later ages, offering a more precise measure of when these improvements occur. It seems that target fixations speed substantially during the late elementary school years, with less change thereafter.

Second, we showed a nearly equivalent pattern in written-word recognition. This has not been observed before but is clearly attested here in a fairly similar developmental time course with rapid changes between the younger and middle groups, and more gradual changes later. This time course is surprising, as the development of written-word recognition

is necessarily delayed relative to that of spoken words—most children don't engage in substantial engagement with text until kindergarten.

Equally importantly, we found no evidence for developmental change in the asymptotic portion of the target fixations for either spoken or written word recognition. Differences in asymptote have previously been associated with developmental language disorder (Dollaghan, 1998; McMurray et al., 2010; McMurray, Klein-Packard, & Tomblin, 2019)—resolution seems to vary with language ability within an age, not with age-group differences. However, this work has largely looked at late adolescents. This study suggests a similar lack of age effects for younger children. However, we did not find a correlation between standard measures of language ability and target asymptote here. This might reflect the reduced range of language ability we sampled; children in this study were generally high performing, so we may not have sampled ranges where language ability would predict substantial changes in resolution.

Our data support the theoretical contrast we've developed linking changes in *activation rate* to development and changes in *resolution* to individual differences in ability (McMurray et al., in press). Here activation rate refers to the ability of the language system to rapidly update itself in response to new information: building initial activation for targets and cohorts and rapidly eliminating competitors when disqualifying information arrives. Computational modelling suggests it is likely supported by changes in the strength of bottom-up connections between input (phonemes or letter representations) and words. In contrast, resolution refers to the ability of the system to stably and fully reject competitors late in processing. It is supported by processes like lateral inhibition (direct inhibitory connections between words; McMurray, Klein-Packard, & Tomblin, 2019) or possibly by domain general cognitive control (Zhang & Samuel, 2018).

Prior studies on this development have not accounted for the potential contribution of non-linguistic abilities. Here, we related these improvements to a non-linguistic VWP task. Results suggest that the improvements in target timing may not reflect specifically age-related changes to word recognition but instead reflect increased cognitive efficiency more broadly. Measures of more basic eye movement dynamics (operationalised as the fixation duration during preview for the non-linguistic task) predicted little variance in the linguistic tasks, whereas the more cognitive measures of target and competitor fixation in a non-linguistic domain were more predictive. This suggests that older children more quickly select and fixate targets in this task regardless of the nature of the stimuli.

However, at the same time, we found that spoken- and written-word recognition were related at an individual level in ways that cannot be accounted for by visuo-cognitive processes. This suggests there may be some core to lexical processing that is individually specific, even if it is not necessarily developing in this age window (or we couldn't detect it).

An intriguing finding was that fixation duration from the non-linguistic VWP was significantly predictive of target timing for written words, but not for spoken words. Although this relationship was small ($B = -.0037$; $R^2 = .057$), it suggests that at least some aspect of target identification in the written modality reflects basic eye movement dynamics.

One possible explanation is that participants with faster fixation durations can more quickly move from fixating the target word at screen centre to fixating response options. In the written modality, participants must fixate the centre to begin the trial by clicking the dot, and then remain there to read the target. The ability to more quickly disengage after reading could lead to more rapid onset of target fixations, and thus faster target timing. In the spoken domain, participants need only fixate the centre dot, after which timing is more dictated by the delay in auditory stimulus presentation.

It is important to point out that the relative simplicity of the words here may have been less sensitive to lexical development than other words. These short words are learned early (in both domains) and are fully decodable (in reading). Consequently, they may reach a “ceiling” in lexical development earlier than longer more complex words. Thus, this issue should be revisited with more complex vocabulary.

However, even if these developments are largely visuocognitive, we would not want to claim that these age-related changes are epiphenomenal to word recognition and reading (e.g., pure oculomotor control). Rather these likely reflect a host of cognitive processes and factors (general speed of processing, decision-making, inhibitory control, etc.) that are brought to bear in language. Further research would do well to try to isolate and identify these components.

The correspondence between spoken- and written-word recognition

The present study showed a striking concordance between spoken- and written-word recognition at a number of levels. First, the dynamics of target activation in both spoken and written words showed a highly similar developmental pattern, with large changes in the timing between the young and middle groups, and smaller changes later. Although this may be driven by domain general cognitive processes, the similarity is striking.

Second, Hendrickson and colleagues (2021) showed that across subjects there was a close correspondence in the profiles of which competitors are active between the modalities. We see the same parallel between modalities here: cohorts and rhymes were active in both modalities but frames were not.

However, the present study also showed qualitative differences in which competitors were active. The prior study revealed stronger cohort competition than rhyme competition in both spoken and written modalities for adults and no evidence for rhyme activation at all. In contrast, the present study showed this only in the spoken modality. For written words, cohorts and rhymes showed similar levels of competition. This pattern held across development; even the older age group showed relatively low levels of written cohort competition. This suggests that the increased cohort competition seen by Hendrickson and colleagues might be a very late developmental pattern.

Third, unlike Hendrickson et al., we directly correlated the dynamics of spoken and written recognition across individuals. We found that the rate of target activation was correlated across domains, and this could not be fully accounted for by visuo-cognitive processes. This suggests that the basic process of building activation for words may be at least

partially similar across domains. At the same time, we found few correlations between competitor fixations across domains. This null effect needs to be taken with a grain of salt, as particularly in the written domain the overall amount of competitor fixations was small, suggesting a reduced range that may limit sensitivity. However, it raises the intriguing possibility that the processes that build activation for words may be similar across domains, but the process of suppressing competitors may be modality specific.

Finally, despite the lack of correlation at the level of individuals, competitor effects were extremely correlated between modalities at the level of individual words. This pattern emerged despite weak links between competition in the two modalities for participants. This suggests that although the degree of competition that a participant exhibits may be fairly independent between modalities, the specific competitors are consistent across modalities—when the target is *goat*, *goal* is activated, whether the word is in spoken or written format. However, this finding should be interpreted cautiously. Several less interesting factors could drive these correlations. In particular, the images used to represent the words were identical between modalities. It is possible that the correlations emerged because of variability in how well the visual image matched the mental representation of the items—participants would be more likely to fixate the *goal* after hearing *goat* if the visual representation is particularly clear. Even more simply, these could reflect variability in visual salience—if the picture of *goal* is particularly salient, it will draw more looks independent of the target presentation modality. As such, the correlations seen here are interesting, but warrant further consideration.

Limitations

Although this study furthered our understanding of the development of word recognition across domains, several limitations suggest avenues of future research. First, as we described the relatively short, simple, and early-learned words may limit our sensitivity to developmental effects as well as our ability to detect correlations with other things.

An additional consideration is the presentation format used in the written-word recognition task. In this task the written targets were presented at screen centre while images were already present, matching the task used by Hendrickson and colleagues (2021). However, this presentation might underestimate the degree of competitor consideration. The need to fixate the centre of the screen to read the target prohibits eye movements to the images until after this fixation is complete. This might slow stimulus-driven looks; indeed, the target fixation curves in Figure 4 show later onset for written words than for spoken words. It is possible that manipulating the presentation format by changing the masking interval or by delaying the presentation of the images until after the target has been masked might result in greater sensitivity to competitor activation. It is also important to point out that the VWP itself introduces visual search demands that can be shaped by the nature of stimulus preview and the geometry in the visual display (Apfelbaum et al., 2021). Even if the listener has activated *goal*, for example, they must find the relevant visual features (or better yet, know where they are before the word) to reveal this activation with a fixation. Without this, unrelated fixations may be higher than necessary, making it difficult to detect competitor effects.

Finally, we found a minimal relationship between individual difference measures and word recognition ability. This was somewhat surprising, particularly for measures of oral language ability. We expected that children with better oral language would also show more efficient word recognition—oral language relies in part on strong word-level processing abilities. However, our convenience sampling approach resulted in low variability of these measures. It is possible that a wider range of ability level would reveal closer links between these abilities and the word recognition measures. In particular, our sample tended to be quite high performing; it is possible that word recognition and oral language ability might be more closely linked for those with weaker abilities.

Conclusion

This study adds to our understanding of the development of word recognition. In both spoken and written modalities, word recognition becomes faster with age, although some of this speed-up may reflect general improvements in cognitive processing, rather than language-specific improvements. Our results revealed links between word recognition in the two domains, but also notable dissociations, particularly in cohort and rhyme competition; these competitor types were highly similar for written words but showed asymmetries for spoken words. Recent data suggest that adults show this asymmetry for written words as well (Hendrickson et al., 2021), this suggests that the development of written-word recognition continues beyond the age range studied here.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors would like to thank the *Growing Words* team for helpful discussions on this project. Stimuli, data, and analyses are available at <https://osf.io/eav72>. This project is part of the larger *Growing Words* project, which includes cross-sectional and longitudinal investigations of the development of word recognition. More information is available at <https://growingwords.lab.uiowa.edu/>.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was funded by R01-DC008089 awarded to B.M. and P50-DC00242 awarded to Bruce Gantz.

Data accessibility statement

The data and materials from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/eav72>.

References

Allopenna PD, Magnuson JS, & Tanenhaus MK (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419–439. 10.1006/jmla.1997.2558

- Apfelbaum KS, Klein-Packard J, & McMurray B (2021). The pictures who shall not be named: Empirical support for benefits of preview in the Visual World Paradigm. *Journal of Memory and Language*, 121, Article 104279. 10.1016/j.jml.2021.104279 [PubMed: 34326570]
- Carreiras M, & Perea M (2002). Masked priming effects with syllabic neighbors in a lexical decision task. *Journal of Experimental Psychology: Human Perception and Performance*, 28(5), 1228–1242. 10.1037/0096-1523.28.5.1228 [PubMed: 12421067]
- Chambers SM (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behavior*, 18(2), 225–241. 10.1016/S0022-5371(79)90136-1
- Connine CM, Blasko DG, & Titone D (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, 32(2), 193–210. 10.1006/jmla.1993.1011
- Dahan D, & Gaskell MG (2007). The temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. *Journal of Memory and Language*, 57(4), 483–501. 10.1016/j.jml.2007.01.001 [PubMed: 18071581]
- Desroches AS, Newman RLL, & Joanisse MF (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, 21(10), 1893–1906. 10.1162/jocn.2008.21142 [PubMed: 18855555]
- Diehl RL, Kluender KR, Foss DJ, Parker EM, & Gernsbacher MA (1987). Vowels as islands of reliability. *Journal of Memory and Language*, 26, 564–573. 10.1016/0749-596X(87)90143-4 [PubMed: 25505818]
- Dollaghan C (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics*, 19(2), 193–207. 10.1017/s0142716400010031
- Dufour S, & Grainger J (2019). Phoneme-order encoding during spoken word recognition: A priming investigation. *Cognitive Science*, 43(10), Article e12785. 10.1111/cogs.12785 [PubMed: 31621125]
- Estes WK, Allmeyer DH, & Reder SM (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, 19(1), 1–15. 10.3758/BF03199379
- Farris-Trimble A, & McMurray B (2013). Test–retest reliability of eye tracking in the Visual World Paradigm for the study of real-time spoken word recognition. *Journal of Speech, Language, and Hearing Research*, 56, 1328–1346. 10.1044/1092-4388
- Fernald A, Perfors A, & Marchman VA (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, 42(1), 98–116. 10.1037/0012-1649.42.1.98 [PubMed: 16420121]
- Fernald A, Pinto JP, Swingley D, Weinberg A, & McRoberts GW (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231. 10.1111/1467-9280.00044
- Frauenfelder UH., Scholten M., & Content A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, 16(5–6), 583–607. 10.1080/01690960143000146
- Galle ME, Klein-Packard J, Schreiber K, & McMurray B (2019). What are you waiting for? Real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cognitive Science*, 43(1), Article e12700. 10.1111/cogs.12700
- Gregg J, Inhoff AW, & Connine CM (2019). Re-reconsidering the role of temporal order in spoken word recognition. *Quarterly Journal of Experimental Psychology*, 72(11), 2574–2583. 10.1177/1747021819849512
- Harm MW, & Seidenberg MS (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, 111(3), 662–720. 10.1037/0033-295X.111.3.662 [PubMed: 15250780]
- Havy M, Bertonecini J, & Nazzi T (2011). Word learning and phonetic processing in preschool-age children. *Journal of Experimental Child Psychology*, 108(1), 25–43. 10.1016/j.jecp.2010.08.002 [PubMed: 20850758]
- Hendrickson K, Apfelbaum K, Goodwin C, Blomquist C, Klein K, & McMurray B (2021). The profile of real-time competition in spoken and written word recognition: More similar

- than different. *Quarterly Journal of Experimental Psychology*. Advance online publication. 10.1177/17470218211056842
- Huestegge L, Radach R, Corbic D, & Huestegge SM (2009). Oculomotor and linguistic determinants of reading development: A longitudinal study. *Vision Research*, 49(24), 2948–2959. 10.1016/j.visres.2009.09.012 [PubMed: 19781566]
- Huetting F, Rommers J, & Meyer AS (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171. 10.1016/j.actpsy.2010.11.003 [PubMed: 21288498]
- Magnuson JS, Dixon JA, Tanenhaus MK, & Aslin RN (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31(1), 133–156. 10.1080/03640210709336987 [PubMed: 21635290]
- Marslen-Wilson WD (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1–2), 71–102. 10.1016/0010-0277(87)90005-9 [PubMed: 3581730]
- Marslen-Wilson WD, Moss HE, & van Halen S (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 22(6), 1376–1392. 10.1037/0096-1523.22.6.1376 [PubMed: 8953227]
- Marslen-Wilson WD, & Warren P (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101(4), 653–675. 10.1037/0033-295X.101.4.653 [PubMed: 7984710]
- Marslen-Wilson WD, & Zwitserlood P (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 576–585. 10.1037/0096-1523.15.3.576
- McClelland JL, & Elman JL (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1–86. 10.1016/0010-0285(86)90015-0 [PubMed: 3753912]
- McClelland JL, & Rumelhart DE (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review*, 88(5), 375–407. 10.1037/0033-295X.88.5.375
- McConkie GW, Kerr PW, Reddix MD, & Zola D (1988). Eye movement control during reading: I. The location of initial eye fixations on words. *Vision Research*, 28(10), 1107–1118. 10.1016/0042-6989(88)90137-X [PubMed: 3257013]
- McMurray B, Apfelbaum KS, & Tomblin JB (in press). The slow development of real-time processing: Spoken Word Recognition as a crucible for new thinking about language acquisition and disorders. *Current Directions in Psychological Science*. <https://psyarxiv.com/uebfc/>
- McMurray B, Danelz A, Rigler H, & Seedorff M (2018). Speech categorization develops slowly through adolescence. *Developmental Psychology*, 54(8), 1472–1491. 10.1037/dev0000542 [PubMed: 29952600]
- McMurray B, Ellis TP, & Apfelbaum KS (2019). How do you deal with uncertainty? Cochlear implant users differ in the dynamics of lexical processing of noncanonical inputs. *Ear and Hearing*, 40(4), 961–980. 10.1097/AUD.0000000000000681 [PubMed: 30531260]
- McMurray B, & Jongman A (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, 118(2), 219–246. 10.1037/a0022325 [PubMed: 21417542]
- McMurray B, Klein-Packard J, & Tomblin JB (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition*, 191, Article 104000. 10.1016/j.cognition.2019.06.012 [PubMed: 31234114]
- McMurray B, Samelson VM, Lee SH, & Tomblin JB (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*, 60(1), 1–39. 10.1016/j.cogpsych.2009.06.003 [PubMed: 19836014]
- Perea M (1998). Orthographic neighbours are not all equal: Evidence using an identification technique. *Language and Cognitive Processes*, 13(1), 77–90. 10.1080/016909698386609
- Perea M, & Lupker SJ (2003). Transposed-letter confusability effects in masked form priming. In Kinoshita S & Lupker SJ (Eds.), *Masked priming: The state of the art* (pp. 97–120). Psychology Press. 10.4324/9780203502846

- Pitt MA, & Samuel AG (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, 32(5), 1120–1135. 10.1037/0096-1523.32.5.1120 [PubMed: 17002526]
- Rastle K, & Coltheart M (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 25(2), 482–503. 10.1037/0096-1523.25.2.482
- Rayner K (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, 41(2), 211–236. 10.1016/0022-0965(86)90037-8 [PubMed: 3701249]
- Rigler H, Farris-Trimble A, Greiner L, Walker J, Tomblin JB, & McMurray B (2015). The slow developmental time course of real-time spoken word recognition. *Developmental Psychology*, 51(12), 1690–1703. 10.1037/dev0000044 [PubMed: 26479544]
- Roembke TC, Hazeltine E, Reed DK, & McMurray B (2019). Automaticity of word recognition is a unique predictor of reading fluency in middle-school students. *Journal of Educational Psychology*, 111(2), 314–330. 10.1037/edu0000279
- Roembke TC, Hazeltine E, Reed DK, & McMurray B (2021). Automaticity as an independent trait in predicting reading outcomes in middle-school. *Developmental Psychology*, 57, 361–375. 10.1037/dev0001153 [PubMed: 33570987]
- Schreiber KE, & McMurray B (2019). Listeners can anticipate future segments before they identify the current one. *Attention, Perception, and Psychophysics*, 81(4), 1147–1166. 10.3758/s13414-019-01712-9
- Seidenberg MS, & McClelland JL (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568. 10.1037/0033-295X.96.4.523 [PubMed: 2798649]
- Sekerina IA, & Brooks PJ (2007). Eye movements during spoken word recognition in Russian children. *Journal of Experimental Child Psychology*, 98(1), 20–45. 10.1016/j.jecp.2007.04.005 [PubMed: 17560596]
- Simmons ES, & Magnuson J (2018). Word length, proportion of overlap, and phonological competition in spoken word recognition. *Proceedings of the Cognitive Science Society*, 40, 1062–1067.
- Storkel HL, & Hoover JR (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, 42(2), 497–506. 10.3758/BRM.42.2.497 [PubMed: 20479181]
- Toscano JC, Anderson ND, & McMurray B (2013). Reconsidering the role of temporal order in spoken word recognition. *Psychonomic Bulletin & Review*, 20, 981–987. 10.3758/s13423-013-0417-0 [PubMed: 23456328]
- Vorstius C, Radach R, & Lonigan CJ (2014). Eye movements in developing readers: A comparison of silent and oral sentence reading. *Visual Cognition*, 22(3), 458–485. 10.1080/13506285.2014.881445
- Weber A, & Scharenborg O (2012). Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3), 387–401. 10.1002/wcs.1178 [PubMed: 26301470]
- White SJ, Johnson RL, Liversedge SP, & Rayner K (2008). Eye movements when reading transposed text: The importance of word-beginning letters. *Journal of Experimental Psychology: Human Perception and Performance*, 34(5), 1261–1276. 10.1037/0096-1523.34.5.1261 [PubMed: 18823209]
- Zhang X, & Samuel AG (2018). Is speech recognition automatic? Lexical competition, but not initial lexical access, requires cognitive resources. *Journal of Memory and Language*, 100, 32–50. 10.1016/j.jml.2018.01.002

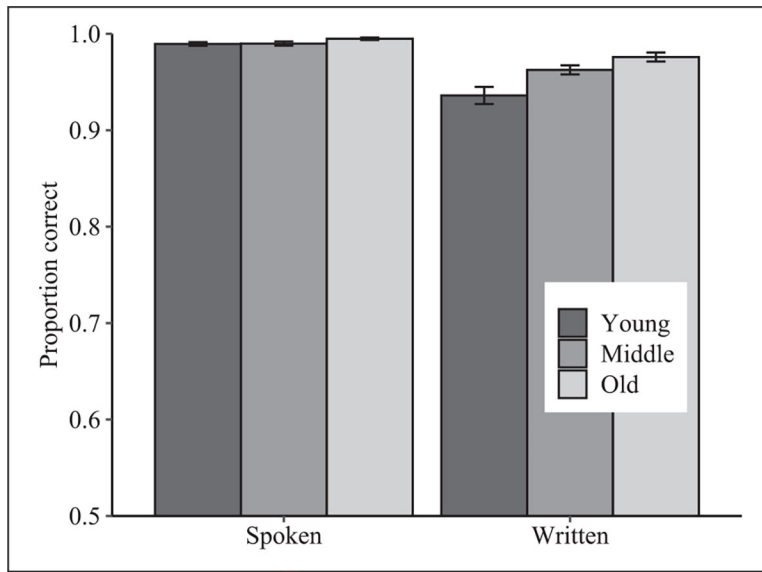


Figure 1. Accuracy during the VWP study by age and modality. This includes only participants who surpassed the 85% accuracy threshold for the written trials. VWP: Visual World Paradigm.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

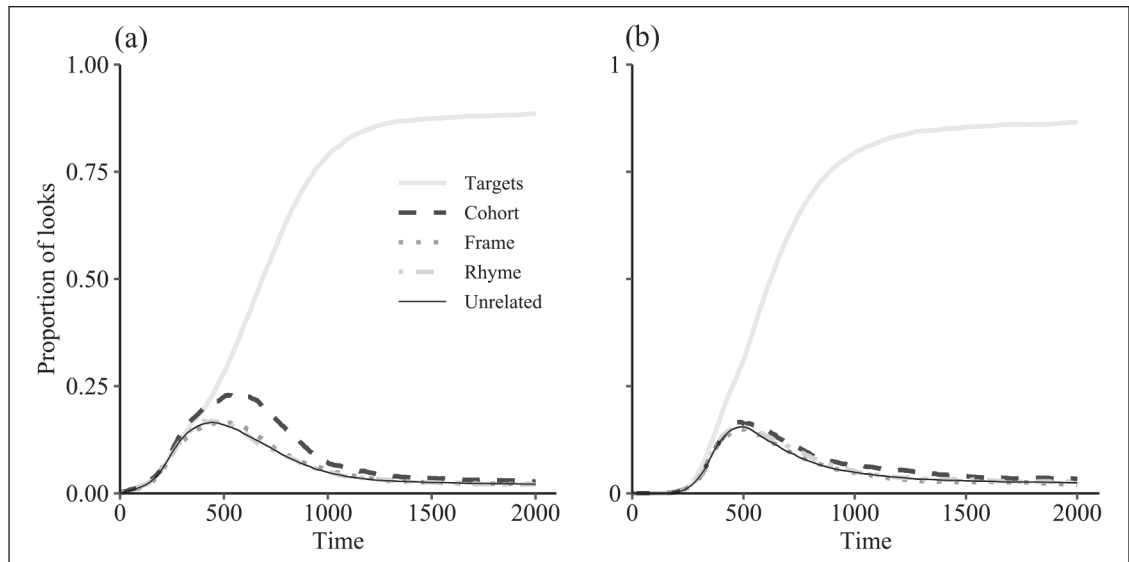


Figure 2. The time course of looks to targets and different competitor types in each modality averaged across all age groups. (a) Spoken trials; (b) Written trials.

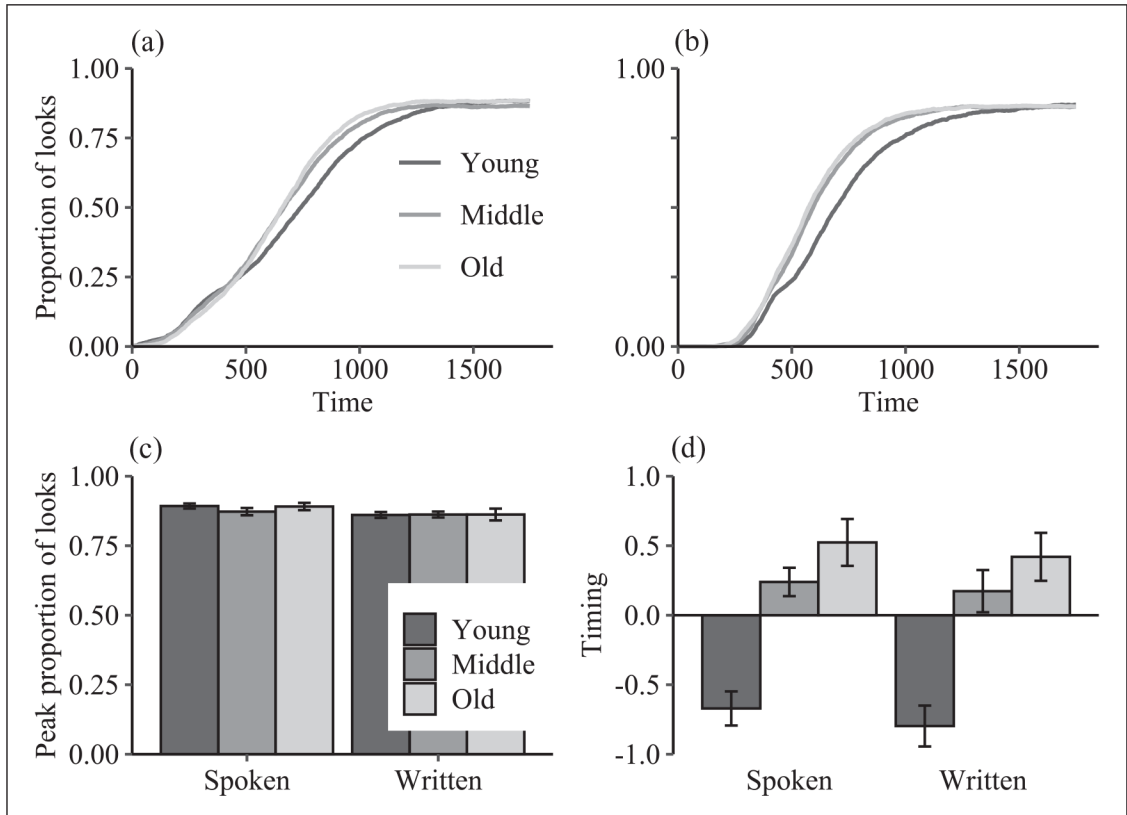


Figure 3. Target fixations. (a) Mean fixations to the target across time by age group in the spoken modality, averaged across all competitor types. (b) Mean fixations to the target across time by age group in the written modality, averaged across all competitor types. (c) Mean maximum curvefit parameters by age and modality. (d) Mean of the timing index by age and modality.

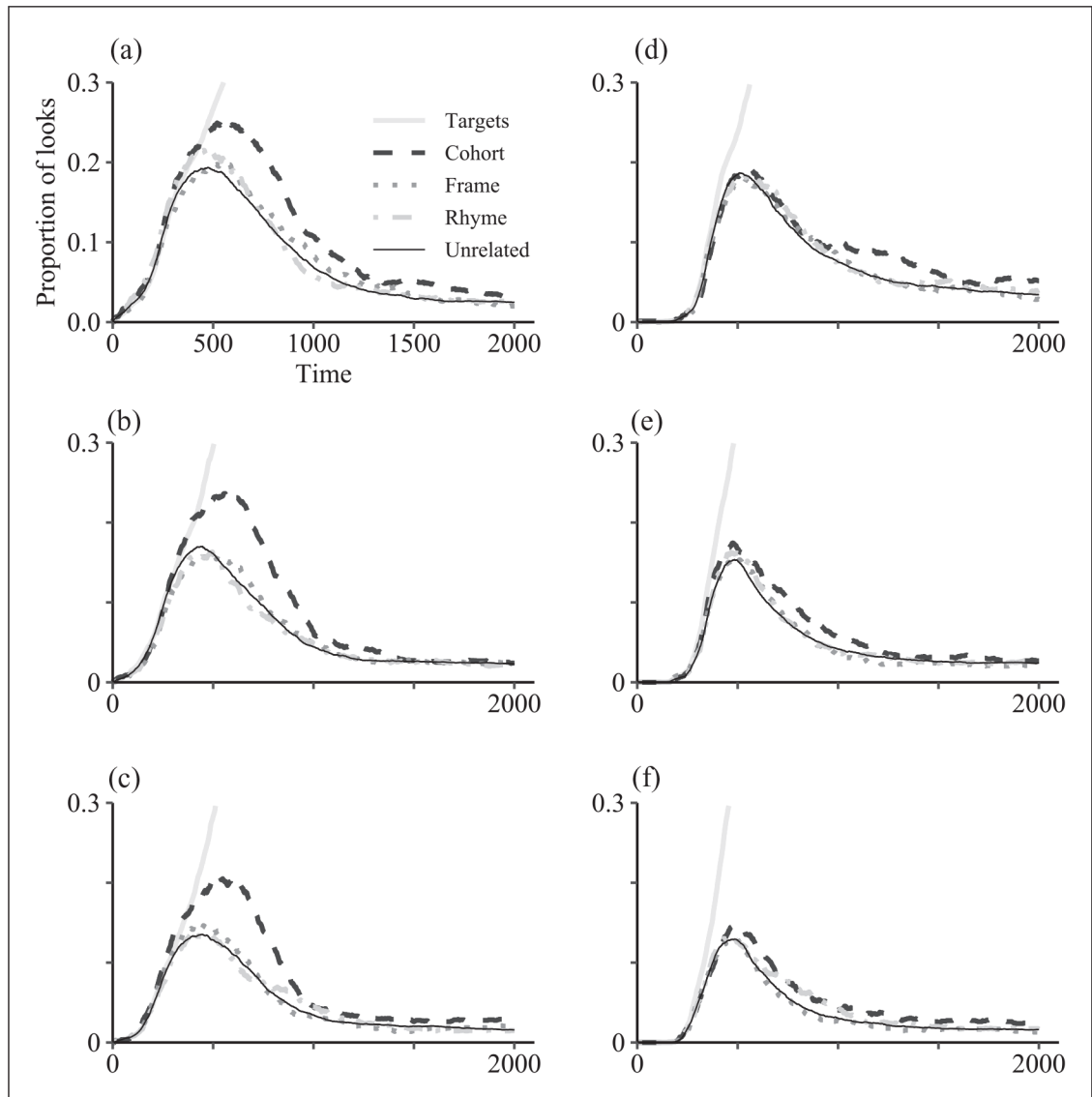


Figure 4. The proportion of looks to each object type across time for each age group and each modality. Note that the figure is capped at a proportion of 0.3; targets rapidly exceed this. (a) Young; Spoken. (b) Middle; Spoken. (c) Old; Spoken. (d) Young; Written. (e) Middle; Written. (f) Old; Written.

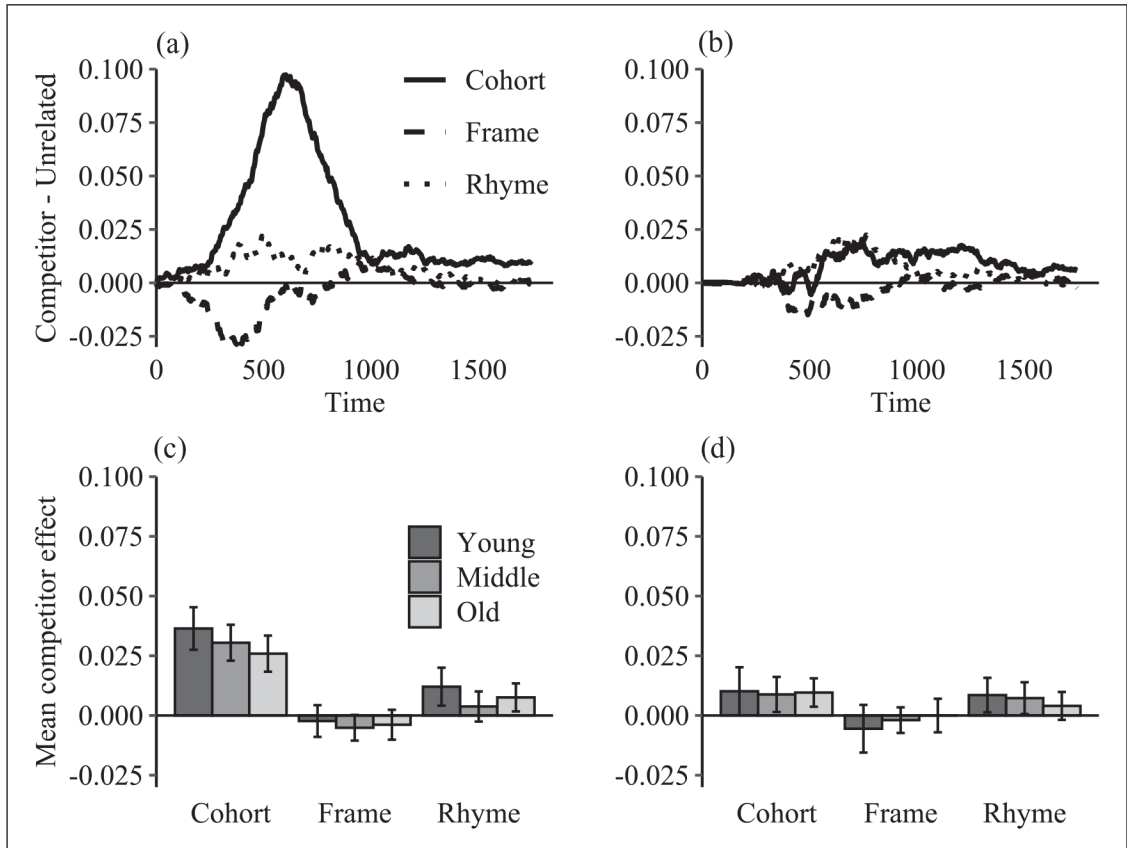


Figure 5. Competitor effects for each competitor type in each modality. These figures plot the difference in looks to the competitor object and looks to the mean of the unrelated objects. (a) The time course of the competitor effect for the spoken modality. (b) The time course of the competitor effect for the written modality. (c) The average competitor effect from 200 to 1000 ms for the spoken modality. (d) The average competitor effect from 200 to 1000 ms for the written modality.

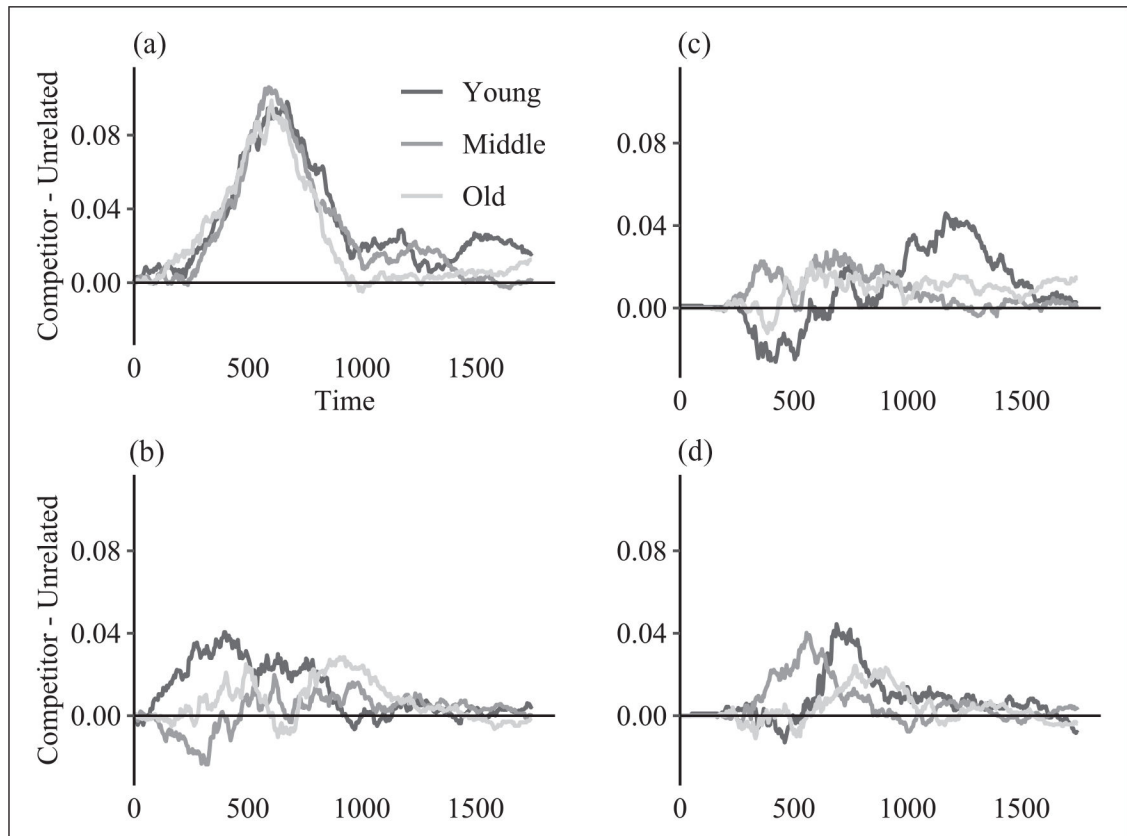


Figure 6. Average competitor effect across time for cohort trials by age. (a) Cohorts in spoken trials. (b) Rhymes in spoken trials. (c) Cohorts in written trials. (d) Rhymes in written trials.

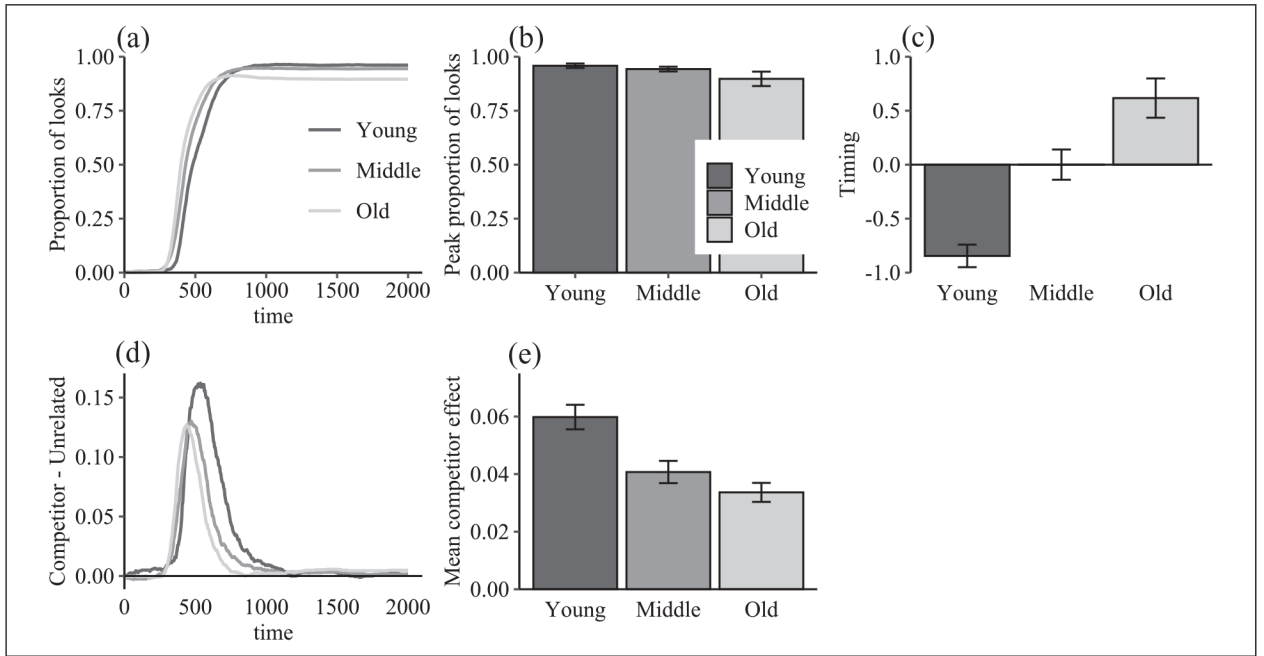


Figure 7. Visual-only VWP experiment data. (a) Time course of target looks by age. (b) Mean of the maximum curvefit parameters by age. (c) Mean of the timing curvefit parameters by age. Error bars display the standard error of the mean. (d) Competitor effect across time. (e) Mean competitor effect from 200–1000 msec post-stimulus onset. VWP: Visual world paradigm.

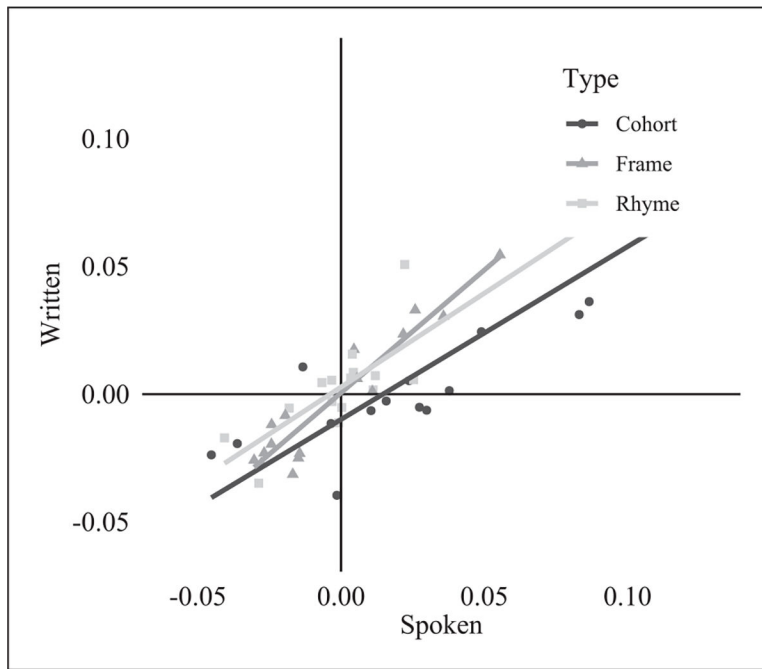


Figure 8. Correlations of competitor effects of items between modalities.

Table 1.

Correlation coefficients of competitor effects between competitor types for spoken trials.

		Cohort	Frame	Rhyme
Spoken	Cohort		-.034	-.11
	Frame	-.034		-.16
	Rhyme	-.11	-.16	
Written	Cohort		-.012	-.15
	Frame	-.012		.091
	Rhyme	-.15	.091	

Hierarchical regression model outputs for competitor effects by competitor type. Orthographic competitor effects were the DV. Model 1 included age and the visual-only task as predictors; model 2 added the auditory competitor effects. *B* and *p* for predictors are from model 2.

Table 2.

	Model 1 <i>R</i> ²	Model 1 <i>p</i>	Model 2 <i>R</i> ²	Model 2 <i>p</i>	Age <i>B</i> ; <i>p</i>	Fix. dur. <i>B</i> ; <i>p</i>	Visual-only <i>B</i> ; <i>p</i>	Auditory <i>B</i> ; <i>p</i>
Cohort	.0053	.95	.0054	.98	.00054; .68	4.8e-6; .91	.082; .57	.010; .93
Frame	.015	.79	.022	.82	.00020; .87	-4.0e-5; .33	-.024; .86	-.092; .48
Rhyme	.082	.11	.082	.20	-.0020; .055	1.2e-5; .75	-.26; .028	-.015; .89

Significant effects are in bold.

Table 3.

Correlation coefficients of competitor effects for items by competitor type.

	<i>r</i>	<i>P</i>
Cohort	.87	.000011
Frame	.95	< .00001
Rhyme	.89	< .00001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Correlation coefficients of individual difference measures with word recognition measures from the VWP tasks.

Table 4.

Modality	Parameter	R^2 of age	Oral language			Reading			Non-verbal intelligence		
			R^2	P	P	R^2	P	P	R^2	P	P
Spoken	Target Max	.00014	.0020	.69	.0017	.71	.021	.18			
	Target Timing	.32	.065	.0043	.011	.24	.0011	.72			
	Cohort CE	.074	.0079	.40	.024	.14	.0013	.29			
	Frame CE	2.1e-7	.0016	.72	.011	.35	.0021	.68			
	Rhyme CE	.00037	.013	.31	.0081	.42	.0052	.52			
Written	Target Max	3.4e-6	2.7e-5	.96	.0018	.72	.017	.27			
	Target Timing	.24	.024	.13	.022	.15	.0054	.48			
	Cohort CE	.00052	.013	.34	.0066	.49	2.9e-5	.96			
	Frame CE	.0021	.028	.15	.023	.19	.014	.31			
	Rhyme CE	.0074	.044	.072	.0064	.50	.0062	.50			

CE signifies competitor effect. Significant effects are in bold.