# What about lexical competition? Exploring the locus of lexical retrieval deficits in adults with developmental dyslexia

Goranova, Jeni; Olson, Andrew; Krott, Andrea

*DOI:*
10.1037/neu0000767



*Citation for published version (Harvard):*
Goranova, J, Olson, A & Krott, A 2021, 'What about lexical competition? Exploring the locus of lexical retrieval deficits in adults with developmental dyslexia', *Neuropsychology*, vol. 35, no. 8, pp. 822-846.
https://doi.org/10.1037/neu0000767

# What about Lexical competition? Exploring the locus of lexical retrieval deficits in adults with Developmental Dyslexia

Zheni O. Goranova, Andrew C. Olson, Andrea Krott

University of Birmingham, United Kingdom

## Abstract

**Objective:** Individuals with dyslexia do not only show deficits with reading but are also less accurate in naming pictures. This has mainly been linked to prevalent phonological deficits. However, deficits in lexical retrieval of picture names could also be due to increased lexical-semantic competition. The present study tested whether adults with dyslexia (AwDs) are more affected by a competitive lexical-semantic context than control participants.

**Method:** Twenty-seven AwD and thirty-four control participants completed the blocked cyclic picture naming paradigm and the Hayling sentence completion task.

**Results**: In the blocked cyclic naming task, AwDs showed a larger semantic interference effect than controls in terms of errors, especially producing competitor errors. In the Hayling sentence completion task, AwDs made more errors than controls when asked to complete sentences with semantically unrelated words, i.e. in the competitive condition. They especially produced semantically related words or antonyms to target words.

**Conclusions:** We found that AwDs experience difficulties with resolving lexical competition that go beyond their phonological deficits. Future studies will need to establish the mechanisms behind the increased lexical competition that AwDs exhibit.


*Keywords: developmental dyslexia, lexical competition, inhibition, blocked cyclic naming task*

# What about Lexical competition? Exploring the locus of lexical retrieval deficits in adults with Developmental Dyslexia

It is well established that individuals with dyslexia experience deficits with literacy skill development commonly in the context of poor phonological processing. However, adults with Dyslexia (AwD) can also be worse than controls in naming pictures. This raises the possibility that AwDs' cognitive components for accessing word forms and meanings are also affected (Castles, Kohnen, Nickels & Brock, 2014) and contribute to reading deficits. That is the issue this paper investigates.

Studies have shown that AwDs can be worse at naming pictured objects (for children see Katz, 1986; Swan & Goswami, 1997; Wolf, 1991 for a review see Nation, 2005; for adults see Raman, 2011). For instance, they make more errors that are phonologically and semantically related to the picture name (Nation et al., 2001). These naming deficits persist even when AwDs are matched to control participants in terms of reading age and vocabulary knowledge (Nation, 2005). Producing the name of a picture involves activation of semantic, lexical and phonological representations. Deficits in any of these processes could lead to poor picture naming.

Similarly, previous picture naming studies of children with dyslexia (CwDs) have reported fewer pictures named correctly (Katz, 1986; Nation et al., 2001; Wolf, 1991) and a large proportion of phonological errors compared to reading matched controls (Swan & Goswami, 1997). For example, Nation et al. (2001) found that the CwDs were less accurate in naming pictures compared to reading-matched control children. Interestingly, the two participant groups also made different types of errors. CwDs made mainly lexical competition errors, producing words semantically related to target words, and their errors were not influenced by the visual similarity with the target (e.g., a coil of rope which is visually similar [but not semantically similar] to a snail). In comparison, control participants' errors were names of objects semantically and visually related to the targets. In addition to the semantic errors, CwDs also made errors that were phonologically similar to the target word, including errors that were both phonologically and semantically related (e.g., "microscope" pronounced instead of "stethoscope"). The authors concluded that CwDs' naming errors were in line with the phonological deficit hypothesis (Snowling, 1981, 2001), which proposes that phonological impairments are the underlying cause of developmental dyslexia. More specifically, individuals with dyslexia are argued to have poor phonological representations and struggle to segment, manipulate, store and retrieve phonemes. They also have difficulty connecting sounds (i.e. phonemes) to letters (i.e. graphemes). These impairments, in turn, affect accumulation of orthographic lexical knowledge, leading to difficulties in reading acquisition (Snowling et. al, 2000). Difficulty manipulating phonological representations might result in increased competition at the phonological level during picture naming and word production and result in more phonological errors. Alternatively, the two types of errors experienced by the CwDs during picture naming could be an indication of increased competition at two different levels: the semantic level and the phonological level of the word production system. Or put differently, the errors could be indicative of a higher intrinsic level of competition in lexical-semantic components used for word production. Speech production models that include competitive mechanisms suggest problems with competition are possible (Dell et al., 1997b; Rapp & Goldrick, 2000).

**Lexical Competition and Speech Production Models**

Lexical Competition in AwDs

Reading models are embedded in a larger architecture for language production that also supports naming, spontaneous speech, repetition and spelling (see Coltheart et al., 2001). They also allow for possible interactions between reading-specific and more general speech production components. For example, a component for grapheme-phoneme conversion that is specific to reading may interact with more general lexical-semantic components for speech (semantic system and phonological output lexicon) to support a self-teaching system for learning new words (see ST-DRC model, Pritchard et al., 2018). The ST_DRC model will be relevant to our results and we will return to a more complete reading model in our discussion. First, however, we start by reviewing speech production models, both because they capture the final stages of reading aloud, and because, for them, the issue of competition has been prominent.

Models of speech production (e.g., Caramazza & Miozzo, 1997; Dell et al., 1997a; Levelt et al., 1999) generally agree that lexical access involves both lexical/semantic and phonological levels of representation. It is also typically accepted that during reading and word production semantically related items are activated in addition to the target. However, speech production models differ as to whether the co-activated items compete for selection or not.

The competitive view of lexical selection assumes that co-activated words/lemmas compete with a target word for production. For instance, in the interactive two-step speech production model (Dell et al., 1997a), activation spreads between semantic, word and phonological levels (in both directions). The spread of the activation thus allows the activation within the semantic level to affect the phonological level and vice versa (see also, e.g., Rapp & Goldrick, 2000). Similarly, in the WEAVER ++ model (Levelt et al., 1999), activation of concept nodes (e.g., *fur* and *tail*) increases the activation levels of both the target lemma (e.g., *cat*) and semantically related representations (e.g., *dog*), even though there is no feedback between levels. The lemma whose activation exceeds that of any competitor is selected for production (according to the "Luce ratio" as defined by Roelofs, 1997). These competitive models can explain why it takes longer to produce a word after a semantically related word has been produced, and why a semantically related word might be produced in error.

Some authors have suggested that lexical-semantic competition is resolved via inhibition of the co-activated competitors, also called lateral inhibition (Feldman & Ballard, 1982; Harley, 1990, 1993; McClelland & Rumelhart, 1981). The involvement of lateral inhibition in resolving lexical competition is implemented in some speech production models (Dell et al., 1997a; Howard et al., 2006; McClelland & Rumelhart, 1981), but not in other competitive models like WEAVER++ (Levelt et al., 1999).

Shao et al., (2013) distinguished two types of inhibition, namely a general, non-selective inhibition, which involves the suppression of planning and execution of any unwanted response, and selective inhibition, operating at the lexical level, which reduces the activation of strong competitors, and allows selection of the target word (Shao et al., 2015). Even though selective and non-selective inhibition are separate components, both act to resolve competition between activated items. This commonality means that some studies consider inhibition to be one general process (Botvinick et al., 2001). Even though some competitive models require inhibition to resolve lexical competition, the mechanism is not crucial, and competition can also be resolved via increased activation (as seen in WEAVER ++).

In contrast to the competitive view of lexical selection, a non-competitive view suggests that the lexical selection occurs when activation of a word passes an absolute threshold, meaning that co-activated words do not affect the selection of a target (Dell, 1986, see Mahon et al., 2007, Oppenheim et. al, 2010, Oppenheim & Balatsou, 2019). Even without competition, however, some models can explain why previously produced semantically related words slow down lexical retrieval. An incremental learning mechanism changes the strength of connections between conceptual and lexical representations based on the speaker's experience. The production of a word strengthens the connection between its conceptual and lexical units, and weakens the connections to co-activated semantically-related items that are not produced. As a result, when a related item needs to be produced later, it is disadvantaged by the weakened connections. In sum, while speech production models differ with regards to the exact mechanisms behind lexical access, all models contain a mechanism that accounts for effects of semantically related words on lexical retrieval. Either through an explicitly competitive process or via incremental learning, these mechanisms produce a common effect: The activation level of the target compared to semantically-related items is less distinct after related items have been recently spoken, either because of long-term (learning) or short-term (direct competition) processes. For the purpose of the current study, we will assume that a more competitive lexicon results whenever the activation difference between the target word and related activated items is small. Regardless of which mechanism is responsible.

Increased competition during picture naming is not the only evidence for a potential lexical retrieval deficit in the people with dyslexia. Other evidence stems from the 'Tip of the Tongue' phenomenon (Faust et al., 2003; Faust & Sharfstein-Friedman, 2003; Hanly & Vandenberg, 2010) and lexical-semantic processing tasks (Jones et al., 2010; Torkildsen et al., 2007).

In the dyslexia literature, the 'Tip of the tongue' (TOT) phenomenon has been investigated with a picture-naming task. CwDs reported more TOT experiences compared to the controls (Faust et al., 2003; Hanly & Vandenberg, 2010), but were able to describe the words successfully, suggesting intact retrieval of semantic representations. The conclusion was that TOT were related to phonological deficits. This is in line with the view that TOT states reflect weakness in name retrieval. However, TOT could also be caused by a more accessible, but incorrect substitute word that spontaneously comes to mind and interferes with the retrieval of the target word because it is a competitor (Logan & Balota, 2003). CwDs's increased TOT states might, therefore, be a result of enhanced lexical competition.

Further evidence of a lexical-semantic dimension to dyslexia can be found in a lexical-semantic priming experiment by Torkildsen et al. (2007). This study compared electrophysiological responses of children with a family risk of dyslexia to those of age-matched controls using three picture-word conditions: congruent pairs (e.g., a picture of a "dog" followed by the auditory presentation of "dog"), incongruent-semantic pairs (e.g., the same picture followed by "cat") and unrelated pairs (e.g., the same picture followed by "car"). While the authors also investigated ERP reflections of phonological abilities, the most interesting results for the present study concern the N400, a negative ERP component peaking around 400 ms after stimulus onset. The N400 component reflects semantic processing demands, with higher amplitudes for more challenging semantic processing (Kutas & Hillyard, 1980). Control children showed an expected N400 increase for unrelated compared to related pairs, but children at risk for dyslexia did not. Authors attributed the results to the highly active cluster of semantic representations resulting in a difficulty singling out the correct phonological representation. In addition, and unlike the controls, children at risk for dyslexia experienced the largest N400 component in the congruent condition. The latter was interpreted by the authors as a result of a possible processing deficit for lexical-semantic information. This semantic processing deficit was confirmed with a subsequent unimodal auditory experiment where cross-modal integration deficits were ruled out. The authors concluded that the deficits in children at-risk for dyslexia affected not only lower-level phonological abilities (measured with earlier ERP components), but also higher order linguistic skills such as lexical and semantic processing.

Further evidence for dyslexic lexical-semantic deficits stems from semantic categorisation tasks. For instance, Jones et al. (2010) asked highly functioning AwDs to either name or classify pictures according to specific categories (e.g., living/non-living). While AwDs were slower in the naming task, which required the retrieval of lexical-phonological codes, AwDs were also slower in the object-categorization task, where only visual processing and access to semantic properties were required, without access to the word’s phonological codes. The authors argued that non-phonological problems contributed to retrieval delays for both semantic and phonological information. A specific deficit resolving lexical competition is a plausible explanation for these findings.

Jones et al. (2016) provide another finding that points at a potential lexical competition deficit in AwDs. They used a rapid automatization naming (RAN) Stroop-like paradigm, where participants were asked to name a sequence of letters as fast as possible, while once in a while the color of the letters changed from black to a different color upon fixation and participants were asked to switch from naming the letter to naming the colour. The study found that both groups showed similar lexical processing and recognition abilities, but AwDs were slower in resolving the lexical competition from the letter name. AwDs continued to be affected by the activation of the competitor for output, even though this activation had been resolved in the control group. The authors propose that inhibiting the previously activated phonological response of the letter name in the output stage. Such a failure to inhibit would lead to an overly competitive lexicon.

Lexical Competition in AwDs

Tasks like picture naming and the TOT paradigm involve lexical selection, supporting the notion of a possible deficit in lexical-semantic processing. Moreover, lexical competition has been considered important in all of these tasks. For instance, lexical competition has been argued to be involved in naming pictures because items semantically related to a target word are activated along with the target (Belke & Stielow, 2013). Similarly, high competition is a possible explanation for the TOT phenomenon that arises when partial information about the target word is recovered (Logan & Balota, 2003). The idea is that this partial information is not sufficient to select the word, but it is sufficient to activate similar words, which creates a competitive situation that is hard to resolve (Maril et al., 2001). In the category fluency task, competition for output is created by the increased activation of the words related to the target category. To perform the task, participants must suppress irrelevant co-activated responses and also avoid repeating names they have already produced (Henry & Crawford, 2004; Hirshorn & Thompson-Schill, 2006). It is proposed that the lexical retrieval errors seen in AwD result from difficulties resolving lexical competition.

**Subtypes of Dyslexia**

The different expressions of literacy deficits observed in individuals with dyslexia suggest two subtypes of the disorder. Individuals with dyslexia can have poor phonological representations and struggle to segment, manipulate, store and retrieve phonemes, together with a difficulty connecting sounds (i.e. phonemes) to letters (i.e. graphemes; a phonological sub-type in Castles, Bates & Coltheart's, 2006 terminology). However, not all participants with dyslexia have poor phonological representations (Castles & Coltheart, 2004). Other individuals have normal phonological processing, but struggle to read irregularly spelled words (e.g. *yacht*; Castles & Coltheart, 1996), where reading strongly depends on orthographic lexical representations (a surface sub-type, Castles, Bates & Coltheart, 2006). Since the current study investigated lexical retrieval deficits, it could be hypothesized that the only relevant participants would be people who struggle with irregular words.. The majority of people with dyslexia, however, present a mixed profile with both non-word and irregular word reading being impaired (Sprenger-Charolles et al., 2011), so lexical retrieval deficits will affect a large subset of individuals. In addition, mixed-profile participants with poor letter-sound conversion skills could end up relying on lexical processing, despite also having some problems in this area, which could reveal issues with lexical competition more clearly (especially when an ambiguous context allows more lexical competitors to be active. See discussion of the ST-DRC model below). Thus, for the present study, which presents a first investigation into lexical competition deficits, we did not distinguish between the two subtypes of dyslexia.

**The Present Study**

The aim of the present study was to explore whether AwDs have lexical-semantic retrieval deficits that could lead to increased lexical competition and its subsequent resolution. More specifically, we investigated lexical retrieval performance under conditions that increase lexical-semantic competition.

Lexical competition in word retrieval has been studied in the neurotypical population using experimental paradigms where participants name pictures in competitive versus non-competitive contexts. Two such tasks are the Colour Stroop task (Stroop, 1935) and the blocked cyclic picture naming task (Belke, 2013; Belke, Meyer, et al., 2005).

In the Colour Stroop task, participants name the colour a word is written in. In some cases, the word’s meaning and the ink colour match (i.e. the word GREEN in green ink, congruent condition), while in other cases they do not (i.e. the word GREEN in blue ink, incongruent condition). The incongruent condition leads to slower responses than the congruent condition. Individuals with dyslexia have been found to have an enhanced Stroop effect compared to a control group (Faccioli et al., 2008; Helland & Asbjornsen, 2000; Kapoula et al., 2010; Protopapas et al., 2007; Proulx & Elmasry, 2015; Reiter et al., 2005; Van der Schoot et al., 2001). It has been hypothesized by Roelofs (2003) that the Stroop effect is a result of the conflict between the activated lexical representation of the word and the representation of its colour. However, another account of the Stroop effect suggests the effect to be caused by participants needing to inhibit (i.e. suppress) the dominant response (i.e. reading the word), in order to name the colour (MacLeod, 1991). The conflict, therefore, might be between different task responses rather than different lexical representations. Furthermore, others (e.g., Posner & Raichle, 1994) have claimed that the Stroop effect involves conflict control at a visual processing level. Given that it is not clear whether the Stroop effect is caused by response inhibition (suppression of the habituated response), by lexical competition or by visual processing mechanisms, it is difficult to tell what produces the enhanced Stroop effect in individuals with dyslexia.

A more appropriate paradigm to study competition during lexical retrieval is the blocked cyclic picture naming task (Belke, 2013; Belke, Meyer, et al., 2005). In this task participants are asked to repeatedly name a small set of pictures that are either all from one semantic category (e.g., all animals = homogeneous condition) or from different categories (e.g., one animal, one piece of furniture, one clothes item and one tool = heterogeneous condition). Response times in the homogeneous condition are longer than response times in the heterogeneous condition and the difference between the two conditions is called the “semantic interference effect” (e.g., Belke et al., 2005). How this effect comes about is debated. Most researchers assume that the increased naming times in the homogeneous block are caused by repeated access to the same semantic category which accumulates activation in a small set of lexical-semantic representations that then compete for selection (Belke, Brysbaert, et al., 2005; Damian et al., 2001). In the heterogeneous condition, semantically unrelated pictures are repeatedly named, so activation is dispersed over different categories and there is little competition. In the homogeneous condition, however, representations of the same semantic category are repeatedly accessed, leading to accumulation of activation within a small set of competitors. In addition, the restriction to a small set of repeated items constitutes a top-down bias. This bias facilitates picture naming in the heterogeneous condition. However, it does not alleviate competition within the semantically related items in the homogenous condition (Belke, 2008).

Not all models, however, locate the competition at the lexical-semantic level. For example, Oppenheim et al. (2010) argue that the effect arises because each trial produces learning that changes the strength of the links between concepts and lexical representations. On each trial, links to selected items are strengthened and links to competitors are weakened. Competition between related items is increased because the target on the current trial is strengthened and the items that will become targets on subsequent trials are weakened. As a result, on a subsequent trial a weakened target competes with a strengthened related item.

The aim of our first experiment was to determine whether or not AwDs show increased interference in a blocked cyclic naming task. If so, either incremental learning or increased lexical competition could offer a plausible explanation.

The blocked cyclic picture naming paradigm is not the only task that has been used to study lexical competition. Other tasks are, for instance, the picture word interference task (Glaser & Düngelhoff, 1984) and the continuous naming task (Howard et al., 2006). An advantage of the blocked cyclic naming task (as well as of the continuous naming task), for the purpose of this study, is that it does not involve word reading which could disadvantage AwDs, unlike the picture word interference paradigm. In contrast to the continuous naming paradigm, it requires the repeated production of a small set of words. This repetition likely leads to a stronger competition than the continuous naming paradigm, which requires participants to produce each word only once.

To probe mechanisms of lexical competition further, the same sample of participants performed a second task that requires selection and suppression of strong lexical competitors: the Hayling task (Burgess & Shallice, 1996). This task was initially designed to test inhibition in patients with frontal lobe lesions (Burgess & Shallice, 1996). It has since been used widely in other patient populations, including groups with Mild Cognitive Impairment (Bélanger & Belleville, 2009), Alzheimer's disease (e.g., Belleville et al., 2006), Parkinson's disease ( e.g., Obeso et al., 2011) as well as in groups of older neurotypical adults ( e.g., Cervera-Crespo & González-Alvarez, 2017) and children ( e.g., Jacobsen et al., 2017). The Hayling task asks participants to complete sentences with missing final words. Each sentence has a highly probable ending (e.g., "dog" for "The cat was chased by the ... "). In the standard form of the task, participants are first asked to complete a sentence with a highly probable word (automatic condition, e.g. "ship" for "The captain wanted to stay with the sinking ... ") and then with an unrelated word (inhibition condition; e.g. "avocado" for "The captain wanted to stay with the sinking ... "). Performance in the automatic condition is linked to lexical-semantic knowledge and response initiation, because the sentence context strongly primes a target word and, thus, limits response possibilities. Performance in the inhibition condition, on the other hand, requires the suppression of previous answers and is therefore, as the name suggests, linked to inhibition as a component of executive function. Importantly, while the task is usually considered to be a measure of general response suppression and linked to executive functions, the task involves the suppression of particular words (the most likely word and any semantically related words) and is, therefore, a measure of selective inhibition and of lexical-semantic competition. The advantage of the Hayling task over other lexical retrieval paradigms such as the Stroop task, in fact, comes from the clear distinction between the response generation and response suppression components associated with its contrasting conditions. A clear distinction between the two conditions is supported by studies using neuro-imaging techniques where distinct cerebral regions were found to be involved in response initiation and response suppression ( e.g., De Zubicaray et al., 2000). In addition, unlike picture naming studies (including the blocked cyclic naming paradigm or the related continuous naming paradigm; Howard et al., 2006), the Hayling task does not involve visual recognition, which could be problematic if individuals with dyslexia have visual processing deficits (Stein & Walsh, 1997). It is an ideal candidate, therefore, to provide corroborating evidence for a lexical-semantic retrieval deficit in AwDs.

It should be kept in mind, though, that while the Hayling task measures suppression, it also requires generation and selection of alternative responses from a large array of competing alternatives (see discussion in De Zubicaray et al., 2000). The suppression condition does not just require the suppression of habitual responses, but also the internal generation of novel responses from a wide array of possibilities. Generation of novel responses could be a problematic area for AwDs, given their problems with semantic fluency tasks which require the production of as many semantic category members as possible in a limited time period (Korhonen, 1995; Levin, 1990; Menghini et al., 2010; Moura et al., 2015; Reiter et al., 2005; Snowling et al., 1997; Varvara et al., 2014). If so, however, they should produce mainly response failures compared to other types of errors and their ratio of response failures to other errors should be larger than that of control participants. This potential word generation difficulty by AwDs can be distinguished from a problem with suppression, which should produce more frequent failures to inhibit the target or members of its semantically related cohort when compared to control participants.

We tested AwDs and controls on both the blocked cyclic picture naming paradigm and the Hayling task. Given previous findings from picture naming, AwDs could name the pictures from the blocked cyclic naming task more slowly than controls in all conditions. Given that participants name the same pictures repeatedly, however, this difference should be minimized. If AwDs have difficulty resolving lexical-semantic competition, instead, we would expect them to show a larger response time difference between semantically homogeneous and heterogeneous conditions than controls, and/or they should produce more semantically related errors. In particular, since all pictures in the homogeneous condition are semantically related, we would expect a higher number of errors and a higher number of competitor errors in this condition.

If AwDs have a deficit in suppressing competing lexical representations, they should also respond more slowly and make more errors than controls in the inhibition condition of the Hayling task. The exact type of errors they make may help to differentiate between different possible suppression mechanisms: task-related response suppression and context-related lexical suppression. If a participant responds with a prepotent response in the inhibition condition and/or if they repeat their response from the inhibition condition in the automatic condition, this implies a failure of task-related response suppression. If a participant responds with a word that completes the sentence context appropriately but is not the prepotent response, this could result, instead, from increased lexical competition between items that are co-activated by the sentence context.

Lexical Competition in AwDs

Both paradigms involve lexical competition. If AwDs have a deficit in lexical competition resolution, then we would expect evidence of such a deficit in both experimental tasks. We might also find positive correlations between measures of this deficit, for instance between the number of erroneously produced words semantically related to the target words in both tasks. But given the different demands of our tasks, such correlations might only be detectable with large participant groups.

Increased lexical competition might affect not only lexical retrieval but might also contribute to reading difficulties in dyslexia. In a dual-route reading context, irregular word and sight-word reading depend on lexical processing. There is reduced support from letter-sound mappings (only partial decoding is possible, e.g. the y and t in *yacht*) and there may be increased reliance on context to discriminate between potential word candidates. If an irregular orthographic input produces more competition among potential outputs in the semantic system and phonological output lexicon, irregular word reading should be slowed (as a selection is made from similarly active entries) or word substitution errors should occur. Thus, given that lexical competition plays a role in both picture naming and reading, we would expect to find correlations between the strength of semantic competition effects in our speech production tasks and performance in tasks that rely heavily on lexical processing, such as sight word and irregular word reading. We, therefore, investigated whether increased semantic competition in our speech production tasks might be related to reading skills by testing participants on a range of literacy tasks (e.g., single word and nonword reading, phonological processing). From these measures we constructed factors (for *lexical skills*, *phonological short-term memory* and *phonological manipulation)* and regressed those onto measures of semantic competition in our speech production tasks.

## Methods

### Participants

Thirty-seven AwDs and forty controls participated in our study. The study was approved by the Ethical Committee of the [Blinded for Review] and informed written consent was obtained from all participants. Participants were either paid £20 compensation or given course credits. All participants were recruited through study-specific recruitment posters displayed around the [Blinded for Review]. For our analyses, we included only data from native monolingual English speakers with normal or corrected-to-normal vision who completed all of our tasks and fit the inclusion criteria based on a neuropsychological assessment (see below). Two participants (one from each group) were excluded due to being bilingual and one participant from the control group failed to complete the second session of the experiment. Participants were categorised into a control and dyslexia group on the basis of a set of verbal and non-verbal tasks described below. Participants were categorised as having dyslexia if: 1) they confirmed they had a formal dyslexia assessment; 2) they had no history of psychological and/or neurological problems/diagnosis; 3) they scored at least 2 SDs below the mean of the control group on two tests from the dyslexia assessment (see below); 4) they scored no more than 2SD below the control means on the non-verbal assessments (non-verbal working memory and non-verbal IQ; see description of assessments below). These criteria are, if anything, more strict than those frequently used, where only a formal diagnosis and group differences are reported, and they are in line with other studies that use multiple measures for criteria (either averaged or individually, see, for example, Kalashnikova & Burnham, 2016; Litt, et al., 2019; Mayringer & Wimmer, 2000). There were significant group differences in tasks important for dyslexia assessment:

phonological STM, phonological awareness, rapid letter naming and word reading (see Table 1). Control participants had to: 1) have no history of psychological and/or neurological problems; 2) score no more than 2 SD below the control mean on no more than one of the dyslexia assessments; 3) score no more than 2 SD below the control mean on the non-verbal assessments (non-verbal working memory and non-verbal IQ).

Four participants from the control group were excluded because they scored more than 2 SDs below controls on at least two of the dyslexia assessments. These participants may have had reading difficulties but had not been formally diagnosed. Nine participants who had a dyslexia diagnosis were excluded because they did not score at least 2 SDs below the control mean on more than one dyslexia assessment. This resulted in a final sample of 27 AwDs (mean age 20.4, SD = 2.3, 6 male) and 34 control participants (mean age 19.1, SD = 1, 6 male).[1]

Participants completed the experiment individually in a quiet testing room [Blinded for Review] over two sessions carried out on two different days, no more than seven days apart. In session one they were assessed with the following dyslexia tests: Gray Silent Reading Test, Comprehensive Test of Phonological Processing, Test of Word Reading Efficiency (TOWRE), and The Irregular Word Reading Efficiency Test (TIWRE). They were also given non-verbal IQ subtests from the Wechsler Adult Intelligence Scale (see Performance IQ from the WAIS-IV), and the Corsi blocks spatial short-term memory test (see details of tests below). Since Dyslexia and Attention Deficit Hyperactivity Disorder (ADHD) have a high rate of comorbidity (e.g., Germanò et al., 2010) and ADHD has been associated with an inhibition deficit (e.g., Barkley,

[1] The male:female ratio is somewhat different from the more typical 3:1 ratio that one might expect in a sample of people with dyslexia. This is most likely because many of our participants were psychology students who in the [Blinded for Review] are predominantly female.

1997), participants completed a self-reported ADHD symptoms checklist **(**Adult ADHD Self-Report Scale**),** to rule out any group differences attributable to ADHD. In session two, participants completed the Cyclic blocking task and the Hayling task. Each testing session lasted between 1.5 and 2 hours. All verbal tasks were tape-recorded for later analysis.

**Assessment of Dyslexia Defining Characteristics**

***1. General Cognition***

We measured perceptual reasoning, processing speed and nonverbal working memory as indicators of participants' general cognition. These non-verbal tests were used to rule out the possibility that lower performance in the group with dyslexia is related to problems in (non-verbal) general cognition. Furthermore, these tests allowed us to test whether any of our participants scored extremely low in terms of general cognition (below 2 DS on any of the tests).

**Non-verbal IQ**. We measured non-verbal IQ with two scales, The Perceptual Reasoning Index Scale and The Processing Speed Index Scale from the WAIS-IV (Wechsler, 2008). From the Perceptual Reasoning Index (PRI), we used Block Design (which requires participants to reproduce abstract designs, using cubes made of red and white parts), Matrix Reasoning (which requires selecting one option that correctly completes an incomplete matrix or series), and Visual Puzzles (which requires selecting three response options that, when combined, reconstruct a previously seen puzzle). For the Processing Speed Index (PSI), we administered the two core subtests: Symbol Search (requires deciding whether one of two varying target symbols appears within a row of distracters) and Coding (requires translation of as many symbols as possible into numbers within a limited timeframe). The WAIS standardised procedure was used for calculating the subtest scores.

**Non-verbal Working Memory.** Spatial short-term working memory was measured with a computerised version of the Corsi-block tapping test (PEBL software; Mueller, 2011). In this task, participants are presented with an arrangement of 9 blocks. Blocks illuminate in a random sequence, starting with 2 blocks and gradually increasing to 9 blocks. Each sequence length is presented twice before it is increased. Participants click on the illuminated blocks to reproduce the order in which blocks were illuminated. The task stops when both sequences of a particular length are reproduced incorrectly. The measure entered into the analyses was the adjusted memory span provided by the PEBL programme and calculated as (start length + total number correct)/(trials per sequence length), where start length and trials per sequence length are set at 2. We used the adjusted memory span instead of the block span because it takes into account the performance on both trials of each sequence length, providing a more sensitive measure (Kessels et al., 2000).

***Literacy***

**Reading Comprehension.** Reading comprehension was measured with the Gray Silent Reading Test (GSRT; Wiederholt & Blalock, 2000). Participants were asked to read, at their own pace, up to 13 brief stories of increasing complexity. They answered five multiple-choice questions for each story probing their comprehension. The test was discontinued if a participant made three incorrect responses for a story. Standard scores are only available for participants up to age 17. We therefore used the raw scores for the group comparison, which included the total number of questions answered correctly. In addition, we compared participant groups on total reading time.

**Phonological Processing.** Subtests from *The Comprehensive Test of Phonological Processing* (CTOPP; Wagner et al., 1999) were used to assess phonological awareness, phonological memory and rapid automatized naming (RAN). Phonological awareness was assessed with the *Elision* and *Phoneme Reversal* subtests. For the *Elision* subtest, participants were asked to omit a sound from a word, thus forming a new word (e.g. “Say *tan* without saying /*t*/” would lead to the production of the word ‘an’). The *Phoneme Reversal* test measured the ability to say phonemes in reversed order to form a real word (e.g., ”Say *ni*, now say *ni* backwards.” answer: ‘in’). Phonological memory was tested with the *Memory for Digits* and the *Non-word Repetition* subtests. Participants were asked to repeat a sequence of digits or non-word stimuli in serial order. *Rapid Letter Naming* required participants to name thirty-six letters arranged in four rows, as fast as possible. There were two trials.

**Single Word and Nonword Reading.** Word and non-word reading fluency were assessed via The Test of Word Reading Efficiency-Second Edition (TOWRE-2; Torgesen et al., 1999). This test required reading either high-frequency words or nonwords as quickly and as accurately as possible within a time limit (45 seconds). The TOWRE score was the number of words pronounced correctly. Irregular word reading was assessed with the Irregular Word Reading Efficiency test (TIWRE; Reynolds & Kamphaus, 2007). The TIWRE score was the number of words pronounced correctly (untimed).

### The Adult ADHD Self-Report Scale (Asrs-v1.1)

The ASRS-v1.1 is a self-reported measure of ADHD symptoms. It is a list of eighteen questions designed by Kessler et al. (2005), asking respondents to indicate how frequently they experience certain behaviours (e.g. “How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?”). Its questions are consistent with the diagnostics criteria from the Diagnostics and Statistical Manual IV (American Psychiatric Association, 2000). The checklist is divided into two parts, A and B. The six questions of part A are most predictive of ADHD (Hines, King, & Curry, 2012) and are a valid and reliable measure for ADHD symptoms in the general population (Kessler et al., 2007). We therefore administered only part A.

Five possible responses are used to indicate the frequency of a specific behaviour (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Very Often). For the first three questions, behaviour experienced *sometimes, often or very often* is considered concerning and thus highlighted in grey. For the last three questions only behaviour experienced *often or very often* is considered concerning and highlighted in grey. The final score is the sum of scores from the grey boxes, ranging from zero to six. A score of four or higher suggests that the individual has symptoms highly consistent with ADHD in adults (Kessler et al., 2005).

**Table 1**

*Means, Standard Deviations and One-Way Analysis of Variance for Demographic Information, Literacy and General Cognitive Abilities scores for AwDs and Controls.*

| | | AwDs Z-scores | | AwDs Raw scores | | Controls Raw scores | | Comparison | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | SD | Mean | SD | Mean | SD | Value | *p* |
| Gender | Male:female | - | - | 6:21 | | 6:28 | | $\chi^2$= 1.9 | n.s |
| Age | Years | - | - | 20.4 | 2.9 | 19.1 | 1.0 | F= 6.0 | * |
| ADHD | | - | - | 3.8 | 1.7 | 1.7 | 1.5 | $\chi^2$= 4169 | *** |
| **Non-verbal IQ** | | | | | | | | | |
| **PRI** | Scaled score | .1 | .8 | 104.7 | 13.5 | 103.2 | 12.1 | F= .2 | n.s |
| **PSI** | Scaled score | -.7 | .8 | 100 | 13.3 | 111.7 | 13.4 | F= 10.1 | ** |
| **Spatial STM** (Corsi-blocks) | Adj.Memory span | -.3 | 1.1 | 5.4 | 0.9 | 5.8 | 0.8 | F= 2.64 | n.s |
| **Reading** | | | | | | | | | |
| Reading Comprehension | N correct | -.3 | 1.1 | 52.4 | 8.3 | 55.2 | 6.5 | F= 2.17 | n.s |
| Reading Comprehension | Time (min.) | .02 | 1.3 | 16.6 | 7.9 | 16.1 | 5.6 | F=.1 | n.s |
| **Phonological STM** | | | | | | | | | |
| Memory digits | N correct | -1.0 | 1.2 | 14.4 | 3.2 | 17.2 | 2.7 | F= 14.7 | *** |
| Non-word repetition | N correct | -1.6 | 1.3 | 12.6 | 2.4 | 15.4 | 1.9 | F= 26.3 | *** |
| **Phonological awareness** | | | | | | | | | |
| Elision | N correct | -1.7 | 2.5 | 16.2 | 3.5 | 18.6 | 1.1 | F= 14.5 | *** |
| Phoneme reversal | N correct | -1.1 | .8 | 9.2 | 2.1 | 11.9 | 2.7 | F= 18.75 | *** |
| TOWRE nonword reading | N correct | -2.1 | 1.2 | 43.2 | 8.7 | 58.6 | 6.2 | F= 65.3 | *** |
| **Rapid Letter Naming** | Time (sec.) | -1.6 | 1.3 | 31.3 | 7.7 | 23.7 | 3.1 | F= 21.7 | *** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Word Reading** | | | | | | | | | |
| TOWRE sight word | N correct | -1.7 | 1.2 | 84.2 | 8.4 | 97.1 | 4.7 | F= 57.9 | *** |
| TIWRE irregular word | N correct | -1 | 1.4 | 44.3 | 3.0 | 46.6 | 2.2 | F= 11.8 | *** |

*Notes*. This table lists group comparisons of demographic information, general cognitive ability and literacy. For the verbal tasks, participants were compared using their raw score. z-scores relative to control means and SD were used for the group selection, the Mean/SD values for the AwDs are shown in the table. The z-scores for the control group are ~0 and are not displayed.

n.s. = $p > .05$;* = $p \leq .05$;** = $p \leq .01$; ***= $p \leq .001$

Table 1 lists the demographic information and the results of our dyslexia assessments for both participant groups. Z-scores were calculated relative to the mean and SD of controls. For the non-verbal IQ measures, a z-score based on the standardised test score was used, in order to control for the age of the participants. For the verbal tasks, participants were compared using their raw scores, as all participants fit in the oldest age group (17+), so standardising for age was not necessary. Controls were recruited from a university community since the participants with dyslexia were a high functioning group of university students.

The participants were all students at the [Blinded for Review], who had just completed or were in the process of completing their undergraduate degrees. As expected, the AwD group did not differ from the control group in terms of gender, non-verbal IQ based on the Perceptual Reasoning Index of the WAIS-IV, and non-verbal spatial memory (Corsi-block test). They did, however, differ in terms of the Processing Speed Index (SPI), with a lower processing speed for AwDs. The latter was not anticipated, but is consistent with earlier findings for children (e.g., Catts et al., 2011). We will return to this difference below. None of our participants scored below 2 SD on two subtests of the non-verbal tasks. The groups also differed in age. AwDs were, on average, a little more than 1 year older. This difference was mainly due to two participants with dyslexia who were older than the remaining participants (aged 27 and 31). However, since AwDs were matched to the controls in terms of education and IQ and all participants were adult readers the two older participants with dyslexia were kept in the analysis. The two groups differed in terms of reported ADHD symptoms. While the average scores on the ASRS-v1.1 were below what would be considered indicative of ADHD for both groups, AwDs reported ADHD related behaviour more frequently compared to the controls. We therefore checked whether any group differences on the experimental tasks could be explained by differences in ADHD scores. Also, we found that age did not correlate with the dependent variables of our experimental tasks and therefore cannot explain any group differences. Furthermore, as expected, AwDs showed severe impairments on word and non-word reading tests, naming speed, verbal short-term memory and sub-lexical phonology tests. In contrast, they did not differ on the text reading comprehension tests. This is in line with what has been found previously for university students with dyslexia (Finucci et al., 1984).

**Experimental Tasks**

***2. Blocked Cyclic Picture Naming Task***

**Stimuli and Design**. The procedure for the blocked cyclic naming task was taken from Damian et al. (2001). The stimuli were taken from previous studies (Belke, Meyer, et al., 2005; Damian et al., 2001). They consisted of sixteen-line drawings, representing daily objects from four different semantic categories: animals (snake, duck, mouse and fish), clothing (tie, coat, boot and skirt), tools (brush, saw, rake and drill) and furniture (lamp, chair, desk and bed). The pictures were approximately 8 x 8 cm. The pictures are available online in the 'Appendix' section of OSF respiratory (Anonymous, 2020). The visual angle was 13.18 degrees at a viewing distance of 60 cm. In each set pictures were controlled for visual similarity and phonological overlap. We constructed four semantically homogeneous stimulus sets and four heterogeneous stimulus sets. Each semantic category formed a homogeneous stimulus set. Heterogeneous sets selected one picture from each of the four homogeneous sets. There were eight blocks in the experiment, one for each stimulus set. In each block, a stimulus set was repeated eight times in varying orders. This resulted in 128 trials per condition. Homogeneous and heterogeneous blocks were alternated. The eight blocks were counterbalanced according to a Greco-Latin square design.

**Procedure**. Each trial started with a fixation cross presented in the centre of the screen for 800 ms, followed by a 100-ms blank interval and the target picture presented for 250 ms. Participants were instructed to name the picture as fast and as accurately as they could. The maximum response time allowed was 2000 ms from the onset of the picture. An intertrial interval of 1500 ms concluded each trial. Responses were audio-recorded, and response times were measured with a Cedrus SV-1 Voice Key (http://www.cedrus.com/sv1/).

***Hayling Test***

**Stimuli**. The stimuli of the Hayling task were unfinished sentences selected from the Bloom and Fischler (1980) pool of 329 sentences. Based on the norms for sentence completion presented in their study, 34 unfinished sentences were selected, with final words being highly predictable from the sentence context (e.g., “The captain wanted to stay with the sinking_”, automatic answer: “ship”). Since the stimuli of Bloom and Fischler had been tested on American English speakers, we checked that the sentences had a highly predictable final word in British English. To do so, we selected 34 sentences with high prediction rates. We asked 10 monolingual native British undergraduate university students to complete the sentences with the most appropriate word. For the large majority of sentences, participants used the words predicted by Bloom and Fischler (1980), being in complete agreement with each other. For three of the sentences some participants used a British equivalent of the American ending (e.g., “You can’t buy anything for a pound/penny” instead of “cent”). We considered both the American and British responses as correct. Sentences were recorded in a soundproof booth by a native British English speaker. They were recorded as complete sentences in order to sound natural. The last word of each sentence was then cut off. This may have left some co-articulatory cues at the end of the stimuli that might have created minor phonological priming in the automatic condition and a small amount of phonological competition in the inhibition condition, but these effects should have been minor. We used 4 of the sentences for the practice session and the remaining 30 sentences for the experimental tasks. These were presented twice, once in the automatic condition and once in the inhibition condition, resulting in 8 and 60 stimuli, respectively, in the practice and experimental sessions. The full list of sentences is available online in the ‘Appendix’ section of OSF respiratory (Anonymous, 2020).

**Types of Errors.**

Following Burgess and Shallice (1996), the following four categories of responses were treated as errors: inhibition failures, perseveration errors, missing responses, and partial (inhibition) failures. Inhibition failures were responses that used the expected sentence endings instead of unrelated words (e.g. saying “ship” in “The captain wanted to stay with the sinking…“). Perseveration errors were responses that were repetitions of earlier responses in the task. Missing responses were failures to produce a response within the time limit of 5 seconds. Partial failures were erroneously produced words in the inhibition condition that fit the context of the sentence but were not the expected words. In other words, they were semantically related to or antonyms of expected words (i.e. “submarine” for “The captain wanted to stay with the sinking...“). Note that we adopted the term “*partial failure*” to capture semantic errors and antonyms, as defined by Burgess and Shallice (1996). For interrater reliability, all responses in the inhibition condition were coded for partial failures by two coders. The coders agreed on 90.5% of the 1562 responses. Disagreements were resolved by a third coder who independently rated the responses. Responses that were coded by two of the three coders as partial failures were accepted as such.

**Procedure.** We adapted the original version of the Hayling task by Burgess and Shallice (1996), whose test-retest reliability is quite high (.76; Burgess & Shallice, 1997). In the original task, automatic and inhibition sentences were presented in a blocked fashion. We used a mixed design, where both sentence types were presented in each block. This design minimizes the use of strategies to boost performance. Stimuli were presented in two blocks. In the first block, half of the 30 sentences were randomly presented as part of the automatic condition and the other half as part of the inhibition condition. The second block included the same sentences, but in the opposite condition and in a different randomised order from the first block. The block order was counterbalanced across participants. Participants listened to the sentences one by one. After each sentence, a cue appeared on the screen (4.25 x 4.25 inches). The use of a cue was taken from the study by Bélanger and Belleville (2009), and was not presented in Burgess and Shallice’ (1996) original task. Depending on the cue shape, participants had to either complete the sentence with the most appropriate word (i.e., automatic condition) or with a totally unrelated word (i.e., inhibition condition). Following Bélanger and Belleville (2009), a blue circle was presented for the automatic condition and a red octagon for the inhibition condition. Each trial started with a fixation cross in the centre of the screen, which was presented for 1500 ms and subsequently remained on the screen during the auditory presentation of the sentence. Fifty milliseconds after the last uttered word, the condition-associated cue replaced the fixation cross. The cue remained on the screen until the participant's verbal response triggered the voice key or until the end of the maximum response period of 5000 ms (see also Bélanger & Belleville, 2009). Participants were instructed to listen carefully to each sentence and to respond as quickly as they could. Both reaction times (RT) and response errors were measured.

**Data Analysis**

For both tasks, response times were analysed with linear mixed effect models employing the lmer function of the *lme4* package (Bates et al., 2015) using R version 3.5.1 (R Core Team, 2018). Accuracy measures were analysed using generalized linear mixed models (*GLMM*) with a binomial link function.

To obtain p-values and degrees of freedom, the "Satterthwaite" approximation from the lmerTest package was used (Kuznetsova et al., 2016). The advantage of this statistical approach is that it takes into account random effects and random slopes for items and subjects. We explored a set of random effects structures including both complex (random intercepts and random slopes) and simpler structures (random intercept for subject). When we explored the random effect for one dimension (e.g. subjects), the random effect for the other dimension (items) and the fixed effects were held constant. The best fitting model for the set of random effects that were explored was identified using the Akaike's Information Criterion (AIC; (Sakamoto et al., 1986; see also Seedorff et al., 2019 PREPRINT). If none of the models with random effect structures converged, the *optimx* package (Nash & Varadhan, 2011) was used as a more sensitive optimizer and then the same procedure for random effect comparison was applied.

Models exploring fixed effects of Condition (Heterogeneous x Homogeneous), Group (Controls x AwDs) and Sentence presentation order (first x second) used dummy coding (0 x 1). The continuous factor Cycle (cycle 2 to 8) was centred using the *scale* function from the *stats* package (R CoreTeam,2018). We assessed main effects and interactions by comparing models with and without critical effects (e.g. a two-way interaction would be assessed by comparing a model with the two-way interaction and both main effects to a model with only the main effects). When interactions were significant, further effects were assessed in separate subsets of the data. Whenever we found group differences, we followed up the analysis with a model that included ADHD (score 0 to 6) as an additional fixed factor, that was centred using the *scale* function. Similarly, whenever we found group differences in reaction times, we followed up the analysis with a model that included processing speed (SPI) as an additional fixed factor. For model comparison we used a maximum likelihood ratio test (Bates et al., 2015). This approach was used for both tasks. For response time analyses, only accurate responses were included. In addition, all responses that were slower than 2.5 SDs from participant means and responses faster than 200ms were considered outliers and removed. The packages *reshape2* (Wickham, 2007) and *tidyverse* (Wickham et al., 2019) were used for data cleaning, processing and visualisation. The percentage of errors and outliers are reported with each task. Due to technical problems with the voice key, one participant from the control group had to be excluded from all of the analyses (removing the participant did not change the group comparison in terms of language assessment; see participant section). When this participant was removed the final sample included 33 control participants and 27 AwDs. The raw data for both experiments is available online in OSF respiratory(Anonymous, 2020).

## Results

**Blocked Cyclic Picture Naming Task**

Since previous studies have demonstrated that the blocking effect in the blocked cyclic naming task starts to emerge from the second cycle of a presentation block (Belke, Meyer, et al., 2005), the first cycle of each block was excluded from the analysis. This also reduces the effect that phonology could have on performance since participants gain initial access to names in the first cycle.

***3. Accuracy***

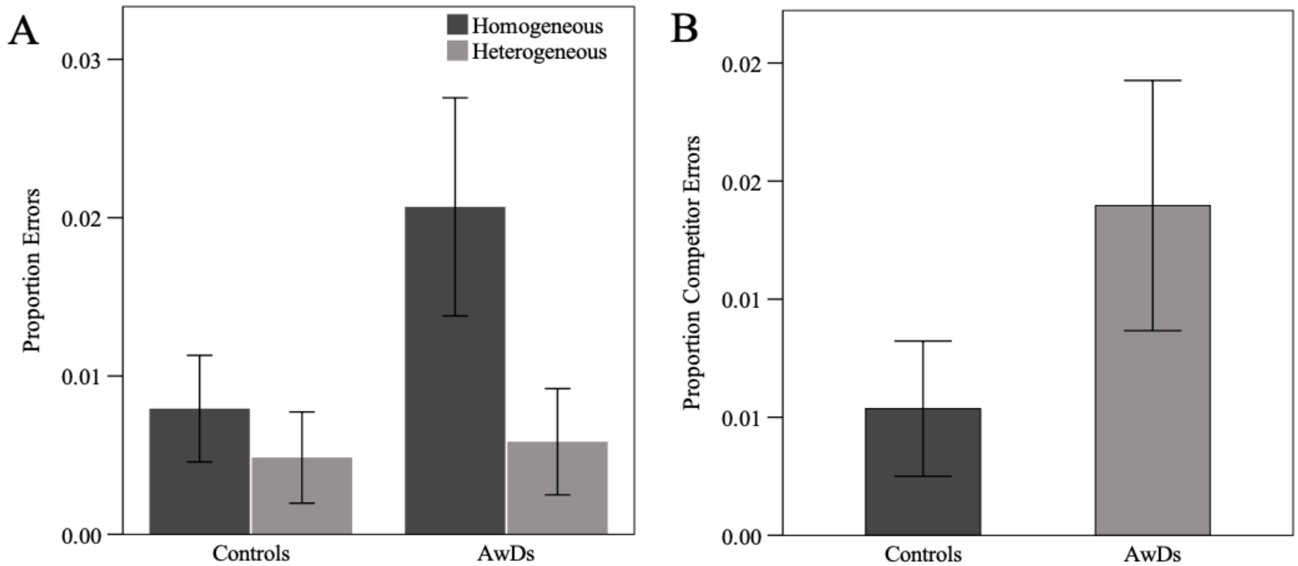
The mean proportion of errors in both heterogeneous and homogeneous conditions for both AwDs and control participants is displayed in Figure 1. Trials where participants failed to pronounce the correct picture name were classed as errors. Types of errors included hesitations (e.g. “uhh” or “aa”), competitor errors (self-corrections = partial repetitions, e.g. “fffi-duck” where fish and duck are in the same block, and full names of other pictures from the same experimental block), stutters, and incorrect object names (e.g. ‘jacket’ instead of ‘coat’). We fit a *GLMM* with accuracy (correct responses vs incorrect responses) as the dependent variable (DV) and Condition (homogeneous versus heterogeneous) and Group (Control versus AwDs) as fixed factors. The best random structure included random intercepts for subjects and items [(1|Subject) + (1|Picture)].

In accordance with previous research, we found a significant main effect of Condition ($G^2(1)$ =25.56, $p$ < .001), with participants making more errors in the homogeneous condition than in the heterogeneous condition. In addition, the effect of Group was significant ($G^2(1)$ =9.09, $p$= .003). AwDs made more errors than controls. Importantly, there was a significant interaction between Group and Condition ($G^2(1)$ =3.98, $p$= .046). We therefore investigated the effect of Group in the homogeneous and heterogeneous conditions separately. We found no significant group difference in the heterogeneous condition ($G^2(1)$ =0.21, $p$=.643), but a significant difference in the homogeneous condition ($G^2(1)$ =12.2, $p$ < .001), where AwDs made more errors than controls.

To investigate if ADHD behaviour in AwDs might explain the group difference, we added ADHD scores as an additional fixed factor to the model. The best random structure included a random intercept for subject and a random intercept for items that is specific to each condition and group, but where only the co-variances are assumed constant [(1|Subject) + (1|Group:Condition:Picture)]. We found a significant Group x Condition x ADHD interaction (G2(1) =4.5, p=.034), no Group x ADHD interaction (G2(1) =1.53, p= .217) and no Condition x ADHD interaction (G2(1) =1.26, p= .262). As for the analysis without the ADHD factor, the post-hoc analysis for the heterogeneous condition showed no main effect of Group ($G^2(1)$ =0.04, p= .846). But it showed a significant Group x ADHD interaction (G2(1) =5.03, p= .025), with a trend for higher ADHD leading to increased errors in controls ($G^2(1)$ =3.60, p= .058), but not in AwDs (G2(1) =1.98, p= .158). The homogeneous condition did not show any Group x ADHD interaction (G2(1) =.01, p= .922), but an effect of ADHD (G2(1) =5.59, p= .018). In comparison to the model without the ADHD factor, the effect of Group was only a trend (G2(1) =11.43, p= .082). Thus, ADHD can explain part of the group difference on overall errors. In addition, we see that ADHD has a general effect on participants' errors, particularly in AwDs, but also in controls, at least in the homogeneous condition.

**Figure 1**

*Proportion of Errors in Blocked Cyclic Picture Naming Task*




*Note.* Mean proportion of overall errors in the blocked cyclic picture naming task (A) and mean proportion of competitor errors only in the homogeneous condition (B) in both groups of participants. Error bars represent 95% confidence intervals.

Overall errors include all kinds of errors (e.g., stutters, hesitations), which might not all be due to a problem of inhibiting competitor words. Instead, the occurrence of competitor errors is a sign that suppression of highly activated competitor words has failed (Belke, 2013; Howard et al., 2006; see also Oppenheim et al., 2010; Oppenheim & Balatsou,2019 for a non-competitive account). We thus investigated whether competitor errors were made more frequently by participants with dyslexia than controls. The audio recordings of three participants, one from the dyslexia group and two from the control group, were corrupted and not usable for the more precise error analysis. Therefore, these three participants were excluded from this particular analysis, reducing the sample to 26 AwDs and 31 controls. Participants made very few errors in the heterogeneous condition overall (N= 34) and even fewer competitor errors (N=22). Given the small number of errors in this condition, we did not analyse these further. In contrast, in the homogeneous condition (see Figure 1 right panel), participants made mostly competitor errors (70%), with higher numbers in the AwD group (N=41) compared to controls (N=18). While these errors were also relatively rare, they occurred frequently enough to analyse. We performed a *GLMM* to test for a group difference. We fit a model with Accuracy (competitor errors versus all other responses) as the Dependent Variable (DV) and Group (Control versus AwDs) as a fixed factor. The random structure included random intercepts for subjects and items [(1|Subject) + (1|Picture)]. We found a significant effect of Group ($G^2(1)$ =9.18, $p$ = .002), confirming that AwDs made more competitor errors than controls in the homogeneous condition.

Adding ADHD scores as a fixed factor to the model to check whether the group difference might be due to ADHD symptoms, confirmed the group effect. The best random structure included a random intercept for subject and items [(1|Subject) + (1|Picture)]. We found no significant Group x ADHD interaction ($G^2(1)$ =1.4, p=.221). The main effect of ADHD was not significant either ($G^2(1)$ =0.8, p=.386). But as in the analysis without ADHD, the main effect of Group was significant ($G^2(1)$ =9.2, p=.002). Thus, while overall errors were affected rather by ADHD than dyslexia, competitor errors which are a sign of competitor suppression, differed between groups, with more competitor errors for AwDs than controls.

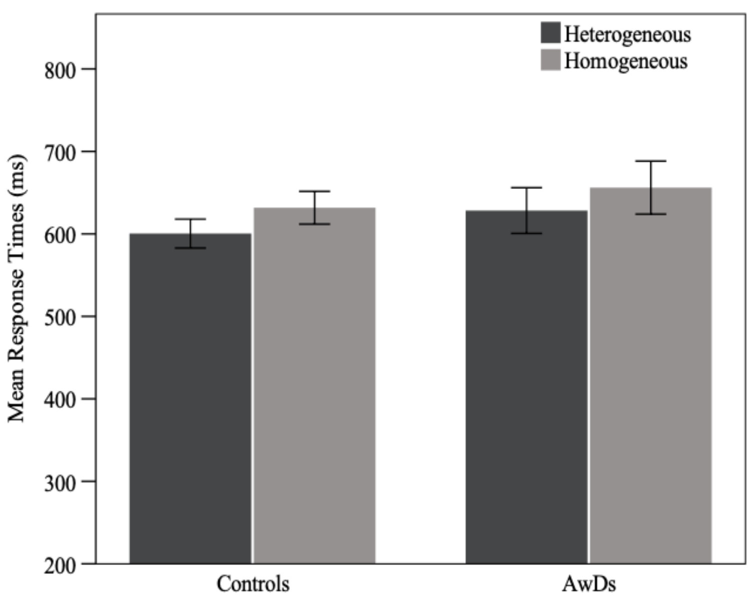
***4. Response Times***

We fit a mixed-effects model with response time as the dependent variable, and Group (Control versus AwDs), Condition (homogeneous versus heterogeneous) and their interaction as fixed factors. The best random structure included a random intercept for subjects and a random intercept for items that is specific to each condition, but where only the variance is modelled and the co-variances are assumed constant [(1|Subject) + (1|Condition:Picture)]; i.e. random slopes, but with constant covariance.

As indicated, response times slower than 2.5 SDs from participant means and faster than 200ms were considered outliers and eliminated. These accounted for 2.7% of the total responses. The results are displayed in Figure 2. We found no significant Group by Condition interaction ($G^2(1) = 0.50$, $p = .479$) and no Group effect ($G^2(1) = 2.57$, $p = .109$). Only the effect of Condition was significant ($G^2(1) = 9.22$, $p = .002$), with the homogeneous condition leading to longer response times than the heterogeneous condition. The competitive semantic context slowed down both groups to a similar degree.

**Figure 2**

*Response Times in Blocked Cyclic Picture Naming Task*




*Note.* Mean response times for controls and AwDs in the blocked cyclic naming task. Error bars represent 95% confidence intervals.

***Accuracy and Response Times Across Cycles***

Previous studies have suggested that additional top-down mechanisms might play a role in the blocked cyclic naming task (Belke & Stielow, 2013; Schnur et al., 2006). Such top-down mechanisms mean that the semantic interference effect is not cumulative, that is it remains pretty stable after the first cycle. In order to investigate if these top-down mechanisms are affected in AwDs, we added the factor *cycle* (7 levels) to both accuracy and response time models. If top-down mechanisms were contributing to the AwDs performance, then a (larger) cumulative semantic interference effect for AwDs was expected, leading to a significant Group x Condition x Cycle interaction in accuracy and/or response times (Belke & Stielow, 2013). The best random structure included a random intercept for subject and a random intercept for items that is specific to each condition and group, but where only the co-variances are assumed constant [(1|Subject) + (1|Group:Condition:Picture)]. We found no significant three-way interaction ($G^2(1)$ =0.06, $p$=.804). There was no evidence of a growing cumulative effect in either group of participants (Condition X Cycle interaction $G^2(1)$ = 0.03, p=.865).

For response times, the best random structure included a random intercept for subjects and random intercept for items that is specific to each condition, but where only the variance is modelled and the co-variances are assumed constant [(1|Subject) + (1|Condition:Picture)]. Similar to the accuracy results, we found no significant three-way interaction ($G^2(1)$ =0.08, $p$=.779) and no evidence across groups of a growing cumulative effect (Condition X Cycle interaction, $G^2(1)$ = 2.3, p=.128). Thus, we found no evidence that interference builds up across cycles in the AwDs group.

**Hayling Task**

The automatic and inhibition conditions of the Hayling task involve very different processes of response initiation and suppression (e.g., De Zubicaray et al., 2000), so we analysed error patterns and response times in the two conditions separately. We removed one participant from the dyslexia group because they used a strategy in the inhibition condition (i.e. naming only body parts as a response) which led to unusually accurate and fast responses. Thus, the final sample included 26 AwDs and 33 control participants.

***5. Errors***

**Automatic Condition.** In the automatic condition, all words that completed the sentences in the expected way (see Bloom and Fischler, 1980, and our pre-test) were counted as correct responses. All other responses were counted as errors. The number of errors in the automatic condition was too small for a meaningful group comparison (9 errors in the dyslexia group and 5 errors in the control group, together accounting for 0.8% of the responses in the automatic condition).

**Inhibition Condition**.

In total, errors accounted for 15.8% of the data in the inhibition condition. In order to test group differences in errors in the inhibition condition we fit a GLMM model (glmer) with errors (all types of errors vs correct responses) as the dependent variable, and Group (Control vs AwDs) as a fixed factor. The best random structure included random intercepts for subjects and items [(1|Subject) + (1|Sentence)]. We found a significant main effect of Group ($G^2(1) = 9.83$, $p = .002$). AwDs made more errors than controls (Figure 3A).

Adding ADHD as a fixed factor to the model confirmed the group difference. The best random structure included random intercepts for subjects and items [(1|Subject) + (1|Sentence)].

Lexical Competition in AwDs

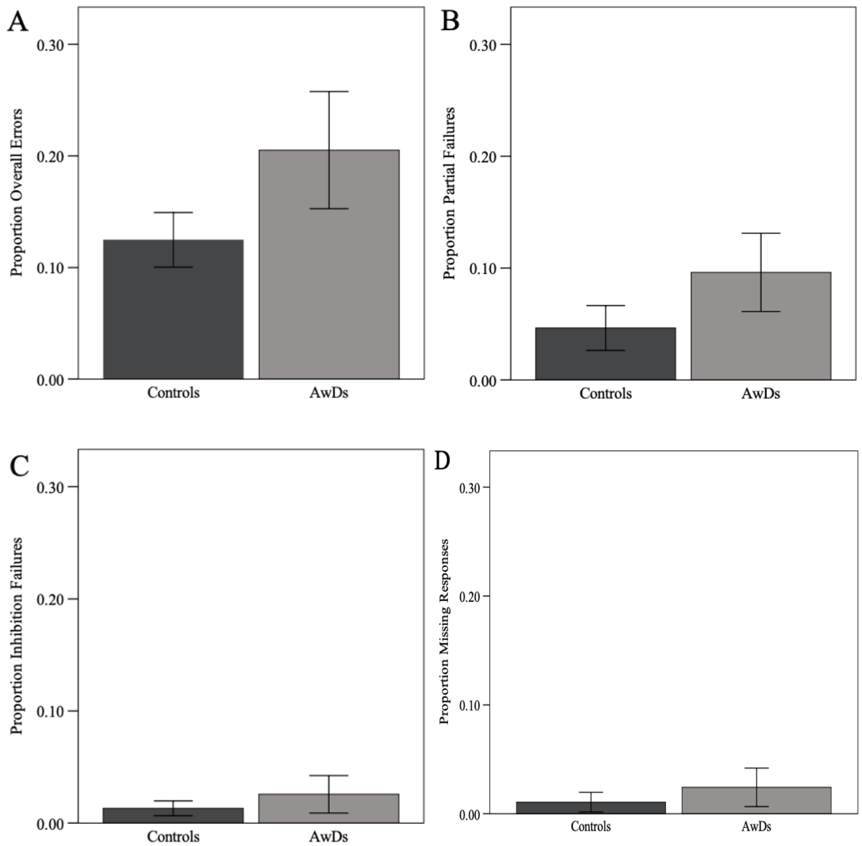
There was no significant Group x ADHD interaction ($G^2(1)$ =0.15, p = .695), no main effect of ADHD ($G^2(1)$ =0.04, p = .852), but, as in the analysis without ADHD, there was a main effect of Group ($G^2(1)$ =9.83, p = .002).

The group difference in overall errors might have been caused by a difference in missing responses. These could represent a wider range of causes, not directly linked to lexical competition. For example, missing responses could stem from problems in the generation of alternative responses from a large array of alternatives. We fitted a model with errors as dependent variable (missing responses versus all other responses) and Group as a fixed factor. The best random structure included only a random intercept for Subjects [(1|Subject)]. There was no significant main effect of Group ($G^2(1)$ =2.13, p=.143), showing no difference in the number of missing responses between the two groups of participants. The group difference in overall errors can, therefore, not have been caused by a larger number of missing responses for AwDs. Figure 3D shows the proportion of missing responses in the inhibition condition.

In contrast, AwDs might have produced an increased number of overall errors in the inhibition condition due to a difficulty suppressing highly activated responses. These might have been expected words (= inhibition failures) or related words that created an appropriate sentence ending (= partial failures). We investigated inhibition and partial errors separately.

**Figure 3**

*Proportion of Errors in Hayling Task*




*Note.* Mean proportion of overall errors (A), partial failures (B), inhibition failures (C) and missing responses (D) in the inhibition condition of the Hayling task. Error bars represent 95% confidence intervals.

**Inhibition Failures.** Figure 3C shows the proportion of inhibition failures in the Inhibition condition. We fitted a model with inhibition failures (versus all other responses) as the dependent variable and Group as a fixed factor. The best random structure included random intercepts for subjects [(1|Subject)]. We found no significant main effect of Group ($G^2(1)$ =1.88, $p$ =.169). Thus, AwDs did not significantly differ from the controls in suppressing prepotent responses.

**Partial Failures.** To investigate if AwDs experienced more partial failures (Figure 3B), we fit a model with partial failures (versus all other responses) as the dependent variable and

Group as the fixed factor. The best random structure included random intercepts for subjects and items [ (1|Subject)+ (1|Sentence)]. We found a significant main effect of Group ($G^2(1) = 7.93$, $p = .005$), with AwDs producing more partial failures than controls.

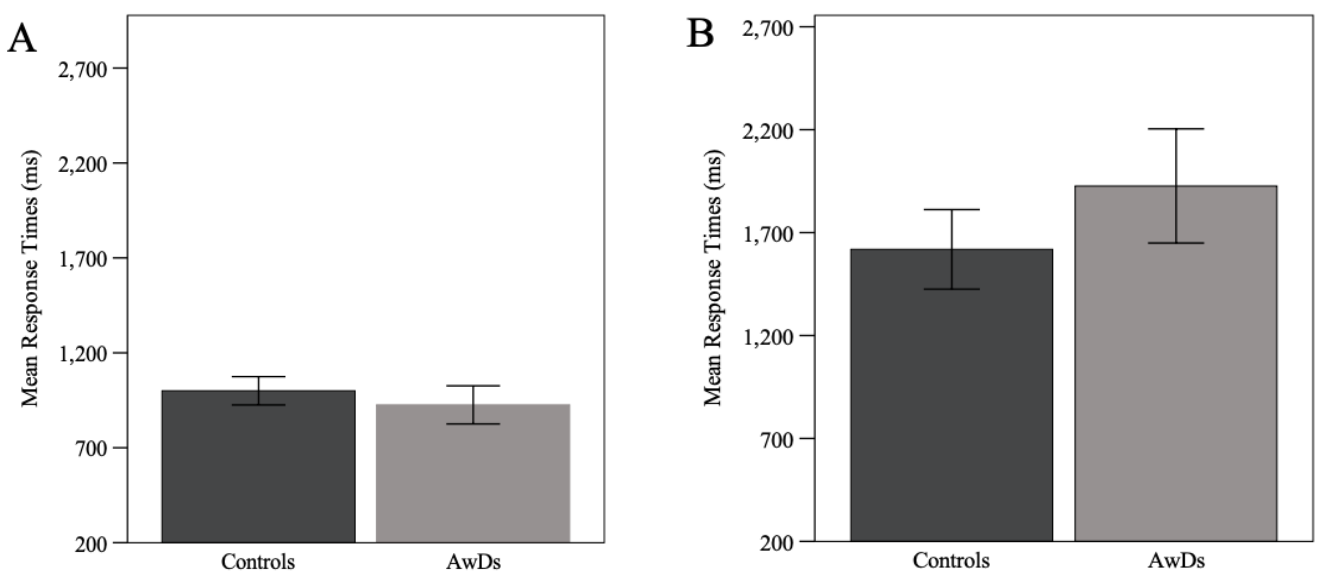
Adding ADHD as a fixed factor to the model confirmed the group difference. The best random structure included random intercepts for subjects and items [ (1|Subject)+ (1|Sentence)]. There was no significant Group x ADHD interaction ($G^2(1) =0.99$, p = .319), no main effect of ADHD ($G^2(1) =0.01$, p = .932), but, as in the analysis without ADHD, there was a main effect of Group ($G^2(1) =7.93$, p = .005).

### *6. Response Times*

As indicated above, response times faster than 200 ms and slower than 2.5 SDs from participant means were considered outliers and excluded from the analysis. Outliers accounted for 2.2% of the total responses. In addition, all errors (missing responses, perseveration errors, inhibition failures and partial failures) were excluded from the RT analysis.

**Figure 4**

*Response Times in Hayling Task*



*Note.* Mean RTs for the automatic condition (A) and the inhibition condition (B) of the Hayling Task. Error bars represent 95% confidence intervals.

Figure 4 shows the response times in both the automatic and inhibition conditions of the Hayling task. The best random effect structure for both of the response time analyses included a random intercept for subjects and items [(1|Subject) + (1|Sentence)]. For the automatic condition, there was no effect of Group ($G^2(1)$ =1.38, *p*= .239). Thus, AwDs performed similarly to controls. For the inhibition condition, instead, there was a marginal main effect of Group ($G^2(1)$ =3.7, *p*= .055), with AwDs being slower than the controls. This is in line with the overall accuracy and partial inhibition analyses above, where AwDs experienced more difficulty with the inhibition condition compared to controls and it rules out a speed/accuracy trade-off as an explanation of the accuracy results.

Adding ADHD as a fixed factor to the model confirmed the group difference. The best random structure included a random intercept for subjects and items [(1|Subject) + (1|Sentence)]. There was no significant Group x ADHD interaction ($G^2(1)$ =1.46, p = .227), no main effect of ADHD ($G^2(1)$ =0.16, p = .693), but, as in the analysis without ADHD, there was a marginal effect of Group ($G^2(1)$ =3.67, p = .055).

Since the two participant groups also differed in terms of processing speed (SPI of the WAIS), we conducted an additional analysis with SPI as a fixed factor in the model. The best random structure still included random intercepts for subjects and items [(1|Subject)+ (1|Sentence)]. There was no significant Group x Processing Speed interaction ($G^2(1)$ =1.0, *p*= .879), neither was there an effect of Processing Speed ($G^2(1)$ =0.16, *p*= .691). The effect of Group was still marginal ($G^2(1)$ =3.67, *p*= .055).

***Sentence Presentation Order***

We also investigated whether AwDs might have found it more challenging than controls to suppress a response that they had given to the same sentence before. For that we investigated, first, the effect of presentation order on overall number of errors in the inhibition condition by adding Sentence Presentation Order (first versus second) as a fixed factor to our *GLMM* model, including the Sentence Presentation Order by Group interaction. We found no significant Group by Sentence Presentation Order interaction ($G^2(1) = 1.27$, $p = .259$), so the presentation order of the sentences did not affect the errors of the groups differently. The main effect of Sentence Presentation Order was also not significant ($G^2(1) = 0.5$, $p = .465$). A lack of an interaction was also observed for inhibition failures ($G^2(1) = 1.77$, $p = .183$), partial failures ($G^2(1) = 0.01$, $p = .912$) and missing responses ($G^2(1) = 0.21$, $p = .644$). A main effect of Sentence Presentation Order was only significant for the partial failures ($G^2(1) = 5.0$, $p = .026$), where all participants experienced more errors when they saw the sentence for the first time in the inhibition condition. For the inhibition failures ($G^2(1) = 2.5$, $p = .115$) and missing responses ($G^2(1) = 0.2$, $p = .690$) the main effect of Sentence Presentation Order was not significant.

We then investigated the effect of presentation order on response times. There was no Group by Sentence Presentation Order interaction in the automatic condition ($G^2(1) = 0.8$, $p = .772$) and in the inhibition condition ($G^2(1) = 0.002$, $p = .961$). The main effect of Sentence Presentation Order was also not significant in either automatic ($G^2(1) = 0.01$, $p = .942$) or inhibition conditions ($G^2(1) = 0.2$, $p = .637$). Thus, presentation order affected response times of the two participant groups in a very similar way and AwDs did not find it more challenging than controls to suppress a previous response to a particular sentence.

In sum, we have found that AwDs showed significantly higher error rates than control participants in both experimental tasks. In the blocked cyclic naming task, AwDs made significantly more overall errors in the homogeneous condition, driven by a higher number of competitor errors. In the Hayling task, they made more overall errors in the inhibition condition, driven by an increase in partial failures. This was reinforced by a trend for AwDs to be slower than controls in the inhibition condition.

**Associations Between Tasks**

We found that AwDs made more errors in conditions with higher lexical competition in both the Hayling and blocked cyclic naming tasks. Next, we asked whether these deficits were caused by the same underlying mechanisms or not. We performed Pearson correlation analyses of the measures that showed group differences in our tasks, correlating proportions of overall errors, partial failures and response times in the inhibition condition of the Hayling task with the proportion of competitor errors in the homogeneous condition of the blocked cyclic naming task. Some of our error variables were positively skewed. We therefore applied an arcsine transformation.

AwDs and control participants might engage different mechanisms to perform the tasks. We therefore analysed the participants not only as a whole but also as two separate groups. Table 2 shows the results of the correlation analyses for the whole group, while Table 3 shows the results split by subgroup.

**Table 2**

*Correlations of proportion of competitor errors in the homogeneous condition of the Blocked Cyclic Picture Naming Task with measures from the inhibition condition of the Hayling Task for all participants.*

| | Proportion of overall errors | | Partial failures | | RT | |
|---|---|---|---|---|---|---|
| | **r** | *p* | **r** | *p* | **r** | *p* |
| Competitor errors homogeneous condition *(N=56)* | .14 | *.307* | .09 | *.531* | -.13 | *.347* |

*Note.* The number in brackets represents the number of participants entered into the correlations.

**Table 3**

*Correlations of proportion of competitor errors in the homogeneous condition of the Blocked Cyclic Picture Naming Task with measures from the inhibition condition of the Hayling Task, split by participant group.*

| | | Proportion of overall errors | | Partial failures | | RT | |
|---|---|---|---|---|---|---|---|
| | | ***r*** | *p* | ***r*** | *p* | ***r*** | *p* |
| *Controls* | | | | | | | |
| | Competitor errors homogeneous condition (N=31) | -.24 | *.194* | -.16 | *.398* | -.27 | *.150* |
| *AwDs* | | | | | | | |
| | Competitor errors homogeneous condition (N=25) | .08 | *.711* | -.01 | *.982* | -.22 | *.284* |

*Note.* The number in brackets represents the number of participants entered into the correlations.

There was no evidence of a relationship between tasks, either for the group as a whole or when splitting the results by group. Thus, we did not find any evidence for a common underlying origin of the competitor errors across the two tasks.

**Dyslexia Severity and Task Related Deficits**

Next, we investigated whether increased semantic competition in our tasks might be related to reading skills and therefore to the severity of dyslexia in AwDs. We used Principal Component Analysis (PCA) to create theoretically determined factors based on the functions probed by our language tests. Following previous research (e.g., Deacon et al., 2012), assessment subtests were grouped into the theoretically defined factors *lexical skills*, *phonological short-term memory* and *phonological manipulation*. Sight word reading and irregular word reading require visual/lexical processing. These were grouped into the principle component *lexical skills.*[2] In the second PCA analysis, we extracted a factor reflecting *phonological STM* from non-word repetition and memory for digits. In the third analysis, we extracted a score for *phonological manipulation* from elision, phoneme reversal and non-word reading tests. The three analyses were performed on all participants together and we extracted a single factor in each analysis. We used the factor loadings derived via Barlett's method. The lexical skills factor accounted for 73% of the variance with loadings from sight word reading (.86) and irregular word reading (.86). The phonological STM factor accounted for 73% of the variance with loadings from memory for digits (.86) and non-word repetition (.86). Finally, the phonological

[2] The lexical retrieval mechanisms involved in rapid letter naming and the Gray Silent Reading Test are debatable, so the two subtests are not included in the factor selection (see Indefrey & Levelt, 2000). However, the direction and the significance of the correlations are preserved if the two tests are added to the lexical skills factor.

manipulation factor accounted for 65% of the variance with factor loadings from phoneme reversal (.83), elision (.74) and non-word reading (.87).

**Table 4**

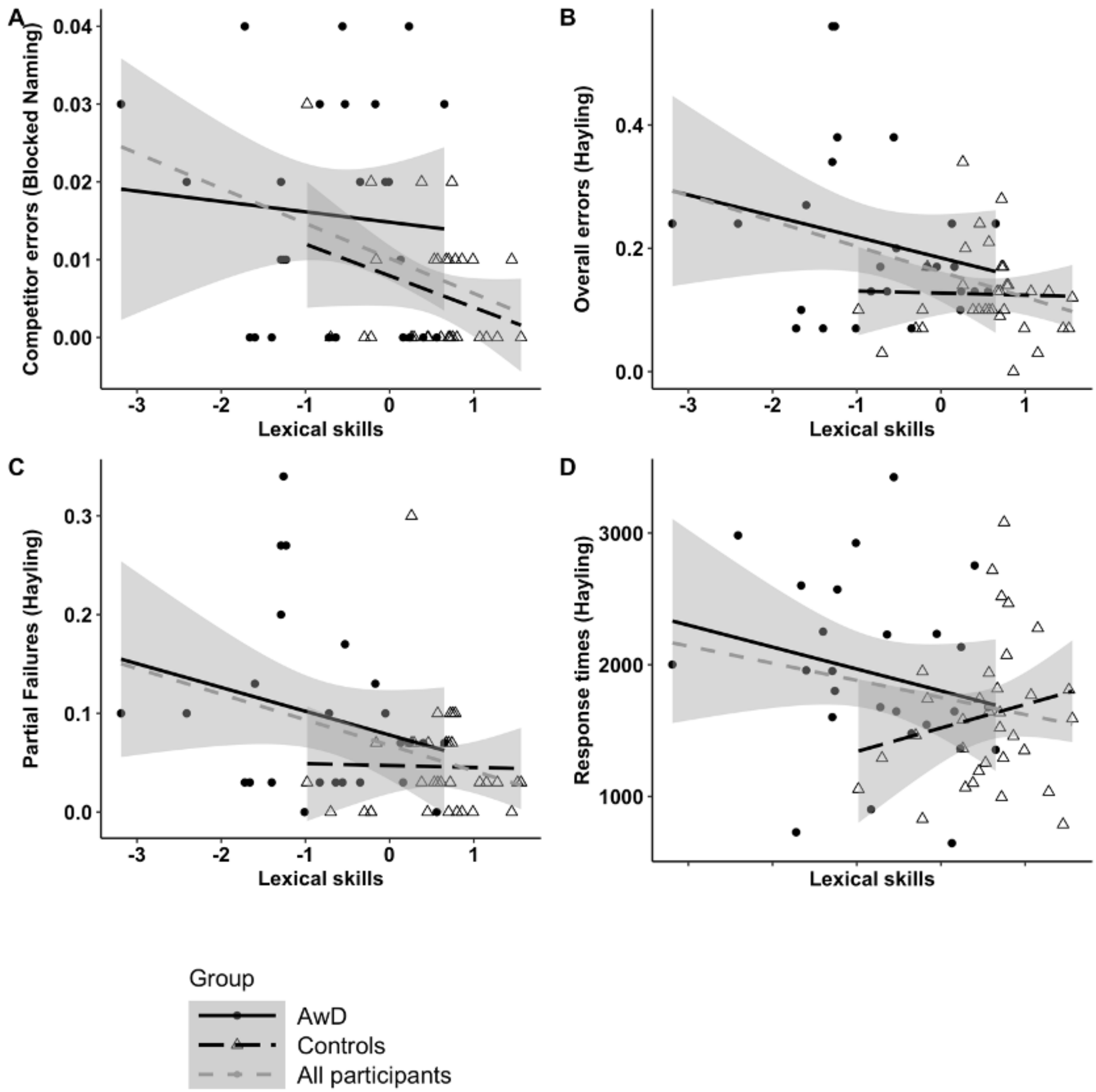
*Correlations Between Measures with Group Differences in the Experimental Tasks and the PCA Factors Lexical Skills, Phonological STM and Phonology Manipulation for all Participants.*

| | | Lexical skills | | Phonological STM | | Phonology manipulation | |
|---|---|---|---|---|---|---|---|
| | | **r** | *p* | **r** | *p* | **r** | *p* |
| *Blocked cyclic naming* | | | | | | | |
| | Competitor errors in the homogeneous condition *(N=57)* | **-.34** | ***.009*** | -.22 | *.095* | **-.27** | ***.043*** |
| *Hayling task (inhibition condition)* | | | | | | | |
| | Proportion of overall errors *(N=59)* | **-.36** | ***.005*** | -.19 | *.139* | **-.46** | ***<.001*** |
| | Partial failures *(N=59)* | **-.35** | ***.007*** | -.24 | *.073* | **-.32** | ***.013*** |
| | RT *(N=59)* | -.21 | *.115* | ***-.40*** | ***.002*** | -.21 | *.105* |

*Note.* The number in brackets represents the number of participants entered into the correlations. Significant correlations are highlighted in bold.

**Figure 5**

*Correlations between lexical competition measures and the factor lexical skills*




*Note.* Blocked cyclic naming: competitor errors (A), Hayling: overall errors (B), partial failures (C) and RTs (D). The correlations are shown for each of the two groups and all participants together. Shaded areas represent 95% pointwise confidence bands.

Lexical Competition in AwDs

Figure 5 and Table 4 show that all lexical competition measures, apart from RTs in the inhibition condition of the Hayling task, significantly correlated with the factor lexical skills for the participants as a whole, suggesting a link between poorer lexical skills and increased lexical competition. Figure 5 also suggests that these relationships were present in both participant groups for the measure of the blocked naming task (panels A), while they appear to be present only for AwDs for the Hayling task (panels B, C).

In addition to correlations with lexical skills, there were relationships with the phonological factors for the participant group as a whole (see Table 4), suggesting an involvement of phonology processing during task performance. Since the factors were related with each other (lexical skills and phonological STM: $r=.56$, $p<.001$; lexical skills and phonological manipulation: $r=.63$, $p<.001$), we tested whether lexical skills independently explained variance in our tasks. We therefore carried out a series of regression analyses, entering the phonological measure(s) found to significantly contribute to each experimental measure in the first step and then lexical skills. Lexical skills did not independently predict performance in the tasks (see Table 5).

**Table 5**

*Regression models showing the contribution of the lexical skills factor to the lexical-competition measures of the Blocked Cycling Picture Naming and Hayling Tasks, after controlling for phonological processing.*

| | | Phonological predictors only | | Phonological predictors + lexical skills | | Overall variance |
|---|---|---|---|---|---|---|
| | Phonological predictor(s) | $R^2$ change | *p* | $R^2$ change | *p* | $R^2$ |
| *Blocked cyclic naming* | | | | | | |
| Competitor errors in the homogeneous condition *(N=57)* | Phonological manipulation | **.07** | *.043* | .05 | *.110* | .12 |
| *Hayling task (inhibition condition)* | | | | | | |
| Proportion of overall errors *(N=59)* | Phonological manipulation | **.21** | *<.001* | .010 | *.400* | .22 |
| Partial failures *(N=59)* | Phonological manipulation | **.11** | *.013* | .03 | *.144* | .14 |
| RT *(N=59)* | Phonological STM | **.16** | *.002* | .001 | *.840* | .13 |

*Note:* The number in brackets represent the number of participants used for the correlations. The significant correlations are highlighted in bold.

While the collinearity of phonological and lexical measures makes it impossible to isolate the contribution of each factor, relationships with phonological factors suggest task requirements could include both phonological and lexical contributions. It is possible that these additional task requirements contributed to the lack of relationship between the two experimental tasks.

## Discussion

The aim of the present study was to explore whether AwDs have lexical-semantic retrieval deficits that involve the resolution of lexical competition, especially in conditions where lexical-semantic competition is increased. We used the blocked cyclic picture naming paradigm and the Hayling sentence completion task. The Hayling task also allowed us to look at task-related suppression of prepotent responses.

For blocked cycling picture naming, we predicted that if AwDs have a deficit in resolving lexical competition, they would have a larger semantic effect in response times and/or errors. The results for semantic competitor errors were as predicted, also when taking into account potential effects of comorbid ADHD in some AwDs. The results for overall errors were less clear, as the difference in overall errors was only a trend once reported ADHD symptoms were taken into account. But overall errors did not just contain competitor errors, but also errors such as stutters, which are not as clearly caused by a suppression deficit than semantic competitor errors. Furthermore, the predicted effect was not present in AwDs' RTs, but there was also no evidence of a speed accuracy trade-off that would influence interpretation of the accuracy result.

For the Hayling task, we predicted that AwDs would be slower and/or make more errors in the inhibition condition compared to controls, but there would be no difference in the automatic condition. In the inhibition condition, AwDs were indeed less accurate overall and also made more partial errors, i.e. semantically related responses. Furthermore, we found a trend for AwDs to be slower than controls in the inhibition condition, but not in the automatic condition. ADHD symptoms in the AwD group did not explain these differences. Importantly, they are in line with our results from the blocked cyclic naming task and suggest increased lexical-semantic competition in AwDs.

The latter conclusion might seem questionable given that subsequent correlation analyses did not show evidence of a relationship between the two tasks. However, the resolution of lexical-semantic competition is clearly only one component of the tasks, evidenced, for instance, by different relationships between task performance and phonological skills. One might, therefore, need a much larger participant sample to understand the nature of the relationships between the tasks and the contributions of factors that are not related to lexical-semantic competition.

We also asked if the lexical-competition deficits experienced by AwDs were related to their reading deficits. If so, we predicted a relationship between lexical skills and increased competition in the lexicon. Consistent with this, we found relationships between all experimental lexical competition measures with AwDs deficits and a lexical skills factor extracted from our dyslexia assessments. Most of these relationships were significant. While we found highly significant correlations of the lexical skills factor with all lexical competition measures, correlations between the lexical skills factor and phonological factors did not allow us to establish whether lexical skills or phonological skills determined performance in our tasks.

In what follows, we will first discuss how our results support the hypothesis of a more competitive lexicon in AwDs and possible underlying processes. We will then discuss how our results can be accounted for by different speech production models. Finally, we will discuss alternative interpretations of the results and evaluate the link to AwDs' reading deficits.

**Adults with Dyslexia Have a More Competitive Lexicon**

Our results in the blocked cyclic naming experiment seem not to be in line with previous research reporting slower RTs for participants with dyslexia when naming pictures (Katz, 1986; Swan & Goswami, 1997). This is probably due, however, to our participants naming the same pictures repeatedly, which reduces the influence of initial access to the picture names and initial phonological encoding. The larger proportion of errors in our dyslexia sample, even if small in terms of numbers, is in line with previous studies of CwDs (e.g., Nation et al., 2001; Swan & Goswami, 1997).

The predominant view of picture naming errors in individuals with dyslexia has been that this originates from an underlying phonological deficit (Nation, 2005). For example, Swan and Goswami (1997) found that CwDs experienced a large proportion of phonological errors only for polysyllabic picture names and when more complex phonological representations were required. Our pictures for the naming task all had monosyllabic names and occurred in both conditions, minimizing item-specific influence on differences. Also, the pictures in both conditions were matched for phonological overlap (Belke, Meyer, et al., 2005). This means that the homogeneous condition was not more demanding than the heterogeneous condition in terms of phonological processing. Still, AwDs made more errors only in the homogeneous condition.

In the Hayling task, we obtained a similar pattern of performance. AwDs produced more errors in the inhibition condition, particularly more partial errors, which were contextually appropriate rather than semantically unrelated to the target word. Thus, AwDs suppressed target words as well as controls, but words *related* to the targets produced stronger competition in AwDs and this led to more failures of the task requirement to produce unrelated words. This finding fits well with the results from blocked cyclic naming because it suggests that AwDs find it difficult to resolve lexical competition within a set of related items. Thus, our results in both the blocked cyclic naming task and the Hayling task are consistent with AwDs' having greater difficulty suppressing lexical competitors.

Our results are also in line with the finding by Jones et al. (2010) that highly functioning AwDs (a similar sample to ours) performed worse in both a picture-naming task and an object-categorization task. The latter result is especially important because object categorization requires only access to semantic and syntactic properties of target words, without access to the words' phonological codes. While they interpreted their results as support for a visual account of dyslexia (Stein & Walsh, 1997), our results raise the possibility that lexical-semantic retrieval deficits are involved. Importantly, our results cannot have been caused by the visual similarity of the pictures because our stimuli were controlled for visual similarity across conditions (Belke, Meyer, et al., 2005). In addition, a visual deficit should have led to slower responses for AwDs compared to controls in both conditions, which we did not find. Finally, we saw evidence for a deficit in competitor suppression also in the Hayling task, which is an auditory task that does not require visual discrimination. Taken together, our results point to increased competition either at the lexical-semantic level or at the interaction between semantic and lexical-phonological representations, depending on the specifics of the speech production model.

Our results are not consistent with previous studies with CwDs that reported deficits in suppressing prepotent responses in Stroop-like paradigms (i.e. reading the word; Faccioli et al., 2008; Jones et al., 2016; Helland & Asbjornsen, 2000; Kapoula et al., 2010; Protopapas et al., 2007). We found no evidence for prepotent inhibition deficits in our high-functioning adult sample. A ceiling effect in the automatic condition and no group difference in inhibition failures in the inhibition condition of the Hayling task suggests no difference in general response suppression abilities. However, as described in the introduction, the Stroop effects could also originate at the lexical level where closely related semantic representations (e.g. red and blue) compete for selection (Roelofs, 2003). AwDs made more errors in the more competitive condition of both of our tasks, suggesting that individuals with dyslexia might perform worse in Stroop-like paradigms because of a deficit in suppressing lexical competitors, which is in line with Roelof's (2003) explanation of the origin of the Stroop-effect.

**Top-Down Control Modulation**

It has been suggested that lexical retrieval is not just affected by a bottom-up spreading activation process, but also by a top-down mechanism (Belke & Stielow, 2013; Roelofs, 2003; Thompson-Schill & Botvinick, 2006). For instance, in case of the cyclic blocked naming task, Belke (2008) argues that participants encode object names from a picture set in the first cycle of the cyclic naming task and use this knowledge in subsequent cycles to restrict lexical activation to only the names in the set. Without this top-down control, lexical competition would increase with each cycle. But this is typically not found in this task. However, it has been reported for aphasic patients (Schnur et al., 2006) and executive function patients (Belke & Stielow, 2013). These studies found an increased semantic interference effect in errors that accumulated over cycles in patients but not in controls. The accumulating effect has been contributed to deficits in top-down mechanisms (Belke, 2017; Belke & Stielow, 2013; Schnur et al., 2006).

To check whether effects were accumulating across cycles in our study, we added the factor *cycle* to our analyses of blocked cyclic naming results. We found equivalent performance (accuracy and RTs) across cycles for both participant groups. Thus, there is no evidence for a top-down control deficit in AwDs.

In summary, we have argued that the blocked cyclic naming accuracy deficit observed in the AwDs group could result from increased lexical competition. Our results suggest this is unlikely to arise from a deficit in top-down control.

**Lexical Competition and AwDs' Co-morbidity with Developmental Language Disorder**

Developmental dyslexia is known to be co-morbid with other disorders and problems with mental health (such as anxiety or low self-concept) (e.g., McArthur & Castles, 2017). Its co-morbidity with Developmental Language Disorder (DLD; Bishop et al., 2009; Bishop & Snowling, 2004) is of particular relevance here. DLD is a language disorder characterised by poor reading comprehension but also poor oral language abilities despite otherwise good hearing abilities and no sign of any neurological disorders. Individuals with DLD have been found to have difficulties resolving lexical competition during language comprehension (Dollaghan, 1998; Helenius et al., 2009; McMurray et al., 2010, 2014, 2019; Tanenhaus et al., 1995). And, similarly to our findings for AwDs, a recent study found reduced accuracy in bilingual children with DLD compared to those without DLD in the blocked cyclic naming task (McMillen et al., 2020). This opens up the possibility that our results might be driven by co-morbid DLD. We cannot tell how many, if any, of our AwDs had co-morbid DLD since we did not test oral language abilities. However, none of the AwDs had a formal diagnosis of DLD. In addition, language ability and reading comprehension are typically highly correlated and there were no differences between the two groups with regards to reading comprehension. Previous studies have reported high comorbidity between the two disorders amongst school children, with 58% of children with dyslexia meeting the criteria for DLD (McArthur et al., 2000; Snowling et al., 2019), suggesting the possibility that some of our AwDs might have an undiagnosed DLD. Whilst we cannot exclude DLD as a contributing factor, there was clear evidence of a deficit related to the resolution of lexical competition in our AwD group. Studies of lexical competition in DLD did not test for reading abilities, so it is not possible, at present, to attribute a lexical competition deficit exclusively to one of the two disorders. We should mention that we have assumed that dyslexia is a distinct disorder, and we used a strict criterion for the inclusion of AwDs into our study. Such an approach is controversial, as it has been suggested that dyslexia is simply the low end of a continuum of reading and writing skills (Shaywitz et al., 1992). A similar argument has been made for DLD (Leonard, 1991, 2010; Tomblin & Christiansen, 2010; Tomblin & Nippold, 2014). However, not all share this view. For instance, dyslexia and DLD have also been argued to be separate conditions with different etiologies (Bishop & Snowling, 2004; Snowling et al., 2020). If dyslexia is not a distinct disorder, the question arises whether lexical competition deficits are also present in individuals with less severely affected literacy skills. Future studies will need to address this question.

**Results in Relation to the Speech Production Models**

We have hypothesized that competitor errors in the blocked naming task and errors and slowing in the Hayling task are due to increased lexical competition in AwDs. However, as we discussed in the Introduction, the assumption that lexical selection is a competitive process is controversial. According to the competitive view (Levelt et al., 1999; Roelofs, 2018), a lexical unit can be selected when its activation exceeds the activation of its semantic competitors (Roelofs, 1997). In contrast, the non-competitive view suggests that the most activated lexical unit is selected regardless of the activation level of co-activated units (see e.g., Oppenheim et al., 2010). In order to achieve this, activation in the semantic network is boosted until one item is selected (Navarrete et al., 2014; Oppenheim et al., 2010).

Lexical Competition in AwDs

The question arises as to how our results can be explained by existing lexical production models. For example, competitive models like WEAVER++ (Levelt et al., 1999) could explain increased semantic errors in the blocked cyclic naming task through increased within-category activation. Belke's (2013) view of a conceptual origin for the semantic interference effect is in line with competitive model predictions and with our results. According to Belke, repeated access to a semantic category causes left over activation to build up at the conceptual level. This activation extends to other lexical items from the same category, leading to higher competition in the homogeneous condition. A similar logic could be applied to the Hayling task. If AwDs generally experience stronger competition among related words in the lexicon, activation of the target by the sentence context will also lead to activation of semantically related responses. Because some of these highly activated related words fit the sentence context, partial failure errors are predicted. The words that appear in partial failure errors would have higher activation compared to unrelated words and, therefore, a higher chance to be selected, in line with lexical competition. An advantage of this account is that it unifies the explanation of blocked cyclic naming and Hayling task results.

What about models that do not include competition (Navarrete et al., 2014; Oppenheim et al., 2010, Oppenheim & Balatsou, 2019)? After production of a word, Oppenheim et al.'s model strengthens the links between the word's semantic features and its lexical representation, while, at the same time, it weakens the links between its semantic features and co-activated non-target lexical representations[3]. It is possible that this incremental learning mechanism is not functioning in AwDs and, as a result, the target is not strengthened sufficiently or the links to co-activated

[3] We should note that the weakening of the links is not "suppression" per se, but the effect is the same.

competitors are not weakened sufficiently, or both. This would lead to a general tendency for targets and competitors to achieve similar levels of activation, with a resulting increase in errors due to natural fluctuations in activation. Weaker learning is consistent with previous reports of word-learning deficits in dyslexia (DiBetta & Romani, 2006, Elbro & Jensen, 2005; Mayringer & Wimmer, 2000; Messbauer& de Jong, 2003; Vellutino et al., 1995), and is even more possible if we consider that our sample could include individuals with co-morbid DLD+dyslexia, as previous studies have linked DLD with deficits in procedural learning (see Hedenius et al., 2011). However, one of the clearest predictions of the learning model is that the semantic interference effect (either in RT or errors), should grow over cycles of the cyclic naming task. We found no evidence that semantic interference effects grew over cycles in the AwD group. Similarly, we should note that the lack of interaction between sentence presentation order and participant group in the Hayling task is an issue for the incremental learning account. This account predicts that it would be harder to produce a different response (whether this is the prepotent response or a different response) the second time a particular sentence is presented, leading to slower responses and/or more errors.

To conclude, both lexical selection by competition and the incremental learning account could, in principle, explain our findings using mechanisms that reduce the normal difference in activation between a target and the words that are semantically related to it. However, the lack of an increased semantic interference effect across cycles in the blocked cyclic naming task is problematic for the incremental learning model, as is the lack of a sentence order effect in the Hayling task. Therefore, a lexical selection by competition account seems a better fit.

**The Underlying Cause of Lexical Competition**

Lexical Competition in AwDs

All speech production models share a mechanism that separates the activation of targets and co-activated words and this mechanism can fail. If the failure involves lexical competition, it could result from several types of mechanism, such as weakened lateral inhibition, a domain-general inhibition deficit, or increased activation of related representations for strategic reasons. If our results are due to changes to incremental learning, instead, this would involve weaker increases in connections when words are selected and weaker decreases when words are activated, but not selected. The interesting possibility that weaker incremental changes raise is that increased competitive effects could go along with deficits in word learning and these have been previously reported in AwDs (Di Betta & Romani, 2006; Romani et al., 2008).

Lateral inhibition can resolve lexical-semantic competition through inhibition of co-activated competitors (Feldman, 2005; Harley, 1990; McClelland & Rumelhart, 1981). It is implemented in some speech production models (Dell et al., 1997b; Howard et al., 2006; McClelland & Rumelhart, 1981), but not in others (Levelt et al., 1999). There is some evidence that deficits in lateral inhibition can increase semantic errors based on naming task results from clinical populations with impaired inhibition abilities (e.g., see Belleville et al., 2006, for Alzheimer disease and Obeso et al., 2011, for Parkinson disease).

Alternatively, an increase in semantic errors could also be explained by a more general inhibition deficit not restricted to interactions within the lexicon. For instance, in the blocked cyclic naming task, less general inhibition applied to semantic competitors (as proposed by Howard et al., 2006) would result in a higher chance of erroneously selecting competitors over the target, due to higher activation of co-activated words. Note that this option of a domain-general suppression deficit is supported by the fact that performance in the Hayling task is typically considered to be a measure of a general inhibition deficit (e.g., Burgess & Shallice, 1996).

Furthermore, it is possible that the increased lexical competition is a side effect of a strategic reading behaviour rather than a deficit in inhibition. As previous research has pointed out, individuals with dyslexia sometimes use strategic behaviour to compensate for their reading deficits (Chiarello et al., 2006; Gelbar et al., 2018). It is possible that individuals with dyslexia learn to allow a higher level of activation of semantically related words as a compensatory mechanism for weaker bottom-up activation of the lexicon during reading. If a person's lexicon is not that reliably activated by the input, either because the input is not of sufficient quality, or because lexical representations are not very robust, AwD might learn to rely more strongly on semantic context to help them select the correct word. While this mechanism would operate to help with written input, AwDs might generally allow higher activation of semantically related words in the lexicon. As a result, even tasks that do not involve written words, like picture naming, could be affected.

Importantly, any of the above-mentioned mechanisms - weakened lateral inhibition, general inhibition deficit, strategic account or the weakened learning mechanism - could explain lexical retrieval problems in people with dyslexia. Similarly, most of the accounts could explain AwDs' performance in the Hayling task, but not the general inhibition deficit account. The Hayling task requires not only resolution of lexical competition, but also suppression of a proponent response, which is reflected in the number of errors where the intended target is not inhibited (= inhibition failures). As pointed out, AwDs did not differ from the controls in the rate of inhibition failures. This finding does not support the general inhibition deficit.

Lexical Competition in AwDs

In order to fully understand the nature of the lexical competition deficit that we found, more research is needed. For instance, it would be useful to see if lexical competition deficits are present in other lexical-semantic tasks. A good candidate is the Continuous Naming Paradigm, a picture naming task where a number of objects from various semantic categories are presented intermixed with a number of unrelated pictures. In this paradigm, naming time, and thus lexical competition, is built up over time, accumulating over each presentation of a member of a semantic category (e.g., Howard et al., 2006). Previous studies have found that, unlike the semantic interference effect, this cumulative effect is not affected by top-down control (Belke & Stielow, 2013). Thus, an increased cumulative effect in individuals with dyslexia in the Continuous Naming Paradigm would strengthen our conclusion that their lexical competition deficit is not due to a deficit in top-down control.

It is also not clear whether AwDs' deficit in lexical selection in our study is related to a general executive function deficit. Shao et al., (2015) have proposed the involvement of a domain-general form of inhibition in lexical selection. Similarly, Calabria et al. (2019) concluded from a blocked cyclic picture naming study with bilingual aphasics that there is a degree of overlap between semantic control and executive control. To properly evaluate a general suppression deficit of prepotent responses, one would need evidence from non-verbal tasks that require the resolution of competition during highly demanding conditions. For instance, the Simon task (Simon & Rudell, 1967) and Erikson Flanker task (Eriksen & Eriksen, 1974) are widely used to measure response competition/conflict resolution. Previous studies have reported a poorer Flanker task performance for CwDs (Bednarek et al., 2004; Buchholz & Davies, 2005; Facoetti et al., 2000; Facoetti & Molteni, 2000; Facoetti & Turatto, 2000) and AwDs (Goldfarb & Shaul, 2013; Mahé et al., 2014). In order to reveal a potential domain-general deficit in suppressing dominant responses in conflict situations, however, it is important to test lexical competition and non-verbal inhibition tasks with the same sample of AwDs and controls.

**Clinical Implications**

The concept of increased lexical competition in individuals with dyslexia has been previously suggested in the dyslexia literature, with CwDs found to produce an increased number of semantically related errors when naming pictures (Nation et al., 2001). Our results extended previous findings to AwDs and showed lexical retrieval deficits in conditions where lexical-semantic competition was increased. Even though our study did not distinguish between the subtypes of dyslexia, lexical retrieval deficits could, in theory, be observed in both surface and phonological profiles. For example, a deficit in adequately managing lexical competition in the lexicon might result from damage to a lexical route (as in the ST-DRC model; Pritchard et al., 2018), leading to stronger reliance on letter-sound mappings, as seen in developmental surface dyslexia. On the other hand, problems with lexical competition could also result if participants rely too heavily on the lexicon, due to impaired grapheme-to-phoneme mappings, as in phonological developmental dyslexia. An overreliance on lexical and contextual information (see mechanisms described in Pritchard et al., 2018) might increase activation in the lexicon and produce increased competition, a prediction that is in line with the strategic account of lexical competition. Problems with lexical competition, therefore, could occur in both subtypes of dyslexia, but for different reasons. It is less clear how these two influences would interact in a mixed profile, where deficits in both lexical and sub-lexical skills are observed, but the effects could be additive (with increased reliance on lexical activation, due to weak mapping, and damaged lexical processing for managing competition both contributing).

Lexical Competition in AwDs

An overly-active lexicon could also influence the effectiveness of self-teaching processes and slow the pace of learning to read (see ST_DRC model, Pritchard et al., 2018; Share, 1995). Self-teaching depends on identifying a single lexical item based on both context and a partial mapping of letters to sounds that will activate the phonological lexicon and restrict the candidates that the semantic context proposes. When a single lexical item can be identified by these converging sources of information it can support learning the letter-sound mappings that have not already been acquired and learning of the new word form in the orthographic lexicon. If a single word cannot be identified in the phonological lexicon, however, or if the wrong target is activated because there is too large a cohort of active items, either self-teaching will not occur or it will occur but reinforce incorrect mappings. Either of these outcomes would slow learning. Lexical competition could affect learning to read on its own in some readers. In others it might combine with poor phonological skills (which will make deriving partial mappings more difficult), eliminating the mutual support between lexical and mapping information that normally allows early readers to acquire familiar spoken words when the orthographic form is unknown.

Note that a lexical-non-semantic route for reading, with direct connections between orthographic and phonological entries that bypass semantic mediation, could influence competition effects. Effects similar to the ones we measured in picture naming would be expected from the lexical-semantic route for reading. The contribution of a lexical-non-semantic route might moderate them. This is because direct connections could help activate a single phonological entry, boosting it above its competitors. Whether competition effects for known words are unaffected, moderated or eliminated would depend on the details of the interaction between lexical-semantic and lexical-non-semantic information. If lexical-non-semantic information has absolute priority, competition should be eliminated. If lexical-semantic information has priority, competition would be unaffected. If information from both routes is summed, competition could be moderated, but perhaps not eliminated.

Note too, however, that a lexical-non-semantic route to reading can only influence outcomes for *known* words. Potentially detrimental effects of competition for learning and self-teaching are not buffered by the lexical-non-semantic route, since it cannot support words that have not yet been stored in the orthographic lexicon.

The influence of competition on reading compared to picture naming is an interesting topic for future work. Comparisons between high-frequency regular words (which have support from both letter-sound mappings and lexical information, but where fast lexical activation is expected), low frequency regular words (where letter-sound mappings should have more influence compared to lexical information) and irregular words (which have greater reliance on word familiarity) could be revealing. For irregular words, the word position where mappings are irregular may also influence outcomes, with irregularities at initial word positions predicted to lead to more competition.Earlier irregular information will mean that the initial portions of words will not get support from regular mappings and, therefore, can't be used to help disambiguate activated entries. Competition effects in relation to self-teaching in younger readers would be another interesting avenue to explore because readers with stronger competition effects would be predicted to benefit less from self-teaching.

Our results indicate that the observed lexical retrieval deficits are in addition to phonological deficits. They are present in adulthood, in line with the view multiple cognitive difficulties can contribute to dyslexia and individuals with multiple difficulties are more likely to suffer from a type of dyslexia that persists later in life (Bishop & Rutter, 2009; Pennington, 2006). Future studies should aim to investigate not only the cause of these competitive deficits but also whether they differ between dyslexia subtypes. All of the potential mechanisms outlined in the present study - weakened lateral inhibition, a general inhibition deficit, strategic changes to activation or a weakened learning mechanism - could potentially be integrated into a wider intervention aimed at improving not only the decoding skills of CwDs, but also additional deficits. Not all children with dyslexia (approximately 30% of the population with reading difficulties; Shanahan & Barr, 1995) respond to "gold standard" dyslexia interventions that concentrate on phonological skills and a small set of sight words (these children are sometimes called *non-responders*). For these children additional mechanisms need to be targeted. Future studies should investigate whether strategies to reduce lexical competition could be one of them.

It is also noteworthy that the observed lexical retrieval deficit did not seem to strongly affect our participants' text comprehension skills as they did not differ from control participants in the Gray Silent Reading Test. However, the Gray Silent Reading Test is untimed. It is possible that the lexical retrieval deficit leads to slower reading. Furthermore, most of our participants were university students, that is individuals with dyslexia with functional reading skills, also evidenced by reading comprehension skills that were equivalent to our control participants. It remains to be seen whether at least some individuals who participated in our study have developed strategies that allow them to overcome their lexical retrieval deficit when reading. Individuals with weaker comprehension skills might be affected more strongly by lexical retrieval problems.

**Conclusion**

In conclusion, we found that AwDs experience difficulties with resolving lexical competition in two verbal production tasks, showing evidence of increased susceptibility to lexical competition. These difficulties could not be accounted for by phonological deficits, at least in blocked cyclic picture naming. We suggest there is an additional deficit in adequately managing lexical competition in the lexicon. This would go beyond  an increase in lexical competition that derives from more basic deficits within the decoding and/or word recognition routes because it is present in a picture naming task. It may, however, interact with deficits in whole-word recognition or letter-sound mapping to influence the persistence and/or severity of reading problems (e.g. through effects on a self-teaching process, or by amplifying the impact of other difficulties). There are various potential mechanisms that could cause problems with lexical competition: a deficit of lateral inhibition in the lexicon, changes to incremental learning or a strategic change in lexical dynamics that would allow competitors to remain highly active. Future studies will be needed to distinguish between specific underlying mechanisms.

**Acknowledgements**

We thank all participants who participated in the study.

**References**

Anonymous. (2020). What about Lexical competition? Exploring the locus of lexical retrieval deficits in adults with Developmental Dyslexia. *Retrieved from Osf.Io/4tdns.*

Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*(1), 65-94. doi:10.1037/0033-2909.121.1.65

Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. *R Package Version 1.1-7.* https://doi.org/citeulike-article-id:1080437

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: linear mixed-effects models using S4 classes. *Journal of Statistical Software*.

Bednarek, D. B., Saldaña, D., Quintero-Gallego, E., García, I., Grabowska, A., & Gómez, C. M. (2004). Attentional deficit in dyslexia: A general or specific impairment? *NeuroReport*, *15*(11), 1787–1790. https://doi.org/10.1097/01.wnr.0000134843.33260.bf

Bélanger, S., & Belleville, S. (2009). Semantic Inhibition Impairment in Mild Cognitive Impairment: A Distinctive Feature of Upcoming Cognitive Decline? *Neuropsychology*, *23*(5), 592–606. https://doi.org/10.1037/a0016152

Belke, E. (2013). Long-lasting inhibitory semantic context effects on object naming are necessarily conceptually mediated: Implications for models of lexical-semantic encoding. *Journal of Memory and Language*, *69*(3), 228–256. https://doi.org/10.1016/j.jml.2013.05.008

Belke, E. (2017). The role of task-specific response strategies in blocked-cyclic naming. *Frontiers in Psychology*, *7*, 1955. https://doi.org/10.3389/fpsyg.2016.01955

Belke, E., Brysbaert, M., Meyer, A. S., & Ghyselinck, M. (2005). Age of acquisition effects in picture naming: Evidence for a lexical-semantic competition hypothesis. *Cognition*, *96*(2), B45–B54. https://doi.org/10.1016/j.cognition.2004.11.006

Belke, E., Meyer, A. S., & Damian, M. F. (2005). Refractory effects in picture naming as assessed in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *58*(4), 667–692. https://doi.org/10.1080/02724980443000142

Belke, E., & Stielow, A. (2013). Cumulative and non-cumulative semantic interference in object naming: Evidence from blocked and continuous manipulations of semantic context. *Quarterly Journal of Experimental Psychology*, *66*(11), 2135–2160. https://doi.org/10.1080/17470218.2013.775318

Belleville, S., Rouleau, N., & Van der Linden, M. (2006). Use of the Hayling task to measure inhibition of prepotent responses in normal aging and Alzheimer's disease. *Brain and Cognition*, *62*(2), 113–119. https://doi.org/10.1016/j.bandc.2006.04.006

Bishop, D., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children who read words accurately despite language impairment: Who are they and how do they do it? *Child Development*, *80*(2), 593–605. https://doi.org/10.1111/j.1467-8624.2009.01281.x

Bishop, D., & Rutter, M. (2009). Neurodevelopmental Disorders: Conceptual Issues. In *Rutter's Child and Adolescent Psychiatry: Fifth Edition* (pp. 32–41). https://doi.org/10.1002/9781444300895.ch3

Bishop, D., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin*. https://doi.org/10.1037/0033-2909.130.6.858

Bloom, P. A., & Fischler, I. (1980). Completion norms for 329 sentence contexts. *Memory & Cognition*, *8*(6), 631–642. https://doi.org/10.3758/BF03213783

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. https://doi.org/10.1037/0033-295X.108.3.624

Brosnan, M., Demetre, J., Hamill, S., Robson, K., Shepherd, H., & Cody, G. (2002). Executive functioning in adults and children with developmental dyslexia. *Neuropsychologia*, *40*(12), 2144–2155. https://doi.org/10.1016/S0028-3932(02)00046-5

Buchholz, J., & Davies, A. A. (2005). Adults with dyslexia demonstrate space-based and object-based covert attention deficits: Shifting attention to the periphery and shifting attention between objects in the left visual field. *Brain and Cognition*. https://doi.org/10.1016/j.bandc.2004.08.017

Burgess, P. W., & Shallice, T. (1996). Response suppression, Initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, *34*(4), 263–272. https://doi.org/10.1016/0028-3932(95)00104-2

Burgess, P. W., & Shallice, T. (1997). The Hayling and Brixton tests. Bury St Edmunds: Thames Valley Test Company.

Calabria, M., Grunden, N., Serra, M., Garcia-Sanchez, C., & Costa, A. (2019). Semantic Processing in Bilingua Aphasia: Evidence of Language Dependency. Frontiers in Human Neuroscience, 13, 15.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from the "tip-of-the-tongue" phenomenon. *Cognition*, *64*(3), 309–343. https://doi.org/10.1016/S0010-0277(97)00031-0

Castles, A., & Coltheart, M. (1996). Cognitive correlates of developmental surface dyslexia: A single case study. *Cognitive Neuropsychology*, *13*(1), 25–50. https://doi.org/10.1080/026432996382051

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Catts, H. W., Gillispie, M., Leonard, L. B., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. Journal of learning disabilities, 35(6), 509–524.

Cervera-Crespo, T., & González-Alvarez, J. (2017). Age and semantic inhibition measured by the hayling task: A meta-analysis. *Archives of Clinical Neuropsychology*, *32*(2), 198–214. https://doi.org/10.1093/arclin/acw088

Chiarello, C., Lombardino, L. J., Kacinik, N. A., Otto, R., & Leonard, C. M. (2006). Neuroanatomical and behavioral asymmetry in an adult compensated dyslexic. *Brain and Language*, *98*(2), 169–181. https://doi.org/10.1016/j.bandl.2006.04.012

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Damian, M. F., Vigliocco, G., & Levelt, W. J. M. (2001). Effects of semantic context in the naming of pictures and words. *Cognition*, *81*(3), B77–B86. https://doi.org/10.1016/S0010-0277(01)00135-4

De Zubicaray, G. I., Zelaya, F. O., Andrew, C., Williams, S. C. R., & Bullmore, E. T. (2000). Cerebral regions associated with verbal response initiation, suppression and strategy use. *Neuropsychologia*, *38*(9), 1292–1304. https://doi.org/10.1016/S0028-3932(00)00026-9

Deacon, S. H., Benere, J., & Castles, A. (2012). Chicken or egg? Untangling the relationship between orthographic processing skill and reading accuracy. *Cognition*, *122*(1), 110–117.

https://doi.org/10.1016/J.COGNITION.2011.09.003

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93(3), 283–321.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997a). Discreteness and interactivity in spoken word production. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997b). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review*, *104*(4), 801–838. https://doi.org/10.1037/0033-295X.104.4.801

Di Betta, A. M., & Romani, C. (2006). Lexical learning and dysgraphia in a group of adults with developmental dyslexia. *Cognitive Neuropsychology*, *23*(3), 376–400. https://doi.org/10.1080/02643290442000545

Dollaghan, C. (1998). Spoken word recognition in children with and without specific language impairment. *Applied Psycholinguistics*. https://doi.org/10.1017/s0142716400010031

Elbro, C., & Jensen, M. N. (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. Scandinavian Journal of Psychology, 46(4), 375–384. doi:10.1111/j.1467- 9450.2005.00468.x

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*. https://doi.org/10.3758/BF03203267

Faccioli, C., Peru, A., Rubini, E., & Tassinari, G. (2008). Poor readers but compelled to read: Stroop effects in developmental dyslexia. *Child Neuropsychology*, *14*(3), 277–283. https://doi.org/10.1080/09297040701290040

Facoetti, A., & Molteni, M. (2000). Is attentional focusing an inhibitory process at distractor location? *Cognitive Brain Research*. https://doi.org/10.1016/S0926-6410(00)00031-8

Facoetti, A., Paganoni, P., & Lorusso, M. L. (2000). The spatial distribution of visual attention in developmental dyslexia. *Experimental Brain Research*. https://doi.org/10.1007/s002219900330

Facoetti, A., & Turatto, M. (2000). Asymmetrical visual fields distribution of attention in dyslexic children: A neuropsychological study. *Neuroscience Letters*. https://doi.org/10.1016/S0304-3940(00)01354-9

Faust, M., Dimitrovsky, L., & Shacht, T. (2003). Naming difficulties in children with dyslexia: Application of the tip-of-the-tongue paradigm. *Journal of Learning Disabilities*. https://doi.org/10.1177/002221940303600301

Faust, M., & Sharfstein-Friedman, S. (2003). Naming difficulties in adolescents with dyslexia: Application of the tip-of-the-tongue paradigm. *Brain and Cognition*, *53*(2), 211–217. https://doi.org/10.1016/S0278-2626(03)00112-X

Feldman, J. A. (2005). Connectionist Models and Their Applications: Introduction. *Cognitive Science*, *9*(1), 1–2. https://doi.org/10.1207/s15516709cog0901_1

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*. https://doi.org/10.1016/S0364-0213(82)80001-3

Finucci, J. M., Whitehouse, C. C., Isaacs, S. D., & Childs, B. (1984). Derivation and Validation of a Quantitative Definition of Specific Reading Disability for Adults. *Developmental Medicine & Child Neurology*, *26*(2), 143–153. https://doi.org/10.1111/j.1469-8749.1984.tb04425.x

Frith, U., Landerl, K., & Frith, C. (1995). Dyslexia and verbal fluency: More evidence for a phonological deficit. Generating words from sound is hard for dyslexics—generating words from meaning is easy. *Dyslexia:An International Journal of Research and Practice*, *1*, 2–11.

Gelbar, N. W., Bray, M., Kehle, T. J., Madaus, J. W., & Makel, C. (2018). Exploring the Nature of Compensation Strategies in Individuals With Dyslexia. *Canadian Journal of School Psychology*, *33*(2), 110–124. https://doi.org/10.1177/0829573516677187

Germanò, E., Gagliano, A., & Curatolo, P. (2010). Comorbidity of ADHD and Dyslexia. *Developmental Neuropsychology, 35*(5), 475-493. doi:10.1080/87565641.2010.494748

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/0096-1523.10.5.640

Goldfarb, L., & Shaul, S. (2013). Abnormal attentional internetwork link in dyslexic readers. *Neuropsychology*, *27*(6), 725–729. https://doi.org/10.1037/a0034422

Hanly, S., & Vandenberg, B. (2010). Tip-of-the-tongue and word retrieval deficits in dyslexia. *Journal of Learning Disabilities*, *43*(1), 15–23. https://doi.org/10.1177/0022219409338744

Harley, T. A. (1990). Paragrammatisms: Syntactic disturbance or breakdown of control? *Cognition*, *34*(1), 85–91. https://doi.org/10.1016/0010-0277(90)90032-F

Harley, T. A. (1993). Phonological Activation of Semantic Competitors During Lexical Access in Speech Production. *LANGUAGE AND COGNITIVE PROCESSES*, *3*, 291–309.

Hedenius, M., Persson, J., Tremblay, A., Adi-Japha, E., Veríssimo, J., Dye, C. D., Alm, P., Jennische, M., Bruce Tomblin, J., & Ullman, M. T. (2011). Grammar predicts procedural learning and consolidation deficits in children with Specific Language Impairment. *Research in Developmental Disabilities*. https://doi.org/10.1016/j.ridd.2011.07.026

Helenius, P., Parviainen, T., Paetau, R., & Salmelin, R. (2009). Neural processing of spoken words in specific language impairment and dyslexia. *Brain*. https://doi.org/10.1093/brain/awp134

Helland, T., & Asbjornsen, A. (2000). Executive functions in dyslexia. *Child Neuropsychology*, *6*(1), 37–48. https://doi.org/10.1076/0929-7049(200003)6:1;1-b;ft037

Henry, J. D., & Crawford, J. R. (2004). Verbal fluency deficits in Parkinson's disease: A meta-analysis. *Journal of the International Neuropsychological Society*, *10*(4), 608–622. https://doi.org/10.1017/S1355617704104141

Hines, J. L., King, T. S., & Curry, W. J. (2012). The adult ADHD self-report scale for screening for adult attention deficit-hyperactivity disorder (ADHD). Journal of the American Board of Family Medicine : JABFM, 25(6), 847–853.

Hirshorn, E. A., & Thompson-Schill, S. L. (2006). Role of the left inferior frontal gyrus in covert word retrieval: Neural correlates of switching during verbal fluency. *Neuropsychologia*, *44*(12), 2547–2557. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2006.03.035

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: experimental and computational studies. *Cognition*, *100*(3), 464–482. https://doi.org/10.1016/j.cognition.2005.02.006

Indefrey, P., & Levelt, W. J. M. (2000). The Neural Correlates of Language Production. In *The new cognitive neurosciences*.

Jacobsen, G. M., de Mello, C. M., Kochhann, R., & Fonseca, R. P. (2017). Executive Functions in School-age Children: Influence of Age, Gender, School Type and Parental Education. *Applied Cognitive Psychology*, *31*(4), 404–413. https://doi.org/10.1002/acp.3338

Jones, M. W., Branigan, H. P., Hatzidaki, A., & Obregón, M. (2010). Is the "naming" deficit in dyslexia a misnomer? *Cognition*, *116*(1), 56–70. https://doi.org/10.1016/j.cognition.2010.03.015

Jones, M. W., Snowling, M. J., & Moll, K. (2016). What Automaticity Deficit? Activation of Lexical Information by Readers With Dyslexia in a Rapid Automatized Naming Stroop-

Switch Task. Journal Of Experimental Psychology-Learning Memory And Cognition, 42(3), 465-474.

Kalashnikova, M. , & Burnham, D. (2016). Novel word learning, reading difficulties, and phonological processing skills. *Dyslexia* , 22(2), 101–119. doi:10.1002/dys.1525

Kapoula, Z., Lê, T. T., Bonnet, A., Bourtoire, P., Demule, E., Fauvel, C., Quilicci, C., & Yang, Q. (2010). Poor stroop performances in 15-year-old dyslexic teenagers. *Experimental Brain Research*, *203*(2), 419–425. https://doi.org/10.1007/s00221-010-2247-x

Katz, R. B. (1986). Phonological deficiencies in children with reading disability: Evidence from an object-naming task. *Cognition*, *22*(3), 225–257. https://doi.org/10.1016/0010-0277(86)90016-8

Kessels, R. P. C., Van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & De Haan, E. H. F. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology*, *7*(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8

Kessler, R. , Adler, L. , Ames, M. , Demler, O. , Faraone, S. , Hiripi, E. , … Walters, E. (2005). The world health organization adult ADHD self-report scale (ASRS): A short screening scale for use in the general population. Psychological Medicine , 35(2), 245–256.

Kessler, R. C. , Adler, L. A. , Gruber, M. J. , Sarawate, C. A. , Spencer, T. , & Van Brunt, D. L. (2007). Validity of the world health organization adult ADHD self-report scale (ASRS) screener in a representative sample of health plan members. International Journal of Methods in Psychiatric Research , 16, 52–65.

Korhonen, T. T. (1995). The Persistence of Rapid Naming Problems in Children with Reading Disabilities. *Journal of Learning Disabilities*, *28*(4), 232–239. https://doi.org/10.1177/002221949502800405

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*(2), 99–116. https://doi.org/10.1016/0301-0511(80)90046-0

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2016). *Tests in Linear Mixed Effects Models*. *R Package*(2).

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, *103*(3), 309–324. https://doi.org/10.1016/j.jecp.2009.03.006

Leonard, L. B. (1991). Specific Language Impairment as a Clinical Category. *Language, Speech, and Hearing Services in Schools*. https://doi.org/10.1044/0161-1461.2202.66

Leonard, L. B. (2010). Is specific language impairment a useful construct? In *Advances in Applied Psycholinguistics*. https://doi.org/10.1017/cbo9780511571374.002

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). Atheory of lexical access in speech production. *Behavioural and Brain Sciences*, *22*, 1–75. https://doi.org/10.1017/S0140525X99001776

Levin, B. E. (1990). Organizational deficits in dyslexia: Possible frontal lobe dysfunction. *Developmental Neuropsychology*, *6*(2), 95–110. https://doi.org/10.1080/87565649009540453

Litt, R. A. , Wang, H. C. , Sailah, J. , Badcock, N. A. , & Castles, A. (2019). Paired associate learning deficits in poor readers: The contribution of phonological input and output processes. *Quarterly Journal of Experimental Psychology* , 72(3), 616–633. doi:10.1177/1747021818762669

Logan, J. M., & Balota, D. A. (2003). Conscious and unconscious lexical retrieval blocking in younger and older adults. *Psychology and Aging*, *18*(3), 537–550. https://doi.org/10.1037/0882-7974.18.3.537

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Mahé, G., Doignon-Camus, N., Dufour, A., & Bonnefond, A. (2014). Conflict control processing in adults with developmental dyslexia: An event related potentials study. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2013.06.005

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 33(3), 503–535.

Mayringer, H., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: evidence for a phonological learning deficit. *Journal of Experimental Child Psychology*, 75, 116–133.

Maril, A., Wagner, A. D., & Schacter, D. L. (2001). On the tip of the tongue: An event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron*, *31*(4), 653–660. https://doi.org/10.1016/S0896-6273(01)00396-8

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the "specifics" of specific reading disability and specific language impairment. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1017/S0021963099006186

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407.

https://doi.org/10.1037/0033-295X.88.5.375

McMurray, B., Klein-Packard, J., & Tomblin, J. B. (2019). A real-time mechanism underlying lexical deficits in developmental language disorder: Between-word inhibition. *Cognition*. https://doi.org/10.1016/j.cognition.2019.06.012

McMurray, B., Munson, C., & Tomblin, J. B. (2014). Individual differences in language ability are related to variation in word recognition, not speech perception: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2014_JSLHR-L-13-0196

McMurray, B., Samelson, V. M., Lee, S. H., & Bruce Tomblin, J. (2010). Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*. https://doi.org/10.1016/j.cogpsych.2009.06.003

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., Ruffino, M., & Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: A comparative study. *Neuropsychologia*, *48*(4), 863–872. https://doi.org/10.1016/j.neuropsychologia.2009.11.003Messbauer, V. C. S., & de Jong, P. F. (2003). Word, nonword, and visual paired associate learning in Dutch dyslexic children. Journal of Experimental Child Psychology, 84(2), 77–96. doi:10.1016/S0022-0965(02)00179-0

Moura, O., Simões, M. R., & Pereira, M. (2015). Executive functioning in children with developmental dyslexia. *Clinical Neuropsychologist*, *28*(S1), 20–41. https://doi.org/10.1080/13854046.2014.964326

Mueller, S. (2011). Tutorial on using the psychology experiment building language (PEBL) in the laboratory, the field, and the classroom. *Proceedings of the Annual Meeting of the Cognitive*

*Science Society*, *33*(33), 54–55.

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v043.i09

Nation, K. (2005). Picture naming and developmental reading disorders. *Journal of Research in Reading*, *28*(1), 28–38. https://doi.org/10.1111/j.1467-9817.2005.00250.x

Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. *Language and Cognitive Processes*, *16*(2–3), 241–259. https://doi.org/10.1080/01690960042000003

Navarrete, E., Del Prato, P., Peressotti, F., & Mahon, B. Z. (2014). Lexical selection is not by competition: Evidence from the blocked naming paradigm. *Journal of Memory and Language*, *76*, 253–272. https://doi.org/10.1016/j.jml.2014.05.003

Obeso, I., Wilkinson, L., Casabona, E., Bringas, M. L., Álvarez, M., Álvarez, L., Pavón, N., Rodríguez-Oroz, M. C., Macías, R., Obeso, J. A., & Jahanshahi, M. (2011). Deficits in inhibitory control and conflict resolution on cognitive and motor tasks in Parkinson's disease. *Experimental Brain Research*, *212*(3), 371–384. https://doi.org/10.1007/s00221-011-2736-6

Oppenheim G.M. & Balatsou, E.(2019) Lexical competition on demand, *Cognitive Neuropsychology*, 36(5-6), 216-219. https://doi.org/10.1080/02643294.2019.1580189

Oppenheim, G. M., Dell, G. S., & Schwartz, M. F. (2010). The dark side of incremental learning: A model of cumulative semantic interference during lexical access in speech production. *Cognition*, *114*(2), 227–252. https://doi.org/10.1016/j.cognition.2009.09.007

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*(2), 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Plaza, M., & Guitton, C. (1997). Working memory limitation, phonological deficit, sequential disorder and syntactic impairment in a child with a severe developmental dyslexia. *Dyslexia*, *3*(2), 93–108. https://doi.org/10.1002/(SICI)1099-0909(199706)3:2<93::AID-DYS79>3.0.CO;2-E

Posner, M. I., & Raichle, M. E. (1994). *Images of mind.* Scientific American Books.

Pritchard, S. C., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive Science*, 42, 722–770.

Protopapas, A., Archonti, A., & Skaloumbakas, C. (2007). Reading ability is negatively related to Stroop interference. *Cognitive Psychology*, *54*(3), 251–282. https://doi.org/10.1016/j.cogpsych.2006.07.003

Proulx, M. J., & Elmasry, H. M. (2015). Stroop interference in adults with dyslexia. *Neurocase*, *21*(4), 413–417. https://doi.org/10.1080/13554794.2014.914544

R Core Team. (2018). A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*.

Raman, I. (2011). The role of age of acquisition in picture and word naming in dyslexic adults. *British Journal of Psychology*, *102*(3), 328–339. https://doi.org/10.1348/000712610X522572

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, *107*(3), 460–499. https://doi.org/10.1037/0033-295X.107.3.460

Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. In *Dyslexia* (Vol. 11, Issue 2, pp. 116–131). https://doi.org/10.1002/dys.289

Reynolds, C., & Kamphaus, R. (2007). Test of Irregular Word Reading Efficiency. *Austin, TX: Pro-Ed.*

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, *64*(3), 249–284. https://doi.org/10.1016/S0010-0277(97)00027-9

Roelofs, A. (2003). Goal-Referenced Selection of Verbal Action: Modeling Attentional Control in the Stroop Task. *Psychological Review*, *110*(1), 88–125. https://doi.org/10.1037/0033-295X.110.1.88

Roelofs, A. (2018). A unified computational account of cumulative semantic, semantic blocking, and semantic distractor effects in picture naming. *Cognition*, *172*, 59–72. https://doi.org/10.1016/J.COGNITION.2017.12.007

Romani, C., Di Betta, A., Tsouknida, E., & Olson, A. (2008). Lexical and nonlexical processing in developmental dyslexia: A case for different resources and different impairments. *Cognitive Neuropsychology*, *25*(6), 798–830. https://doi.org/10.1080/02643290802347183

Sakamoto, Y., Ishiguro, M., & Kitagawa, G. (1986). Akaike information criterion statistics. *KTK Scientific Publishers ; Dordrecht ; Boston : D. Reidel ; Hingham, MA*.

Schnur, T. T., Schwartz, M. F., Brecher, A., & Hodgson, C. (2006). Semantic interference during blocked-cyclic naming: Evidence from aphasia. *Journal of Memory and Language*, *54*(2), 199–227. https://doi.org/10.1016/j.jml.2005.10.002

Schwartz, S., Baldo, J., Graves, R. E., & Brugger, P. (2003). Pervasive influence of semantics in letter and category fluency: A multidimensional approach. *Brain and Language*, *87*(3), 400–411. https://doi.org/10.1016/S0093-934X(03)00141-X

Seedorff, M., Oleson, J., & McMurray, B. (2019). Maybe maximal: Good enough mixed models optimize power while controlling Type I error. *Preprints.Https://Doi.Org/10.31234/Osf.Io/Xmhfr*.

Shanahan, T., & Barr, R. (1995). Reading Recovery: An Independent Evaluation of the Effects of an Early Instructional Intervention for At-Risk Learners. *Reading Research Quarterly*. https://doi.org/10.2307/748206

Shao, Z., Meyer, A. S., & Roelofs, A. (2013). Selective and nonselective inhibition of competitors in picture naming. *Memory and Cognition*, *41*(8), 1200–1211. https://doi.org/10.3758/s13421-013-0332-7

Shao, Z., Roelofs, A., Martin, R. C., & Meyer, A. S. (2015). Selective inhibition and naming performance in semantic blocking, picture-word interference, and color-word stroop tasks. *Journal of Experimental Psychology: Learning Memory and Cognition*, *41*(6), 1806–1819. https://doi.org/10.1037/a0039363

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151–218.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence That Dyslexia May Represent the Lower Tail of a Normal Distribution of Reading Ability. *New England Journal of Medicine*. https://doi.org/10.1056/nejm199201163260301

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effects of an irrelevant directional cue on human information processing. *Journal of Applied Psychology*. https://doi.org/10.1016/S0166-4115(08)61218-2

Snowling, M. J., Hayiou-Thomas, M. E., Nash, H. M., & Hulme, C. (2020). Dyslexia and Developmental Language Disorder: comorbid disorders with distinct effects on reading comprehension. *Journal of Child Psychology and Psychiatry and Allied Disciplines*. https://doi.org/10.1111/jcpp.13140

Snowling, M. J., Nash, H. M., Gooch, D. C., Hayiou-Thomas, M. E., & Hulme, C. (2019). Developmental Outcomes for Children at High Risk of Dyslexia and Children With

Developmental Language Disorder. *Child Development*. https://doi.org/10.1111/cdev.13216

Snowling, M. J., Nation, K., Moxham, P., Gallagher, A., & Frith, U. (1997). Phonological processing skills of dyslexic students in higher education: A preliminary report. *Journal of Research in Reading*, *20*(1), 31–41. https://doi.org/10.1111/1467-9817.00018

Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences*, *20*(4), 147–152. https://doi.org/10.1016/S0166-2236(96)01005-3

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Swan, D., & Goswami, U. (1997). Picture naming deficits in developmental dyslexia: the phonological representations hypothesis. *Brain and Language*, *56*(56), 334–353. https://doi.org/10.1111/1467-7687.00054

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*. https://doi.org/10.1126/science.7777863

Thompson-Schill, S. L., & Botvinick, M. M. (2006). Resolving conflict: A response to Martin and Cheng (2006). *Psychonomic Bulletin and Review*, *13*(3), 402–408. https://doi.org/10.3758/BF03193860

Tomblin, J. B., & Christiansen, M. H. (2010). Explaining developmental communication disorders. *Speech Sound Disorders in Children: In Honor of Lawrence D. Shriberg*, 35–49.

Tomblin, J. B., & Nippold, M. A. (2014). *Understanding individual differences in language development across the school years*. Psychology Press.

Torgesen, J., Rashotte, C., & Wagner, R. (1999). TOWRE-2: Test of Word Reading Efficiency, 2nd Ed. In *Austin, TX: Pro-Ed*.

Torkildsen, J. von K., Syversen, G., Simonsen, H. G., Moen, I., & Lindgren, M. (2007). Brain responses to lexical-semantic priming in children at-risk for dyslexia. *Brain and Language*, *102*(3), 243–261. https://doi.org/10.1016/j.bandl.2006.11.010

Van der Schoot, M., Licht, R., Horsley, T. M., & Sergeant, J. A. (2001). Inhibitory deficits in reading disability depend on subtype: Guessers but not spellers. *Child Neuropsychology*, *6*(4), 297–312. https://doi.org/10.1076/chin.6.4.297.3139

Varvara, P., Varuzza, C., Sorrentino, A. C. P., Vicari, S., & Menghini, D. (2014). Executive functions in developmental dyslexia. *Frontiers in Human Neuroscience*, *8*, 120. https://doi.org/10.3389/fnhum.2014.00120

Vellutino, F. R., Scanlon, D. M., & Spearing, D. (1995). Semantic and phonological coding in poor and normal readers. Journal of Experimental Child Psychology, 59(1), 76–123.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). CTOPP: Comprehensive Test of Phonological Processing. *Austin, TX: PRO-ED*.

Wechsler, D. (2008). Wechsler adult intelligence scale. *San Antonio, TX: Pearson*, *4*.

Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal Fluency: Language or Executive Function Measure? *Applied Neuropsychology:Adult*, *23*(1), 29–34. https://doi.org/10.1080/23279095.2015.1004574

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*. https://doi.org/10.18637/jss.v021.i12

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*. https://doi.org/10.21105/joss.01686

Wiederholt, J., & Blalock, G. (2000). Gray Silent Reading Test. *Pro-Ed, Austin*.

Wolf, M. (1991). *"Word-wraiths": The unique contribution of the naming system to reading prediction and intervention in developmental dyslexia*.