Semantic fluency difficulties in Greek children with dyslexia and/or Developmental Language Disorder: Poor lexical-semantic structure or slower retrieval processes?

Published 7 November 2019: *International Journal of Language and Communication Disorders,* doi: 10.1111/1460-6984.12512

Maria Mengisidou, PhD

UCL Institute of Education, University College London

20 Bedford Way

London WC1H 0AL, UK

E-mail: maria.mengisidou.14@ucl.ac.uk

Chloë R. Marshall, PhD

Professor of Psychology, Language and Education

UCL Institute of Education, University College London

20 Bedford Way

London WC1H 0AL, UK

E-mail: chloe.marshall@ucl.ac.uk

Stavroula Stavrakaki, PhD

Associate Professor of Neurolinguistics

School of Philosophy, Aristotle University of Thessaloniki

University Campus

Thessaloniki 541 24, Greece

Email: svoula@itl.auth.gr

Abstract

Background: Children with dyslexia and/or Developmental Language Disorder (hereafter children with DDLD) have been reported to retrieve fewer words than their typically-developing (TD) peers in semantic fluency tasks. It is not known whether this retrieval difficulty can be attributed to the semantic structure of their lexicon being poor or, alternatively, to words being retrieved more slowly despite semantic structure being intact.

Aims: The current study was designed to test two theoretical models that could potentially account for retrieval difficulties in semantic fluency tasks, namely the Poor Lexical-Semantic Structure Model and the Slow-Retrieval Model. Both models predict that children with DDLD will retrieve fewer items compared to TD children. However, while the Poor Lexical-Semantic Structure Model predicts a less sophisticated network of semantic connections between words in the lexicon, as evidenced by smaller clusters of related items in children with DDLD, the Slow-Retrieval Model predicts intact inter-item associations in the lexicon, as evidenced by the two groups' clusters being of a similar size. The groups' semantic fluency performance was therefore compared. How semantic fluency performance related to children's language, literacy, and phonological skills was also investigated.

Methods & Procedures: Sixty-six children with DDLD aged 7-12 years and 83 TD children aged 6-12 years, all monolingual Greek speakers, were tested on semantic fluency, using the categories 'animals', 'foods', and 'objects from around the house'. The number of correct and incorrect responses, the number of clusters and switches, and the average cluster size were computed. Children were also assessed on nonverbal IQ, language, literacy, and phonological tasks.

Outcomes & Results: In both groups, productivity in semantic fluency tasks correlated strongly with the number of clusters and the number of switches, but not with average cluster size. The DDLD group produced significantly fewer correct responses and fewer clusters compared to the TD group, but the two groups showed similar switching and average cluster size. Children's language, literacy, and phonological skills significantly predicted the number of correct responses produced, beyond the significant effect of age.

Conclusions & Implications: We conclude that poorer semantic fluency performance in children with DDLD results not from a lexicon with poor semantic structure, but rather from slower retrieval processes from a lexicon with intact semantic structure. The underlying causes of slow lexical retrieval still need further investigation.

What this paper adds

What is already known on the subject

Children with dyslexia and DLD show poorer semantic fluency performance compared to age-matched TD children. It remains an open question whether this poor performance can be better explained by impoverished semantic structure or by slower retrieval processes despite intact inter-item associations in the mental lexicon. Whether poorer language, literacy, and phonological skills influence semantic fluency performance in these groups is also unexplored.

What this paper adds to existing knowledge

This is the first study to assess semantic fluency in Greek children with dyslexia and/or DLD, and to this end children with dyslexia and DLD were combined in a single group. Language, literacy, and phonological skills were a significant predictor of semantic fluency performance after controlling for age. This is also the first developmental study designed to test two theoretical models accounting for lexical difficulties in semantic fluency tasks. The findings support the Slow-Retrieval Model, which attributes poorer semantic fluency performance to the slower retrieval of lexical items from the mental lexicon while the semantic structure is intact.

What are the potential or actual clinical implications of this work?

Insight into children's lexical retrieval difficulties can inform theory, and theory can inform treatment and training of children with dyslexia and DLD in clinical and educational settings. As a first step towards this direction, intervention studies designed to improve children's retrieval processes are needed to investigate any potential gains on semantic fluency performance.

Introduction

In a seminal publication, Collins and Loftus (1975) claimed that words are stored in a mental lexicon which is organized in a semantic network. Decades of research since then have investigated the structure of this network. One characteristic feature is considerable local structure in the form of clusters of words that are highly interconnected by semantic relatedness (e.g., Rogers and McClelland 2008, Hills *et al.* 2009). Individuals differ in this connectivity, and such individual differences are associated with individual differences in language development. For example, Beckage *et al.* (2011) investigated the structure of the semantic networks in children aged 15 to 36 months who had either typical vocabulary size for their age or whose vocabulary size was small relative to their age (i.e., children who were late talkers). They found that the semantic networks of typical language learners had a greater degree of connectivity and clustering compared to late talkers, even when their overall vocabulary size was the same.

The semantic network is important for many aspects of language processing, and different experimental tasks can be used to investigate its structure. One such task – which can be used across the age span, from children to older adults, and is widely administered in clinical settings – is semantic fluency. Semantic fluency tasks require people to produce as many words as they can which belong to certain categories, such as ‘animals’, in a limited period of time (e.g., 60 s). Productivity declines as the test period increases, and especially after the first 15 s have elapsed (e.g., Henry *et al.* 2015), suggesting that retrieval becomes harder during the course of the test period. In semantic fluency tasks, responses are often produced in clusters of semantically-related items (e.g., Troyer *et al.* 1997). Semantic clusters provide a measure of lexical organization on the basis that producing successive semantically-related items reveals relations between items in a semantic network. It also reveals that lexical retrieval is facilitated by inter-item associations between a given item and other items whose semantic representations partly overlap. For example, in the ‘animals’ category, the retrieval of ‘cat’ might facilitate the retrieval of ‘dog’ since both items are ‘mammals’ and ‘pets’. From individuals’ responses, both thematic (e.g., ‘pets’ in the animals category; 'breakfast foods' in the food category) and taxonomic (e.g., 'birds' in the animals category; 'fruits' in the food category) semantic clusters can be identified. Given the time pressure of the task, once the retrieval of items within a cluster slows down, it is most efficient to quickly move to another cluster, from which other items will be available for more rapid retrieval. This process is referred to as switching. Thus, response patterns in semantic fluency tasks reflect lexical organization and lexical retrieval processes (e.g., Marshall *et al.* 2018).

Dyslexia and Developmental Language Disorder (DLD) are two neurodevelopmental disorders affecting the development of literacy and oral language skills, respectively. With respect to semantic fluency performance in dyslexia, inconsistent findings have been reported, with studies reporting that children with dyslexia scored significantly lower on semantic fluency tasks compared to their age-matched peers (e.g., in German: Reiter *et al.* 2005; in Italian: Varvara *et al.* 2014), and other studies reporting no group differences (e.g., in English: Griffiths 1991, Brosnan *et al.* 2002; in German: Frith *et al.* 1995, Landerl *et al.* 2009; in Italian: Marzocchi *et al.* 2008). The findings originating from children with DLD, however, show consistently poorer performance compared to their age-matched peers (in English: Weckerly *et al.* 2001, Henry *et al.* 2012, 2015). It has been reported that children with DLD know fewer words, show a limited depth of word knowledge, and have difficulty learning new words (see Jackson et al. 2019 and Marshall 2014 for reviews), so their poorer semantic fluency performance is not surprising.

Understanding the underlying cause of lower semantic fluency scores of children with dyslexia and DLD would indicate whether lexical difficulties can be explained by poor semantic structure or slower retrieval processes. To our knowledge, the only published study which has investigated clustering and switching in children with dyslexia reported that Polish adolescents with dyslexia aged 16-18 years produced fewer clusters and switched fewer times compared to controls (Mielnik *et al.* 2015). There was no difference in the size of the clusters produced by the two groups, however. With respect to DLD, Weckerly *et al.* (2001) found that children with DLD aged 8-12 years produced significantly fewer clusters compared to TD children, but the two groups did not differ on the number of switches or on cluster size. The authors interpreted these findings as evidence that the organization of the lexicon did not differ between the TD group and the clinical groups. Similar findings have been reported by Henry *et al.* (2015) for English-speaking children with DLD and by Marshall *et al.* (2013) for deaf children with DLD who used British sign language.

We therefore argue that there is only limited data originating from child samples of dyslexia and DLD reporting patterns of lexical retrieval in semantic fluency tasks in languages other than English. There is also no published study in the Greek language assessing children with dyslexia and/or DLD in semantic fluency tasks. There is an adult study by Kosmidis *et al.* (2004) reporting that word productivity in healthy adults was strongly associated with the number of switches but only weakly associated with the size of clusters. The current study aims to fill this gap in the developmental literature by testing semantic fluency performance in children with dyslexia and DLD speaking Greek.

Greek has a shallow orthography, which means that it is characterized by consistent grapheme-to-phoneme mappings (Seymour *et al.* 2003), estimated to be 95% consistent for reading and 80% consistent for spelling (Protopapas and Vlahou 2009). Considering this high level of orthographic consistency for reading, it is not surprising that reading difficulties are evident primarily in poor reading fluency rather than poor reading accuracy (Nikolopoulos *et al.* 2003). Having said that, reading accuracy difficulties are evident in children with dyslexia even in Grade 7 (Protopapas and Skaloumbakas 2007, Protopapas *et al.* 2008, 2012). With respect to phonological difficulties, children with dyslexia and DLD have been reported to show phonological deficits in tasks measuring phonological awareness, phonological short-term memory and rapid automatic naming skills (e.g., Talli *et al.* 2016, Diamanti *et al.* 2018, Spanoudis *et al.* 2018). Research in the semantic processing skills of Greek children with dyslexia and/or DLD is, however, almost entirely lacking. Spanoudis *et al.* (2018) argued that Greek children with DLD experience difficulties in assimilating new information through printed text because of poor vocabulary and semantic skills. Further research is needed to investigate the profile of Greek children with dyslexia and DLD in tasks that assess lexical organization and patterns of lexical retrieval, such as semantic fluency.

Two models accounting for poorer semantic fluency performance, namely the Structure-Loss Model and the Retrieval-Slowing Model (Rohrer *et al.* 1995), have been developed on the basis of *adult* data and have been tested only in adult samples. We argue that developmentally, the concept of loss of structure should be replaced by the concept of a less sophisticated network of semantic connections, as proposed by Messer and Dockrell (2006). This implies that in the course of development, children with dyslexia and DLD have experienced difficulties in establishing lexical-semantic representations, resulting in an impoverished network of semantic connections. For the purpose of the current study, this model is called the Poor Lexical-Semantic Structure Model. Further, in the current study, the Retrieval-Slowing Model is called instead the Slow-Retrieval Model. In the context of adult samples, 'slowing' means that retrieval speeds were once normal but have slowed down. In developmental disorders, however, the model implies that retrieval was never at normal speed. The Slow-Retrieval Model, adapted from the Retrieval-Slowing Model, attributes lower semantic fluency performance to an access deficit which renders lexical items difficult to retrieve. The access deficit originates from slow retrieval processes of lexical items stored in the lexicon, while the structure of the semantic memory retrieval processes is preserved (Lenio *et al.* 2016).

Turning now to the predictions of the current study, both models predict that children with dyslexia and DLD will show lexical retrieval difficulties exemplified in the retrieval of fewer items and in the production of fewer clusters compared to TD children. The pattern of retrieval, however, is predicted to be different for the two models. The Poor Lexical-Semantic Structure Model predicts that children with dyslexia and DLD will cluster their responses around subcategories of a smaller average cluster size than TD children. In contrast, the Slow-Retrieval

Model predicts a similarly-sized average cluster in TD children and children with dyslexia and DLD.

The assumption is that in a sophisticated network of semantic connections, as illustrated in figure 1a for TD children, individuals are able to cluster their responses around subcategories based on the inter-item associations between items whose semantic representations partly overlap. In the 'animals' category, for example, an individual could produce a cluster of 'lion-tiger-giraffe-elephant' since all four items might be connected to each other under the subcategory of 'safari animals' in their semantic network. The richer those semantic connections are – in terms of their number and their strength – the easier it will be for individuals to retrieve items belonging to a subcategory, and therefore the greater the number of items produced belonging to that subcategory (i.e., cluster size) will be. Thus, lexical retrieval is facilitated by the inter-item associations between words. Even if, according to the Slow-Retrieval Model in figure 1b, lexical retrieval is slower in children with dyslexia and DLD (DDLD) compared to TD children, cluster size should not differ between the TD group and the DDLD group. However, the number of items retrieved overall in the task and the number of clusters will differ. This is because slower retrieval means that children with DDLD will retrieve a smaller number of items in the limited time permitted and will not make it as far through the semantic network in their search for lexical items as TD children. According to the Poor Lexical-Semantic Structure Model, however, in an impoverished network of semantic connections, as illustrated in figure 1c, individuals with DDLD are predicted to have fewer and weaker semantic connections between semantically-related items, which will result in fewer clusters and in clusters of a smaller size being produced, and in fewer words being retrieved overall. The size of clusters retrieved during the semantic fluency task is therefore considered to reflect the structure of the semantic lexicon, and the two models make different predictions with respect to cluster size. The legend for figures 1a-1c attempts to explain more clearly how these predictions arise from the models.

Figures 1a-lc to be inserted about here

Another issue concerns whether children's language, literacy, and phonological skills have an impact on how efficiently they retrieve lexical items from the mental lexicon. In the current study, the impact of language, literacy, and phonological skills on word productivity in semantic fluency tasks is considered. If it is the case that children's word productivity in semantic fluency tasks is predicted by language, literacy, and phonological skills, then poorer semantic fluency performance in children with DDLD will be partly attributed to their inferior language, literacy, and phonological skills. To this end, the associations between a range of language, literacy, and phonological tasks to semantic fluency performance will be first investigated, and next, how much of the variance in semantic fluency is explained by children's language, literacy, and phonological skills will be explored in a regression model. The rationale for this investigation is as follows. Nation (2017) proposed the Lexical Legacy Hypothesis which argues that reading experience allows a reader to read words in different semantic contexts, and this leads to a rich and nuanced database about a word and its connections to other words. This hypothesis therefore states that word knowledge is based on lexical co-occurrence in the sense that a word is known as it is related in meaning with other words. In a similar vein, the Lexical Quality Hypothesis claims that "a lexical representation has high quality to the extent that it has a fully specified orthographic representation (a spelling) and redundant phonological representations (one from spoken language and one recoverable from orthographic-to-phonological mappings)" (Perfetti and Hart 2001, p. 68). Therefore, lexical quality concerns the knowledge of the form and the meaning of the word, leading to rapid processing (Perfetti and Hart 2001). The origin of high-quality representations may be therefore sought in the amount of experience with both oral and written language. Concomitantly, this suggests a relation between children's lexical-semantic representations and their language, literacy, and phonological skills. On the basis of this argument, a valid index of lexical quality is performance on semantic fluency tasks. In support of the Lexical Quality Hypothesis, Dyson *et al.*'s (2017) intervention study suggested that it is accessing the meaning of a word after pronouncing it correctly that improves children's ability to learn to read.

Overall, the current study sets out to answer the following research questions about semantic fluency in Greek-speaking children with DDLD:

- What is the structure of the lexicon in children with DDLD compared to TD children? Is poorer semantic fluency performance in children with DDLD better explained by impoverished semantic structure or slower retrieval processes?
- What drives productivity in semantic fluency tasks in TD children and children with DDLD?
- Does semantic fluency performance relate to children's language, literacy, and phonological skills?

Methods

Participants

Participants were 149 children (66 children with dyslexia and/or DLD (43 males), and 83 TD children (35 males)), who were monolingual Greek speakers. We recruited children with dyslexia and/or DLD based on previous clinical diagnoses: they were selected on the basis that they had received a diagnosis because of persistent and specific reading or language problems. Thirty children with dyslexia and/or DLD had co-existing difficulties accompanying the diagnosis of persistent and specific reading or language problems, such as attention-deficit/hyperactivity disorder, developmental disorder of motor skills, articulation disorder, specific disorder in speech fluency, or dysgraphia. In line with the CATALISE consortium (Bishop *et al.* 2017), children with additional disorders were not excluded from the study given that additional disorders are considered as descriptors of a child's profile. Bishop (2017) argues that it is misleading to assume that co-occurring conditions are causes of language disorder, but that DLD should be distinguished from cases of language disorder associated with 'differentiating conditions' that have a known or likely biomedical origin, including brain injury, sensorineural hearing loss, genetic syndromes, intellectual disability and autism spectrum disorder. None of the children recruited to the current study had any of these conditions. Considering this conceptualization of DLD, it was not considered appropriate to control for co-occurring conditions in the analyses. Furthermore, the nonverbal IQ inclusion criterion for both the DDLD and TD groups was a score of 70 or above, following the CATALISE consortium (Bishop *et al.* 2017) and Norbury *et al.*'s (2016) population study which reported that children with a lower nonverbal ability (i.e., a standard score between 70 and 85) did not differ significantly in their language profile from children with an average nonverbal ability (i.e., a standard score > 85). In the current study, five children with DDLD had lower nonverbal ability (all of whom had a standard score of 75). Including in the clinical group children who had comorbid disorders and/or a nonverbal IQ standardised score between 70 and 85 can be considered as a strength of the current study on the basis that the sample is more representative of the children seen in clinics.

When recruiting children for the TD group, we excluded those who achieved a percentile score of 10 or lower on a standard text-reading fluency measure, or substantial difficulties with the language and literacy tasks (e.g., children who had substantial difficulty understanding instructions and/or whose response times were extremely slow). Nine children who had a percentile score of 10 or lower on a text-reading fluency measure, and another child who also had substantial difficulties with the language and literacy tasks, were therefore excluded from the study. None of the TD children included in the study had a current or prior history of hearing or visual deficit, neurological disease, or medication for any neurological, psychiatric, or behavioural disorder. Although our inclusion criteria permitted TD children who had a nonverbal IQ score as low as 70, none actually scored lower than 80.

Traditionally, dyslexia and DLD are viewed as separate disorders. In the current study, however, the children with dyslexia and DLD were combined in one group, the DDLD group, as proposed for this same group of participants by Mengisidou and Marshall (2019). In fact, literacy difficulties are very common in children with DLD (e.g., Conti-Ramsden *et al.* 2001), and it is the case that approximately 50% of children who fit the criteria for dyslexia also fit the criteria for DLD, and vice versa (e.g., Messaoud-Galusi and Marshall 2010, Spanoudis *et al.* 2018). In addition, there are currently no gold-standard assessments of diagnosing dyslexia and DLD with adequate psychometric properties, namely, valid and reliable assessments with diagnostic or prognostic value. In this context, Dockrell and Marshall (2015) argue that screening measures to date do not meet psychometric properties to identify language problems, and also that the interpretation of language assessments is challenged by a range of factors, including socioeconomic status, multilingualism, hearing impairment, and even the characteristics of the assessment. The last of these factors is particularly relevant to the current study.

In Greece, while there has been some progress in the development of psychometric materials over recent years (e.g., Sideridis *et al.* 2008; Vogindroukas *et al.* 2009), standardised clinical tools for the diagnosis of dyslexia and DLD for preschool- and school-aged children are still lacking. Dyslexia is therefore often diagnosed on the basis of non-standardized measures of reading and spelling ability (Anastasiou and Polychronopoulou 2009), and the same is also the case for DLD. This raises the issue of how accurately children with dyslexia, children with DLD, and children with dyslexia plus DLD can be differentiated; this might not be as easy as in studies of English-speaking children (e.g., Catts *et al.* 2005, Ramus *et al.* 2013). Further, previous research in Greek has explored the overlap between dyslexia and DLD, and reported that dyslexia and DLD show common deficits in tasks measuring reading skills and reading-related phonological skills (Talli *et al.* 2016, Spanoudis *et al.* 2018), even though they do not completely overlap. With respect to semantic fluency in particular, evidence originates from the study of Hall *et al.* (2017) who reported that participants with just a reading impairment were more fluent than participants with both language and reading impairments, while participants with language impairment did not differ from either subgroup. Hall *et al.*'s study therefore suggests that dyslexia and DLD should not be treated separately.

In the light of this evidence, a Principal Component Analysis (PCA) with oblique rotation was carried out on language and literacy scores within just the children with dyslexia and/or DLD in order to determine whether there were separate loadings onto different components that might justify keeping these children separate. If there are different components for children's language and literacy skills, it can be argued that dividing the children into separate subgroups would be appropriate. If not, then this finding would offer strong evidence suggesting that combining children with dyslexia and/or DLD in a single group is appropriate. The PCA revealed that language and literacy variables did not load on different components, which suggests that it is appropriate to combine the children with dyslexia and/or DLD into one DDLD group. This PCA is presented in detail in the Appendix.

This DDLD group had a mean age (*SD*, range) of 9.51 (1.46, 7;4-12;2) years and the TD group had a mean age of 8.37 (1.77, 6;3-12;4) years. The DDLD group was significantly older than the TD group, $t_{(147)} = -4.30$, $p < 0.001$. On the Greek standardization of the nonverbal IQ task (Sideridis *et al.* 2015) of the Raven's Coloured Progressive Matrices (CPM; Raven 2008), the mean standard score of the DDLD group was 96.74 (*SD* = 15.12) and of the TD group was 104.75 (12.94). The TD group significantly outperformed the DDLD group, $t_{(147)} = 3.48$, $p = 0.001$, as has been found in previous studies of children with literacy and language disorders (e.g., Ramus *et al.* 2013). Nonverbal IQ was not statistically controlled in the analyses, however, following Dennis *et al.* (2009) who argued that using IQ scores as a covariate is misguided and unjustified in cognitive studies with children with neurodevelopmental disorders.

In order to better appreciate the DDLD group's overall performance on language, literacy, and phonological tasks, analyses of covariance (ANCOVAs) were carried out, with the score of each task as a dependent variable, group as a fixed factor, and age in months as a covariate variable. Since semantic fluency performance was related to age in each group (see the "Results" section), age was controlled in the analyses, and estimated marginal means and estimated standard error are presented. These tasks are described later, but the data are presented here in order to provide information about the language, literacy, and phonological profile of the DDLD group compared to the TD group. Table 1 shows that the TD group significantly outperformed the DDLD group in all language tasks, in all literacy tasks except for the syllable reading task, and in all phonological tasks except for the phoneme deletion task of CVC items (C: consonant; V: vowel).

Table 1 to be inserted about here

Procedure

Ethical approval for the study was obtained from the Departmental Research Ethics Committee of the UCL Institute of Education, and from the Hellenic Ministry of Education, Research and Religious Affairs. Parents gave informed written consent on behalf of the participating children. Children were individually tested in one session of approximately 90 min in their schools, or in the referral centre where they were receiving speech and language therapy. The first author, a native Greek speaker, assessed all the children. Responses were recorded when needed, using Audacity for Windows 7 and a microphone for later transcription.

Tasks

Language, literacy, and phonological skills were assessed using a wide range of tasks in order to profile the DDLD group's language, literacy, and phonological difficulties. This paragraph explains our strategy for task selection. In the overall sample (n = 149), language skills were assessed with a widely used task of receptive vocabulary, in addition to tasks drawing upon a range of language processing skills, namely verbal comprehension, syntax comprehension, and sentence repetition. Literacy skills were assessed with reading accuracy, text-reading fluency and spelling tasks. Reading accuracy and reading fluency are sensitive measures and can reveal reading difficulties in children who are reading in the Greek consistent orthography (Diamanti *et al.* 2018). Spelling performance is another sensitive index of reading difficulty in Greek (Porpodas 1999, Protopapas and Skaloumbakas 2007). Two literacy tasks were used in the overall sample, namely text-reading fluency, as measured by the Alouette task, and spelling. However, given that there were no Greek standardised reading accuracy tasks designed for the age range of the study, first and second Graders (n = 47) were assessed with syllable and nonword reading tasks, and third to sixth Graders (n = 102) were assessed with reading accuracy and text-reading fluency tasks. In the overall sample, phonological skills were assessed with widely-used tasks assessing reading-related phonological skills, namely phoneme deletion, nonword repetition and rapid automatic naming tasks, which reveal the typical phonological deficit in children with dyslexia and DLD (e.g., Ramus *et al.* 2013).

Semantic fluency tasks

The semantic categories 'animals', 'foods', and 'objects from around the house' were used, in that order. Children were instructed to produce as many different words belonging to the target category as possible, allowing 60 s for each category. No examples were given, but 'countries' was used as a practice category. The number of correct responses retrieved for the three semantic categories was combined to create a composite semantic fluency score.

Nonverbal IQ task

The Greek standardization (Sideridis *et al.* 2015) of the Raven's CPM (Raven 2008) was used to assess children's nonverbal ability. Children were instructed to look carefully a visual design with one missing part and to choose the image that fits the missing part of the design. One point was scored for each correct answer, and a total raw score was computed which was then converted to a standardised score.

Language ability tasks

Verbal comprehension subtasks

Children's verbal comprehension was assessed with the Similarities and Vocabulary subtasks of the Greek standardization (Georgas *et al.* 1997) of the Wechsler Intelligence Scale for Children (WISC-III; Wechsler 1991). For the similarities subtask, children had to identify how two words are alike (maximum score = 33). For the vocabulary subtask, children were asked to define words (maximum score = 60). In both subtasks, responses scored two, one, or zero points, with the difference in scores reflecting the quality (accuracy and detail) of the response given.

Syntax comprehension and sentence repetition subtasks

Two subtasks of the Diagnostic Verbal Intelligence (DVIQ) Test (Stavrakaki and Tsimpli 2000) were used, namely syntax comprehension and sentence repetition. For syntax comprehension, children were required to listen carefully to a sentence and to select the picture that best depicted the meaning of the sentence. The child's score was the number of correctly-selected pictures (maximum = 17). For sentence repetition, children heard a sentence and were asked to repeat it as accurately as possible (maximum = 30).

Peabody Picture Vocabulary Test-Revised (PPVT-R)

The Greek non-standardised version (Simos *et al.* 2011) of the PPVT-R (Dunn and Dunn 1981) was used to assess children's receptive vocabulary. Children were provided orally with a word and were instructed to decide which of the pictures provided best represented its meaning. The child's score was the number of correctly selected pictures (maximum = 173).

Literacy ability tasks

The Alouette task

In the overall sample, text-reading fluency was assessed with the Alouette task (Lefavrais 1967), which has been adapted into Greek by Talli *et al.* (2016). Children were required to read as accurately and as fast as possible a 271-word text bearing no meaning. The number of words read correctly within 3 min was recorded for each child.

Reading accuracy and text-reading fluency subtasks (for children from Grade 3 to 6)

The reading accuracy and text-reading fluency subtasks of the Reading Test Alpha (Panteliadou and Antoniou 2007) were used to assess, respectively, reading accuracy and reading fluency in children from Grade 3 to 6. In the reading accuracy subtasks, children were asked to read as accurately as possible the presented words and nonwords, and to report aloud only the real words of the lexical decision subtask. The reading accuracy score was the number of words and nonwords read correctly (maximum = 77), alongside the number of words and nonwords identified as such in the lexical decision subtask (maximum = 36). In the text-reading fluency subtask, children were required to read as accurately and fast as possible a text of 279 words for 60 s. The reading fluency score was the number of words read correctly within the time limit.

Syllable and nonword reading subtasks (for children from Grade 1 to 2)

The syllable and the nonword reading subtasks of the Test of Detection and Investigation of Reading Difficulties (Porpodas 2007) contained 24 syllables and 24 nonwords, respectively. A child's score for each subtask was the number of syllables and nonwords read correctly (maximum = 24 for each subtask).

Spelling-to-dictation task

The spelling-to-dictation task (Sideridis *et al.* 2008) was another literacy task which was administered to the entire sample. The task consisted of 60 words presented orally in the context of a short sentence. First, the word is read aloud, then the sentence including the target word, and then the word is read aloud again. Any word with correct spelling scored one point and the task was discontinued after six consecutive incorrect spellings.

Phonological ability tasks

Phoneme deletion tasks

Children's phonological awareness skills were assessed with three phoneme deletion tasks of the computerised battery Evaluation de la Lecture (EVALEC; Sprenger-Charolles *et al.* 2005), which has been adapted into Greek by Talli *et al.* (2016). The first two tasks contained 12 monosyllabic items each of CVC or CCV syllable structure, respectively, and the third task contained 10 trisyllabic items of CVCVCV syllable structure. Children had to produce the nonword without the initial consonant or consonant cluster. Three nonwords were given as practice trials for each task and children asked to produce them without their first sound. A child's score was the total number of correct responses in each task.

Nonword repetition task

The EVALEC's nonword repetition task has been adapted from Sprenger-Charolles *et al.* (2005; see Talli *et al.* 2016) and required children to repeat 24 nonwords. Nonwords increased in length from three to six syllables. Three nonwords were given as practice trials and children asked to repeat them. The number of nonwords repeated correctly was the child's score.

Rapid automatic naming task

Rapid automatic naming was assessed with the picture naming subtask of the Phonological Assessment Battery (PhAB; Frederickson *et al.* 1997). The measure contains two cards of five pictures repeated ten times on each card. Prior to testing, the experimenter named the pictures to familiarize children with the task. Children were required to name the pictures as fast as possible while trying not to make any mistakes. Children's score was the average naming time (in seconds) taken for the two cards.

Coding of semantic fluency responses

Responses were timed (i.e., the total number of responses in the first 15 s and in the subsequent 45 s of the test period was calculated) and coded as correct or incorrect. There were three types of incorrect responses scoring zero points, namely repetitions, intrusions (i.e., real but irrelevant words to the target category, e.g., 'balcony', 'wall' in the category of 'objects from around the house'), and unintelligible responses (i.e., made-up words or words which could not be transcribed). All other responses were considered as correct and scored one point each. Aside from those items repeated exactly as before, different noun forms (e.g., 'apple' followed by 'apples') were counted as correct. The rationale for this is that children were asked to try to avoid producing the same word, but they were not instructed to avoid different forms of the same word.

The number of semantic clusters was computed for each semantic category. For example, in the category of 'animals', PIG followed by CHICKEN followed by GOAT belonged to the subcategory of 'farm animals'. Subcategories were driven by the data and were non-exclusive, as suggested by Marshall *et al.* (2013). Repeated responses, if any, were counted in computing the number and the size of clusters. The rationale is that even repeated responses might have aided children's semantic clustering. Given that neither intrusions nor unintelligible responses could contribute to a cluster, they were not relevant for computing the number and the size of clusters. Cluster size was also computed for each semantic category, and it was counted beginning with the first item in a cluster (i.e., a two-item cluster was given a size of 2). An average cluster size was then computed based on clusters sizes of all three semantic categories. Switches were counted as the number of transitions between semantic clusters but also between non-clustered responses.

Results

Statistical analyses were carried out using statistical package SPSS 24. We present the results in two parts. The first part considers the groups' performance and group differences on semantic fluency tasks including patterns of lexical retrieval (clustering, switching, and average cluster size), types of incorrect responses, and responses in the first 15 s and in the subsequent 45 s of the test period. In the second part, the relationship between semantic fluency and children's language, literacy, and phonological skills is investigated.

Groups' performance and group differences on semantic fluency tasks

The number of correct responses correlated strongly with age in the TD group, $r_{(83)} = 0.635$, $p < 0.001$, and moderately in the DDLD group, $r_{(66)} = 0.422$, $p < 0.001$, as presented in figure 2. Nonverbal IQ performance correlated weakly with semantic fluency in the TD group, $r_{(80)} = 0.238$, $p = 0.031$, but not in the DDLD group, $r_{(63)} = 0.189$, $p = 0.132$.

Figure 2 to be inserted about here

In order to understand whether semantic fluency performance in each group was related to the production of a greater number of clusters or to the production of bigger clusters, correlations between the number of correct responses and the number of clusters, the number of switches, and average cluster size were investigated. In the TD group, the number of correct responses correlated strongly with the number of clusters, $r_{(80)} = 0.647$, $p < 0.001$, and the number of switches, $r_{(80)} = 0.728$, $p < 0.001$, but not with average cluster size, $r_{(80)} = 0.108$, $p = 0.336$. Likewise, in the DDLD group, the number of correct responses correlated with cluster number, $r_{(63)} = 0.779$, $p < 0.001$, and the number of switches, $r_{(63)} = 0.762$, $p < 0.001$, but not with average cluster size, $r_{(63)} = 0.006$, $p = 0.964$. Thus, in both groups, productivity in semantic fluency tasks is driven by the production of more clusters and more switches, but not by bigger clusters. ANCOVAs (table 2) were carried out to assess group differences with semantic fluency variables as dependent variables, group as a fixed factor, and age in months as a covariate variable. Since the DDLD group was significantly older than the TD group, estimated marginal means and estimated standard error are presented.

Table 2 to be inserted about here

ANCOVAs revealed that the TD group significantly outperformed the DDLD group with respect to the mean total number of responses, mean number of correct responses, mean number of incorrect responses, mean number of total responses in the first 15 s, mean number of total responses in the subsequent 45 s, and mean number of clusters. There were no group differences for any of the types of incorrect responses, for the mean number of switches or for average cluster size.

Relationship between semantic fluency and language, literacy, and phonological skills

Table 3 presents the partial correlations (controlling for age) between the number of correct responses produced in the semantic fluency tasks and the scores for the language, literacy, and phonological tasks. Correlations are reported for the overall sample and for the DDLD and TD groups separately. In the overall sample, performance on WISC Similarities and Vocabulary subtasks correlated most strongly with semantic fluency. Syntax comprehension, sentence repetition, receptive vocabulary, Alouette, spelling-to-dictation, nonword repetition, and rapid automatic naming also correlated with semantic fluency. Text-reading fluency (as measured by Reading Test Alpha), reading accuracy, syllable reading, nonword reading, and all three phoneme deletion tasks did not correlate with semantic fluency.

Table 3 to be inserted about here

In order to investigate further the relationship between semantic fluency and language, literacy, and phonological tasks in the overall sample, raw scores of all five language, the two literacy, and the two phonological tasks that correlated significantly with semantic fluency in the overall sample were converted to *z*-scores. *Z*-scores were computed relative to the TD group's mean and *SD* for each task, with the mean *z*-score being equal to 0 and *SD* equal to 1 for all tasks. *Z*-scores of all 9 tasks associated significantly with semantic fluency were entered into the PCA with oblique rotation. The PCA was carried out to explore the number of components to enter in the regression analyses models presented next. A component consists of tasks that are correlated, with each component accounting for an amount of variance in the dataset. The amount of variance explained by a component is expressed by its eigenvalue. The dataset was suitable for the PCA: Kaiser-Meyer-Olkin Measure of Sampling Adequacy value was 0.876, meeting Kaiser's (1974) criterion for this value, Bartlett's Test of Sphericity value was significant ($p < 0.001$; Bartlett 1954), and most of the intercorrelations observed among all 9 tasks of interest had a value of 0.30 and above.

The PCA revealed that five language tasks (WISC Similarities, WISC Vocabulary, syntax comprehension, sentence repetition, and receptive vocabulary), two literacy tasks (Alouette and spelling-to-dictation), and two phonological tasks (nonword repetition and rapid automatic naming) loaded onto component 1, and that sentence repetition and nonword repetition also loaded onto component 2. Table 4 presents each task's contribution to components 1 and 2, which is expressed by its loading value. WISC Vocabulary, spelling-to-dictation and Alouette had the highest loadings onto the first component, while nonword repetition and sentence repetition had the lowest loadings onto this component.

Table 4 to be inserted about here

As illustrated in figure 3, only the first component had an eigenvalue larger than 1 (i.e., 5.27), meeting Kaiser's (1974) criterion. The second component had an eigenvalue of 1.0 and accounted for 11% of the variance in all 9 tasks, while the remaining components had an eigenvalue lower than 1.0. The PCA was therefore repeated, and a one-factor solution was selected. Component 1 had an eigenvalue of 5.27 and explained 58.65% of the variance in all 9 tasks.

Figure 3 to be inserted about here

Next, a regression analysis was carried out in the overall sample with semantic fluency performance as the dependent variable, and age and component 1 as the predictors. Age was entered in the first block, and component 1 in the second block. This model was significant, $F_{(2, 143)} = 39.090$, $p < 0.001$, accounting for 35.3% of the variance in semantic fluency performance. Age and component 1 were significant predictors; age: *Beta* = 0.509, $t = 7.095$, $p < 0.001$; component 1: *Beta* = 0.378, $t = 4.571$, $p < 0.001$. Component 1 accounted for 9.4% of the variance in semantic fluency performance. The results demonstrate that children's language, literacy, and phonological skills significantly predict semantic fluency performance after controlling for age.

Discussion

The objectives of the current study were to investigate semantic fluency in Greek children with DDLD aged 7-12 by comparing DDLD and TD children's clustering strategies, towards teasing apart two theoretical models accounting for lexical difficulties in semantic fluency tasks. Children with dyslexia and/or DLD were combined in one group, the DDLD group, based on strong evidence originating from a Principal Component Analysis carried out on language and literacy scores within just the children with dyslexia and/or DLD. This analysis revealed that language and literacy variables did not load on different components in these children, which suggests that it is appropriate to combine them into a single DDLD group. We also investigated how semantic fluency performance is related to DDLD and TD children's language, literacy, and phonological skills.

Three semantic categories, namely 'animals', 'foods', and 'objects', were used, and a composite semantic fluency score was computed. In both groups, semantic fluency performance was driven by the number of clusters and the number of times children switched between clustered- and/or non-clustered responses, but not by bigger clusters. Children with DDLD produced fewer correct responses than TD children after controlling for age. This finding is consistent with previous studies in children with dyslexia in languages using transparent orthographies other than Greek (e.g., in German: Reiter *et al.* 2005; in Italian: Varvara *et al.* 2014), and in English-speaking children with DLD (Weckerly *et al.* 2001, Henry *et al.* 2012, 2015). Although the TD group produced significantly fewer incorrect responses than the DDLD group, both groups produced low numbers of incorrect responses (repetitions, intrusions, and unintelligible responses).

Further, children with DDLD produced significantly fewer clusters than TD children, but the two groups did not differ significantly on the number of switches or on average cluster size. The findings suggest that children with DDLD produced fewer correct responses compared to TD children because they did not retrieve as many clusters, and not because their clusters were smaller. The findings in turn imply that children with DDLD have difficulty accessing lexical items and that they do not make it as far through the semantic network in their search for lexical items as TD children in this time-constrained task, but that their semantic representations are typical given that they cluster their responses around subcategories of a similar size to those of TD children.

This is the first study which considered cluster size in children with dyslexia and/or DLD, and we found a similar average cluster size between children with and without DDLD. This is despite the DDLD group having poorer vocabularies as measured by a task of receptive vocabulary, a task of word similarity, and a word definitions task. Previous evidence reported no significant difference in cluster size between adolescents with dyslexia aged 16-18 years and controls (Mielnik *et al.* 2015). The finding also replicates the findings of Weckerly *et al.* (2001) and Marshall *et al.* (2013) that cluster size did not differ in English-speaking children with DLD and deaf children with DLD who used British sign language, respectively, compared with age-matched TD children. We interpret the data as indicating that the semantic network in the lexicon of children with DDLD is established age-appropriately, but that lexical items are retrieved more slowly, as proposed by the Slow-Retrieval Model, which explains why they produce fewer correct responses in semantic fluency tasks.

Children's language, literacy, and phonological skills were associated with semantic fluency scores. The PCA revealed one component defined by tasks of verbal comprehension, syntax comprehension, sentence repetition, receptive vocabulary, text-reading fluency, spelling, nonword repetition, and rapid automatic naming. This component was a significant predictor of semantic fluency performance after controlling for age, explaining 9.4 and 8.4% of the variance in semantic fluency performance in the overall sample and in the DDLD group, respectively. This suggests that children with DDLD achieved lower semantic fluency scores than TD children partly because they had poorer language, literacy, and phonological skills. This finding in turn suggests that the retrieval of fewer responses in semantic fluency tasks originating from slow lexical retrieval in children with DDLD is partly attributed to DDLD children's inferior language, literacy, and phonological skills.

An important strength of the study is its inclusion of a large overall sample, which covers a wide range of scores, and large numbers of children in each participating group to study the relationship between measures of interest. This allows us to investigate the associations between measures of interest without over-estimating the size of any association and avoids results of low statistical power which yield many false positive results (West *et al.* 2018). Limitations are that although the Slow-Retrieval Model is supported by the empirical evidence we have presented regarding cluster size, we have not been able to identify exactly what the reason(s) for this slow retrieval might be. As Messer and Dockrell (2006) have argued in the context of children with word-finding difficulties, word-retrieval problems can potentially be caused by impairments in phonology, semantics, or processing speed. Distinguishing between these possibilities is not straightforward: attempting to do so requires both well-designed tasks and appropriate developmental models of lexical retrieval (Messer and Dockrell 2006). Similarly, Marshall (2014) has argued in the context of semantic fluency performance in deaf children with DLD who use British sign language, that it is not clear whether slow retrieval is due to slower access to the sign's semantic representation, or to less efficient mapping between the semantic representation and the phonological form of the sign.

The battery of language and phonological tasks that we administered offers some clues as to what the reasons for slow retrieval might be, but strong conclusions cannot be drawn. That children with DDLD had poorer performance on the word similarity and word definitions tasks, and performance on these tasks correlated with semantic fluency scores, suggests poor semantic representations. The DDLD group also performed more poorly on the nonword repetition and rapid automatic naming tasks, and they too correlated with semantic fluency scores, which suggests a phonological impairment. And of course, the underlying problem might be different for different children. A limitation of the design of the study is that we recruited children with DDLD based on previous clinical diagnoses. However, we did not have access to the full record of the types of assessments used for each child, along with their scores, and as such we were not in a position to distinguish children with dyslexia from those with DLD. That limitation led us to combine children with different diagnoses together in a single group, as justified in the *Participants* section. However, careful selection of children who fall into more distinct groups (e.g., Ramus *et al.* 2013) – assuming that adequate assessments of language and literacy skills exist in the language to differentiate these groups – might reveal that children with different profiles of literacy and language impairments have different loci for their retrieval difficulties. Ultimately, of course, only well-designed intervention studies are able to offer firm conclusions about causal links. To this end, Ebbels *et al.* (2012) revealed a positive effect of a semantic therapy on word-finding difficulties in children with language disorders aged 9-15, as exemplified in a word-finding test. It remains to be investigated though whether a semantic therapy designed to improve children's lexical retrieval processes has a positive effect on semantic fluency tasks in particular. Similarly, using either phonological or semantic approaches to intervention for word-finding difficulties, as devised by Best *et al.* (2017), might help identify where the locus of the word retrieval difficulty is and whether focused intervention improves semantic fluency performance.

Conclusions

The objective of the current study was to test two theoretical models accounting for lexical difficulties in semantic fluency tasks by investigating the structure of semantic fluency in a sample of Greek children with DDLD. Clustering and switching behaviour was associated with productivity in the TD and DDLD groups, revealing that the semantic lexicon is organized in a similar way in both groups. Given that a similarly-sized average cluster was found in the two groups, the findings indicated that slower retrieval processes of lexical items from the mental lexicon resulted in poorer semantic fluency performance in children with DDLD relative to their TD peers. We conclude that the lexical retrieval difficulties experienced by children with DDLD in semantic fluency tasks are better explained by the Slow-Retrieval Model than by the Poor Lexical-Semantic Structure Model.

Acknowledgements

We would like to thank all participating children, parents, teachers and speech and language therapists. We also thank Dr Ioanna Talli for her support with the recruitment of participating children, and for her permission to use tasks of the EVALEC battery adapted into Greek.

References

ARÁN-FILIPPETTI, V. A. and ALLEGRI, R. F., 2011, Verbal fluency in Spanish-speaking children: analysis model according to task type, clustering, and switching strategies and performance over time. *Clinical Neuropsychology*, **25**, 413-436.

BARTLETT, M. S., 1954, A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society*, **16**, 296-298.

BECKAGE, N., SMITH, L. and HILLS, T., 2011, Small worlds and semantic network growth in typical and late talkers. *PLoS One*, **6**, e19348.

BEST, W., HUGHES, L. M., MASTERSON, J., THOMAS, M., FEDOR, A., RONCOLI, S., FERN-POLLAK, L., SHEPHERD, D. L., HOWARD, D., SHOBBROOK, K. and KAPIKIAN, A., 2017, Intervention for children with word-finding difficulties: a parallel group randomised control trial. *International Journal of Speech-Language Pathology*, **31**, 1-12.

BISHOP, D. V. M. (2017). Why is it so hard to reach agreement on terminology? The case of developmental language disorder (DLD). *International Journal of Language & Communication Disorders*, **52**, 671-680.

BISHOP, D. V., SNOWLING, M. J., THOMPSON, P. A., GREENHALGH, T. and CATALISE-2 consortium., 2017, Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, **58**, 1068-1080.

BROSNAN, M., DEMETRE, J., HAMILL, S., ROBSON, K., SHEPHERD, H. and CODY, G., 2002, Executive functioning in adults and children with developmental dyslexia. *Neuropsychologia*, **40**, 2144-2155.

CATTS, H. W., ADLOF, S. M., HOGAN, T., and WEISMER, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, **48**, 1378-1396.

COLLINS, A. M. and LOFTUS, E. F., 1975, A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407-428.

DENNIS, M., FRANCIS, D. J., CIRINO, P. T., SCHACHAR, R., BARNES, M. A. and FLETCHER, J. M., 2009, Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, **15**, 331-343.

DIAMANTI, V., GOULANDRIS, N., CAMPBELL, R. and PROTOPAPAS, A., 2018, Dyslexia profiles across orthographies differing in transparency: An evaluation of theoretical predictions contrasting English and Greek. *Scientific Studies of Reading*, **22**, 55-69.

DUNN, L. M. and DUNN, D. M., 1981, *Peabody Picture Vocabulary Test-Revised* (Circle Pines, MN: American Guidance Service).

DYSON, H., BEST, W., SOLITY, J. and HULME. C., 2017, Training mispronunciation correction and word meanings improves children's ability to learn to read words. *Scientific Studies of Reading*, **21**, 392-407.

EBBELS, S. H., NICOLL, H., CLARK, B., EACHUS, B., GALLAGHER, A. L., HORNIMAN, K., JENNINGS, M., MCEVOY, K., NIMMO, L. and TURNER, G., 2012, Effectiveness of semantic therapy for word-finding difficulties in pupils with persistent language impairments: a randomized control trial. *International Journal of Language & Communication Disorders*, **47**, 35-51.

FREDERICKSON, N., FRITH, U. and REASON, R., 1997, *Phonological Assessment Battery* (Windsor, UK: NFER-Nelson).

FRITH, U., LANDERL, K. and FRITH, C., 1995, Dyslexia and verbal fluency: More evidence for a phonological deficit. *Dyslexia*, **1**, 2-11.

GEORGAS, D., PARASKEVOPOULOS, I., MPEZEVEGKIS, I. and GIANNITSAS, N., 1997, *Greek WISC-III: Wechsler Intelligence Scales for Children* (Athens, Greece: Ellinika Grammata).

GRIFFITHS, P., 1991, Word-finding ability and design fluency in developmental dyslexia. *British Journal of Clinical Psychology*, **30**, 47-60.

HALL, J., MCGREGOR, K. K. and OLESON, J., 2017, Weaknesses in lexical-semantic knowledge among college students with specific learning disabilities: evidence from a semantic fluency task. *Journal of Speech, Language, and Hearing Research*, **60**, 640-653.

HENRY, L. A., MESSER, D. J. and NASH, G., 2015, Executive functioning and verbal fluency in children with language difficulties. *Learning and Instruction*, **39**, 137-147.

HENRY, L. A., MESSER, D. J. and NASH, G., 2012, Executive functioning in children with specific language impairment. *Journal of Child Psychology and Psychiatry*, **53**, 37-45.

HILLS, T. T., MAOUENE, M., MAOUENE, J., SHEYA, A. and SMITH, L. B., 2009, Categorical structure among shared features in networks of early-learned nouns. *Cognition*, **112**, 381-396.

JACKSON, E., LEITÃO, S., CLAESSEN, M. and BOYES, M., 2019, The evaluation of word-learning abilities in people with developmental language disorder: A scoping review. *International Journal of Language & Communication Disorders*, doi.org/10.1111/1460-6984.12490

KAISER, H., 1974, An index of factorial simplicity. *Psychometrika*, **39**, 31-36.

KOSMIDIS, M. H., VLAHOU, C. H., PANAGIOTAKI, P. and KIOSSEOGLOU, P., 2004, The verbal fluency task in the Greek population: Normative data, and clustering and switching strategies. *Journal of the International Neuropsychological Society*, **10**, 164-172.

LANDERL, K., FUSSENEGGER, B., MOLL, K. and WILLBURGER, E., 2009, Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, **103**, 309-324.

LEFAVRAIS, P., 1967, *Test de l'Alouette* (Paris, France: Les Editions du Centre de Psychologie Appliquée).

LENIO, S., LISSEMORE, F. M., SAJATOVIC, M., SMYTH, K. A., TATSUOKA, C., WOYCZYNSKI, W. A. and LERNER, A. J., 2016, Detrending changes the temporal dynamics of a semantic fluency task. *Frontiers in Aging Neuroscience*, **8**, 1-9. https://doi.org/10.3389/fnagi.2016.00252

MARSHALL, C. R., 2014, Word production errors in children with developmental language impairments. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **369(1634)**, 1-8.

MARSHALL, C. R., JONES, A., FASTELLI, A., ATKINSON, J., BOTTING, N. and MORGAN, G., 2018, Semantic fluency in deaf children who use spoken and signed language in comparison with hearing peers. *International Journal of Language & Communication Disorders*, **53**, 157-170.

MARSHALL, C. R., ROWLEY, K., MASON, K., HERMAN, R. and MORGAN, G., 2013, Lexical organization in deaf children who use British sign language: Evidence from a semantic fluency task. *Journal of Child Language*, **40**, 193-220.

MARZOCCHI, G. M., OOSTERLAAN, J., ZUDDAS, A., CAVOLINA, P., GEURTS, H., REDIGOLO, D., VIO, C. and SERGEANT, J. A., 2008, Contrasting deficits on executive functions between ADHD and reading disabled children. *Journal of Child Psychology and Psychiatry*, **49**, 543-552.

MENGISIDOU, M. and MARSHALL, C. R. (2019). Deficient explicit access to phonological representations explains phonological fluency difficulties in Greek children with dyslexia and/or developmental language disorder. *Frontiers in Psychology*, **10:638**. doi: 10.3389/fpsyg.2019.00638

MESSER, D. and DOCKRELL, J. E., 2006, Children's naming and word-finding difficulties: descriptions and explanations. *Journal of Speech, Language, and Hearing Research*, **49**, 309-324.

MIELNIK, A., LOCKIEWICZ, M. and BOGDANOWICZ, M., 2015, Semantic and phonological verbal fluency in students with dyslexia. *Acta Neuropsychologica*, **13**, 253-266.

NATION, K., 2017, Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning*, **2**, Article: 3.

NIKOLOPOULOS, D., GOULANDRIS, N. and SNOWLING, M. J., 2003, Developmental dyslexia in Greek. In N. Goulandris (Ed.), *Dyslexia in different languages: Cross-linguistic comparisons* (pp. 53-67). London: Whurr.

NORBURY, C. F., GOOCH, D., WRAY, C., BAIRD, G., CHARMAN, T., SIMONOFF, E., VAMVAKAS, G. and PICKLES, A., 2016, The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry*, **57**, 1247-1257.

PANTELIADOU, S. and ANTONIOU, F., 2007, *Test Alpha* (Athens, Greece: YPEPQ-EPEAEK).

PERFETTI, C. A. and HART, L., 2002, The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.

PORPODAS, C. D., 1999, Patterns of phonological and memory processing in beginning readers and spellers of Greek. *Journal of Learning Disabilities*, **32**, 406-416.

PORPODAS, C. D., 2007, *Test of Detection and Investigation of Reading Difficulties in Kindergarten and Grades 1 and 2* (Athens, Greece: YPEPQ-EPEAEK).

PROTOPAPAS, A. and SKALOUMBAKAS, C., 2007, Traditional and computer-based screening and diagnosis of reading disabilities in Greek. *Journal of Learning Disabilities*, **40**, 15-36.

PROTOPAPAS, A., SKALOUMBAKAS, C. and BALI, P., 2008, Validation of unsupervised computer-based screening for reading disability in the Greek elementary Grades 3 and 4. *Learning Disabilities: A Contemporary Journal*, **6**, 45-69.

PROTOPAPAS, A. and VLAHOU, E. L., 2009, A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*, **41**, 991-1008.

PROTOPAPAS, A., SIMOS, P. G., SIDERIDIS, G. D. and MOUZAKI, A., 2012, The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology*, **33**, 217-240.

RAMUS, F., MARSHALL, C. R., ROSEN, S. and VAN DER LELY, H. K. J., 2013, Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, **136**, 630-645.

RAVEN, J. C., 2008, *Raven's Coloured Progressive Matrices and Vocabulary Scales* (London, UK: Pearson).

REITER, A., TUCHA, O. and LANGE, K. W., 2005, Executive functions in children with dyslexia. *Dyslexia*, **11**, 116-131.

ROGERS, T. T. and MCCLELLAND, J. L., 2008, Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, **31**, 689-714.

ROHRER, D., WIXTED, J. T., SALMON, D. P. and BUTTERS, N., 1995, Retrieval from semantic memory and its implications for Alzheimer's disease. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 1127-1139.

SEYMOUR, P. H. K., ARO, M. and ERSKINE, J. M., 2003, Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, **94**(Pt 2), 143-174.

SIDERIDIS, G., ANTONIOU, F., MOUZAKI, A. and SIMOS, P., 2015, *Raven's Educational CPM/CVS* (Athens, Greece: Motivo).

SIDERIDIS, G., MOUZAKI, A., PROTOPAPAS, A. and SIMOS, P., 2008, Psychometric evaluation of a spelling test for elementary students, *Psychologia*, **15**, 290-315.

SIMOS, P. G., SIDERIDIS, G. D., PROTOPAPAS, A. and MOUZAKI, A., 2011, Psychometric evaluation of a receptive vocabulary test for Greek elementary students. *Assessment for Effective Intervention*, **37**, 1-17.

SPANOUDIS, G. C., PAPADOPOULOS, T. C. and SPYROU, S., 2018, Specific language impairment and reading disability: Categorical distinction or continuum? *Journal of Learning Disabilities*, **52**, 3-14.

SPRENGER-CHAROLLES, L., COLE, P., BECHENNEC, D. and KIPFFER-PIQUARD, A., 2005, French normative data on reading and related skills from EVALEC, a new computerized battery of tests (end Grade 1, Grade 2, Grade 3, and Grade 4). *Revue Europeenne de Psychologie Appliquee*, **55**, 157-186.

STAVRAKAKI, S. and TSIMPLI, I. M., 2000, Diagnostic Verbal IQ Test for Greek preschool and school age children: Standardization, statistical analysis, psychometric properties. In Proceedings of the 8th Symposium of the Panhellenic Association of Logopedists (Athens, Greece: Ellinika Grammata).

TALLI, I., SPRENGER-CHAROLLES, L. and STAVRAKAKI, S., 2016, Specific language impairment and developmental dyslexia: What are the boundaries? Data from Greek children. *Research in Developmental Disabilities*, **49-50**, 339-353.

TROYER, A. K., MOSCOVITCH, M. and WINOCUR, G., 1997, Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology*, **11**, 138-146.

VARVARA, P., VARUZZA, C., SORRENTINO, A. C. P., VICARI, S. and MENGHINI, D., 2014, Executive functions in developmental dyslexia. *Frontiers in Human Neuroscience*, **8**, 1-8. https://doi.org/10.3389/fnhum.2014.00120

VOGINDROUKAS, I., PROTOPAPAS, A. and STAVRAKAKI, S., 2009, *Action pictures. Test of grammar and information accuracy* (Chania, Greece: Glafki).

WECHSLER, D., 1991, *WISC-III: Wechsler Intelligence Scale for Children* (San Antonio, TX: Psychological Corporation).

WECKERLY, J., WULFECK, B. and REILLY, J., 2001, Verbal fluency deficits in children with specific language impairment: slow rapid naming or slow to name? *Child Neuropsychology*, **7**, 142-152.

WEST, G., VADILLO, M. A., SHANKS, D. R. and HULME, C., 2018, The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*, **21**, e12552.

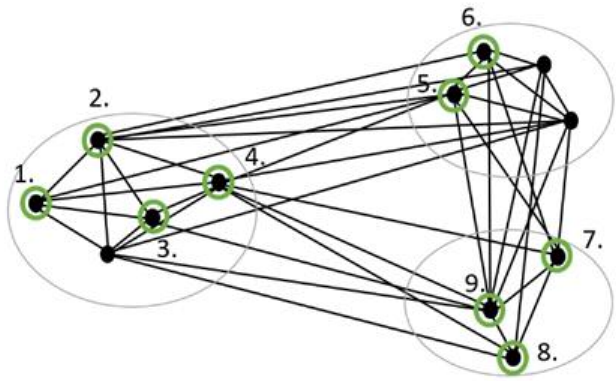

Figure 1a

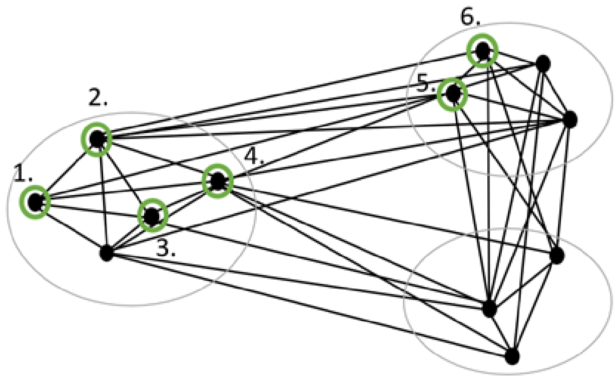

Figure 1b

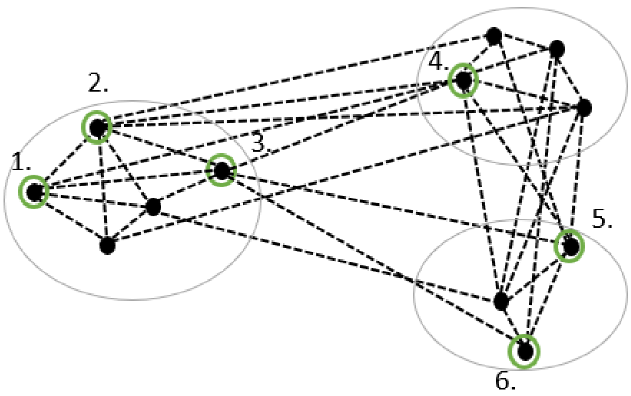

Figure 1c

Figures 1a-1c. Hypothesised semantic networks and lexical retrieval during the semantic fluency task in 1a typical development, 1b children with DDLD according to the Slow-Retrieval Model, and 1c children with DDLD according to the Poor Lexical-Semantic Structure Model.

In all figures, words are the nodes of the network, indicated by black dots. Lines between words indicate the semantic associations between words. Shorter lines indicate that words share more semantic features and are therefore closer together in semantic space. Clustering, as indicated by the grey circles, emerges from this architecture. The lines are dotted rather than solid in figure 1c, to represent the hypothesised weakness of the associations, and there are fewer lines to indicate that there are fewer associations between words. The "direction of travel" through this semantic network during the course of the semantic fluency task is shown by the green numbered circles, which indicate the order in which words are retrieved. Words retrieved from within clusters give rise to the clustering of responses that can be identified in the child's spoken output. Importantly for the predictions of the two hypotheses, clustering occurs in both figure 1b and 1c, but the average size of clusters is smaller in figure 1c ((3+2)/2=2.5) compared to figures 1a ((4+2+3)/3=3) and 1b ((4+2)/2=3). This is because the poorer semantic structure results in children being less likely to retrieve a word from within the same cluster as the previous word than they would if the semantic structure was developmentally-appropriate: the overall result is the retrieval of fewer items than is typical. Meanwhile, fewer items are retrieved in figure 1b compared to figure 1a because retrieval, while following an age-appropriate pattern retrieval, is slower than is typical.

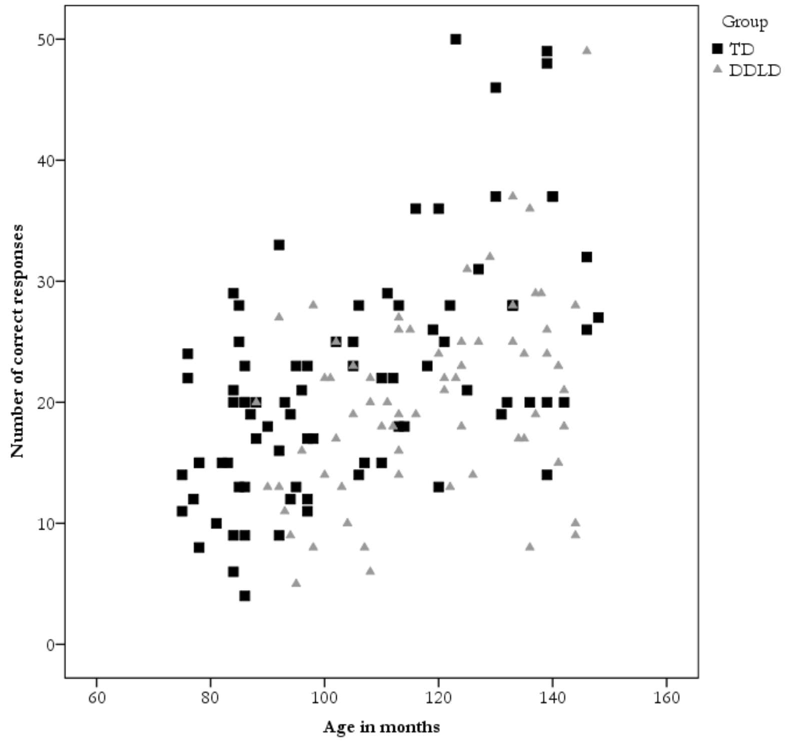

Figure 2. Scatterplot showing the association between the number of correct responses and age for the TD and DDLD groups

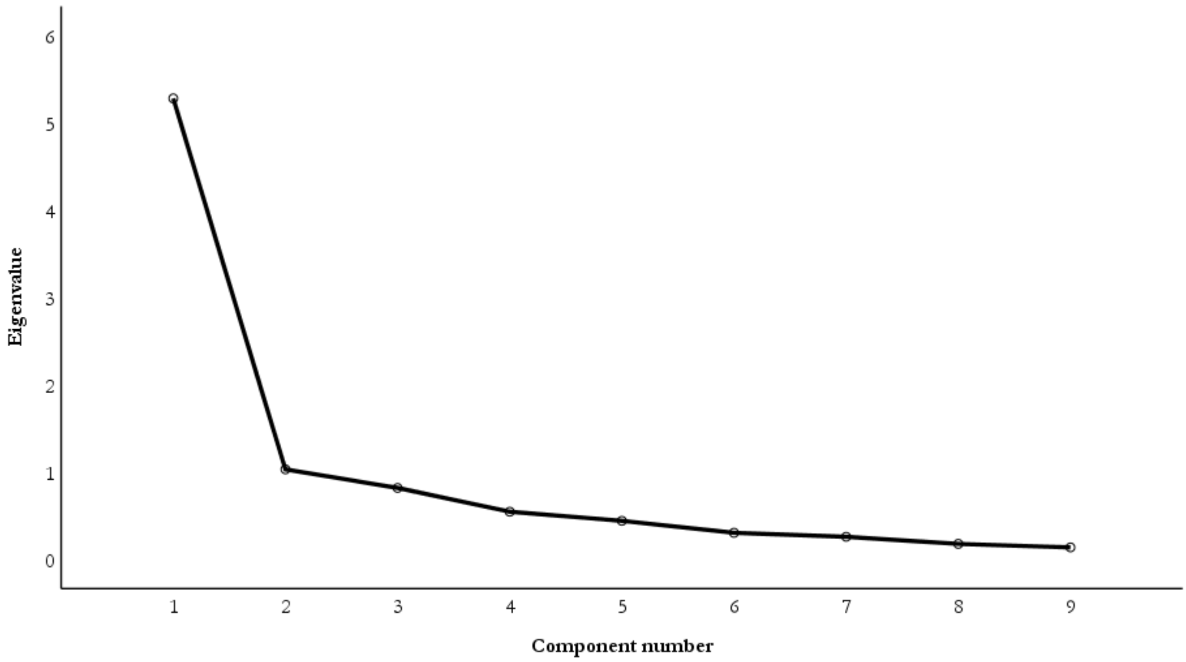

Figure 3. Scree plot showing the number of components generated by the Principal Component Analysis (PCA) with oblique rotation and each component's eigenvalue in the overall sample

Table 1. Groups' performance and group differences on language, literacy, and phonological tasks

	DDLD group		TD group					
Tasks	*e. m.* mean	*e.* SE	*e. m.* mean	*e.* SE	*n*	*F*	*p*	ηp^2
Verbal comprehension: WISC Similarities	8.16	0.42	11.13	0.37	149	26.23***	< 0.001	0.152
Verbal comprehension: WISC Vocabulary	14.79	0.59	21.61	0.52	149	69.91***	< 0.001	0.324
Syntax comprehension: DVIQ Test	12.71	0.26	13.80	0.23	149	8.37**	0.004	0.054
Sentence repetition: DVIQ Test	23.44	0.48	27.09	0.43	149	29.93***	< 0.001	0.170
Receptive vocabulary: PPVT-R	108.60	1.68	121.41	1.48	149	30.79***	< 0.001	0.174
Text-reading fluency: Alouette	105.51	5.26	177.60	4.66	149	99.62***	< 0.001	0.406
Text-reading fluency: Reading Test Alpha	62.57	3.14	92.58	3.47	102	40.86***	< 0.001	0.292
Reading accuracy: Reading Test Alpha	95.07	1.36	105.06	1.50	102	24.11***	< 0.001	0.196
Syllable reading: Test of DIRD	20.00	1.07	22.45	0.49	47	3.82	0.057	0.080
Nonword reading: Test of DIRD	16.97	1.28	21.33	0.59	47	8.40**	0.006	0.160
Spelling ability: Spelling-to-dictation	14.96	0.91	28.68	0.81	149	118.74***	< 0.001	0.449
Phoneme deletion of CVCVCV items: EVALEC	6.51	0.26	8.33	0.23	149	25.51***	< 0.001	0.143
Phoneme deletion of CVC items: EVALEC	10.93	0.17	11.03	0.15	149	0.17	0.675	0.001

Phoneme deletion of CCV items: EVALEC	8.79	0.28	10.57	0.25	149	20.12***	< 0.001	0.130
Nonword repetition: EVALEC	13.60	0.49	18.14	0.42	149	45.48***	< 0.001	0.241
Rapid automatic naming: PhAB	151.69	4.34	109.87	3.75	149	49.99***	< 0.001	0.259

Notes: WISC, Wechsler Intelligence Scale for Children; DVIQ, Diagnostic Verbal Intelligence Test; PPVT-R, Peabody Picture Vocabulary Test-Revised; Test of DIRD, Test of Detection and Investigation of Reading Difficulties; EVALEC, Evaluation de la Lecture; PhAB, Phonological Assessment Battery.

**$p < 0.01$, ***$p < 0.001$; ηp^2, partial eta-squared, respectively, 0.01, 0.06, and 0.14 small, medium, and large effect size.

Table 2. Groups' performance and group differences on semantic fluency tasks

Variables		DDLD group		TD group				
		e. m. mean	*e.* SE	*e. m.* mean	*e.* SE	F	p	ηp^2
Total number of responses		41.03	1.41	46.80	1.25	8.827**	0.004	0.057
Number of correct responses		38.86	1.33	45.40	1.17	12.895***	< 0.001	0.081
Total incorrect responses		2.17	0.23	1.40	0.20	5.854*	0.017	0.039
Types of incorrect responses	Repetitions	0.89	0.14	0.52	0.12	3.669	0.057	0.025
	Intrusions	1.06	0.15	0.74	0.13	2.247	0.136	0.015
	Unintelligible	0.22	0.10	0.14	0.09	0.406	0.525	0.003
Total responses in 0-15 s		17.51	0.51	20.91	0.45	23.028***	< 0.001	0.136
Total responses in 16-60 s		23.52	1.02	25.89	0.90	4.453*	0.037	0.030
Clusters	Number of switches	23.08	0.75	23.28	0.66	0.037	0.849	0.000
	Number of clusters	8.73	0.42	11.32	0.41	16.025***	< 0.001	0.099
	Average cluster size	2.80	0.07	2.95	0.06	1.981	0.161	0.013

Notes: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$; ηp^2, partial eta-squared, respectively, 0.01, 0.06, and 0.14 small, medium, and large effect size.

Table 3. Partial correlations (controlling for age) between semantic fluency (number of correct responses) and language, literacy, and phonological tasks in the overall sample and in the DDLD and TD groups

	Overall sample		DDLD group		TD group	
Tasks	*r*	*p*	*r*	*p*	*r*	*p*
Verbal comprehension: WISC Similarities	0.313^{***}	< 0.001	0.336^{**}	0.006	0.075	0.503
Verbal comprehension: WISC Vocabulary	0.323^{***}	< 0.001	0.254^{*}	0.041	0.124	0.266
Syntax comprehension: DVIQ Test	0.248^{**}	0.002	0.238	0.056	0.142	0.202
Sentence repetition: DVIQ Test	0.248^{**}	0.002	0.255^{*}	0.040	-0.008	0.942
Receptive vocabulary: PPVT-R	0.280^{**}	0.001	0.162	0.198	0.200	0.072
Text-reading fluency: Alouette	-0.263^{**}	0.001	0.092	0.468	0.069	0.538
Text-reading fluency: Reading Test Alpha	0.183	0.067	0.107	0.438	-0.103	0.501
Reading accuracy: Reading Test Alpha	0.097	0.335	0.007	0.960	-0.101	0.509
Syllable reading: Test of DIRD	0.015	0.922	-0.522	0.149	0.002	0.990
Nonword reading: Test of DIRD	-0.071	0.638	-0.731^{*}	0.025	-0.031	0.857
Spelling ability: Spelling-to-dictation	0.268^{**}	0.001	0.191	0.128	-0.005	0.962
Phoneme deletion of CVCVCV items: EVALEC	0.162	0.051	-0.022	0.867	0.124	0.265
Phoneme deletion of CVC items: EVALEC	-0.051	0.542	-0.017	0.895	-0.105	0.348
Phoneme deletion of CCV items: EVALEC	0.052	0.533	-0.073	0.571	-0.043	0.702
Nonword repetition: EVALEC	0.210^{*}	0.011	0.097	0.453	0.044	0.693
Rapid automatic naming: PhAB	-0.273^{**}	0.001	-0.126	0.330	-0.188	0.091

Notes: WISC, Wechsler Intelligence Scale for Children; DVIQ Test, Diagnostic Verbal Intelligence Test; PPVT-R, Peabody Picture Vocabulary Test-Revised; Test of DIRD, Test of Detection and Investigation of Reading Difficulties; EVALEC, Evaluation de la Lecture; PhAB, Phonological Assessment Battery.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 4. The loadings onto components 1 and 2 for each task generated by the Principal Component Analysis (PCA) with oblique rotation in the overall sample

Tasks	Component 1	Component 2
WISC Vocabulary	0.89	-0.04
Spelling-to-dictation	0.86	-0.16
Alouette	0.83	-0.23
WISC Similarities	0.82	-0.13
Receptive vocabulary	0.81	-0.15
Rapid automatic naming	-0.75	0.14
Syntax comprehension	0.64	-0.06
Nonword repetition	0.62	0.60
Sentence repetition	0.54	0.71

Appendix

Principal Component Analysis (PCA) within the DDLD group

The dataset was suitable for the PCA: Kaiser-Meyer-Olkin Measure of Sampling Adequacy value was 0.787, meeting Kaiser's (1974) criterion for this value, Bartlett's Test of Sphericity value was significant ($p < 0.001$; Bartlett 1954), and most of the intercorrelations observed among all 7 tasks of interest had a value of 0.30 and above. The PCA revealed that five language tasks (WISC Similarities, WISC Vocabulary, syntax comprehension, sentence repetition, and receptive vocabulary) and two literacy tasks (Alouette and spelling-to-dictation) used in the overall sample to profile children with dyslexia and/or DLD loaded onto component 1. Table 5 presents each task's contribution to components 1 and 2, which is expressed by its loading value. WISC Vocabulary, receptive vocabulary and WISC Similarities had the highest loadings onto the first component, while Alouette and sentence repetition had the lowest loadings onto this component.

Components 1 and 2 had an eigenvalue larger than 1 (i.e., 3.4), meeting Kaiser's (1974) criterion. The first component had, however, by far the largest eigenvalue of all 7 components generated by the PCA. The second component had an eigenvalue of 1.2 and accounted for 18% of the variance in all 7 tasks, while the remaining components had an eigenvalue lower than 1, and as such, they were not considered further. Even though components 1 and 2 had an eigenvalue larger than 1, a one-factor solution was selected. This selection was based on the scree plot generated by the PCA illustrating a clear split between component 1 and the remaining components. The PCA was therefore repeated, and a one-factor solution was selected. This analysis revealed that component 1 had an eigenvalue of 3.4 and explained 49.68% of the variance in all 7 tasks. The result of the PCA suggests that it is appropriate to combine the children with dyslexia and/or DLD into a combined DDLD group.

Table 5. The loadings onto components 1 and 2 for each task generated by the Principal Component Analysis (PCA) with oblique rotation in the DDLD group

Tasks	Component 1	Component 2
WISC Vocabulary	0.83	0.10
Receptive vocabulary	0.78	0.13
WISC Similarities	0.80	-0.02
Spelling-to-dictation	0.70	-0.52
Syntax comprehension	0.68	0.48
Alouette	0.60	-0.62
Sentence repetition	0.41	0.57