
**Executive Functioning in Children with Developmental Dyslexia**

**Authors:**

Octávio Moura¹

Mário R. Simões¹

Marcelino Pereira¹

¹ *Faculty of Psychology and Education Sciences, University of Coimbra, Coimbra, Portugal.*

**Correspondence concerning this article should be addressed to:**


Email: octaviomoura@gmail.com
Executive Functioning in Children with Developmental Dyslexia

Abstract
The term executive function has been used to describe several higher-order cognitive processes. This study examined the processing speed, shifting, planning, and verbal fluency of a sample of 50 Portuguese children with developmental dyslexia (DD) and 50 typically developing children (TDC; chronological-age-matched controls) between 8 and 12 years of age to evaluate the children’s executive functioning. Compared to TDC, children with DD revealed significant processing speed, shifting, and verbal fluency deficits. After controlling for differences in the general intellectual ability, significant group differences remained for shifting, verbal fluency and marginally for processing speed. No significant differences in planning ability were observed between the groups. No significant interaction of group, gender, and age was found for any of the executive functions measures studied. Word productivity in both semantic and phonemic verbal fluency tasks decreased significantly over the 60 seconds for both groups. Shifting was the only significant predictor of DD in the binary logistic regression analysis and yielded the highest area under the curve value (receiver operating characteristics curve analysis). Therefore, although these findings highlight the presence of specific executive functions deficits in children with DD, they should not be interpreted as indicative of the presence or absence of this learning disorder.

Keywords: processing speed, planning, shifting, verbal fluency, developmental dyslexia

Introduction
Developmental dyslexia (DD) is a specific learning disability that is neurobiological in origin and characterized by difficulty with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These traits typically result from a phonological deficit (Lyon,
Shaywitz, & Shaywitz, 2003) and are not a consequence of sensory impairments, low intelligence or a lack of educational opportunities (American Psychiatric Association, 2013).

A large number of studies have supported the hypothesis that phonological processing is the most relevant neurocognitive phenotype of DD in opaque and transparent orthographies (Landerl et al., 2013; Ramus, Marshall, Rosen, & van der Lely, 2013). Although deficits are most pronounced in measures of phonological processing, other studies suggest that individuals with DD also have weaknesses in other neurocognitive domains. Traditionally, neuropsychological models of neurodevelopmental disorders have typically proposed that a single primary neurocognitive deficit was sufficient to explain all of the symptoms observed for a disorder. Recently, some researchers have challenged the validity of single-deficit models and suggested the presence of a multiple cognitive deficit model for understanding “complex” neurodevelopmental disorders, such as DD, attention-deficit hyperactivity disorder (ADHD), dyscalculia, and other disorders (Pennington, 2006; Willcutt et al., 2013; Willcutt, Sonuga-Barke, Nigg, & Sergeant, 2008). For instance, Willcutt, Pennington, Olson, Chhabildas, and Hulslander (2005) found evidence of a cognitive overlap between DD and ADHD, in which both neurodevelopmental disorders were associated with weaknesses on most cognitive measures [more pronounced in measures of processing speed (PS)]. Similarly, shared neuropsychological weaknesses were observed between children with DD and dyscalculia (Willcutt et al., 2013).

Therefore, impairment in executive functions (EF) is ubiquitous across neurodevelopmental disorders, although distinct profiles emerge from various aspects of EF (Willcutt et al., 2008). Many studies have consistently found that children with DD exhibit weaknesses on a range of EF measures (Altemeier, Abbott, & Berninger, 2008; Brosnan et al., 2002; Helland & Asbjørnsen, 2000; Moura, Simões, & Pereira, 2014b; Reiter, Tucha, & Lange, 2005; Varvara, Varuzza, Sorrentino, Vicari, & Menghini, 2014), which are not simply
secondary consequence of a deficit in another domain (Willcutt et al., 2008). Nonetheless, the
literature has been discordant concerning which executive processes are compromised in DD.
Therefore, the present study examined the presence of specific deficits in the executive
functioning of children with DD who were native speakers of an orthography of intermediate
depth (European Portuguese orthography). We also investigated the diagnostic accuracy of EF
measures to correctly discriminate between typically developing children (TDC) and children
with DD.

EF is a shorthand description of a complex set of processes associated with the
metacognitive capacities that allow an individual to perceive stimuli in his or her
environment, respond adaptively, flexibly change direction, anticipate future goals, consider
consequences, and respond in an integrated way (Baron, 2004). Studies of brain-damaged
patients and neuroimaging studies have located EF in the frontal (particularly the prefrontal
cortex) and parietal lobes (Collette, Hogge, Salmon, & Van der Linden, 2006; Demakis, 2004;
Wager & Smith, 2003). For example, the ability to maintain verbal information in working
memory has been found to rely primarily on the lateral prefrontal cortex (Narayanan et al.,
2005), switching ability has been associated with the medial prefrontal cortex and posterior
parietal cortex (Collette et al., 2006; Crone, Wendelken, Donohue, & Bunge, 2006), the
ability to inhibit responses was found to rely on the right inferior frontal cortex (Aron,
Robbins, & Poldrack, 2004), and updating was associated with cerebral activity in the
prefrontal (dorsolateral, inferior and cingulate) and parietal (posterior and superior) areas
(Collette et al., 2006). Although the frontal and parietal lobes play an important role in the
mediation of EF, researchers also agree that the integrity of the entire brain is necessary for
efficient executive functioning (Stuss & Alexander, 2000; Tamnes et al., 2010).

So, the current conceptualizations support the idea of a fronto-parietal network
supporting executive processes, which is relevant in light of the recent findings about the
involvement of frontal and parietal areas in DD (Bloom, Garcia-Barrera, Miller, Miller, & Hynd, 2013; Boets et al., 2013). Reading development requires the coordination of many aspects of cognition; therefore, it is not surprising that early reading skills (Foy & Mann, 2013), reading comprehension (Borella & de Ribaupierre, 2014; Sesma, Mahone, Levine, Eason, & Cutting, 2009) and reading decoding (Altemeier et al., 2008; Bental & Tirosh, 2007) have been associated with specific executive processes, particularly working memory, inhibition and shifting. For example, working memory plays an important role in reading comprehension because it enables readers to process and access text information to build a coherent representation of the text’s meaning. Cognitive inhibition has also frequently been considered in reading comprehension to contribute to selecting of relevant items, to enable individuals to form a coherent representation of the text (Borella & de Ribaupierre, 2014).

Neurodevelopmental studies have shown that executive functioning emerges in early childhood, develops significantly throughout childhood and adolescence, and that adult-level performance on the most complex EF tasks does not occur until adolescence or even early adulthood (V. Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Best & Miller, 2010; Davidson, Amso, Anderson, & Diamond, 2006). Indeed, executive processes are subject to distinct developmental trajectories. Anderson (2002) found that attentional control appears to emerge in infancy and develops rapidly in early childhood, whereas cognitive flexibility, goal setting, and information processing experience a critical period of development between 7 and 9 years of age, and are relatively mature by 12 years of age. Additionally, working memory capacity has been found to gradually develop throughout childhood and into young-adulthood, shifting attained mature levels during adolescence (Huizinga, Dolan, & van der Molen, 2006), whereas inhibition was found to reach adult-level performance in late childhood or adolescence (Bedard et al., 2002; van den Wildenberg & van der Molen, 2004). These findings about the influence of age on EF task performance have
also been supported by neuroimaging studies examining the maturation of frontal lobe 
(Blakemore & Choudhury, 2006; Tamnes et al., 2010). Studies about the influence of gender 
differences on EF task performance have reported inconsistent findings. Though some studies 
have indicated that boys and girls develop executive functioning in similar ways during 
childhood (Davidson et al., 2006; Marzocchi et al., 2008), others have observed gender 
differences on specific tasks (V. Anderson et al., 2001; Rosselli, Ardila, Bateman, & Guzman, 
2001). These differences may be related to gender-specific differences in brain development 
(De Bellis et al., 2001; Giedd et al., 1996).

Despite its wide acceptance, conceptually defining EF has been difficult. There is no 
consensus among researchers about the executive components involved (for a review, see 
Chan, Shum, Toulopoulou, & Chen, 2008; Jurado & Rosselli, 2007; Wasserman & 
Wasserman, 2013). Some researchers have conceptualized EF as a single construct (Sala, 
Gray, Spinnler, & Trivelli, 1998), but others view it as comprising multiple process-related 
is probably more accurate given that global executive impairment is rare, specific executive 
processes are thought to be associated with distinct frontal systems, and executive processes 
demonstrate variable developmental profiles”. Factor analytic-studies have identified multiple 
EF components. For instance, Welsh, et al. (1991) identified three factors reflecting speeded 
responding, set maintenance, and planning. Miyake et al. (2000) examined three often-
postulated aspects of EF (shifting, inhibition, and updating ) through a confirmatory factor 
analysis and found that, although they are distinguishable, they share some underlying 
commonality. Anderson (2002) proposed four distinct domains: attentional control, 
information processing, cognitive flexibility, and goal setting.

Another problem affecting the measurement of executive functioning is the “task 
impurity problem” (Miyake et al., 2000; van der Sluis, de Jong, & van der Leij, 2007). EF
regulates other cognitive processes, and assessing them requires other non-executive
cognitive abilities to be considered (e.g., verbal and visual-spatial abilities, motor speed, or
attention). Furthermore, executive tasks often require more than one executive function and
the intercorrelations among EF tasks are low to moderate (Lehto, Juujärvi, Kooistra, &
Pulkkinen, 2003; Miyake et al., 2000).

Despite these methodological issues, there exists a relative agreement in terms of the
complexity and importance of executive functioning to human adaptive behavior. EF
measures are widely used in clinical neuropsychological assessment and typically include (but
are not limited to) PS, planning, shifting, verbal fluency (VF), inhibition, updating, divided
attention, and working memory tasks. PS, shifting, planning, and VF tasks are the measures
most often used in studies of children with DD (Brosnan et al., 2002; Reiter et al., 2005;
Shanahan et al., 2006) and in clinical evaluations. Therefore, these four tasks were used to
explore EF deficits in Portuguese-speaking children with DD in the present study.

**Processing Speed**

PS is the ability to automatically and fluently perform relatively easy or over-learned
elementary cognitive tasks, especially when high mental efficiency is required (McGrew,
2009). The Coding and Symbol Search subtests from the Processing Speed Index of the
Wechsler Intelligence Scale for Children (WISC) are two of the most common tasks used to
measure PS performance among children and adolescents. These subtests also measure visual-
motor coordination, scanning ability and visual perception (Kaufman & Lichtenberger, 2000;

Several studies have found that children with DD showed deficits on both WISC-III
Processing Speed Index subtests (Moura, Simões, & Pereira, 2014a; Thomson, 2003).
Shanahan et al. (2006) performed a detailed study that examined the presence of PS deficits in
children and adolescents with DD and ADHD using a wide range of PS tasks. The results suggested that, compared to TDC, a general PS deficit exists in both clinical groups but that children with DD showed greater PS deficits than children with ADHD. Likewise, Willcutt et al. (2005) also found that children with DD or ADHD performed worse than TDC on five PS tasks. More recently, Peng, Sha, and Li (2013) also observed that TDC outperformed children with DD on all PS tasks in a sample of Chinese children.

Catts, Gillispie, Leonard, Kail, and Miller (2002) found that PS explained unique variance in reading comprehension and word recognition even when Full Scale IQ (FSIQ) and rapid naming were introduced into the regression model first. Rapid naming, however, did not predict the additional variance in these two reading measures after FSIQ and PS were taken into account. The authors hypothesized that PS deficit may be an extra-phonological factor in some reading disabilities.

**Shifting**

Shifting has been conceptualized as the ability to flexibly switch between multiple tasks, strategies, or mental sets (Miyake et al., 2000; van der Sluis, de Jong, & van der Leij, 2004; van der Sluis et al., 2007). Miyake et al. (2000) suggested that shifting is a basic underlying component of executive functioning, which is implicated in the performance of more complex executive tasks. A recent meta-analytic study found that shifting was significantly associated with children's performance in both reading and math (Yeniad, Malda, Mesman, van Ijzendoorn, & Pieper, 2013). While some studies have observed that children with DD have difficulty performing tasks that rely on shifting (Helland & Ashjørnsen, 2000; Horowitz-Kraus, 2012; Menghini et al., 2010), others did not find significant differences between children with DD and TDC (Bental & Tirosh, 2007; Reiter et al., 2005).
The Wisconsin Card Sorting Test (WCST) and the Trail Making Test (TMT) – Part B are often used to measure shifting ability. Willcutt and colleagues (2005) found that children with DD scored significantly lower than controls on TMT-B and WCST perseverative errors scores. These main effects did not remain significant after controlling for FSIQ, suggesting that shifting difficulties associated with DD may be explained by group differences in general intelligence. Other studies that used the WCST revealed that individuals with DD committed more perseverative errors (Marzocchi et al., 2008) and more non-perseverative errors (Helland & Asbjørnsen, 2000) and completed fewer categories (Helland & Asbjørnsen, 2000; Menghini et al., 2010) than typically developing individuals. Narhi and colleagues (1997) found that children with DD performed worse on the TMT-B but not on the TMT-A than TDC. They hypothesized that the poorer performance of children with DD on the TMT-B might reflect the difficulty those with DD have in following the alphabetical series. In the studies of both Reiter et al. (2005) and van der Sluis et al. (2004), the results of TMT-B showed non-significant differences between children with DD and chronological-age controls.

Planning

Planning ability is one of the major aspects of executive functioning and has been described as the ability to identify and organize the steps and elements that are required to achieve a goal (Lezak, Howieson, & Loring, 2004). In clinical neuropsychology, planning ability is assessed most often using the Tower of London (ToL) and the Tower of Hanoi (ToH) tests or one of their variants.

Studies testing the planning ability of children with DD have yielded inconsistent findings. Condor, Anderson, and Saling (1995) found that young TDC require significantly fewer trials to reach a successful solution to five-problem variations of the ToH than children with DD, but no significant differences were observed among older children. DD and typical
readers did not obtain significantly different scores for number of errors, initial thinking time, or subsequent thinking time in Brosnan et al.’s (2002) study. Reiter et al. (2005) used the ToL to measure differences in planning abilities between children with DD and TDC. They found that the groups did not differ in the number of problems solved but that the planning time was significantly longer in the DD group. Marzocchi et al. (2008), who also used the ToL, did not find significant group differences in total score, planning time, or execution time.

**Verbal Fluency**

VF tests require participants to retrieve words based on semantic (subjects should produce as many different words as possible within a particular semantic category, e.g., animals, food, names) and phonemic (subjects should produce as many different words as possible that begin with a particular letter, e.g., the letters F, A, or S) criteria within a time constraint (Lezak et al., 2004). VF tests have been used to measure specific aspects of EF, memory, and language. Several neuroimaging studies have suggested that although both semantic and phonemic fluency tasks are associated with frontal and temporal lobe processes, phonemic tasks are more dependent on the frontal lobe and semantic tasks on the temporal lobe (Baldo, Schwartz, Wilkins, & Dronkers, 2006; Birn et al., 2010).

Empirical research has shown that children with DD generate significantly fewer words than TDC in phonemic VF tasks; for semantic VF tasks, however, inconsistent findings have been reported (Landerl, Fussenegger, Moll, & Willburger, 2009; Marzocchi et al., 2008; Reiter et al., 2005). Cohen and colleagues (1999) found that phonemic VF tasks were clinically useful in differentiating two subgroups of children with DD (dysphonetic and dyseidetic) and that the performance of dysphonetic children was significantly lower than that of children with ADHD. Furthermore, semantic VF tasks have been shown to be easier than phonemic VF tasks for TDC (Filippetti & Allegri, 2011; Martins, Vieira, Loureiro, & Santos,
Researchers have hypothesized that semantic tasks are easier because phonemic tasks depend more on the maturation of the frontal lobe; to retrieve words beginning with a letter, an individual must explore more category subsets than is required to retrieve words within a semantic category (Riva et al., 2000).

Troyer (2000) and Hurks et al. (2004; 2006) argued that the total number of words an individual can generate in 60 seconds does not provide sufficient information about the specific cognitive mechanisms that underlie poor performance on VF tasks. They suggested other scoring methods that measured (i) word productivity as a function of time and/or (ii) systematic organization of information, such as clustering (i.e., the production of two or more words within the same semantic or phonemic subcategory) and switching (i.e., the ability to shift between subcategories). The few studies that have analyzed word productivity as a function of time in children found that word production decreased significantly over time (Filippetti & Allegri, 2011; Hurks, 2012; Hurks et al., 2006; Moura et al., 2013; Takács, Kóbor, Tárnok, & Csépe, 2014). No studies have analyzed children with DD. Using the second alternative scoring method, Troyer et al. (2000; Troyer, Moskovitch, & Winocur, 1997) and other authors (Hurks, 2012; Unsworth, Spillers, & Brewer, 2010) demonstrated that clustering and switching are dissociable components of VF performance. Both skills were equally important in semantic VF tasks, but switching made a greater contribution to phonemic VF than did clustering, possibly because switching is more related to frontal lobe functioning.

Although an increasing number of studies about EF in DD have been published recently, inconsistent findings have been obtained. Therefore, the present study has two main objectives: (i) to examine the presence of specific deficits in the executive functioning of Portuguese-speaking children with DD and (ii) to analyze the ability of four different EF
measures (PS, shifting, planning, and VF) to accurately discriminate between children with DD and TDC. Based on the existing literature from different languages spanning a large range of orthographic complexities (e.g., Norwegian: Helland & Asbjørnsen, 2000; Italian: Marzocchi et al., 2008; German: Reiter et al., 2005; English: Willcutt et al., 2005), we expected that Portuguese-speaking children with DD would show significant impairment in the EF measures. We also expected that EF tasks would be accurate measures for distinguished children with DD from TDC. Currently, no studies have analyzed the diagnostic accuracy of EF measures for discriminating between subjects (DD vs. TDC) or have analyzed the executive functioning in Portuguese-speaking children with DD (the European Portuguese orthography is considered to be an intermediate depth). The large body of research about EF deficits in DD has been conducted in English-speaking samples (opaque orthography).

Method

Participants

The participants included 100 Portuguese children between the ages of 8 and 12 ($M = 9.81$; $SD = 1.34$) in grades 3 through 6. In the DD group ($N = 50$), 74% were male and 26% were female, with a mean age of 9.80 years ($SD = 1.38$). Among the children with DD, 26% had undergone school retention, 36% were included in special education system, and 30% had relatives with reading difficulties. In the TDC group ($N = 50$), 64% were male and 36% were female, with a mean age of 9.82 years ($SD = 1.32$). Only 2% had experienced school retention, and 4% had relatives with reading difficulties. The children in the DD and TDC groups were matched for age $\chi^2(4) = 0.487$, $p = .975$, yielding non-significant differences in gender $\chi^2(1) = 1.169$, $p = .387$ and grade $\chi^2(3) = 1.776$, $p = .620$. 
Criteria for inclusion. For both groups, only children who met the following criteria were included in the study: (i) WISC-III FSIQ ≥ 90; (ii) native speakers of European Portuguese; (iii) at least two years of school attendance; (iv) absence of a visual, hearing, or motor handicap; (v) never diagnosed with a language impairment, emotional disturbance, dyscalculia, disruptive behavior disorder (ADHD, oppositional defiant disorder, and conduct disorder), neurological impairment, or other psychiatric disorder. Children with special educational needs were excluded from the TDC group.

All subjects attended regular classes in public and private Portuguese schools. Children with DD were recruited for participation through contact with school psychologists and special education teachers, and referrals from the medical, psychological and other educational/clinical professions (e.g., teachers and speech therapists). The TDC group was recruited through contact with teachers, parents and other participants using a snowball sampling strategy. In the DD group, only children who had previously been diagnosed with DD by a psychologist, child psychiatrist, developmental pediatrician, or child neurologist and had received a score lower than or equal to the 15th percentile on a reading fluency and accuracy measure ("O Rei"; Carvalho & Pereira, 2009) administered during the testing session were included. These cut-off criteria (WISC-III FSIQ ≥ 90 and reading fluency and accuracy measures ≤ 15th percentile) are similar (and in some cases stricter than) the inclusion criteria used by other studies (e.g., Bental & Tirosh, 2007; Frijters et al., 2011; Gooch, Snowling, & Hulme, 2011; Reiter et al., 2005; Swanson, 2011). For the TDC group, only children with a score greater than the 40th percentile on both reading measures were included.

Measures and Procedures

Intellectual ability. The Portuguese version of the WISC-III (Wechsler, 2003) was administered to measure general intellectual ability. The General Ability Index (GAI) scores
were analyzed and used as a covariate in the inferential analysis. The WISC-III GAI is a composite score, which is derived from the four Verbal Comprehension Index subtests and the four Perceptual Organization Index subtests (Prifitera, Weiss, & Saklofske, 1998). We used GAI (rather than FSIQ) because it excludes subtests that are related to EF (i.e., PS and working memory). As suggested by Saklofske, Prifitera, Weiss, Rolfhus, and Zhu (2005), in some special educational cases (e.g., children with learning disability and ADHD), the GAI may be a slightly higher estimate of overall intellectual ability than the FSIQ.

Processing Speed. The Coding and Symbol Search subtests from the WISC-III and the Trail-A test from the Coimbra Neuropsychological Assessment Battery (BANC; Simões et al., in press) were used to measure PS. The Coding (Form B) subtest requires that the child rapidly copy (in two minutes) nine types of symbols, each paired with a number, using a key provided at the top of the page. The Symbol Search (Form B) subtest requires that the child match a specific symbol to an identical target that is displayed among several distracter stimuli. This test also lasts for two minutes. Age-scaled scores ($M = 10, SD = 3$) from the Portuguese version of the WISC–III (Wechsler, 2003) were used for both tasks. The Trail-A test requires the child to draw a line sequentially connecting 25 encircled numbers (1 through 25) randomly distributed on a sheet of paper (similar to the TMT-A). The raw score of the Trail-A test represents the amount of time (in seconds) taken to complete the task.

Shifting. The Trail-B test from the BANC (Simões et al., in press) was administered to examine participants’ shifting ability (similar to the TMT-B). This test requires the child to draw a line connecting 25 circles containing numbers or letters randomly distributed on a sheet of paper, alternating between numbers and letters (1, A, 2, B, etc.). The Trail-B is more complex than the Trail-A because it makes greater demands on an individual’s rapid visual


scanning and visuospatial sequencing capacities and involves cognitive shifting, flexibility, and divided attention. The raw score of the Trail-B represents the amount of time (in seconds) taken to complete the task.

Planning. The Tower test from the BANC (Simões et al., in press) was used to assess planning and problem solving abilities (similar to the ToL). The test comprises 14 models that the child is asked to reproduce by creating a tower using three balls of different colors (red, blue, and green) and three pegs (large, medium, and small). The child must move the three colored balls to specific positions on the three pegs in a specific number of moves (starting with one move and gradually increasing to five moves). The child has four trials in which to correctly solve each of the 14 models. Three raw scores were analyzed: Correct First Trials (i.e., the total number of models correctly solved on the first trial; range = 0-14), Correct Models (i.e., the total number of models correctly solved; range = 0-14), and Total Trials (i.e., the total number of trials taken to solve the 14 models; range = 14-56).

Verbal Fluency. The Semantic and Phonemic Verbal Fluency test from the BANC (Simões et al., in press) comprises three semantic (Animals, Names, and Food) and three phonemic (letters P, M, and R) tasks. VF tests have been used extensively in neuropsychological assessments to measure executive functioning, executive aspects of language processing, and semantic memory. For each of the semantic and phonemic tasks, the child was asked to generate as many words as possible within a time constraint of 60 seconds. The raw score was the total number of correct words (different forms of the same word were excluded) generated within the time limit for the three semantic or phonemic tasks. Additionally, to analyze word productivity as a function of time, the number of words generated by the child were recorded
over four time intervals (0-15 seconds, 16-30 seconds, 31-45 seconds, and 46-60 seconds), as recommended by Hurks et al. (2004; 2006).

The administration of these tasks was part of a broad neuropsychological protocol that also included measures of intelligence, memory, attention, language, reading, and spelling. The children were tested in two sessions separated by a 10- to 15-day interval. The sessions were approximately 90-minutes long and took place in a clinic or school setting during a regular day.

**Statistical Analyses**

The statistical analyses were performed using IBM SPSS Statistics 19. Group differences were analyzed using a multi-factorial multivariate analysis of variance (MANOVA) and covariance (MANCOVA). Group, gender, and age were included as fixed factors, and the executive functions measures were used as dependent variables. If the multivariate analysis (Pillai’s trace) indicated a significant overall difference ($p < .05$), then a univariate test was applied to determine which dependent variables were responsible for the multivariate difference. In specific cases, univariate analysis of covariance (ANCOVA), repeated measures ANOVAs and independent- and paired-samples t-tests were also used. Cohen’s $d$ or partial eta-squared ($\eta^2_p$) was also calculated to determine the effect size of the differences between groups.

A receiver operating characteristics (ROC) curve and binary logistic regression analysis were also performed to examine the accuracy with which EF tasks were able to discriminate between children in the DD and TDC groups. A ROC curve analysis systematically sweeps across all possible true positive (sensitivity) and false positive (1-specificity) values of a diagnostic test and calculates the area under the curve (AUC), which provides an accuracy index of the test (Fawcett, 2006). An AUC of .5 to .7 indicates low test
accuracy, .7 to .9 moderate accuracy, and .9 to 1.0 high accuracy (Swets, 1988). For the binary logistic regression analysis, the fit of the model (Hosmer-Lemeshow test, Cox and Snell $R^2$, and Nagelkerke $R^2$) and the statistical tests of individual predictors were analyzed (regression coefficient, Wald’s $\chi^2$, and odds ratio).

Results

Processing Speed

A 2 X 2 X 5 (group X gender X age) MANOVA was performed and a significant main effect was observed for group, $F(3, 78) = 4.073, p = .010, \eta^2_p = .135$. The univariate analysis revealed significant effects in Coding, $F(1, 80) = 4.823, p = .031, \eta^2_p = .057$), Symbol Search, $F(1, 80) = 7.269, p = .009, \eta^2_p = .083$), and Trail-A, $F(1, 80) = 6.274, p = .014; \eta^2_p = .073$). Children with DD scored significantly lower than TDC (see Table 1).

No significant group X gender, $F(3, 78) = 0.330, p = .804, \eta^2_p = .013$, group X age, $F(12, 240) = 0.824, p = .625, \eta^2_p = .040$, or group X gender X age interactions, $F(9, 240) = 0.604, p = .793, \eta^2_p = .022$, were found.

(Table 1 about here)

Shifting

A 2 X 2 X 5 (group X gender X age) ANCOVA was performed with Trail-B as a dependent variable and Trail-A as a covariate in order to “isolate” the shifting effect on the Trail-B. A significant main effect for group was observed, $F(1, 80) = 10.371, p = .002, \eta^2_p = .115$. Children with DD took more time than TDC to complete the Trail-B (see Table 1). No significant interactions were observed for group X gender, $F(1, 80) = 0.004, p = .953, \eta^2_p < .001$, group X age, $F(4, 80) = 0.559, p = .693, \eta^2_p = .027$, or group X gender X age, $F(3, 80) = 0.149, p = .930, \eta^2_p = .006$. 
As previously noted, Trail-B is a more complex task than Trail-A because it makes greater cognitive demands. To examine this hypothesis, two paired-samples t-tests were performed for each group. The results indicated that TDC, \( t(49) = 13.773, p < .001, d = 2.27 \), and children with DD, \( t(49) = 15.191, p < .001, d = 2.54 \), take more time to complete the Trail-B than the Trail-A.

**Planning**

A multi-factorial MANOVA performed on the three Tower scores yielded a non-significant main effect for group, \( F(3, 79) = 0.915, p = .438, \eta^2_p = .034 \), and for the group X gender, \( F(3, 79) = 2.034, p = .116, \eta^2_p = .072 \), group X age, \( F(12, 243) = 1.297, p = .221, \eta^2_p = .060 \), and group X gender X age interactions, \( F(9,243) = 0.825, p = .593, \eta^2_p = .030 \) (see Table 1).

**Verbal Fluency**

The performance scores of TDC and children with DD on Semantic and Phonemic VF tests are shown in Table 1. The scores on the two tasks tapping VF were entered into a MANOVA as dependent variables and group, gender and age as fixed factor. The multivariate main effect of group proved to be significant, \( F(2, 80) = 7.975, p = .001, \eta^2_p = .166 \). At the univariate level, significant group differences were observed for Semantic VF, \( F(1, 81) = 10.479, p = .002, \eta^2_p = .115 \), and Phonemic VF, \( F(1, 81) = 12.579, p = .001, \eta^2_p = .134 \). Children with DD produced significantly fewer words within the 60-second time limit than TDC on both VF tests. No significant interactions were observed for group X gender, \( F(2, 80) = 0.516, p = .599, \eta^2_p = .013 \), group X age, \( F(8, 162) = 1.525, p = .152, \eta^2_p = .070 \), and group x gender x age, \( F(6, 162) = 1.372, p = .229, \eta^2_p = .048 \). For the TDC and DD groups [TDC: \( t(49) = 21.033, p < .001, d = 3.02 \); DD: \( t(49) = 25.170, p < .001, d = 3.22 \)], the higher number of words produced within the time limit were observed on the Semantic VF (see Table 1).
To analyze the performance of both groups over four time intervals (0-15 seconds; 16-30 seconds; 31-45 seconds; and 46-60 seconds) on the Semantic VF and Phonemic VF tests, we performed four repeated measures ANOVAs. As shown in Table 2, the number of words produced in each of the four time intervals differed significantly, with word production decreasing over time in both VF measures. The Bonferroni adjustment for multiple comparisons revealed the presence of significant differences among all the time intervals in the Semantic VF task and almost all the time intervals in the Phonemic VF task for both groups. As expected, children tended to produce more words in the first 15 seconds than in the remaining three time intervals. Additional independent-samples t-tests revealed statistically significant differences between the TDC and DD groups in the first two time intervals of both VF tasks, 0-15 seconds: $t_{SVF}(98) = 3.986, p < .001, d = 0.79$; $t_{PVF}(98) = 2.582, p = .011, d = 0.51$; 16-30 seconds: $t_{SVF}(98) = 2.576, p = .011, d = 0.51$; $t_{PVF}(98) = 2.824, p = .006, d = 0.56$.

(Table 2 about here)

**Group Differences on Executive Functions after controlling for WISC-III GAI**

The WISC-III GAI scores differed significantly, $t(98) = 3.569, p < .001, d = 0.71$, between the TDC and the children with DD (sum of the eight age-scaled scores that enter the Verbal Comprehension Index and the Perceptual Organization Index; TDC group: $M = 88.76 \pm 13.02$, and DD group: $M = 80.31 \pm 10.36$). Therefore, we additionally examined whether GAI scores could explain the group differences on EF tasks. A series of 2 X 2 X 5 (group X gender X age) MANCOVAs and ANCOVAs, covarying WISC-III GAI, were conducted on all EF tasks. After controlling for differences in general intellectual ability, the main effect of group remained significant for shifting, $F(1, 79) = 7.616, p = .007, \eta^2_p = .089$, for VF, $F(2, 78) = 3.901, p = .024, \eta^2_p = .091$ (univariate analysis: Semantic VF, $p = .022$; Phonemic VF, $p =$
0.022), and marginally significant for PS, $F(3, 77) = 2.727, p = .050, \eta^2_p = .096$ (univariate analysis: Coding, $p = .113$; Symbol Search, $p = .020$; Trail-A, $p = .063$). In contrast, none of the interactions or the main effect of group for planning were significant.

**ROC Curve and Binary Logistic Regression Analysis**

Although the findings presented above report the presence of significant differences in EF between TDC and children with DD (except in the Tower results), it is not certain that these tasks can successfully discriminate between subjects. Therefore, a ROC curve analysis and a binary logistic regression analysis were also performed to determine which EF independently contributed to distinguishing between children with DD and TDC. As shown in Table 3, only the Trail-B test showed moderate accuracy (ROC curve analysis), with an AUC of .730 (i.e., a randomly selected child with DD will take more time to complete the Trail-B than a randomly selected child from the TDC group approximately 73% of the time), while the remaining tasks showed low accuracy.

The goodness-of-fit test of the binary logistic regression analysis yielded a Hosmer-Lemeshow $\chi^2(8) = 5.495, p = .704$, suggesting that the model fit the data well. A Cox and Snell $R^2 = .241$ and a Nagelkerke $R^2 = .322$ were also found. This binary logistic regression model of the four EF tasks correctly classified 71.7% of the participants according to their DD diagnosis: 69.4% true-positive (sensitivity), 74% true-negative (specificity), 26% false-positive, and 30.6% false-negative. As shown in Table 3, only the Trail-B score was a significant predictor, with an odds ratio of 1.015 ($= e^{0.015}$). This result indicates that each one-second increase of the Trail-B score increased a child’s odds of being in the DD group by 1.5%. For example, an increase of 10 seconds on the Trail-B test increases the odds from 1 to 1.161 ($= e^{10*0.015}$).

(Table 3 about here)
Discussion

EF encompasses a set of inter-related processes necessary for goal-directed behavior. These processes develop throughout childhood and adolescence, are largely mediated by the prefrontal and the temporal cortex of the brain, and regulate other cognitive processes. Unsurprisingly, some aspects of EF have been associated with academic achievement (Clair-Thompson & Gathercole, 2006; Thorell, Veleiro, Siu, & Mohammadi, 2012) and reading ability (Foy & Mann, 2013; Sesma et al., 2009) and may therefore play an important role in DD (Altemeier et al., 2008; Booth, Boyle, & Kelly, 2010).

The first main objective of the present study was to analyze the performance of Portuguese TDC and children with DD on EF tasks. As expected, based on previous studies from other orthographies, our findings showed the presence of specific EF deficits in children with DD; the results revealed significant differences on PS, shifting, and VF tasks. Larger effect sizes were observed in analyses of the Trail-B, Semantic and Phonemic VF results, suggesting that children with DD may exhibit more deficits on EF tasks that place greater demands on switching abilities and verbal skills. The finding that DD is associated with slower PS and shifting replicates other studies that used the same measures (Narhi et al., 1997; Willcutt et al., 2005) and with those that used different measures (Boets et al., 2010; Shanahan et al., 2006) that incorporated a verbal component of PS (rapid automatized naming) and shifting (rapid alternating stimulus). Non-significant differences were found for all ToL scores, indicating that planning and problem-solving abilities are not compromised in children with DD. This finding is consistent with previous studies examining children with DD (Brosnan et al., 2002; Marzocchi et al., 2008; Reiter et al., 2005) or reading difficulties (Sikora, Haley, Edwards, & Butler, 2002). This non-significant group difference in planning ability may also be related to the presence of a ceiling effect in two of the three ToL scores.
(Correct Models score and Total Trials score). No interaction of group and gender and/or age was found for any of the EF tasks.

Because the mean WISC-III GAI scores of TDC and DD were significantly different, we additionally examined whether general intellectual ability could explain group differences on EF tasks. The main effect of group remained significant for shifting, VF and marginally significant for PS. The significant main effect on two of the three PS tasks was eliminated after controlling for WISC-III GAI, suggesting that Coding and Trail-A difficulties associated with DD are explained by group differences in general intellectual ability. Whereas some researchers suggest that general intellectual ability should be statistically controlled in cognitive studies of neurodevelopmental disorders, other researchers propose that this approach is misguided and unjustified (for a review, see Dennis et al., 2009; Willcutt et al., 2013).

A more detailed analysis was performed on the results of the two VF tasks. Despite the existence of statistically significant differences between groups (TDC > children with DD) on both the semantic and phonemic VF tasks (as observed in other studies: Landerl et al., 2009; Marzocchi et al., 2008), both groups scored significantly higher on the semantic than the phonemic VF task. This confirms the results of previous studies (Filippetti & Allegri, 2011; Martins et al., 2007; Reiter et al., 2005), corroborating the consensus that the phonemic VF task is more difficult, possibly because it requires the exploration of more category subsets, relies more on the central executive component of working memory, and it is more dependent on the frontal lobe (Birn et al., 2010). As suggested by Troyer (2000) and Hurks et al. (2004; 2006), the pattern of word production over time is relevant to understanding the specific cognitive mechanisms that underlie poor performance on VF tasks. Our results revealed that there is a significant decrease in the number of words produced among both groups (children with DD and TDC) and on both tasks (semantic and phonemic) as a function of time (over
four time intervals), which is congruent with the model of lexical organization proposed by Crowe (1998). This model states that in the first period, a ready pool of frequently used words is available and is automatically active for production (automatic processing), but as time passes, the pool becomes exhausted and the search for new words becomes both more effortful and less productive (controlled processing). Notably, significant group differences were only observed in the first two time intervals (TDC > children with DD), suggesting that poor performance on VF tasks among children with DD was particularly related to deficits in automatic processing. Recently, Takács et al. (2014) also found that TDC and children with ADHD generated the largest number of correct responses during the first two time intervals and that significant group differences were only observed in the first quarter. Similarly, Hurks et al. (2004) also observed that children with ADHD generated fewer words (phonemic VF) in the first 15 seconds than did healthy controls and children with other psychopathologies. The authors suggested that children with ADHD may have a developmental delay in automatic processing of abstract verbal information.

Because the presence of a significant difference alone does not indicate that a test can discriminate among subjects with sufficient accuracy for clinical use, the second main objective of the study was to analyze the accuracy with which the EF measures under study discriminate between children with DD and TDC. The results of the ROC curve analysis yielded low diagnostic accuracy for all the tests except Trail-B. The binary logistic regression model, however, yielded an accuracy rate of 71.7% in classifying children into their correct group (Trail-B was the only significant predictor). No previous studies appear to have analyzed the utility of the different EF processes in diagnosing DD. Although the results highlighted the presence of specific EF deficits in children with DD, they should not be interpreted as indicative of the presence or absence of this learning disorder. As Willcutt et al. (2008, p. 202) stated “EF weakness are neither necessary or sufficient to cause any of the
disorders (…), and are instead one important component of the complex neuropsychology of childhood disorders”. Indeed, the information obtained from EF measures should only be a component of the neuropsychological evaluation and decision-making process and need to be viewed in the context of a more comprehensive assessment that includes other measures, such as phonological awareness, rapid naming, working memory, reading, and spelling measures.

Notwithstanding the uniqueness of the present study, it had several limitations that should be addressed in future studies. First, some of the EFs were assessed only by one task. Clearly the inclusion of more tasks per component would have increased the construct validity and interpretability of the results. Second, the inclusion of other EF tasks (e.g., inhibition, updating, working memory) would also contribute to a better understanding of executive functioning deficits in children with DD. Third, the performance of children with DD on EF tasks was only compared to a TDC group (chronological-age-matched controls) and did not include other clinical samples or a reading-level-matched control group. The literature has clearly demonstrated that children with ADHD also exhibit deficits in a wide range of EF measures (Frazier, Demaree, & Youngstrom, 2004; Fuggetta, 2006), and that DD and ADHD co-occur more frequently than would be expected by chance (15% to 40% of individuals with DD meet criteria for ADHD) (Willcutt & Pennington, 2000). Furthermore, recent studies proposed a multiple cognitive deficit model of neurodevelopmental disorders and found that DD and ADHD shared neurocognitive deficits (McGrath et al., 2011; Willcutt et al., 2005). Thus, future studies should include ADHD children with and without comorbidity with DD in order to increase the generalizability of the findings.

References


Table 1. Means and Standard Deviations of Executive Functions for Typically Developing Children and Children with Developmental Dyslexia

<table>
<thead>
<tr>
<th>Processing Speed</th>
<th>Typically Developing Children</th>
<th>Children with Developmental Dyslexia</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Male</td>
</tr>
<tr>
<td><strong>Coding</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.70)</td>
<td>(2.64)</td>
</tr>
<tr>
<td><strong>Symbol Search</strong>&lt;sup&gt;a&lt;/sup&gt;</td>
<td>10.96</td>
<td>10.94</td>
</tr>
<tr>
<td></td>
<td>(3.12)</td>
<td>(3.41)</td>
</tr>
<tr>
<td><strong>Trail-A</strong></td>
<td>37.14</td>
<td>38.06</td>
</tr>
<tr>
<td><strong>Shifting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Trail-B</strong></td>
<td>91.12</td>
<td>94.78</td>
</tr>
<tr>
<td></td>
<td>(31.53)</td>
<td>(33.14)</td>
</tr>
<tr>
<td><strong>Planning (Tower)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.72)</td>
<td>(1.87)</td>
</tr>
<tr>
<td></td>
<td>(0.56)</td>
<td>(0.55)</td>
</tr>
<tr>
<td></td>
<td>(3.25)</td>
<td>(3.15)</td>
</tr>
</tbody>
</table>
### Verbal Fluency

<table>
<thead>
<tr>
<th></th>
<th>Semantic</th>
<th>Phonemic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50.72</td>
<td>22.22</td>
</tr>
<tr>
<td></td>
<td>50.06</td>
<td>22.99</td>
</tr>
<tr>
<td></td>
<td>51.89</td>
<td>21.56</td>
</tr>
<tr>
<td></td>
<td>52.00</td>
<td>18.86</td>
</tr>
<tr>
<td></td>
<td>43.84</td>
<td>18.11</td>
</tr>
<tr>
<td></td>
<td>57.13</td>
<td>24.25</td>
</tr>
<tr>
<td></td>
<td>52.75</td>
<td>29.88</td>
</tr>
<tr>
<td></td>
<td>57.50</td>
<td>25.25</td>
</tr>
<tr>
<td></td>
<td>44.70</td>
<td>18.12</td>
</tr>
<tr>
<td></td>
<td>43.86</td>
<td>18.30</td>
</tr>
<tr>
<td></td>
<td>47.08</td>
<td>17.62</td>
</tr>
<tr>
<td></td>
<td>38.33</td>
<td>13.67</td>
</tr>
<tr>
<td></td>
<td>42.82</td>
<td>16.53</td>
</tr>
<tr>
<td></td>
<td>42.25</td>
<td>18.13</td>
</tr>
<tr>
<td></td>
<td>51.71</td>
<td>21.86</td>
</tr>
<tr>
<td></td>
<td>51.33</td>
<td>22.67</td>
</tr>
</tbody>
</table>

Note. ss = age-scaled score \((M = 10, SD = 3)\). All other scores are raw scores. Standard deviations in parentheses.
Table 2. Repeated Measures ANOVA of Verbal Fluency Over Four Time Intervals

<table>
<thead>
<tr>
<th></th>
<th>(1) 0-15s</th>
<th>(2) 16-30s</th>
<th>(3) 31-45s</th>
<th>(4) 46-60s</th>
<th>Repeated Measures ANOVA</th>
<th>Pairwise comparisons*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Semantic VF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDC</td>
<td>24.38 (4.44)</td>
<td>12.36 (3.00)</td>
<td>8.30 (2.60)</td>
<td>5.66 (3.15)</td>
<td>$F(3, 147) = 535.845$</td>
<td>$1 &gt; 2 &gt; 3 &gt; 4$</td>
</tr>
<tr>
<td>DD</td>
<td>20.96 (4.12)</td>
<td>10.74 (3.27)</td>
<td>7.84 (3.08)</td>
<td>5.24 (2.42)</td>
<td>$p &lt; .001, \eta^2_p = .916$</td>
<td></td>
</tr>
<tr>
<td><strong>Phonemic VF</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDC</td>
<td>10.96 (3.30)</td>
<td>4.80 (2.42)</td>
<td>3.44 (2.20)</td>
<td>3.02 (2.36)</td>
<td>$F(3, 147) = 211.141$</td>
<td>$1 &gt; 2 &gt; 3, 4$</td>
</tr>
<tr>
<td>DD</td>
<td>9.30 (3.11)</td>
<td>3.56 (1.93)</td>
<td>2.86 (1.78)</td>
<td>2.34 (1.61)</td>
<td>$p &lt; .001, \eta^2_p = .812$</td>
<td></td>
</tr>
</tbody>
</table>

* Bonferroni adjustment for multiple comparisons ($p < .05$); TDC = typically developing children; DD = children with developmental dyslexia; VF = verbal fluency.
Table 3. Receiver Operating Characteristics Curve Analysis and Binary Logistic Regression

<table>
<thead>
<tr>
<th></th>
<th>ROC Curve Analysis</th>
<th>Binary Logistic Regression Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AUC (95% CI)</td>
<td>SE</td>
</tr>
<tr>
<td><strong>Processing Speed</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coding</td>
<td>.670 (.563 – .777)**</td>
<td>.055</td>
</tr>
<tr>
<td>Symbol Search</td>
<td>.663 (.556 – .770)**</td>
<td>.055</td>
</tr>
<tr>
<td>Trail-A</td>
<td>.651 (.542 – .759)**</td>
<td>.055</td>
</tr>
<tr>
<td><strong>Shifting</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail-B</td>
<td>.730 (.631 – .829)***</td>
<td>.051</td>
</tr>
<tr>
<td><strong>Planning (Tower)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct 1st Trials</td>
<td>.592 (.480 – .704)</td>
<td>.057</td>
</tr>
<tr>
<td>Correct Models</td>
<td>.539 (.425 – .652)</td>
<td>.058</td>
</tr>
<tr>
<td>Total Trials</td>
<td>.594 (.482 – .706)</td>
<td>.057</td>
</tr>
<tr>
<td><strong>Verbal Fluency</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic</td>
<td>.660 (.554 – .766)**</td>
<td>.054</td>
</tr>
<tr>
<td>Phonemic</td>
<td>.644 (.536 – .753)*</td>
<td>.055</td>
</tr>
</tbody>
</table>

*Note: *p < .05, **p < .01, ***p < .001; ROC = receiver operating characteristics; AUC = area under the curve; CI = confidence interval; SE = standard error.