



ELSEVIER

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp



Language balance and switching ability in children acquiring English as a second language

Claire Goriot^{a,*}, Mirjam Broersma^a, James M. McQueen^{b,c}, Sharon Unsworth^a, Roeland van Hout^a

^a Centre for Language Studies, Radboud University, 6500 HD Nijmegen, The Netherlands

^b Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, Radboud University, 6500 HB Nijmegen, The Netherlands

^c Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands

ARTICLE INFO

Article history:

Received 14 July 2017

Revised 9 February 2018

Keywords:

Child second-language acquisition

Language balance

Switching

Early-English education

Vocabulary development

Executive functions

ABSTRACT

This study investigated whether relative lexical proficiency in Dutch and English in child second language (L2) learners is related to executive functioning. Participants were Dutch primary school pupils of three different age groups (4–5, 8–9, and 11–12 years) who either were enrolled in an early-English schooling program or were age-matched controls not on that early-English program. Participants performed tasks that measured switching, inhibition, and working memory. Early-English program pupils had greater knowledge of English vocabulary and more balanced Dutch–English lexicons. In both groups, lexical balance, a ratio measure obtained by dividing vocabulary scores in English by those in Dutch, was related to switching but not to inhibition or working memory performance. These results show that for children who are learning an L2 in an instructional setting, and for whom managing two languages is not yet an automatized process, language balance may be more important than L2 proficiency in influencing the relation between childhood bilingualism and switching abilities.

© 2018 Elsevier Inc. All rights reserved.

* Corresponding author.

E-mail address: c.goriot@let.ru.nl (C. Goriot).

Introduction

Learning two languages instead of one language might affect not only language acquisition and processing but also cognitive development, especially in the domain of executive functions. It has been shown, for instance, that bilingual children outperform monolinguals on tasks requiring cognitive flexibility and inhibition skills (Barac, Moreno, & Bialystok, 2016; Poarch & van Hell, 2012). Here we examined whether the balance in first language (L1) and second language (L2) proficiency in child L2 learners is related to executive functioning.

Executive functioning is an umbrella term for a set of processes that together foster cognitive skills that are needed for goal-directed behavior itself as well as for reflection on one's own behavior (Diamond, 2013). In general, three key processes are identified (Miyake et al., 2000): (a) inhibition, the ability to control one's attention in order to replace preliminary responses by more deliberate ones; (b) switching (or shifting), the process that fosters flexibility and adaptation to changed circumstances, including the ability to behave according to different rules or demands; and (c) verbal and nonverbal working memory, the capacity to hold information in mind and to manipulate it. Executive functions develop during childhood and do not fully mature until early adulthood (Diamond, 2013). The development of executive functions may be positively influenced by different factors, including socioeconomic background, intelligence (Diamond, 2013), and, according to some, bilingualism (Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2009).

One of the most widely accepted theories about the relation between bilingualism and cognitive development is the inhibitory control model (Green, 1998). This model proposes that control of the lexico-semantic system is more demanding for bilinguals than for monolinguals because bilinguals need to control two languages that are simultaneously active. Monitoring the competing semantic activation between words requires goal maintenance, conflict monitoring, and, specifically for bilinguals, suppression of interference from the other language. According to this model, dealing with this linguistic competition strengthens bilinguals' linguistic control as well as behavioral-related control processes, in particular executive functions (Green, 1998; Green & Abutalebi, 2013).

Recently, however, the latter has been challenged (de Bruin, Treccani, & Della Sala, 2015; Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2015). Whereas some studies have failed to find any bilingual advantages in inhibition (Duñabeitia et al., 2014), switching (Paap et al., 2017), working memory (Ratiu & Azuma, 2014), or attentional control mechanisms (Antón et al., 2014; Duñabeitia et al., 2014), others have reported that children growing up bilingually from birth do show more developed inhibitory skills (Barac et al., 2016), conflict resolution (Poarch & van Hell, 2012), working memory (Morales, Calvo, & Bialystok, 2014), and attentional control (Poarch & Bialystok, 2015). These inconclusive findings indicate that it is still not clear which specific bilingual groups show advantages in executive functioning, when such differences manifest themselves, and in which specific components. This study aimed to address all three of these issues.

Executive functioning in L2 learners

Gathercole et al. (2014) presented simultaneous bilingual children and adults with card-sorting and Simon tasks. Participants were either monolingual or bilingual (English dominant, Welsh dominant, or balanced in language use). For kindergartners and primary school children, no general bilingual advantage was found. The authors suggested that for simultaneous bilinguals, language switching may be an automatic and effortless process. They theorized that this may be different for L2 learners (i.e., sequential bilinguals) because linguistic selection requires a greater level of control in this group, which in turn strengthens their executive functions.

Indeed, in contrast to Gathercole et al.'s (2014) findings for early bilinguals, findings with bilinguals who are in the process of learning an L2 indicate that there is a relation between language balance and executive functioning performance (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Thomas-Sunesson, Hakuta, & Bialystok, 2018). Thomas-Sunesson et al. (2018) suggested that, if we assume that managing ongoing linguistic competition between two languages results in executive functioning benefits, those who are more equally proficient in both languages, and hence have the most experi-

ence in managing two languages, would show the greatest benefits. Thus, it may be that for L2 learners who are relatively proficient in both languages, but for whom language monitoring is not yet automatized, controlling two languages places demands on executive functioning. If this is the case, enhanced executive functioning performance should hold for these bilinguals in particular. Consequently, given that managing linguistic competition is dependent on proficiency in the two languages, language balance, rather than L2 proficiency alone, should be an important predictor of executive functioning.

Language balance

Several studies have investigated executive functioning performance in L2 learners or sequential bilingual children. Studies that have compared L2 learners, monolinguals, and sometimes also early bilinguals have assumed that differences between the groups could be attributed to differences in exposure and proficiency; those who are most exposed to the L2 are likely to have greater proficiency in that language and more practice in monitoring two languages at the same time, which in turn enhances their performance on executive functioning measures.

These studies, however, have yielded variable results. [Carlson and Meltzoff \(2008\)](#), for example, found that Spanish–English bilingual children outperformed both English monolingual children and children enrolled in immersion education on tasks involving conflicting attention. The authors suggested that by attending an immersion program for 6 months only, immersion pupils might not have had enough exposure to the additional language to show the same advantages as the bilingual group.

In another study ([Poarch & van Hell, 2012](#)), German–English bilinguals outperformed German monolingual children on conflict resolution in a Simon task, but German children enrolled in English immersion education did not significantly differ from either the monolingual group or the bilingual group. The pupils had attended immersion education for less than 2 years and, thus, had more limited experience in controlling two languages than early bilinguals, which could, according to the authors, explain why the immersion group also had less developed control skills.

[Purić, Vuksanović, and Chondrogianni \(2017\)](#) compared monolingual children with a group of pupils enrolled in foreign language education for 5 h a day (high-exposure group) and a group exposed to the L2 for 1.5 h a day (low-exposure group). The high-exposure group outperformed the other groups on working memory tasks, whereas the low-exposure and monolingual groups' scores did not significantly differ. Purić et al. concluded that the high-exposure group had more practice in working memory performance as a result of needing to continuously monitor the vocabulary and grammar of both languages.

All of the aforementioned studies focused on L2 development, assuming that as the L2 developed, children would become more balanced in proficiency across their two languages. However, none of these studies included a measure of L1 proficiency or of language balance despite large individual variation in L1 vocabulary scores ([Carlson & Meltzoff, 2008](#); [Poarch & van Hell, 2012](#)). In this article, we argue that executive functioning depends on children's proficiencies in the L2 and the L1 and, hence, depends on language balance. Language balance accounts for the fact that two children who are equally proficient in the L2 do not need to be equally balanced bilinguals; someone with low proficiency in the L2 and the L1 is balanced, whereas someone with low proficiency in the L2 but high proficiency in the L1 is unbalanced. Assessing language balance in relation to executive functions, thus, requires a measure that comprises both the L1 and the L2.

Several researchers have indeed directly investigated the relation between relative L1 and L2 proficiency and executive functions. For example, [Vega and Fernandez \(2011\)](#) classified sequential Spanish–English bilingual children as “balanced” or “unbalanced” based on their scores on Spanish and English vocabulary tests. The balanced bilinguals performed significantly better than the unbalanced bilinguals on a card-sorting task measuring switching abilities. [Blom et al. \(2014\)](#) presented Turkish–Dutch children with verbal and nonverbal working memory tasks. Language balance was operationalized as the difference in lexical proficiency between Dutch and Turkish by dividing children's highest vocabulary score by their lowest vocabulary score. A significant positive association was revealed between language balance and performance on verbal working memory tasks. In a study on executive functioning performance of Spanish–English bilingual children enrolled in English monolingual or

Spanish–English immersion education, [Thomas-Sunesson et al. \(2018\)](#) operationalized language balance as the difference between the (standardized) English and Spanish vocabulary scores. A positive relation was found between balance and both executive control and working memory but not inhibition. It should be noted that previous research failed to find a relation between language balance and executive functioning in (young) adults ([Paap, Johnson, & Sawi, 2014](#); [Paap et al., 2017](#)). Such a relation, thus, might exist only in children.

In previous studies, children acquired their L2 in (relatively) naturalistic language environments, often being immersed in the L2 for at least 50% of the time. It remains unclear whether there is a similar relation between language balance and executive functioning for children who acquire the L2 via instruction in which there is more limited input of the L2. That question was the focus of this study, which examined the relation between executive functioning and language balance in L2 learners, operationalizing balance using lexical proficiency but now examining pupils in early-English primary schools in The Netherlands where the input in English was much more limited compared with earlier studies. In previous studies, immersion pupils received much more input in English than the pupils in our study, but some pupils in our study had been exposed to the L2 for a longer time because we also included pupils in the highest grade of primary school.

English education in The Netherlands

In The Netherlands, English has been an obligatory subject from the penultimate grade of primary school (Grade 7; 10- and 11-year-olds) since 1986. In this type of education, pupils have approximately 1 h of English instruction per week during the final 2 years of primary school only. Currently, approximately 18% of the primary schools provide children with English lessons from Grade 1 (i.e., kindergarten) onward ([EP–Nuffic, n.d.](#)). Schools are allowed to use English for up to 4 h of teaching time ([EP–Nuffic, 2015](#)), but most do so for less than 1 h per week ([Thijs, Trimbos, Tuin, Bodde, & de Graaff, 2011](#)).

Previous studies showed that despite the limited exposure to English, children make significant progress in their English vocabulary ([Goorhuis-Brouwer & de Bot, 2010](#); [Lobo, 2013](#); [Unsworth, Persson, Prins, & de Bot, 2014](#); [van der Leij, Bekebrede, & Kotterink, 2010](#)), pronunciation ([Lobo, 2013](#)), and grammar ([de Graaff, 2015](#); [Unsworth et al., 2014](#)). All these studies, however, tested children in only one age group. In most studies, pupils had had maximally 2 years of English lessons; in one study ([de Graaff, 2015](#)), participants had had either 2 or 5 to 8 years of English lessons. In our study, we included pupils who had just started, were halfway, or were at the end of primary school. By doing so, we also aimed to shed more light on the effects of early-English education on the vocabulary knowledge of pupils after different amounts of English education.

The previous studies on early-English education show that there is considerable variation in the extent to which pupils have acquired the L2. Considering language balance, only two studies took pupils' L1 development into account. In both studies, early-English pupils' Dutch vocabulary development was comparable to that of same-aged peers ([Goorhuis-Brouwer & de Bot, 2010](#); [van der Leij et al., 2010](#)). Despite this, there was individual variation in Dutch vocabulary development. Primary school pupils' L1 vocabularies are still developing. As was hypothesized by [Cummins \(1979\)](#), the development of an L1 and the development of an L2 are related, and so, once again, it is important to take into account not only the development of the L2 but also that of the L1, as we did in this study.

Previous studies showed advantages for early-English pupils in English language skills. Pupils from control schools, however, were also found to have some knowledge of English ([de Graaff, 2015](#); [Goorhuis-Brouwer & de Bot, 2010](#); [Lobo, 2013](#); [Unsworth et al., 2014](#)), and some individual control school pupils had greater knowledge of English than some enrolled in early-English education ([de Graaff, 2015](#)). Out-of-school exposure may have played a role in these findings. Although English is not an official language of The Netherlands, it is very present in everyday life; for example, advertisements are often in English, and movies are subtitled rather than dubbed ([Kuppens, 2010](#)). Indeed, primary school pupils' knowledge of English is influenced by watching subtitled movies and playing English computer games ([Kuppens, 2010](#)). Given these findings, we assumed that there will be considerable variability in pupils' Dutch–English language balance for both early-English and control pupils.

The current study

This study investigated whether differences in lexical balance are related to executive functioning measures in children who have limited exposure to the L2. Such findings should contribute to the ongoing debate on the relation between bilingualism and cognitive development by helping to specify the conditions of bilingual experience that are most likely to stimulate these functions. We chose to focus on Dutch children who are learning English as an L2 either because of contact with English in everyday life or by means of a special educational program. Although the first group was not completely monolingual, these children were functionally monolingual because their knowledge of English was very limited and they did not actively use the language. We included children of different age groups to be able to evaluate the development of vocabulary in the L1 and L2 as well as executive functions as children get older: 4- and 5-year-olds (Grade 1, i.e., kindergarten), 8- and 9-year-olds (Grade 5), and 11- and 12-year-olds (Grade 8, i.e., final grade of primary school).

Our main research question was whether individual differences in lexical balance between the L1 (Dutch) and the L2 (English) are related to individual differences in executive functions. We expected that more balanced L2 learners would perform better on executive functioning tasks. Whereas previous research mostly took into account one or two specific executive functioning processes (Blom et al., 2014; Carlson & Meltzoff, 2008; Poarch & van Hell, 2012; Purić et al., 2017; Vega & Fernandez, 2011), we included all key factors in Miyake et al.'s (2000) model (switching, inhibition, and verbal and non-verbal working memory). Given that previous research showed that balance is an important factor in L2 learners' executive functioning performance (Blom et al., 2014; Thomas-Sunesson et al., 2018), we expected that also for this group of L2 learners there would be a relation between lexical balance and executive functions.

Lexical balance was expected to be a better predictor of executive functioning than L1 or L2 development alone because, as argued above, it is assumed that differences between L2 learners and monolinguals stem from the fact that L2 learners need to control two languages that are simultaneously active (Green, 1998; Green & Abutalebi, 2013). This becomes more demanding when both languages are mastered and the interplay between the two languages gets stronger.

Our second research question was whether children enrolled in an early-English school would differ in development from pupils who were not enrolled in such a program. Given previous studies' results on vocabulary development (Goorhuis-Brouwer & de Bot, 2010; Unsworth et al., 2014), we expected early-English pupils to have greater knowledge of English vocabulary than pupils from control schools and, therefore, to have more balanced lexicons.

Previous studies showed that a bilingual immersion program can positively influence executive functioning development (Poarch & van Hell, 2012; Purić et al., 2017). The participants in our study were exposed to the L2 to a lesser extent than previous participants in terms of hours per week. Some of them were, however, exposed for a longer time given that we also included 11- and 12-year-olds who had been exposed to the L2 for 8 years. Because sufficient exposure to the L2 seems to be a prerequisite for executive functioning advantages to show (Carlson & Meltzoff, 2008), we expected that 11- and 12-year-old early-English pupils would outperform their peers from control schools.

In summary, this study examined three hypotheses, namely that (a) there is a positive relation between lexical balance and executive functioning performance, (b) early-English pupils have on average a greater vocabulary in English and are more balanced across their two lexicons, and (c) older early-English pupils (i.e., the 11- and 12-year-old group) have better developed executive functions than control pupils of the same age.

Method

Participants

Four early-English schools in The Netherlands that had at least 8 years of experience with teaching English were recruited. Four control schools were matched to the early-English schools on neighborhood (average income), area (urbanized or rural), religious denomination, and educational philosophy.

We asked the head teacher of each school to select 10 children in each age group to participate. Criteria for selection were that children should not have any developmental disorders or any severe motor, sight, or hearing impairments and should not be exposed to another language at home. Data were collected from 241 primary school pupils. Participants were part of one of the following age groups: 4- and 5-year-olds (Grade 1, i.e., kindergarten), 8- and 9-year-olds (Grade 5), or 11- and 12-year-olds (Grade 8, i.e., final year of primary school). On average, the 11- and 12-year-olds in early-English education had had 640 h of English lessons as opposed to 80 h for control pupils who started their English lessons in the penultimate grade (Grade 7). Of the 241 participants, 37 were removed from the analyses because they did not complete enough tasks ($n = 8$), were enrolled in extracurricular language learning activities ($n = 1$), were clearly not concentrating during testing ($n = 1$), were exposed to another language at home ($n = 23$), or had a diagnosis of dyslexia ($n = 4$). Parents gave informed consent for their children's participation. Table 1 provides the number of participants per group, their mean age, and the standard deviation. The exposure to English measure was based on the estimates that parents provided in a questionnaire. There was a response rate of 37.2% on the questionnaire, which is a lower completion rate than in previous research (e.g., Unsworth et al., 2014).

Executive function measures

Switching

In the Dimensional Change Card Sort task (DCCS; Zelazo, 2006), children were first presented with two cards: a red boat and a blue rabbit. These cards remained visible throughout the task. Experimental cards depicted a blue boat ($n = 12$) or a red rabbit ($n = 12$). In the *pre-switch phase*, participants sorted 6 cards on color. In the *post-switch phase*, they sorted 6 cards on shape. If children responded correctly to at least five trials, they passed on to the *border phase*, where children sorted cards with a black border ($n = 6$) on color and sorted cards without a border ($n = 6$) on shape. Instructions were repeated halfway. The order of the cards was the same for all children. The pre-switch and border phases were preceded by two explanation trials. No more than 2 cards with a border or with the same picture were shown after each other, and cards of blue boats and red rabbits were equally distributed in each phase. The switching score was computed as the total number of correctly performed trials over all three conditions.

Inhibition

In this version of the Simon task (Hedge & Marsh, 1975; Simon & Small, 1967), children saw a blue triangle or red square on either the left or right side of a computer screen. Instructions were counter-balanced; half of the children were asked to press the left button on the button box when they saw a square and to press the right one when they saw a triangle, and the other half were asked the opposite. Buttons were marked with a sticker in the corresponding color. Response location, thus, was either congruent or incongruent with stimulus location. The trial was terminated by a response, which immediately triggered the start of a new trial. Otherwise, the trial was terminated after 2500 ms for kindergartners or 1000 ms for older children. Stimuli were preceded by a fixation cross presented

Table 1
Participants: Descriptive statistics per age group.

	Control schools			Early-English schools		
	4- and 5-year-olds	8- and 9-year-olds	11- and 12-year-olds	4- and 5-year-olds	8- and 9-year-olds	11- and 12-year-olds
<i>n</i>	38	34	26	40	38	28
Girls (<i>n</i>)	21	17	16	23	19	19
Boys (<i>n</i>)	17	17	10	17	19	9
Age in years [<i>M</i> (<i>SD</i>)]	4.90 (0.27)	9.08 (0.40)	12.15 (0.53)	4.81 (0.35)	8.95 (0.41)	11.97 (0.38)
Exposure to English at home in hours per week [<i>M</i> (<i>SD</i>)]	5.09 (6.08)	9.27 (6.35)	16.10 (14.11)	9.14 (7.97)	7.44 (6.41)	22.5 (16.43)

in the middle of the screen for 500 ms. Children were told to react accurately and as quickly as possible. The task started with 8 practice trials, which were repeated until children understood the task, and were followed by 240 experimental trials.

Working memory

Two subtests of the Automated Working Memory Assessment (AWMA; Alloway, Gathercole, Kirkwood, & Elliott, 2008), a PC-based test for working memory skills, were used. Nonverbal working memory was assessed by the *Odd One Out* subtest. A trial started with a set of three shapes in a 3×3 matrix being presented together in a row. Children needed to indicate the odd one out that had a different shape. The shapes then disappeared, leaving only the boxes. Children needed to recall the location of the shape identified as the odd one out. The test started with 4 practice trials. The test phase started with a block of trials in which the location of one shape needed to be remembered, building up to a block in which the location of seven shapes needed to be remembered in order.

In the *Backward Digit Recall* subtest, which taps into verbal working memory, children orally repeated a string of spoken digits in reverse order. After 4 practice trials, the test phase started with a string of two digits up to a maximum of nine digits. In both subtests, a block consisted of 4 trials. When a trial was performed incorrectly, additional trials of the same length were administered, with a maximum of 2 trials. The tests ended after 3 incorrect trials. The test–retest reliability coefficients range between .86 for the Backward Digit Recall and .88 for the Odd One Out (Alloway et al., 2008).

Language measures

English and Dutch vocabularies

The Peabody Picture Vocabulary Task–Fourth Edition (PPVT-4; Dunn & Dunn, 2007) and the PPVT-III Dutch (Dunn, Dunn, & Schlichting, 2005) were administered to assess English and Dutch vocabularies, respectively. The PPVT is a standardized receptive vocabulary task in which children hear a word and need to choose the corresponding picture from a set of four pictures. The PPVT-4 consists of 228 words. The PPVT-III Dutch consists of 204 words. The test–retest reliability coefficients for children between the ages of 4 years 0 months and 13 years 0 months range between .91 and .94 for the English version (Dunn & Dunn, 2007) and between .91 and .96 for the Dutch version (Dunn et al., 2005). The score was determined by the number of correctly identified words. Raw scores were used instead of age-related norm scores because norm scores are based on native speakers and, therefore, are inappropriate for L2 learners.

Balance

Lexical balance was the natural logarithm of the proportion correct on the English vocabulary test divided by the proportion correct on the Dutch vocabulary test. An outcome of 0 indicated a turning point from being more proficient in Dutch (negative value) to being more proficient in English (positive value). A comparable calculation of language balance can be found in Blom et al. (2014).

Intelligence measure

Intelligence was measured with the brief version of the Wechsler Nonverbal Scale of Ability (WNV; Wechsler & Naglieri, 2008). All children performed the *Matrices* subtest; they needed to complete an incomplete figural matrix by choosing the correct piece out of four or five alternatives. The total number of trials was 41. As for the *Recognition* subtest described below, testing ended when 4 of 5 consecutive trials were incorrect.

Kindergartners performed the *Recognition* subtest, in which they were presented with a geometric design for 3 s. Subsequently, they needed to identify the design they had seen out of four or five designs given on a new page. The total number of trials was 21.

Older children performed the *Spatial Span* subtest. The examiner tapped a pattern on 10 irregularly placed blocks. The task consisted of a forward phase and a backward phase, in which children repeated the sequence in the same order and in reverse order, respectively. Testing ended after two failed attempts at sequences of the same length (max = 32 trials).

For all subscales, raw scores were computed as the total number of correctly performed trials. The raw score for intelligence was determined by the total score of both subtests. The internal consistency of the various subtests ranges from $\alpha = .63$ to $\alpha = .78$ (Wechsler & Naglieri, 2008).

Procedure

Children were tested individually in a quiet room at school. Testing took place in two sessions of approximately 30 min. The first session started with the DCCS, followed by the AWMA (Odd One Out and then Backward Digit Recall) and the PPVT-4. The second session began with the WNV (Matrix Reasoning and then Recognition or Spatial Span), followed by the PPVT-III and the Simon task. All tests were carried out following the procedure outlined in the task manuals except the DCCS and the Simon task, for which general descriptions were used (Hedge & Marsh, 1975; Zelazo, 2006). The two sessions were separated by at least 1 day and no more than 28 days ($M = 5.87$ days, $SD = 4.63$) except for 1 participant who performed the two sessions on the same day but separated from one another by more than 4 h. Responses were registered on a laptop (AWMA and Simon task) and/or noted and written down by the experimenter (all tasks except the Simon task).

Analyses

Given the data's hierarchical (multilevel) structure, with child level gathered in the context of a specific school, the linear mixed models function in SPSS 22.0 was used for the analysis. In all models, school was entered as random factor.

First, we investigated whether there were relations between each of the language measures and intelligence with age group (4–5, 8–9, or 11–12 years), type of education (early-English or control), and the interaction between age group and type of education. Thereafter, we investigated which variables could explain executive functioning development. We started with a base model with age group, type of education, and their interaction. Subsequently, we compared this model with one in which we added intelligence and the language measure balance. The aim was to investigate whether individual differences in these latter two variables influenced executive functioning outcomes. If balance was included in the model, we checked whether there were interaction effects of this measure with type of education and/or age group. The overall aim was to find the most parsimonious model—a model that included the variables that could explain the variance in the executive functioning measure with the least degrees of freedom. We performed all of the aforementioned steps for each executive functioning measure separately (switching, inhibition, and verbal and nonverbal working memory). To determine whether a newer model was a better fit than a previous one, the deviance score ($-2 * \log$ likelihood ratio [$-2LL$]) was used as a goodness of fit measure (Heck, Thomas, & Tabata, 2014). All models were tested against a .05 significance level. The direction and size of the effect parameters were checked only after the goodness of fit had been checked. We used the MuMIn package in R (Version 3.4.1) to determine the R^2 for each model. We report both the marginal and conditional R^2 (Nakagawa & Schielzeth, 2013). The marginal R^2 (R_m^2) shows the variance that is explained by fixed factors, whereas the conditional R^2 (R_c^2) is concerned with the variance explained by both fixed and random factors. We followed the general rule of thumb that R^2 values of .10, .30, and .50 are small, moderate, and large effects, respectively.

Results

General results

The results for all measures are presented in Table 2. This table reveals that pupils in the highest groups obtained the best scores on all measures. Descriptively, there is little difference in scores between early-English and control pupils except that early-English pupils seem to have higher scores on the English vocabulary test and more balanced lexicons than pupils from control schools, although differences are very small for 8- and 9-year-olds. Note that a higher score on the English vocabulary

Table 2

Raw scores on executive functioning measures, the intelligence measure, and vocabulary measures.

		4- and 5-year-olds		8- and 9-year-olds		11- and 12-year-olds		
		Control	Early-English	Control	Early-English	Control	Early-English	
Switching (accuracy; max = 24)	<i>n</i>	29	30	33	37	26	28	
	<i>M</i>	18.45	18.10	19.70	19.32	20.77	21.25	
	(<i>SD</i>)	(1.76)	(1.47)	(3.03)	(3.09)	(2.86)	(2.29)	
Inhibition	<i>n</i>	36	35	34	37	25	27	
	Congruent	<i>M</i>	105.39	105.97	108.06	107.51	113.56	113.41
	(accuracy; max = 120)	(<i>SD</i>)	(11.15)	(8.71)	(5.78)	(6.74)	(4.08)	(4.30)
	Incongruent	<i>M</i>	99.56	90.54	99.97	99.43	105.76	105.00
	(accuracy; max = 120)	(<i>SD</i>)	(13.77)	(13.77)	(9.20)	(9.26)	(9.26)	(8.20)
Simon effect (RT)	<i>M</i>	64.00	72.03	39.81	45.51	36.51	39.77	
	(<i>SD</i>)	(41.04)	(60.93)	(18.71)	(0.94)	(20.03)	(17.35)	
Verbal working memory (accuracy; max = 162)	<i>n</i>	37	39	32	37	26	28	
	<i>M</i>	4.08	3.46	12.38	11.73	16.38	15.07	
Nonverbal working memory (accuracy; max = 36)	(<i>SD</i>)	(3.44)	(2.61)	(3.70)	(3.44)	(4.22)	(3.64)	
	<i>n</i>	37	39	32	37	26	28	
Intelligence (raw score; max = 62 for 4- and 5-year-olds, max = 73 for older pupils)	<i>M</i>	11.62	10.28	20.00	19.89	25.00	23.93	
	(<i>SD</i>)	(3.56)	(2.80)	(4.02)	(3.96)	(4.72)	(4.40)	
Dutch vocabulary (raw score; max = 204)	<i>n</i>	38	40	34	38	26	28	
	<i>M</i>	72.05	72.58	110.68	111.13	133.65	134.86	
English vocabulary (raw score; max = 228)	(<i>SD</i>)	(10.28)	(12.21)	(7.53)	(7.56)	(10.14)	(9.80)	
	<i>n</i>	38	40	34	38	26	28	
Balance	<i>M</i>	13.39	18.18	55.94	56.13	93.77	111.50	
	(<i>SD</i>)	(8.33)	(10.47)	(26.67)	(16.77)	(32.00)	(27.14)	
	<i>n</i>	38	40	34	38	26	28	
	<i>M</i>	-1.97	-1.65	-0.89	-0.84	-0.52	-0.33	
	(<i>SD</i>)	(0.61)	(0.56)	(0.44)	(0.32)	(0.29)	(0.22)	

Note. A higher score means poorer performance for inhibition and better performance for all other measures.

test than on the Dutch one does not necessarily indicate a larger vocabulary in English due to differences in the number of items in the two tests. When comparing norm scores for both vocabulary tests, all pupils except 1 had a higher norm score for Dutch than for English.

A mixed-models analysis on age with schools as random variable and age group, type of education, and the interaction between age group and type of education as fixed variables revealed that only age group was significantly related to age ($p < .001$). Thus, there were no differences in age between the groups of pupils from the two types of schools. An analysis of variance (ANOVA) with age group and type of education as independent variables and out-of-school exposure to English as dependent variable showed a significant relation between age group and out-of-school exposure ($p < .001$), indicating that the pupils in the higher grades were more exposed to English than the pupils in the lower grades. There were again no differences between the pupils from the two types of schools.

Correlations between measures

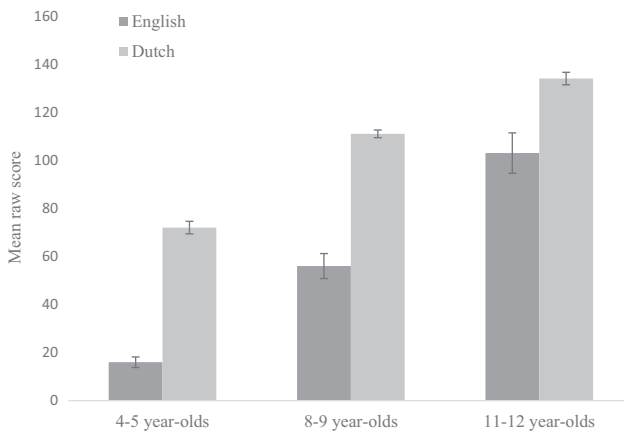
First, we examined the partial correlations between the measures, controlling for school-specific effects. Table 3 shows that there was a very strong positive correlation among balance, English vocabulary, and Dutch vocabulary. For the theoretical reasons outlined in the Introduction, balance rather than English vocabulary is used as the primary measure in the following analyses.

Figs. 1 and 2 give a graphical overview of the relations between the raw scores on the English and Dutch vocabulary tests. As shown in Fig. 1, the difference between Dutch and English vocabularies becomes smaller for children in the higher age groups. Fig. 2 shows the relation between developments of both vocabularies, showing that the development of Dutch vocabulary is less prone to indi-

Table 3

Partial correlations among age group, type of education, intelligence, language measures, and executive functions.

	1	2	3	4	5	6	7	8	9	10
Age group	1									
Type of education	-.100	1								
Intelligence	.770**	-.080	1							
Balance	.764**	.150	.620**	1						
English vocabulary	.817**	.070	.686**	.878**	1					
Dutch vocabulary	.929**	-.011	.807**	.750**	.832**	1				
Switching	.359**	-.077	.346**	.336**	.401**	.390**	1			
Inhibition	-.377**	.086	-.267**	-.218**	-.303**	-.341**	-.060	1		
Nonverbal working memory	.798**	-.080	.716**	.624**	.680**	.796**	.387**	-.342**	1	
Verbal working memory	.790**	-.074	.765**	.626**	.647**	.804**	.355**	-.263**	.787**	1

* $p < .05$.** $p < .01$ (two-tailed).**Fig. 1.** Average scores for English and Dutch vocabularies for the different age groups.

vidual variation than that of English. The benefit of balance over the other two vocabulary measures is that it captures the relation between English and Dutch vocabularies. That is, it takes into account the differences between L1 and L2 proficiency shown in these figures.

All three language measures correlated with the executive functioning measures (cf. Table 3), with inhibition showing the weakest correlations of the four measures in general. Note that the correlations for inhibition are negative because lower scores indicate better inhibitory control.

Intelligence and language measures

First, we used a linear mixed model with schools as random factor and age group, type of education, and the interaction between the two variables as fixed factors to investigate whether there were any significant differences between the early-English and control pupils in intelligence or any of the language measures. The results are shown in Table 4.

For all measures, age group was a significant predictor, with children in the higher age groups obtaining significantly higher scores on the intelligence tasks and language measures. For balance, type of education was a significant predictor as well; children from early-English schools had on average more balanced lexicons than children from control schools. Finally, there was a marginally significant effect of type of education ($p = .056$) and a marginally significant interaction between age group and type of education ($p = .062$) on English vocabulary. The model parameter estimates show that

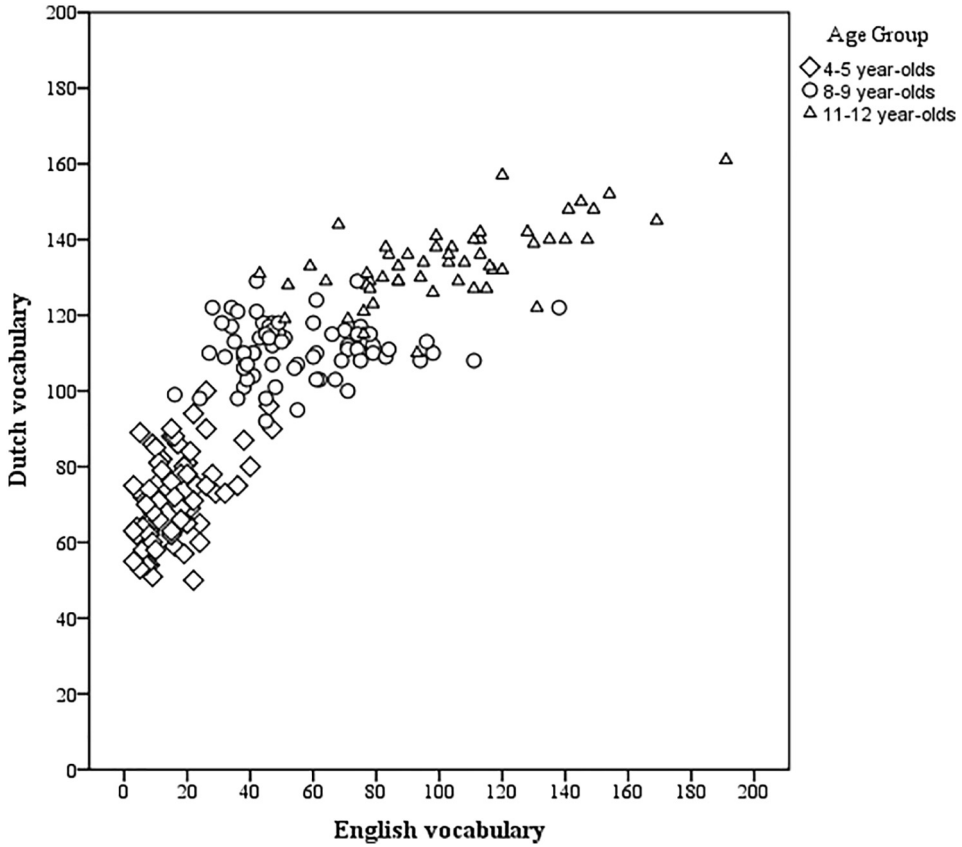


Fig. 2. Relationship between Dutch and English vocabularies for the different age groups.

Table 4

F values of fixed effects for intelligence and language measures.

	Intelligence	Dutch vocabulary	English vocabulary	Balance
Intercept	2130.04 ^{***}	17630.22 ^{***}	1477.99 ^{***}	1064.33 ^{***}
Age group	173.37 ^{***}	683.22 ^{***}	276.88 ^{***}	171.30 ^{***}
Type of education	0.26	0.24	6.24 [*]	8.65 ^{**}
Age group * type of education	0.67	0.03	2.82 [*]	1.61
Variance components				
Schools	2.40	1.21	0.73	0.00
Residual	22.56	94.39	434.80	0.20
R_m^2	.616	.871	.737	.636
R_c^2	.653	.872	.737	.636

^{*} $p < .10$.

^{**} $p < .01$.

^{***} $p < .001$.

there was a significant effect of type of education (beta = -17.69 , $SD = 5.72$, $p = .004$ for control schools compared with early-English schools) and a significant interaction effect for 8- and 9-year-olds and control schools (beta = 17.48 , $SD = 7.52$, $p = .021$). This shows that the relation between type of education and English vocabulary is not the same for all age groups (see Fig. 3). Differences between control and early-English pupils in English vocabulary were found only in the highest grade of primary school.

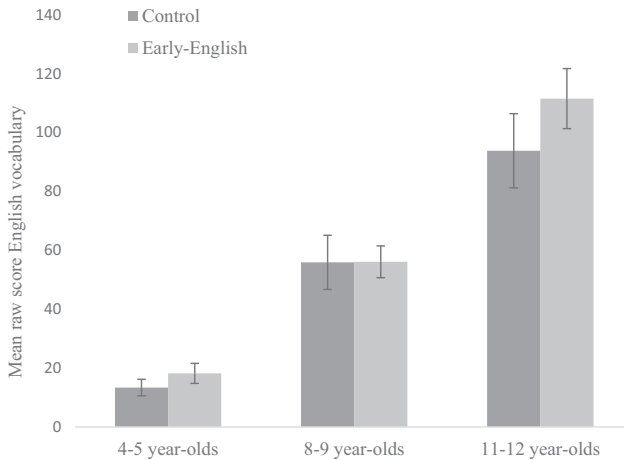


Fig. 3. Mean scores on the English vocabulary test split out by age group and type of education.

Executive functions

Switching

We assume that all children who were not able to sort the cards correctly after the sorting feature in the DCCS had changed from color to shape did not understand the task (19 4- and 5-year-olds and 2 8- and 9-year-olds; 10 early-English pupils and 11 control pupils). Their scores on the DCCS, therefore, were excluded from the analyses.¹

We started with the basic model in which age group, type of education, and the interaction between the two variables were included as fixed factors. This model revealed that only age group was significantly related to switching outcomes ($-2LL = 850.03$, $df = 13$, $R_m^2 = .159$, $R_c^2 = .186$).

Subsequently, we ran the model adding either intelligence or balance. The model with intelligence was not significantly different from the basic model ($-2LL = 851.96$, $df = 14$, $R_m^2 = .171$, $R_c^2 = .186$). Adding balance as a fixed effect to the basic model resulted in a significant improvement of fit ($-2LL = 844.92$, $df = 14$, $R_m^2 = .181$, $R_c^2 = .215$). Therefore, we proceeded with balance rather than intelligence in the following analyses. We checked whether there were any significant interactions between balance and age group or type of education. Although the model improvement was just significant ($-2LL = 823.58$, $df = 25$, $R_m^2 = .212$, $R_c^2 = .236$), the p values of the interactions all were $p > .05$. Therefore, we continued with the model without the interactions.

In the model with balance as well as in the basic model, type of education showed no significant relation with switching scores. Therefore, we removed type of education and the interaction from the model ($-2LL = 849.88$, $df = 6$, $R_m^2 = .177$, $R_c^2 = .208$). This latter model was not significantly different from the basic model or from the model with type of education and balance both included, but the model with age group and balance as predictors did have considerably fewer degrees of freedom and, thus, was considered the most parsimonious model. In this model, balance showed a significant and positive relation with switching. Age group reached borderline significance, but removing this factor from the model resulted in a significant deterioration in goodness of fit ($-2LL = 857.09$, $df = 3$, $R_m^2 = .151$, $R_c^2 = .184$), which is why it was retained.

Continuing with the model that included age group and balance, we investigated whether Dutch vocabulary, English vocabulary, or scores on both vocabulary tests could replace balance. Replacing balance by scores on both vocabulary tests ($-2LL = 856.59$, $df = 7$, $R_m^2 = .207$, $R_c^2 = .223$), English vocabulary only ($-2LL = 853.45$, $df = 6$, $R_m^2 = .196$, $R_c^2 = .212$), or Dutch Vocabulary only ($-2LL = 853.28$, $df = 6$,

¹ Including these children in the analyses did not change the pattern of results.

Table 5

Model parameter estimations for executive functioning measures.

	Switching	Inhibition	Verbal working memory	Nonverbal working memory
Intercept	21.34 (0.41) ^{***}	–	–1.32 (3.18)	17.56 (2.08) ^{***}
4- and 5-year-olds	–1.53 (0.75) [*]	–	–7.35 (0.90) ^{***}	–10.71(1.07) ^{***}
8- and 9-year-olds	–1.15 (0.49) [*]	–	–2.05 (0.62) ^{**}	–3.61 (0.74) ^{***}
Balance	0.92 (0.43) [*]	–	–	–
Intelligence	–	–	0.30 (0.046) ^{***}	0.19 (0.05) ^{***}
Variance components				
Schools	0.230	–	0.252	0.366
Residual	6.04	–	9.75	13.832

Note. Standard errors are in parentheses.

^{*} $p < .05$.

^{**} $p < .01$.

^{***} $p < .001$.

$R_m^2 = .188$, $R_c^2 = .207$) resulted in a significant deterioration in model fit. For both models, removing age group did not result in a better model fit ($-2LL = 855.34$, $df = 3$, $R_m^2 = .196$, $R_c^2 = .210$ and $-2LL = 857.48$, $df = 3$, $R_m^2 = .188$, $R_c^2 = .207$ for English and Dutch vocabularies, respectively).

In summary, switching scores seem to be best predicted in a model that includes both age group and balance. In this model, balance showed a positive relation with switching scores; children with more balanced Dutch and English lexicons showed better switching abilities. Age group also showed a positive relation with switching scores; children in the higher age groups scored better on the switching task than children in the lower age groups. Table 5 shows the model parameter estimations for the variables. For age group, 11- and 12-year-olds are the reference category.

Inhibition

Two children were excluded from the analyses; one child did not complete the Simon task due to technical errors with the equipment, and the other child was not concentrating during the task, according to the experimenter's assessment. All trials to which children gave no response, or gave an incorrect or anticipatory response (response time [RT] < 200 ms), were excluded from the analyses (4.3% removed). Thereafter, all responses with RTs above or below 2.5 standard deviations from the participant's mean were removed (1.6%), comparable to procedures used in previous studies (Davidson, Amso, Anderson, & Diamond, 2006; Duñabeitia et al., 2014). For 8 participants, less than 60 trials (1 block in the Simon task) remained. Their scores were left out of the analyses. The Simon effect for the remaining 194 children was calculated as the difference between RTs on incongruent trials minus the RTs on congruent trials. Table 3 shows the correlations between inhibition scores and age group, type of education, language measures, and intelligence. All measures except type of education correlated negatively with inhibition; in general, the children in the higher age groups, those with more balanced lexicons, a larger vocabulary in Dutch, and/or a larger vocabulary in English, and those obtaining higher scores on the intelligence tasks showed a smaller Simon effect. Except for the correlation with balance, these results converge with previous findings with monolinguals and highly proficient bilinguals (Davidson et al., 2006; Diamond, 2013; Tse & Altarriba, 2014).

We started again with the basic model, first adding age group, type of education, and the interaction as fixed factors to the model ($-2LL = 1895.38$, $df = 13$, $R_m^2 = .128$, $R_c^2 = .142$). Only age group turned out to be a significant predictor of inhibition scores ($p < .001$). Thereafter, we added intelligence to the model. The model fit was not significantly different from the previous model ($-2LL = 1892.00$, $df = 14$, $R_m^2 = .140$, $R_c^2 = .152$), and intelligence was not significant. We removed intelligence from the model and added balance. This resulted in an improvement in model fit compared with the basic model ($-2LL = 1886.21$, $df = 14$, $R_m^2 = .145$, $R_c^2 = .150$). There was a significant and positive effect of balance ($p = .049$); children with more balanced lexicons showed larger Simon effects (indicating less inhibition). This result is difficult to interpret.² Fig. 4 shows that it is probably due to the positive relation between balance and inhibition in 4- and 5-year-olds. The cause might be that the 4- and 5-year-olds had more

² A negative effect was expected because children with more balanced lexicons were predicted to show a smaller Simon effect.

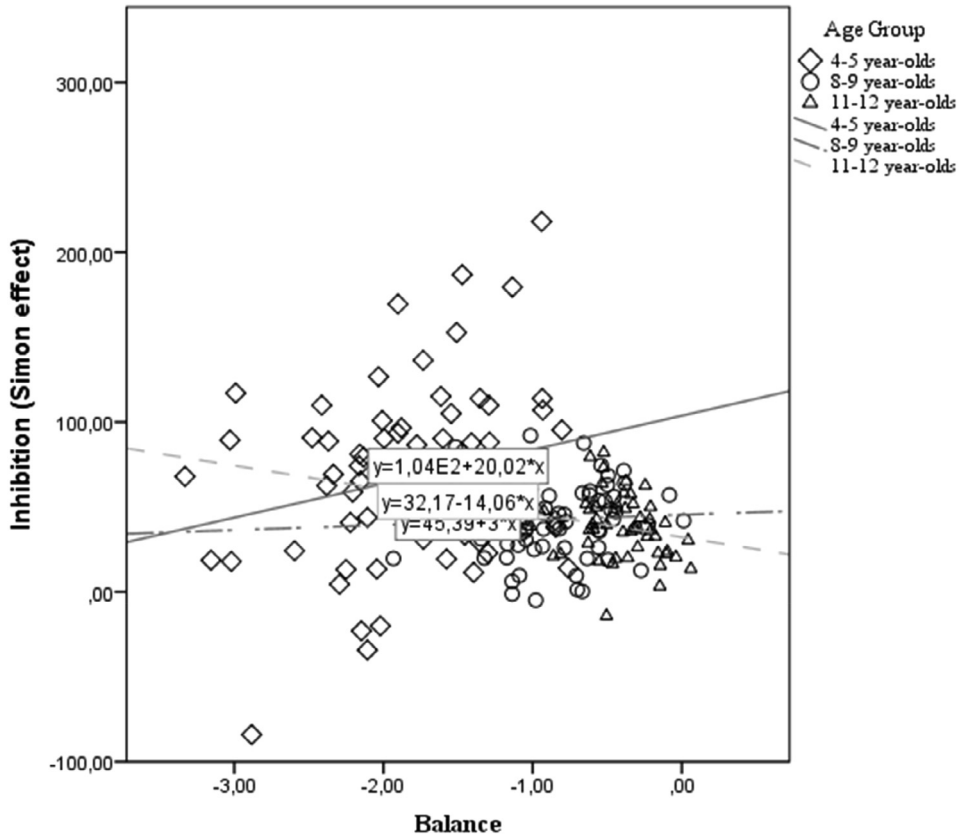


Fig. 4. Relation between balance and inhibition split out by age group.

time to respond. From Table 2, it is clear that the average Simon effect is substantially larger for 4- and 5-year-olds than for 8- and 9-year-olds and 11- and 12-year-olds. Therefore, we decided to leave the 4- and 5-year-olds out of the analyses. The basic model ($-2LL = 1068.62$, $df = 4$, $R_m^2 = .020$, $R_c^2 = .105$) was significantly different from the model with balance ($-2LL = 1063.27$, $df = 5$, $R_m^2 = .021$, $R_c^2 = .104$), but balance was no longer significant. Balance was first replaced by English vocabulary ($-2LL = 1070.59$, $df = 11$, $R_m^2 = .032$, $R_c^2 = .107$) and then by Dutch vocabulary ($-2LL = 1067.88$, $df = 11$, $R_m^2 = .037$, $R_c^2 = .099$), but both language measures showed a nonsignificant relation with inhibition ($p > .05$). The analysis was stopped at this point because there was no optimal model.

Working memory

Five children did not complete the working memory measures due to technical errors. Their scores were excluded from the analyses. Table 3 shows the correlations between working memory measures and age group, type of education, language measures, and intelligence. Except for type of education, working memory measures are highly and positively correlated with all measures.

Verbal working memory. We started with the same basic model again ($-2LL = 1039.68$, $df = 13$, $R_m^2 = .676$, $R_c^2 = .681$). Thereafter, we checked whether the individual measure intelligence or balance could improve the basic model. Adding intelligence significantly improved model fit ($-2LL = 1015.55$, $df = 14$, $R_m^2 = .733$, $R_c^2 = .738$). Including balance in the basic model did not result in a better $-2LL$ fit ($-2LL = 1047.41$, $df = 14$, $R_m^2 = .677$, $R_c^2 = .683$). Therefore, we continued with the model with intelligence as the only individual differences variable. Adding interactions between intelligence, age group, and type

of education did not improve the model ($-2LL = 1015.6$, $df = 25$, $R_m^2 = .743$, $R_c^2 = .745$). In the basic model, type of education was not significant. Removing this factor from the model containing intelligence resulted in a model that did not differ significantly from the previous one but had fewer degrees of freedom ($-2LL = 1021.83$, $df = 6$, $R_m^2 = .732$, $R_c^2 = .739$); therefore, this was considered to be a more parsimonious model. Including English vocabulary resulted in significantly decreased model fit ($-2LL = 1028.84$, $df = 7$, $R_m^2 = .732$, $R_c^2 = .738$). Although including Dutch vocabulary in the model revealed a significant relation between Dutch vocabulary and verbal working memory ($p = .026$), the model fit did not improve ($-2LL = 1022.44$, $df = 7$, $R_m^2 = .738$, $R_c^2 = .745$). Therefore, the model with age group and intelligence was considered the best model (see Table 5). In summary, verbal working memory was best predicted by age group and intelligence; older children and children with higher scores on the intelligence test generally performed better on the verbal working memory task.

Nonverbal working memory. Again, we started with the basic model ($-2LL = 1090.28$, $df = 13$, $R_m^2 = .681$, $R_c^2 = .692$). Adding Intelligence resulted in a significantly better model fit ($-2LL = 1082.00$, $df = 14$, $R_m^2 = .700$, $R_c^2 = .706$). Adding balance to the basic model did not improve model fit ($-2LL = 1088.28$, $df = 14$, $R_m^2 = .681$, $R_c^2 = .695$). Because the $-2LL$ was not significantly better than in the basic model, and balance was not significant, we rejected this model. We continued with the model including age group, type of education, and intelligence. Adding the interactions between intelligence, age group, and type of education did not improve the model ($-2LL = 1088.8$, $df = 25$, $R_m^2 = .698$, $R_c^2 = .700$). Type of education showed no significant relation with nonverbal working memory. Removing type of education did not improve model fit compared with the basic model plus intelligence, but this model had less degrees of freedom ($-2LL = 1090.10$, $df = 6$, $R_m^2 = .670$, $R_c^2 = .707$) and, therefore, was considered the most parsimonious model. Both age group and intelligence showed a significant relation with nonverbal working memory. Neither including Dutch vocabulary ($-2LL = 1093.03$, $df = 7$, $R_m^2 = .702$, $R_c^2 = .710$) nor including English vocabulary ($-2LL = 1096.55$, $df = 7$, $R_m^2 = .699$, $R_c^2 = .708$) resulted in a better model fit. In conclusion, nonverbal working memory performance is best predicted by age group and intelligence, with older children and children with higher intelligence scores performing better on the nonverbal working memory assessment (see Table 5).

Discussion and conclusion

This study investigated the relation between lexical balance and executive functioning in Dutch primary school pupils learning English as an L2. Pupils, who were 4 or 5 years old (Grade 1, i.e., kindergarten), 8 or 9 years old (Grade 5), or 11 or 12 years old (Grade 8, i.e., final grade of primary school), were enrolled in an early-English educational program or not, and both groups were exposed to English in everyday life, for example, via media. Children performed executive functioning tasks and Dutch and English vocabulary tasks. We investigated whether individual differences in Dutch–English lexical balance were related to differences in executive functioning and whether early-English children performed differently on the tasks than children from control schools.

We hypothesized that there would be a positive relation between lexical balance and executive functioning performance because lexical competition would be more demanding for L2 learners with more balanced language proficiencies (Blom et al., 2014; Vega & Fernandez, 2011). This hypothesis was confirmed, but only for switching. Our results are in line with previous findings showing that children who are more balanced in language proficiency show advantages in executive functioning (Blom et al., 2014; Thomas-Sunesson et al., 2018) and, in particular, with the results of Vega and Fernandez (2011), who found that child bilinguals who are balanced in language proficiency perform better than less balanced bilinguals in switching but not in inhibition. Our findings demonstrate, for the first time, that even for children who are exposed to the L2 in an instructed setting rather than a naturalistic setting, and who also have minimal input in the L2, lexical balance is related to executive functioning performance and to switching in particular. Our results, on the other hand, contradict the findings of previous studies with children and (young) adults (Gathercole et al., 2014; Paap et al., 2014; Paap et al., 2017) that found no relation between balanced bilingualism and switching RTs in card-sorting tasks. The participants in those studies, unlike those in our study, were either early bilinguals

or bilinguals who started L2 learning many years previously. A relation between L1/L2 proficiency and switching might exist only for individuals who, like our participants, are in the process of learning an L2. Such a relation may fade when managing two languages becomes automatized.

All tasks were chosen because they are considered good measures of executive functions (Diamond, 2013) and because they were suitable for use with 4- to 12-year-old children. For switching, the average performance was high: 76.1% in the youngest group and 87.6% in the oldest group. Our analyses revealed an effect of age group for all measures, thereby indicating that the tasks were suitable for measuring developmental differences in executive functioning. Despite careful selection of the tasks, we did not find a relation between either inhibition or working memory and any of the language measures. One possible reason for this is our measure of inhibition—RTs in the Simon task. Vega and Fernandez (2011) suggested that only older participants show a bilingual advantage on a timed inhibition task because in young children brain processes needed for optimal performance on such a task are not yet mature. Two studies provide evidence for this. Mohades et al. (2014) found that 8- to 11-year-old bilinguals and L2 learners showed over-recruitment of brain areas such as the bilateral cingulate cortex and larger Simon effects than monolinguals, whereas Yow and Li (2015) found a relation between lexical balance and inhibition in adult participants. In line with those studies, we found a trend for the expected relation between language balance and inhibition for pupils only in the highest age group ($r = -.205, p = .146$; i.e., more balanced pupils show a smaller Simon effect, indicating better inhibition), but not for younger children ($r = .052, p = .665$ for 8- and 9-year-olds; $r = .234, p = .049$ for 4- and 5-year-olds). Furthermore, following the standard procedure, participants were instructed to respond accurately and as quickly as possible in the Simon task. Young children tend to rush responses on timed tasks, thereby making errors (Diamond, 2013). Altogether, it may be that a relation between lexical balance and inhibition does exist but is observable only in younger participants when using a nontimed task such as the Day–Night test (Gerstadt, Hong, & Diamond, 1994).

Our results also tie in with previous research (Blom et al., 2014), showing no relation between lexical balance and nonverbal working memory. However, contrary to Blom et al. (2014), we did not find any relation between lexical balance and verbal working memory. These diverging results may be attributable to differences in language balance in both samples. Blom et al.'s sample was quite balanced; only two children showed a difference between Dutch and Turkish vocabularies of 10 or more points on scales ranging from 0 to 30 (Turkish) and 0 to 45 (Dutch) (Blom et al., 2014), whereas in our sample the average difference in the scores on the Dutch and English vocabulary tests was 56.47 points ($SD = 11.12$) on scales of 0 to 228 (English) and 0 to 204 (Dutch). It is worth noting that previous research has not consistently observed a bilingual advantage in working memory (Barac, Bialystok, Castro, & Sanchez, 2014). If concurrently managing two languages indeed requires greater working memory demands (Adesope et al., 2010), such a relation might exist only in more advanced L2 learners.

We expected that lexical balance would be a better predictor of executive functioning than L1 or L2 development alone because lexical competition should be more effortful once L2 learners are more proficient in two languages (Blom et al., 2014; Vega & Fernandez, 2011). Our results show that, indeed, lexical balance rather than L1 or L2 vocabulary knowledge best explained differences in switching. Our results thereby suggest that rather than L1 or L2 differences, differences in lexical balance, even if they are small, can explain variation in switching outcomes.

Our results confirmed our second hypothesis, namely that pupils enrolled in early-English programs would have larger English vocabularies and more balanced Dutch–English vocabularies than pupils from control schools. This suggests that pupils in early-English schools expand their knowledge of English at greater speed than their peers at control schools. When pupils start learning English, their level of English is much lower than their (already more greatly developed) Dutch. They can make larger improvements in English, and in theory eventually English may approximate the level of Dutch.

Previous research investigating pupils' knowledge of English at either the start or end of primary school (de Graaff, 2015; Goorhuis-Brouwer & de Bot, 2010; Lobo, 2013; Unsworth et al., 2014) showed that pupils enrolled in an early-English program outperform children from control schools. We extended previous studies by also including pupils who were halfway through their primary school career. However, we found a significant difference in English vocabulary between the two types of schools only for the oldest pupils. There are many factors that could influence English vocabulary size,

including the amount of classroom exposure to English, the teachers' English proficiency (Unsworth et al., 2014), and out-of-school exposure to English (de Graaff, 2015), and these may also have an effect on the development of executive functions. All early-English schools reported teaching English for at least 60 min per week, but personal communications with teachers revealed that they sometimes switched to Dutch during those lessons, so the actual time devoted to English may be less. Even though the available data from the questionnaire did not show differences in out-of-school exposure between early-English and control pupils, differences may actually exist because, unfortunately, data for the majority of the pupils (62.8%) are unavailable.

Our third hypothesis was that the group of 11- and 12-year-old early-English pupils would outperform control pupils of the same age on executive functioning tasks. Previous research (Purić et al., 2017) found advantages in working memory for pupils enrolled in a foreign-language learning program for 5 h per day, but the group that received only 1.5 h of instruction per day scored similarly to the monolingual group. Our participants' exposure to the L2 was more limited (in terms of hours per week), although overall length of exposure was longer (8 years for the oldest groups). No significant differences in executive functioning performance were found between the early-English and control pupils in any of the age groups. Our results show that devoting 15% of the teaching time or less to foreign language learning does not foster advantages in executive functioning or, at least if it does, these were not strong enough to be detected using the tasks employed in the current study. By including children from different age groups, we investigated linguistic and cognitive development over time. We made sure that children in the highest age groups had been enrolled in an early-English program since the start of primary school, but several factors may have influenced their exposure to English over that period; for example, the program may have changed over time, and children were educated by different teachers who likely differed in proficiency. Future research should follow children over time, thereby taking into account factors that influence the development of L1 and L2 vocabulary and, probably, also of executive functions in L2 learners. We found a relation between lexical balance and switching, arguing that those who are more balanced in their lexicons show better switching abilities. It could equally well be the case that children who have better switching abilities are better in mastering an L2. Again, a longitudinal study could allow for stronger claims about the causality of this relation. In addition, future research with a similar population but with higher exposure to the L2 could investigate whether the relation between language balance and executive functioning also holds for this group. Despite these limitations, our study with its cross-sectional design is a first and important exploration of L2 development in children enrolled in early-English programs.

Research on bilingual advantages in executive functioning has been highly controversial during recent years. Some authors claim that there is no convincing evidence of such advantages or that they may appear only under specific conditions (de Bruin et al., 2015; Paap & Greenberg, 2013; Paap et al., 2015). It has been suggested that studies should include at least two measures for each process of executive functioning to show that bilingual advantages are not task specific (Paap et al., 2015). Like in most studies with primary school pupils and teachers, time constraints made it impossible to include more tasks. Future research could include multiple tasks, especially for switching, to reveal whether the relation between lexical balance and switching can be generalized to other tasks.

In conclusion, the results of our study show that being exposed to an L2, even for a limited amount of time, is beneficial for pupils' vocabulary development in the L2 and, consequently, for their balance in lexical proficiency. In turn, language balance is related to switching abilities, but not to inhibition or working memory. We did not observe an overall global positive effect of bilingualism on executive functions. Our findings contribute to the ongoing debate about executive functioning development in (emerging) bilingual children, and we also show that language balance may be a more important predictor of executive functioning than L2 proficiency alone. These findings support the view that specific relations between bilingualism and executive functions may be determined by various variables such as age, type of bilingual setting, and stage of acquisition. Our study also supports the use of a relative proficiency measure that balances L1 and L2 vocabulary knowledge because it reflects the potential competition between the two languages involved. This measure seems particularly relevant when managing two languages is not yet automatized. It shows that, for children at that stage of development who are learning their L2 in an instructional setting, language balance is associated with the cognitive flexibility to switch between tasks.

References

- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*, 207–245.
- Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, *28*, 725–734.
- Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., ... Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*. <https://doi.org/10.3389/fpsyg.2014.00398>.
- Barac, R., Bialystok, E., Castro, D. C., & Sanchez, M. (2014). The cognitive development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*, 699–714.
- Barac, R., Moreno, S., & Bialystok, E. (2016). Behavioral and electrophysiological differences in executive control between monolingual and bilingual children. *Child Development*, *87*, 1277–1290.
- Bialystok, E. (2009). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition*, *12* (1). <http://doi.org/10.1017/S1366728908003477>.
- Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*, *128*, 105–119.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*, 282–298.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, *49*, 222–251.
- Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*, 2037–2078.
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*, 99–107.
- de Graaff, R. (2015). Vroeg of laat Engels in het basisonderwijs; Wat levert het op? *Levende Talen Tijdschrift*, *16*(2), 3–15.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168.
- Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2014). The inhibitory advantage in bilingual children revisited: Myth or reality? *Experimental Psychology*, *61*, 234–251.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4 manual*. Bloomington, MN: Pearson Assessments.
- Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody picture vocabulary test-III-NL* (2nd ed.). Amsterdam: Pearson Assessment and Information.
- EP-Nuffic, *Stap in de wereld: Starten met vroeg vreemdetalenonderwijs: Stappenplan*, 2015, The Hague, Netherlands.
- EP-Nuffic. (n.d.). *Locaties vvto-scholen*. Retrieved September 8, 2016, from <https://www.epnuffic.nl/primair-onderwijs/talenonderwijs/vroeg-vreemdetalenonderwijs-vvto/locaties-vvto-scholen>.
- Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., ... Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*. <https://doi.org/10.3389/fpsyg.2014.00011>.
- Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3½–7 years old on a Stroop-like day–night test. *Cognition*, *53*, 129–153.
- Goorhuis-Brouwer, S., & de Bot, K. (2010). Impact of early English language teaching on L1 and L2 development in children in Dutch schools. *International Journal of Bilingualism*, *14*, 289–302.
- Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81.
- Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*, 515–530.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS* (2nd ed.). New York: Taylor & Francis.
- Hedge, A., & Marsh, N. W. A. (1975). The effect of irrelevant spatial correspondences on two-choice response time. *Acta Psychologica*, *39*, 427–439.
- Kuppens, A. H. (2010). Incidental foreign language acquisition from media exposure. *Learning, Media and Technology*, *35*(1), 65–85.
- Lobo, V. R. (2013). *Teaching L2 English at a very early age: A study of Dutch schools*. Utrecht: Netherlands Graduate School of Linguistics. Retrieved from <http://www.lotschool.nl>.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100.
- Mohades, S. G., Struys, E., Van Schuerbeek, P., Baeken, C., Van De Craen, P., & Luypaert, R. (2014). Age of second language acquisition affects nonverbal conflict processing in children: An fMRI study. *Brain and Behavior*, *4*, 626–642.
- Morales, J., Calvo, A., & Bialystok, E. (2014). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114*, 187–202.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*, 615–639.
- Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278.
- Paap, K. R., Myuz, H. A., Anders, R. T., Bockelman, M. F., Mikulinsky, R., & Sawi, O. M. (2017). No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost. *Journal of Cognitive Psychology*, *29*, 89–112.

- Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review, 35*, 113–124.
- Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 113*, 535–551.
- Purić, D., Vuksanović, J., & Chondrogianni, V. (2017). Cognitive advantages of immersion education after one year: Effects of amount of exposure. *Journal of Experimental Child Psychology, 159*, 296–309.
- Ratiu, I., & Azuma, T. (2014). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology, 27*, 1–11.
- Simon, J. R., & Small, A. M. (1967). Processing auditory information: Interference. *Journal of Applied Psychology, 53*, 433–435.
- Thijs, A., Trimbois, B., Tuin, D., Bodde, M., & de Graaff, R. (2011). *Engels in het basisonderwijs*. Enschede, Netherlands: SLO.
- Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism, 21*, 197–206.
- Tse, C. S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: Evidence from Simon, Simon switching, and working memory tasks. *Frontiers in Psychology, 5*. <https://doi.org/10.3389/fpsyg.2014.00954>.
- Unsworth, S., Persson, L., Prins, T., & de Bot, K. (2014). An investigation of factors affecting early foreign language learning in The Netherlands. *Applied Linguistics, 36*, 527–548.
- van der Leij, A., Bekebrede, J., & Kotterink, M. (2010). Acquiring reading and vocabulary in Dutch and English: The effect of concurrent instruction. *Reading and Writing, 23*, 415–434.
- Vega, C., & Fernandez, M. (2011). Errors on the WCST correlate with language proficiency scores in Spanish-English bilingual children. *Archives of Clinical Neuropsychology, 26*, 158–164.
- Wechsler, D., & Naglieri, J. A. (2008). *WNV NL: wechsler nonverbal scale of ability: Nederlandstalige bewerking*. Amsterdam: Pearson.
- Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanisms of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in Psychology, 6*. <https://doi.org/10.3389/fpsyg.2015.00164>.
- Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297–301.