The relationship between second language acquisition and non-verbal cognitive abilities

Evy Woumans[1], Sofie Ameloot[1], Emmanuel Keuleers[2], & Eva Van Assche[1]

[1]Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Gent, Belgium

[2]Department of Cognitive Science and Artificial Intelligence, Tilburg University, Warandelaan 2, 5037 AB Tilburg, The Netherlands

Address for correspondence:

Evy Woumans

Department of Experimental Psychology

Ghent University

Henri Dunantlaan 2, 9000 Gent, Belgium

Telephone: +32 474 54 83 46

Email: evy.woumans@ugent.be

Abstract

We monitored the progress of 40 children when they first started to acquire a second language (L2) implicitly through immersion. Employing a longitudinal design, we tested them before they had any notions of an L2 (T0) and after one school year of L2 exposure (T1) to determine whether cognitive abilities can predict the success of L2 learning. Task administration included measures of intelligence, cognitive control, and language skills. Initial scores on measures of inhibitory control seemed predictive of L2 Dutch vocabulary acquisition. At the same time, progress on IQ, inhibitory control, attentional shifting, and working memory were also identified as contributing factors, suggesting a more intricate relationship between cognitive abilities and L2 learning than previously assumed. Furthermore, L1 development was mainly predicted by performance on inhibitory control and working memory.

Keywords: second language acquisition; immersion; cognitive control; cognitive development; language development

1. Introduction

Knowing more than one language is becoming increasingly important in today's society. As a consequence, an exceptional form of foreign language instruction is progressively making its way into current educational programmes, namely second language (L2) immersion schooling. In this type of education, L2 is the medium of classroom instruction for part of the curriculum. Naturally, parents may wonder whether their child will be able to succeed in acquiring an L2 in this context of implicit learning. Hence, this study set out to investigate whether it is possible to predict how proficient children will become in their new language by looking at their initial cognitive control skills.

Cognitive control is thought to be made up of three main executive functions; inhibition of prepotent responses, mental set shifting, and updating and monitoring of working memory representations (Miyake et al., 2000). Previous research into cognitive development has already demonstrated that these executive skills provide a critical foundation for school readiness (e.g. Blair, 2002; McClelland, et al., 2007). Specifically with regard to literacy, Blair and Razza (2007) established that inhibitory control was positively associated with letter knowledge in kindergarten. Furthermore, St Clair-Thompson and Gathercole (2006) exposed a correlation between literacy (measured through reading, writing, spelling, and handwriting) and both inhibitory and working memory skills, whereas Fitzpatrick and Pagani (2012) determined a positive link between working memory performance and receptive vocabulary. In addition, among elementary school children, attentional shifting seems to be predictive of spelling competencies (Lubin, Regrin, Boulc'h, Pacton, & Lanoë, 2016). Intuitively, one would assume it to be logical that measures such as verbal inhibition and verbal working memory correlate with other aspects of language proficiency. It is therefore important to note that these studies employed non-verbal measures of these executive functions.

Seemingly, language learning depends on non-verbal cognitive control; or at least, the above-mentioned research shows that executive skills could be predictive of success in first language (L1) acquisition. This may, however, not necessarily the case for L2 acquisition. In a study among children enrolled in L2 immersion kindergarten, Nicolay and Poncelet (2013) ascertained that verbal cognitive measures, such as phonological processing as well as auditory attention and flexibility, determined the success of vocabulary learning during the first stages of L2 acquisition. They were, however, unable to detect any relationship with non-verbal inhibitory skills. In contrast, Kapa and Colombo (2014) did show that inhibition together with WM span were significant predictors of artificial language acquisition in adults, with higher skills leading to better acquisition. Yet, in a similar experiment with children, the authors found that not inhibitory control, but attention shifting abilities together with WM span were determining factors.

Although research on the predictive relationship between cognitive control and L2 acquisition is scarce, there is a wealth of studies associating bilingualism with enhanced executive skills (see Bialystok, 2017 for a review; and de Bruin, Treccani, & Della Sala, 2014 and Lehtonen et al., 2018 for meta-analyses). The so-called bilingual cognitive advantage has been reported in various contexts and within different bilingual populations (e.g. Bialystok, Martin, & Viswanathan, 2005; Calvo & Bialystok, 2014; Crivello et al., 2016; Prior & Gollan; 2011; Woumans, Ceuleers, Van der Linden, Szmalec, & Duyck, 2015). Also in an immersion context, L2 acquisition has been shown to improve cognitive abilities (e.g. Bialystok & Barac, 2012; Nicolay & Poncelet, 2015; Woumans, Surmont, Struys, & Duyck, 2016). It is proposed that the mental juggling of two languages makes the bilingual's brain a more flexible and adaptable organ, and Green's (1998) model of inhibitory control (IC) can be used to explain these beneficial effects. It states that because the two languages are always simultaneously activated (e.g. Marian, Spivey, & Hirsh, 2003; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009), bilinguals experience a constant need to select one language for production while inhibiting the other. Due to this inhibitory practice and the need to be able to switch from one language to another, advantages have been found in children for inhibition (Poarch & van Hell, 2012) and shifting (Bialystok, 1999; Bialystok & Martin, 2004), as well as for WM (Kaushanskaya, Gross, & Buac, 2014; Morales, Calvo, & Bialystok, 2013).

Still, if superior cognition is a result of becoming bilingual, it does not necessarily convey anything about the initial cognitive capacities children must possess to acquire an L2 and succeed in immersion education. The current study's aim was therefore to determine whether specific CC components (inhibition, shifting, and WM) affect L2 learning in the same way as they do L1. In addition, it was our aspiration to predict success in acquiring a language through L2 immersion. We therefore monitored children's L2 receptive vocabulary in an immersion setting, when they first started learning their new form of speech. Importantly, our design was longitudinal, which entails that we had CC measures both at baseline (T0) before any L2 acquisition took place, and after one school year of L2 exposure (T1). This also gave us the opportunity to take into account children's initial L1 proficiency and its development in an immersion environment. Based on previous research, we anticipated that baseline CC performance would be predictive of L2 acquisition in the sense that higher performance leads to more acquisition. Analogously, we assumed that superior CC abilities would predict L1 development (e.g. McClelland et al., 2007). Lastly, we expected CC to improve between T0 and T1 due to normal age-related cognitive development (e.g. Best & Miller, 2010).

2. Method

2.1 Participants

Our participants consisted of 47 children (28 female, *M* = 58.53 months, *SD* = 3.46) attending their second year of kindergarten in May/June 2015 at one of the five selected second language (L2) immersion schools, all located within a small region of Wallonia, the French-speaking part of Belgium. In September 2015, these children started an immersion programme in their third year of kindergarten. In the case of our schools, Dutch was used as the L2 medium for instruction approximately 50% of the time. Approval for the study was obtained from our faculty's ethical committee. All subjects were healthy children and their participation was entirely voluntarily. Parental consent was obtained through an information letter and a document of informed consent, which were distributed via the schools. These documents were read and signed before task administration commenced. The description of the study's goal was kept vague to minimise the effects of confounding factors (such as parents devoting more time to practicing their child's L2 skills).

The children were tested for the first time at the end of their second year of kindergarten (T0) at the age of 4 to 5 years old (range: 55 to 65 months), before the start of immersion. Prior to the first testing, parents were asked to complete a questionnaire including questions about both the child's and parents' linguistic background and SES. It was imperative that children had so far only been exposed to their native language. Parents were also asked to report possible learning disorders or problems with language development, comprehension, and sight. No problems were indicated for any of our participants. At the second time point (T1) in March - April 2016 our sample included 40 children (21 female, *M* = 68.7 months, *SD* = 3.5) out of the initial 47; one was not retested due to computer failure at T0, four were sick at home, one changed schools, and another was considered not to be a native French speaker.

2.2 Materials

All children were tested individually for CC (inhibition, shifting, and WM), intelligence, and French language proficiency (receptive vocabulary) at baseline (T0) and after one year of immersion schooling (T1), At T1, the test battery also included a test of Dutch language proficiency (receptive vocabulary). The total duration for the test battery was between 30 and 45 minutes per child. All computer tasks were programmed in Tscope (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006) and presented on a laptop with a 15-inch monitor running Windows 7 32 bit, and reaction times as well as accuracy scores were recorded.

Échelle de Vocabulaire en Images Peabody (EVIP). This French translation of the Peabody Picture Vocabulary Task – Revised (Dunn, Theriault-Whalen, & Dunn, 1993) is a norm-referenced language assessment, which can be used to measure receptive vocabulary from the age of 2.5 to 18. The complete test is made up of 170 items, but only 25 to 50 items were needed to obtain accurate proficiency scoring for each of our subjects. Each item consisted of four black and white pictures presented on a card in multiple-choice format. Participants were asked to choose the picture that best depicted the word read aloud by the experiment leader. Test administration lasted approximately 15 minutes. Percentile scores calculated according to the EVIP manual were used for analyses.

Peabody Picture Vocabulary Task III – Dutch (PPVT). This test (Dunn & Dunn, 2005) - similar to EVIP – measures receptive vocabulary and is norm-referenced for participants between 2.3 and 90 years old. It can also be used with L2 learners. Administration of the test lasted between 10 and 15 minutes. Since the participants just started learning Dutch, test administration started with item 1 instead of the age-based start item. Raw scores were used for analyses, as the test is not normed for L2 learners.

Flanker. To measure inhibition, a flanker task was used. The task was based on the experiment developed by Eriksen and Eriksen (1974). Here, a central stimulus, < or >, is surrounded by four more 'flankers' (two on each side), which can be congruent or incongruent with the central stimulus. Conflict resolution (i.e. inhibition) is quantified as the congruency effect or 'Flanker effect' (i.e. the difference in performance between incongruent and congruent trials). To make the task suitable for children, each arrowhead was replaced by an image of a fish. The subjects were told that the central fish was called 'Jacques' and that they had to indicate the direction in which Jacques was swimming by pressing the corresponding button. These buttons (Q and M on an AZERTY keyboard) were labelled with stickers depicting a fish swimming to the left (Q) and to the right (M). Half of all trials were congruent (i.e. the central fish swimming in the same direction as the four flanker fish), the other half incongruent (i.e. the central fish swimming into the other direction). Each trial started with a fixation cross remaining on the screen for 500 ms, followed by the stimuli and an interval of 750 ms. The stimuli disappeared when the participant responded or until 3,000 ms had passed. Participants first completed a practice block of 10 trials. The experimental block consisted of 68 trials. After 34 trials, instructions reappeared and a break was suggested.

Dimensional Change Card Sorting (DCCS). This task was adapted from Zelazo (2006) and is a measure of attentional shifting. In our computerised version, we employed four different stimuli (red rabbit, blue rabbit, red boat, and blue boat) randomised across participants. In the pre-switch condition, children were asked to sort blue rabbits and red boats according to one dimension (colour or shape) by pressing a left or right button on the keyboard, which corresponded to a red rabbit and a blue boat. In the post-switch condition, the same stimuli had to be sorted according to the other dimension. The target and test dimensions were counterbalanced across participants. Each condition consisted of 16 trials. The starting dimension (i.e. colour or shape) and response mapping were counterbalanced across participants. Furthermore, we included a third condition, in which sorting rules were randomised. This condition also consisted of 16 trials, with a rule switch occurring randomly eight times. A border around the image implied it had to be sorted by shape; no border meant the colour rule was in order. A fixation cross appeared on screen for 1400 ms, after which the image was shown and followed by an interval of 750 ms. No time limit was set for responses.

Instructions were given at the beginning of a trial and the relevant dimension was repeated for each stimulus. Administration of the task took approximately five minutes.

Working Memory task. Based on Morales et al. (2013), this task was designed to manipulate working memory demands by comparing conditions based on two rules and four rules, which had to be kept in mind at all times. It consisted of six images in total, each requiring either a left or right response. As we realised that some children may not yet be able to distinguish between these two spatial responses, we assigned specific response colours to the targets during the instructions. The children learnt the association between the two key colours (yellow and purple; corresponding to the left and right Shift key on an AZERTY keyboard) and the six targets (red heart, turquoise flower, brown star, orange sun, green tree, blue cloud). Background colours were chosen to be different from the colours of the stimuli and response mapping was counterbalanced across participants. The circular yellow or purple background drawn around the targets were only present in the practice phase until children had learnt the association and were no longer visible in the testing phase. Each stimulus was presented either in the centre (neutral trials) or left and right of the screen (congruent and incongruent). When location and response mapping elicited the same response, a trial was considered congruent, if not, it was considered incongruent. The task was divided into four blocks of 40 trials each. Each block stood for one level of the 2 x 2 design: (2 stimuli vs. 4 stimuli) x (central presentation vs. side presentation). The first and second block consisted of two different stimuli (low WM load), while the third and fourth block consisted of four different stimuli (high WM load). The first and third block consisted exclusively of neutral trials, whereas the second and fourth were made up of congruent and incongruent trials. Block 1 and 3 started with four practice trials, Block 2 and 4 with eight practice trials. Each trial started with a fixation cross visible for 500 ms, followed by the stimulus that remained on the screen for a maximum of 3,000 ms and an interval of 500 ms. Instruction screens appeared before the start of a new block.

Raven's Coloured Progressive Matrices. To measure fluid intelligence, Raven's Coloured Progressive Matrices was administered (Raven, 2000). This test is a non-verbal measurement of fluid intelligence and can be used to evaluate general cognitive development, independently of linguistic development. The test consists of three different sets (A, AB, and B), each made up of twelve geometrical patterns with one piece missing. Participants were asked to complete the pattern. Standardised scores were calculated from the raw scores according to the manual (Raven, Court, & Raven, 1998) and used for further analysis.

Table 1. Demographics and task scores (with 95% confidence intervals between brackets). RT stand for reactions times and ACC for accuracy rates.

	T0	**T1**
N	40	40
Age (in months)	59.1 [57.9; 60.2]	68.7 [67.6; 69.8]
Male/Female ratio	19/21	19/21
Mother's education[a]	2.9 [2.7; 3.0]	2.9 [2.7; 3.0]
Father's education[a]	2.6 [2.4; 2.8]	2.6 [2.4; 2.8]
L1 vocabulary (Peabody French)[b]	60.7 [52.7; 68.6]	70.8 [62.9; 78.7]
L2 vocabulary (Peabody Dutch)[c]	N/A	32.6 [27.7; 37.5]
Raven's Matrices[b]	52.0 [42.7; 61.2]	73.3 [65.4; 81.2]
Flanker (RT in ms)		
Congruent	1413 [1314; 1511]	1190 [1104; 1275]
Incongruent	1578 [1452; 1703]	1275 [1176; 1375]
Flanker (ACC in %)		
Congruent	77 [72; 82]	91 [87; 95]
Incongruent	62 [54; 70]	85 [80; 90]
Working Memory (RT in ms)		
Block 1	1190 [1107; 1264]	971 [920; 1022]
Block 2		
Congruent	1240 [1168; 1311]	1043 [991; 1095]
Incongruent	1257 [1169; 1344]	1077 [1026; 1139]
Block 3	1286 [1202; 1370]	1080 [1021; 1141]
Block 4		
Congruent	1252 [1140; 1365]	1104 [1051; 1156]
Incongruent	1311 [1183; 1438]	1145 [1097; 1194]
Working Memory (ACC in %)		

Block 1	83 [78; 88]	91 [87; 94]
Block 2		
Congruent	81 [77; 88]	90 [86; 93]
Incongruent	83 [79; 87]	88 [84; 91]
Block 3	80 [75; 85]	89 [85; 93]
Block 4		
Congruent	82 [77; 86]	88 [84; 91]
Incongruent	76 [71; 81]	84 [79; 88]
Dimensional Card Sorting (RT in ms)	3005 [2441; 3569]	1677 [1574; 1779]
Dimensional Card Sorting (ACC in %)	84 [80; 88]	93 [90; 96]

Note. [a] Parents could indicate three options: 1 = first half of secondary school 2 = second half of secondary school, 3 = university or university college; [b] Percentile scores are reported; [c] Raw scores are reported.

3. Results

Our final sample size included 40 children, a sample equivalent to that of Kapa and Colombo (2014) who adopted a similar design and employed multiple regression analyses. We conducted an a priori power analysis using the correlation coefficients from Kapa and Colombo's models (Model 1: $R^2 = .424$, Model 2: $R^2 = .360$) to calculate the effect size, and included the number of predictors proposed by these authors (i.e. 7 and 4). Our analyses yielded sample sizes of 38 and 39 with an actual power of .95 and .96.

Table 1 reports mean reaction times (RT) from accurate trials only as well as accuracy rates (ACC). Analyses were performed on these results. Outlier RTs were trimmed individually by calculating a mean RT across all trials and excluding any response 2.5 SD of this mean. This procedure eliminated 2.2% of all flanker data, 2.6% of all WM data, and 4.0% of all DCCS data.

3.1 Progress on CC, IQ, and L1 French

Flanker task. The flanker task was analysed with a 2 x 2 (Time Point: T0, T1 x Congruency: congruent, incongruent) ANOVA. With regard to RTs, significant main effects were found for Time Point ($F(1,39) = 27.93$, $p < .001$, $\eta_p^2 = .417$) and Congruency ($F(1,39) =$

30.07, $p < .001$, $\eta_p^2 = .435$). Children were overall faster at T1 than at T0, and on congruent trials compared to incongruent trials. The Time Point*Congruency interaction was only marginally significant ($F(1,39) = 3.20$, $p = .081$, $\eta_p^2 = .076$), with the congruency effect being smaller at T1. ACC analyses resulted in a main effect of Time Point ($F(1,39) = 47{,}41$, $p < .001$, $\eta_p^2 = .549$) and Congruency ($F(1,39) = 14.50$, $p < .001$, $\eta_p^2 = .271$). Participants were more accurate at T1 and on congruent trials. The Time Point*Congruency interaction was also significant ($F(1,39) = 20.89$, $p < .001$, $\eta_p^2 = .349$), indicating smaller congruency effects at T1.

*Dimensional Change Card Sorting***.** Two paired *t*-tests showed RTs on the pre-switch ($t(39) = 6.26$, $p < .001$) and the post-switch ($t(39) = 4.51$, $p < .001$) condition were faster at T1 compared to T0. Another two paired *t*-tests indicated higher accuracy scores at T1 for both the pre-switch ($t(39) = -4.50$, $p < .001$) and the post-switch condition ($t(39) = -2.58$, $p = .014$).

*Working Memory task***.** WM RTs were analysed using a 2 x 2 design (Time Point: T0, T1 x WM Load: Block 1 = low, Block 3 =high). This resulted in a main effect of Time Point ($F(1,39) = 27.91$, $p < .001$, $\eta_p^2 = .417$) and of WM Load ($F(1,39) = 22{,}32$, $p < .001$, $\eta_p^2 = .364$). Participants were faster at T1 and in the low WM load condition. The interaction between Time Point and WM Load was not significant ($F(1,39) = 0.05$, $p = .818$, $\eta_p^2 = .001$). A 2 x 2 x 2 (Time Point: T0, T1 x WM Load: low, high x Congruency: congruent, incongruent) ANOVA showed a significant main effect of Time Point ($F(1,39) = 19.77$, $p < .001$, $\eta_p^2 = .336$) and Congruency ($F(1,39) = 13.12$, $p = .001$, $\eta_p^2 = .252$), but only a marginal effect of WM Load ($F(1,39) = 3.72$, $p = .061$, $\eta_p^2 = .087$). Participants were faster at T1 and for congruent trials. No significant interaction effects were found. For ACC, the 2 x 2 (Time Point x WM Load) ANOVA resulted in a significant main effect of Time Point ($F(1,39) = 10.82$, $p = .002$, $\eta_p^2 = .217$), but not of WM Load ($F(1,39) = 2.98$, $p = .092$, $\eta_p^2 = .071$). Participants were more accurate at T1. No significant interaction effect was found ($F(1,39) =$

0.40, $p = .529$). The 2 x 2 x 2 design (Time Point x WM Load x Congruency) ANOVA showed a significant main effect of Time Point ($F(1,39) = 11.66$, $p = .002$, $\eta_p^2 = .230$), of WM Load ($F(1,39) = 6.05$, $p = .018$, $\eta_p^2 = .134$), and of Congruency ($F(1,39) = 6.02$, $p = .019$, $\eta_p^2 = .139$). Participants were more accurate at T1, on congruent trials, and for low WM load. The WM Load*Congruency interaction was also significant ($F(1,39) = 6.30$, $p = .016$, $\eta_p^2 = .139$). Participants made more errors on incongruent trials when WM load was high.

*Raven Progressive Matrices***.** IQ was analysed using a paired samples *t*-test. IQ percentiles differed significantly for the two time points ($t(39) = -4.81$, $p < .001$), with higher scores at T1.

*Échelle de Vocabulaire en Images Peabody***.** A paired *t*-test showed that participants scored higher on L1 vocabulary at T1 compared to T0 ($t(39) = -2.38$, $p = .022$).

3.2 Influence of CC and IQ on L2 acquisition and L1 progress

Since mothers' education levels were uniformly high, we opted to use only father's education level as an indication of SES. The flanker congruency effect was computed comparing performance on congruent and incongruent trials for RT and ACC. Furthermore, RT and ACC progression scores for L1 French, IQ, and CC were computed by comparing performance on T0 and T1.

Due to the many variables, we employed a Principal Component Analysis (PCA) for both baseline and progress results. For the WM task, a correlation scatter plot indicated that WM ACC scores were concentrated at the high end of the range, so we applied a logit transformation. Using the Kaiser-Meyer-Olkin measure, anti-image correlations and Bartlett's test of sphericity, the number of variables included in the PCA on T0 measures was reduced from 15 to six WM predictors. These six variables loaded on two components. Component 1, which explained 39.9% of the variance, was associated with the RT measures of block 1, 3, and 4. Component 2, explaining 31.8% of the total variance, was associated with the ACC scores of Block 2, 3, and 4. The internal consistency of each component was assessed with Cronbach's α, which was .838 for Component 1 and .713 for Component 2. New WM RT and ACC regression variables were computed based on these two components. A second PCA was conducted on progression scores. The same method was used as with the previous PCA and resulted in one WM component, which explained 49.7% of the total variance. The component consisted of ACC progression in Block 2 and 4 of the WM task, and RT progressions in Block 1, 3, and 4. Cronbach's α = .730. A new regression variable for WM was computed based on this component.

Linear regression modelling was used to assess the influence of baseline IQ, CC, and SES on participants' L1 French progress (difference between T0 and T1 on French Peabody scores). Since the French Peabody score at T0 is already present in the outcome measure, it was not included in the analysis. Using the backward stepping method, 10 models were computed, where one variable at a time was removed. Durbin-Watson measure was 2.06. All models were significant, with Model 10 offering the best fit ($F = 9.28$, $p < .001$). Models 1 and 10 are reported in Table 2. Greater progress on L1 French vocabulary was associated with faster flanker RT, lower flanker ACC, and faster WM RT. It must be noted that the correlation between flanker RT and ACC was negative ($r^2 = -.527$, $p < .001$), indicating that these results were not an artefact due to a speed-accuracy trade-off.

The same linear regression modelling method was used to assess the influence of baseline IQ, CC, SES, and L1 French on L2 Dutch acquisition (i.e. Dutch Peabody scores). The backward stepping method provided 11 models. All variables that were included are reported in Model 1, presented in Table 2. The Durbin-Watson measure was 2.03 and Models 10 and 11 were significant. Model 11 offered the best fit for predicting the outcome measure L2 Dutch ($F = 3.26$, $p = .034$). This model together with the first is reported in Table 2. Higher performance on L2 Dutch vocabulary was associated with faster flanker RT, smaller flanker RT congruency effects, and slower DCCS Pre-switch RT. The lack of any correlation between DCCS RT and ACC measures ruled out effects of any potential time/accuracy trade-offs.

Another linear regression was employed to evaluate progress on IQ, CC, SES, and L1 French on L2 Dutch acquisition, which resulted in 7 models, all significant. The Durbin-Watson measure was 2.27 and Model 7 offered the best fit ($F = 7.09$, $p < .001$). This model together with the first is reported in Table 2. Higher performance on L2 Dutch vocabulary was associated with higher progress on IQ (increased percentiles), flanker ACC (increased ACC), flanker ACC congruency effects (smaller effects), DCCS Pre-switch RT (decreased RT), and WM RT (decreased RT).

Table 2. Regression models for dependent variables 'Progress L1 French' and 'L2 Dutch'.

	R^2	ΔR^2	β	*t*
Initial CC and progress L1 French				
Model 1	.597	.597		
Father's education level			.021	.134
T0 IQ			.246	1.280
T0 Flanker Overall RT			-.600	-3.104**
T0 Flanker Overall ACC			-.488	-2.724*
T0 Flanker Effect RT			-.135	-.790
T0 Flanker Effect ACC			-.190	-1.174
T0 DCCS Pre-switch RT			.035	.160
T0 DCCS Post-switch RT			-.358	-1.497
T0 DCCS Pre-switch ACC			-.020	-.090
T0 DCCS Post-switch ACC			-.166	-.800
PCA Working Memory RT			.515	3.104**
PCA Working Memory ACC			-.258	-1.416
Model 10	.465	-.023		
T0 Flanker Overall RT			-.705	-4.572***
T0 Flanker Overall ACC			-.550	-3.617**
PCA Working Memory RT			.452	3.378**
Initial CC and L2 Dutch				
Model 1	.403	.403		
Father's education level			.139	.674
T0 L1 French			.135	.491
T0 IQ			-.333	-1.432
T0 Flanker Overall RT			-.308	-1.026
T0 Flanker Overall ACC			.031	.124

T0 Flanker Effect RT			.292	1.365
T0 Flanker Effect ACC			.117	.536
T0 DCCS Pre-switch RT			.279	.935
T0 DCCS Post-switch RT			.171	.571
T0 DCCS Pre-switch ACC			-.268	-.955
T0 DCCS Post-switch ACC			.384	1.480
PCA Working Memory RT			-.305	-1.300
PCA Working Memory ACC			.025	.104
Model 11	.234	-.035		
T0 Flanker Overall RT			-.350	-2.128*
T0 Flanker Effect RT			.296	1.700
T0 DCCS Pre-switch RT			.356	2.135*
Progress CC and L2 Dutch				
Model 1	.620	.620		
Progress L1 French			.135	.763
Progress IQ			.489	3.396**
Progress Flanker Overall RT			.019	.104
Progress Flanker Overall ACC			.331	2.341*
Progress Flanker Effect RT			.143	.896
Progress Flanker Effect ACC			.336	2.417*
Progress DCCS Pre-switch RT			.393	2.101*
Progress DCCS Post-switch RT			.113	.673
Progress DCCS Pre-switch ACC			-.164	-1.164
Progress DCCS Post-switch ACC			-.075	-.488
Progress PCA Working Memory			-.358	-2.384*
Model 7	.550	-.034		
Progress IQ			.502	3.785**
Progress Flanker Overall ACC			.294	2.259*
Progress Flanker Effect ACC			.345	2.717*
Progress DCCS Pre-switch RT			.412	3.123**
PCA Progress Working Memory			-.307	-2.321*

Note. *** $p < .001$, ** $p < .01$, * $p < .05$

4. Discussion

At present, second language (L2) immersion programmes are quickly gaining popularity. The principal purpose of this method of schooling is to foster bilingualism in an implicit and natural manner. Nevertheless, parents may wonder whether their children will in fact be able to acquire this new language and whether native language development will suffer in the process, raising the question of whether success in this type of education can be anticipated. Indeed, studies into school readiness and first language (L1) acquisition determined that specific cognitive control (CC) functions, such as inhibition, attentional shifting, and working memory, are predictive of language learning success. This is however less clear for L2 learning. A study by Kapa and Colombo (2014) suggests that cognitive capacities (most notably shifting and working memory) may predict artificial language acquisition. Still, this may not necessarily be the case for natural L2 learning.

Hence, the present study set up a longitudinal protocol, charting the cognitive and linguistic development of 40 children enrolled in L2 immersion. At baseline, none of the children had any notions of an L2. They were tested at the end of second kindergarten (T0), before the start of the immersion programme in third kindergarten. Task administration included measures of fluid intelligence, attentional shifting, inhibitory control, working memory, and L1 receptive vocabulary. The second moment of testing (T1) took place at the end of third kindergarten, i.e. after one school year of L2 immersion, using the same test battery and adding a measure of L2 receptive vocabulary.

4.1 Progress on CC, IQ, and L1 French

Children's performance between T0 and T1 on CC improved over time. They were faster and more accurate in the flanker, the DCCS, and the WM picture task, indicating advanced processing skills. They also demonstrated a smaller congruency effect for accuracy in the flanker task, and equally, a marginally smaller effect for reaction times, which signifies better conflict resolution. Overall, the progress on these tasks was expected, because cognitive functions, upon which these measures rely, mainly develop between the ages of three and six (Best & Miller, 2010; Cragg, 2016; Garon, Bryson, & Smith, 2008).

Intelligence scores also differed significantly between both time points; percentiles incremented from the 52 to the 73. Since, Raven's Progressive Matrices is an age-normed test, the increase in IQ is unlikely to be an artefact, but rather a result of rising bilingualism. In fact, the same outcome was found by Woumans et al. (2016), where the increase in IQ scores took place only in a similar sample of immersion children and not in their monolingual peers. This may be evidence for a cognitive advantage yielded by L2 acquisition and is in line with studies documenting this type of bilingual benefits (e.g. Bialystok et al., 2005; Costa, Hernández, & Sebastián-Gallés, 2008).

With regard to L1 French vocabulary, results were analogous. The average percentile at T0 was 61, which increased to 71 at T1. Again, test scores were age-normed and should therefore be expected to remain steady. The EVIP has been tested extensively on validity and reliability, and it is therefore likely that this outcome reflects an increase in L1 vocabulary that cannot be attributed to age. Speculatively, the difference may be a result of immersion schooling, which was shown to improve metalinguistic awareness (Bialystok & Barac, 2012; Bialystok, Peets, & Moreno, 2014) and could also benefit L1 proficiency. However, as this study lacks a non-immersion control group, interpretation of these results should be treated with caution. In any case, we can safely deduce that French vocabulary acquisition was positive and immersion was therefore not detrimental to L1 development.

4.2 Influence of CC and IQ on L1 progress and L2 acquisition

Looking at progress for L1 French, we determined that better performance working memory task reaction times and on flanker reaction times was related to more extensive L1 vocabulary development. An association with the flanker congruency effect was absent, indicating that not necessarily inhibitory control but rather general processing speed may influence language development. Still, we also determined a relation between L1 vocabulary and lower accuracy scores on the flanker. This was quite unexpected, but a correlation analysis revealed that this observation was not the result of a speed accuracy trade-off. We therefore treat this particular result with caution. All in all, our findings are partly in line with the vast amount of research supporting an association between literacy and both inhibition and working memory (e.g. Blair & Razza, 2007; Fitzpatrick & Pagani, 2012; Peng et al., 2018; St Clair-Thompson & Gathercole, 2006). For instance, regarding the involvement of working memory, it has been demonstrated that novel word learning in L1 consists of learning both a sequence of sounds or letters and the correct order of the sounds, similar to learning sequences in a span task, which is considered to be a measure of working memory (Page & Norris, 1998). Still, many studies have also reported the involvement of these CC components in other types of academic achievement, such as mathematics (see Allan, Hume, Allan, Farrington, & Lonigan, 2014 and Yeniad, Malda, Mesman, van IJzendoorn, & Pieper, 2013 for meta-analyses). This may raise the question of whether these functions are task specific and directly required for performing academic tasks such as language learning, or rather domain-general (Yeniad et al., 2013). It may be that cognitive control facilitates behavioural regulation, increasing a child's ability to pay attention and inhibit inappropriate behaviour, and hereby generating optimal learning conditions.

Regarding the question of whether these different components of CC are predictive of L2 acquisition, our findings indicate that to a certain degree this is indeed the case. Analyses revealed that higher performance on baseline inhibitory control (flanker processing speed in general as well as flanker conflict resolution) was related to more extensive Dutch vocabulary knowledge. It thus seems that L2 acquisition – like L1 – benefits from the indirect impact of cognitive processing, which facilitates learning in general. However, the additional relationship between Dutch vocabulary and the flanker congruency effect implies a relationship between inhibitory control and L2 acquisition, which may be cultivated by the L2 learner's necessity to inhibit interference from L1. This view is supported by Green's Inhibitory Control model (1998), which states that inhibition is necessary when controlling two languages. We also found an association with slower reaction times on the DCCS. This was in strong contrast to the study of Kapa and Colombo (2014) that actually reported positive association between attentional shifting and artificial language learning, using the same task. However, this study only recorded accuracy scores, whereas in our computerised task, we were able to take into account precise reaction time measures. And although not significant, the relation between L2 Dutch and DCCS accuracy was positive in our sample as well. Although no trade-off was found between DCCS reaction times and accuracy, it is possible that children – especially at this age – generally focus more on being correct than being fast to respond. To support this assumption, we note that studies employing this task commonly measure only accuracy and not reaction speed (see, for instance, Zelazo, 2006).

Furthermore, the lack of SES as a significant predictor of L2 acquisition is also noteworthy. In essence, this implies that L2 learning does not necessarily depend on social status (e.g. Hoff, 2006), which in turn suggests that immersion schooling does not only benefit the 'elite'. Nevertheless, it should be taken into account that our measure of SES was only quantified by parental educational level. As a result, we cannot rule out a possible role of other socioeconomic aspects.

At the same time, L2 acquisition and progress on IQ, inhibitory control, attentional shifting, and working memory were associated. This suggests that the relation between cognitive capacities and L2 proficiency is not as straightforward as previously assumed. Rate of cognitive development, quantified as progress on intelligence and CC, seemed to determine the pace of L2 learning. Conversely, it is also possible that acquisition of an L2 actually fostered cognitive development, with more L2 learning leading to better progress for these different cognitive skills. This hypothesis is supported by the bilingual advantage theory, which states that controlling two languages leads to enhanced cognitive functioning. These advantages have been found for inhibitory control (e.g. Poarch & van Hell, 2012), shifting (e.g. Bialystok, 1999), and working memory (e.g. Morales et al., 2013) as well as intelligence (Woumans et al., 2016). In fact, these results shed light on the causal relationship between bilingualism and cognitive control, suggesting that it is indeed the acquisition of an L2 that leads to enhanced cognitive functioning, and not the other way around. It should, however, be noted that evidence supporting such an advantage is not clear-cut (see de Bruin, Treccani, &

Della Sala, 2014; Lehtonen et al., 2018), so these implications should be interpreted with caution.

5. Limitations and future directions

While the current study demonstrates a relationship between cognitive development and language acquisition (both L1 and L2), we acknowledge that L1 and L2 language proficiency only consisted of receptive vocabulary. Future research should consider including different measures, such as productive vocabulary and syntax. Regarding typology, the languages employed in this study were French and Dutch, which are similar in some ways but differ in others. Although they are derived from different language families (Romance versus Germanic), they still contain a fair amount of cognates as well as homonyms, and also share the same script. A study by Coderre and van Heuven (2014) showed that executive functioning is enhanced in similar-script bilinguals compared to different-script bilinguals, presumably because high orthographic overlap creates more cross-linguistic activation and hence demands more cognitive control, and most notably, inhibition (cf. Green, 1998). It is therefore possible that more similar or more dissimilar language pairs require different levels of baseline cognitive control and also affect cognitive progress in a different manner. Additional research is necessary to confirm this theory.

Furthermore, performance on cognitive tasks was quantified by accuracy as well as response times, but our conclusions are mainly based on results from the latter. We therefore feel it necessary to point out the variability of reaction time measures in children. Equally important is the manner in which individual differences other than cognitive capacities and additional variables may influence language development. To illustrate, recent research has demonstrated that different strategies for learning an L2 may match or mismatch with what works best for a child, and that genetic background as well as the age of acquisition may affect linguistic outcome (e.g. Vaugh & Hernandez, 2018).

6. Conclusion

All in all, this study provides evidence that performance on measures of inhibitory control and working memory are predictive of L1 proficiency, whereas only inhibitory control is associated with L2 acquisition. Equally important, is that all our participants were able to gain proficiency in L2, which indicates that enrolment in L2 immersion can benefit any child, and not only those with enhanced inhibitory skills. Crucially, L1 vocabulary also progressed (even more than would be expected looking at norm scores), demonstrating that acquiring an L2 via immersion does not impede L1 development. In addition, we found that the importance of IQ and socioeconomic status was limited, once more suggesting that immersion schooling is not only suited for the privileged few. In contrast, it may actually help to foster cognitive development.

Acknowledgements

This study was made possible by LEMMA (Language, Education, and Memory in Multilingualism and Academia), a Concerted Research Actions fund (GOA – BOF13/GOA/032) of Ghent University, and the Research Foundation - Flanders (FWO Belgium – FWO 1255214N).

Ethical considerations

All subjects were healthy children. Their participation was entirely voluntarily and parental consent was obtained through an information letter and a document of informed consent. These documents were read and signed before task administration commenced.

Author note

The data presented in this article have previously been disseminated in the form of a poster at the Conference on Multilingualism 2016.

Context

The idea for this study was a logical step following a previous study, in which we compared a similar group of second language immersion children to their peers in traditional monolingual education and found higher IQ scores for the former after a year of second language acquisition (Woumans, Surmont, Struys, & Duyck, 2016). Having demonstrated that learning a language may affect general cognition, we were wondering whether the opposite was also a possibility, i.e. whether cognitive capabilities can predict language learning skills. Relying on the methodology employed by Kapa and Colombo (2014) and using it in a context of natural second language acquisition is how we set up the current study.

References

Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology, 50* (10), 2368-2379. doi:10.1037/a0037493

Arsenian, S. (1937). Bilingualism and mental development. *The Teachers College Record*, *39* (3), 242-243.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development*, *81* (6), 1641-1660. doi:10.1111/j.1467-8624.2010.01499.x

Bialystok, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development, 70*, 636-644.

Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143* (3), 233-262. doi:10.1037/bul0000099

Bialystok, E. & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science, 7*, 325-339.

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism, 9* (1), 103-119.

Bialystok, E. & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition, 122*, 67-73.

Bialystok, E., Peets, K. F., & Moreno, S. (2014). Producing bilinguals through immersion education: Development of metalinguistic awareness. *Applied Psycholinguistics, 35* (1), 177-191. doi: 10.1017/S0142716412000288

Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57* (2),111-127.

Blair, C. & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78* (2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x

Calvo, A. & Bialystok, E. (2014). Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. *Cognition, 130*, 278-288.

Coderre, E. L. & van Heuven, W. J. B. (2014). The effect of script similarity on executive control in bilinguals. *Frontiers in Psychology, 5*, 1070. doi: 10.3389/fpsyg.2014.01070

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition, 106*, 59-86.

Cragg, L. (2016). The development of stimulus and response interference control in midchildhood. *Developmental Psychology, 52* (2), 242-252. doi:10.1037/dev0000074

Crivello, C., Kuzyk, O., Rodrigues, M., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2016). The effects of bilingual growth on toddlers' executive function. *Journal of Experimental Child Psychology, 141*, 121-132. doi:10.1016/j.jecp.2015.08.004

de Bruin, A., Treccani, B., & Della Sala, S. (2014). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science, 26* (1), 99-107. doi:10.1177/0956797614557866

Dunn, L. M., & Dunn, L. M. (2005). *Peabody Picture Vocabulary Test-III-NL, Nederlandse versie door Liesbeth Schlichting.* Amsterdam: Harcourt Assessment B.V.

Dunn, L., Theriault-Whalen, C., & Dunn, L. (1993). *Echelle de vocabulaire en images peabody: Adaptation française du peabody picture vocabulary test-revised.* Toronto, ON: PsyCan.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*, 143149. doi:10.3758/bf03203267

Fitzpatrick, C. & Pagani, L. S. (2012). Toddler working memory skills predict kindergarten school readiness. *Intelligence, 40*, 205-212. doi:10.1016/j.intell. 2011.11.007

Garon, N., Bryson, S. E., & Smith, M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin, 134* (1), 31-60. doi:10.1037/0033-2909.134.1.31

Green, D. (1998). Mental control of the bilingual lexico-semantic system*. Bilingualism: Language and Cognition, 1,* 67-81. doi:10.1017/s1366728998000133

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55-88. doi:10.1016/j.dr.2005.11.002

Kapa, L. L. & Colombo, J. (2014). Executive function predicts artificial language learning. *Journal of Memory and Language*, *76*, 237-252. doi:10.1016/j.jml.2014.07.004

Kaushanskaya, M., Gross, M., & Buac, M. (2014). Effects of classroom bilingualism on task-shifting, verbal memory, and word learning in children. *Developmental Science*, *17* (4), 564-583. doi:10.1111/desc.12142

Lehtonen, M., Soveri, A., Laine, A., Järvenpää, J., de Bruin, A., & Antfolk, J. (2018). Is bilingualism associated with enhanced executive functioning in adults? A meta-analytic review. *Psychological Bulletin, 144* (4), 394-425. doi: 10.1037/bul0000142

Lubin, A., Regrin, E., Boulc'h, L., Pacton, S., & Lanoë, C. (2016). Executive functions differentially contribute to fourth graders' mathematics, reading, and spelling skills. *Journal of Cognitive Education and Psychology, 15* (3), 444-463. doi:10.1891/1945-8959.15.3.444

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eye-tracking and brain imaging. *Brain and Language*, *86* (1), 70-82. doi:10.1016/S0093-934X(02)00535-7

McClelland, M. M., Cameron, C. E., Connor, C. M., Farris, C. L., Jewkes, A. M., & Morrison, F. J. (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. *Developmental Psychology, 43*, 947-959. doi:10.1037/0012-1649.43.4.947

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wagner, 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.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology*, *114* (2), 187-202. doi:10.1016/j.jecp.2012.09.002

Nicolay, A.-C., & Poncelet, M. (2013). Cognitive abilities underlying second-language vocabulary acquisition in an early second-language immersion education context: A longitudinal study. *Journal of Experimental Child Psychology*, *115* (4), 655-671. doi:10.1016/j.jecp.2013.04.002

Nicolay, A.-C. & Poncelet, M. (2015). Cognitive benefits in children enrolled in an early bilingual immersion school: A follow up study. *Bilingualism: Language and Cognition, 18* (4), 789-795. doi:10.1017/S1366728914000868

Page, M. P. A. & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105* (4)*,* 761-781.

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., William, D., & Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin, 144* (1), 48-76. doi:10.1037/bul0000124

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, 133*, 535-551.

Prior, A. & Gollan, T. H. (2011). Good language-switchers are good task-switchers: evidence from Spanish-English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society, 17*, 1-10.

Raven, J. (2000). The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cognitive Psychology*, *41* (1), 1-48. doi:10.1006/cogp.1999.0735

Raven, J. C., Court, J. H., & Raven, J. (1998). *Progressive Coloured Matrices*. Oxford: Oxford Psychologists Press.

St Clair-Thompson, H. L. & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, *59* (4), 745-759. doi:10.1080/17470210500162854

Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). Tscope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods*, *38* (2), 280-286. doi:10.3758/BF03192779

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, *20* (8), 923-927. doi:10.1111/j.1467-9280.2009.02389.x

Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck W. (2015). Verbal and non-verbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41* (5), 1579-1586.

Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The longitudinal effect of bilingual immersion schooling on cognitive control and intelligence. *Language Learning, 66* (2), 76-91. doi:10.1111/lang.12171

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences, 23*, 1-9. doi:10.1016/j.lindif.2012.10.004

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols*, *1* (1), 297-301. doi:10.1038/nprot.2006.46