LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

The Longitudinal Effect of Bilingual Immersion Schooling on Cognitive Control and Intelligence*

Evy Woumans,[a] Jill Surmont,[b] Esli Struys,[b] and Wouter Duyck[a]

[a]Department of Experimental Psychology, Ghent University and [b]Centre for Linguistics, Vrije Universiteit Brussels

Throughout the past century, the effects of bilingualism on general cognition have been extensively explored. Studies evolved from a negative to a more positive perspective, but longitudinal assessments of effects of bilingualism are scarce. This study investigated the long-term effect of becoming a bilingual on the development of general intelligence and cognitive control. We followed 27 five-year-old children initiating bilingual kindergarten and 27 age-matched controls enrolled in monolingual kindergarten. The two groups were similar with regard to socioeconomic status. At baseline, both groups spoke only French and performed equally on measures of intelligence, cognitive control, and verbal fluency. One year later, all children were tested again. Results revealed that, after 1 year, both groups improved similarly on verbal fluency and cognitive control. However, only children attending bilingual kindergarten improved significantly on intelligence, indicating that cognitive practice gained from acquiring a second language may improve general cognitive abilities assessed by intelligence tests, outside the verbal domain.

Keywords bilingualism; intelligence; cognitive control; cognitive development; second language acquisition

Introduction

Does becoming bilingual impair cognitive development or does it make children smarter? This question dates back to the first half of the 20th century, when there was a consensus that bilingualism was detrimental and that bilinguals performed worse on measures of intelligence (Darcy, 1946). This view

*We gratefully acknowledge support from Ghent University's Special Research Fund.

Correspondence concerning this article should be addressed to Evy Woumans, Henri Dunantlaan 2, B9000, Gent, Belgium. E-mail: evy.woumans@ugent.be

DOI: 10.1111/lang.12171

remained dominant until the sixties, when Peal and Lambert (1962) reported for the first time that bilinguals actually outperformed their monolingual peers on tests of intellectual reasoning. They argued that the constant switching between languages enhanced mental flexibility, yielding benefits for nonlinguistic mental abilities. This outcome was later confirmed by Ben-Zeev (1977). The difference between the earliest and later studies was that the former failed to control for confounding between-group variables. As pointed out by McCarthy (1930), bilingual children often had a lower socioeconomic status (SES). They also did not take into account children's degree of bilingualism (Brunner, 1929). Conversely, other studies employed such strict measures of control that they were matching the groups for the same abilities as those underlying the tasks for which they were trying to find differences (Hill, 1935).

Contemporary psycholinguistic research has now abandoned the broad concept of intelligence when measuring the effects of bilingualism on cognition. Instead, it focuses on the more specific concept of cognitive control (or executive functioning). This shift of interest evolved from the recent consensus that a bilingual's languages are always simultaneously active and interacting (Dijkstra, Grainger, & Van Heuven, 1999; Martin, Dering, Thomas, & Thierry, 2009) both during production (Costa, Caramazza, & Sebastián-Gallés, 2000) and comprehension (Van Assche, Duyck, Hartsuiker, & Diependaele, 2009). So, when bilinguals are reading or speaking in a given language (even their native language), the other irrelevant language is always also active to a certain degree. Hence, bilingualism implies constant cognitive conflict and requires monitoring each situation, activating the appropriate language, while resisting interference from the irrelevant language (Green, 1998).

Interestingly, this practice of language control is now assumed to transfer into, and improve, domain-general processes of cognitive control. For instance, Bialystok and colleagues showed that bilinguals show smaller Simon effects than monolinguals (Bialystok, Martin, & Viswanathan, 2005). Ample studies have now reported superior performance of bilinguals, relative to monolinguals, for several types of cognitive control tasks, throughout the entire lifespan. These studies included accelerated development of cognitive control in bilingual children (Bialystok & Barac, 2012; Carlson & Meltzoff, 2008; Kovács & Mehler, 2009a, 2009b), advanced cognitive control for bilingual young (Costa, Hernández, & Sebastián-Gallés, 2008) and middle-aged adults (Bialystok, Klein, Craik, & Viswanathan, 2004), improved cognitive reserve in ageing bilingual adults (Bialystok, Craik, & Luk, 2008), and delayed clinical manifestation of neurodegenerative diseases like Alzheimer's disease (Bialystok, Craik, & Freedman, 2007). Recently, however, these bilingual effects on cognitive control have failed to replicate consistently. Several authors have now reported null effects in different age groups (Antón et al., 2014; Duñabeitia et al., 2014; Kousaie & Phillips, 2012; Morton & Harper, 2007; Paap & Greenberg, 2013), and confounding variables other than bilingualism have been suggested as alternative explanations for between-group differences in studies claiming a bilingual advantage. So, following the research tradition started by Peal and Lambert (1962), history has repeated itself, and the much more recent bilingual cognitive control studies also resulted in disagreement, without reference to the much earlier bilingual intelligence debate.

As both research lines evolved into a debate about confounding variables, we propose that the only way to answer the question about the cognitive effects of bilingualism properly is to conduct longitudinal field studies. Moreover, we would like to extend the scope of the current-day research to once more include other cognitive aspects, such as general intelligence, as was the case in the early 1900s. As it happens, there is a significant amount of research focusing on how foreign language exposure affects reasoning and abstract thinking, which is almost entirely ignored in the current literature on the bilingual cognitive advantage. Some of these studies have found that bilinguals display, for instance, enhanced spatial reasoning skills (e.g., Greenberg, Bellana, & Bialystok, 2013) and improved analytical abilities (e.g., Marinova-Todd, 2012). Still, most of this type of research is also cross-sectional. The only previous study to follow our train of thought and perform a longitudinal study focusing on general intelligence reported higher scores in bilinguals (Hakuta & Diaz, 1985). Nevertheless, the bilinguals here were Hispanics with Spanish as their first language (L1), living in a second language (L2) dominant (English) environment (New Haven, CT). Also, this study did not include any other measures of cognitive control.

In the current study, we wanted to start from two monolingual groups, speaking the same L1, matched at baseline for intelligence as well as cognitive control, and for which bilingualism became an additional factor over time. Therefore, we followed a group of 54 monolingual French-speaking 5-year-olds in two types of kindergartens; for the last kindergarten year, half of children were about to enrol in a traditional monolingual and half in a bilingual program. Because it was impossible, and maybe even ethically unacceptable, to assign children randomly to educational programs independently from the parents' preference, we opted to match children's cognitive profile and background at baseline between the two groups that would result from the parent's choice. In this last kindergarten year there was no formal education; the bilingual option just implied that 50% of everyday classroom communication occurred in a new

L2, Dutch. Our main focus was the development of intelligence, as this remains the most frequently investigated and valid concept of cognitive ability. Given the evolution in the literature toward the more specific concept of cognitive control, we also included such a measure, even though these paradigms are primarily developed to study functional processes, and are not designed as equally reliable and normed measures of individual cognitive ability differences (Miyake & Friedman, 2012).

We employed Raven's Colored Matrices (Raven, Court, & Raven, 1998), a test of analytic reasoning, generally accepted as a good measure of fluid intelligence (Daley, Whaley, Sigman, Espinosa, & Neumann, 2003; Mani, Mullainathan, Shafir, & Zhao, 2013). Importantly, because this is a nonverbal test, it allows the assessment of general cognitive effects of bilingualism, independent of linguistic development, which may be influenced (either for the better or the worse) by becoming a bilingual. As a measure of cognitive control, we implemented the Simon task (Simon & Rudell, 1967), commonly used in the psycholinguistic literature discussed above. This is a spatial incompatibility task requiring rapid responses, sometimes contrary to initial impulses. At baseline, before the start of the last kindergarten year, both groups were matched for these measures. They were tested again one school year later to see whether attending bilingual kindergarten, and hence becoming a bilingual, had influenced development.

Materials and Methods

Participants

At the beginning of the school year (September 2012—T0), we started monitoring 54 preschool children who had only attended French-speaking kindergarten, and who spoke no other language at home. Initially, 64 participants were considered for participation from schools offering either monolingual ($N = 29$) and bilingual ($N = 35$) school programs. However, controlling for intelligence, SES, verbal fluency, and cognitive control left us to exclude 10 children with deviant (low or high) scores on any of these pre-matching variables (which could confound the effect of interest), so that the largest-possible ($N = 54$) subset of 2 equally-large groups remained, for which independent samples t-tests between groups yielded $p > .30$ for all matching variables. Hence, we selected two groups of maximally comparable pupils at T0; 27 children that would enrol in a bilingual final kindergarten year, with Dutch as a L2, and 27 in a traditional monolingual (L1) French program. Only these children were then followed for a year.

Participants were recruited from six different schools in the same region of the Walloon Community, Belgium. The pupil/teacher ration varied from 15 to 20 pupils per teacher, with one exception; a monolingual classroom with only six pupils. Kindergartens do not yet have final attainment levels or formal education, but they do have “developmental goals with regard to skills and social competences, which every school should pursue” (Portal Belgian Government, 2012). These were the same for both monolingual and bilingual programs, with the difference that in the bilingual programs 50% of the skills and competences were taught in a L2. These skills and competences were acquired through guided activities, such as songs, poems, stories, and different types of craft (e.g., drawing, coloring, painting, making things out of clay or building blocks). Depending on the school, the 50% L2 immersion (with L2 Dutch) was filled in by either alternating between a half-day of French instruction and a half-day of Dutch instruction or one day of French instruction and another of Dutch. During free activities (e.g., games on the playground), the teachers ensured that the children were speaking Dutch when they were having their Dutch day or half-day.

We contacted the schools before the children started their final (third) kindergarten year. For the bilingual schools, this is the year when pupils first come in contact with the L2 (Dutch). The two prior years are the same as in monolingual schools. We obtained consent from parents through an information letter distributed by the schools. From the three monolingual schools, we recruited, 5, 6, and 11 participants; from the two bilingual schools, 8 and 10 participants. One school offered both traditional and bilingual education. Here, there were 5 monolingual and 9 bilingual participants. Both the school and parents were blind to the actual purpose of the study. Instead, the aim was kept very vague in all communication, stating our intention was to record children’s development. This way, we were able to inform parents and schools about our test battery without revealing that afterwards, a comparison between monolingual and bilingual school programs was going to take place.

Consenting parents completed a questionnaire assessing the child’s and parents’ linguistic background and SES. They confirmed that their child did not have learning disorders or language development, comprehension, or sight problems. SES was a composite score of the parents’ educational (elementary, lower secondary, higher secondary, or higher education) and occupational levels, based on the Programme for International Student Assessment (PISA) classification (Organisation for Economic Co-Operation and Development, 2014), which is Europe’s most developed and comprehensive educational monitor.

All parents were monolinguals and none of the children had any exposure to other languages prior to the study. Our two groups were matched for intelligence, SES, verbal fluency, and cognitive control and also did not differ on age (months) or male/female ratio.

Design and Procedure

All tasks were administered two times for all 54 participants, once at the beginning (T0) and once at the end of the school year (T1). An entire session usually lasted no longer than 30 minutes per child and entailed a verbal fluency task, Raven's Coloured Matrices, and a Simon task. Initially the test battery also consisted of an Attention Network Test (ANT, Rueda et al., 2004), but this task appeared to be too difficult and was discarded for analyses below. At baseline, 25% of the children performed around chance level (<55% accuracy) and 45% performed below levels that are generally considered to be acceptable accuracy in this task (75%).[1] Testing took place in one of the empty classrooms of the children's respective schools. All participants received a present for participation.

Semantic Verbal Fluency

This was applied as a measure of L1 proficiency. Participants were given 60 seconds twice to verbally produce as many word exemplars as possible, within the categories "animals" and "things you can eat or drink." The order in which the categories were administered was counterbalanced across participants.

Raven's Coloured Progressive Matrices

Raven's Matrices is a non-verbal test of analytic reasoning, generally accepted as a good measure of fluid intelligence. The colored version is designed for children aged five to eleven (Raven et al., 1998). The test consists of 36 items over three sets (A, Ab, B). Within each set, colored items are ordered in terms of increasing difficulty. All items have a missing segment with six possible options for completion. Participants were asked to indicate which of the six pieces would fit the drawing best. Percentile scores were used for analyses and calculated from the raw scores, following the Raven's instruction manual (Raven, Raven, & Court, 1998). The manual offers a chart with normative scores for children from the ages of 48 to 120 months, for every 6 months. Extrapolations for ages in between (per month) were done according to the equations provided by the manual (Table 1). A percentile score of 50 is equivalent to an IQ of 100, the population average.

Table 1 Demographic data and verbal fluency scores by group (standard deviation in parentheses)

	Monolinguals	Bilinguals	Test	p
N	27	27		
Male/female ratio	15/12	13/14	$\text{Chi}^2(1) = 0.30$	.586
Age (months)	63.6 (4.2)	62.5 (3.8)	$t_{52} = 1.20$	.235
SES	3.1 (0.5)	3.3 (0.6)	$t_{52} = -0.90$	.375
Verbal Fluency at T0 (words)	16.9 (4.2)	15.6 (4.8)	$t_{52} = 1.00$	.320
Verbal Fluency at T1 (words)	20.6 (5.7)	20.3 (4.6)	$t_{52} = 0.49$	.627

Simon Task

This task was adapted from Simon and Rudell (1967). Colored dots appeared either on the left or right side of the screen. The children were asked to press the left (right) key on the keyboard when a green dot appeared, and the right (left) key when the red dot appeared as quickly and as accurately as possible. The response keys were marked with red and green stickers. Response mapping was counterbalanced across participants according to parity of participant number. Each trial began with a fixation of 800 milliseconds, then the colored dot appeared until the participant's response or up until 5,000 milliseconds. There was a 500-millisecond blank interval before the next fixation period. The task consisted of 10 randomized practice trials and four blocks of 25 randomized experimental trials. Half of all trials presented the colored dot on the same side of the associated response key (congruent trials) and half on the opposite side (incongruent trials). Participants could take a break between experimental blocks. They viewed the screen from a distance of approximately 60 centimeters. Stimuli were presented via Tscope software (Stevens, Lammertyn, Verbruggen, & Vandierendonck, 2006) on an IBM-compatible laptop with 15-inch screen, running XP.

Results

Analyses revealed that at baseline, the two groups did not differ for age, gender, SES, or verbal fluency (tests and results reported in Table 1), which confirmed their similar background. Crucially, they performed equally on Raven's intelligence test, $t_{52} = -.032, p = .975$ (age-controlled percentiles). Cognitive control tasks were analyzed by mean reaction times (RTs) and accuracy scores (ACC). Outlier RTs were trimmed for individual participants by calculating the mean across all trials and excluding any response deviating by more than 2.5 standard deviations of the mean. This procedure eliminated 3.58% of all Simon data. At baseline, groups did not differ for overall performance on the Simon task (RT: $t_{52} = .591, p = .407$; ACC: $t_{52} = -.118, p = .907$).

Verbal Fluency

To compare performance of the two groups over time, repeated measures analyses of variance were performed with "Time" as within-subject factor and "Group" as between-subject factor. Results revealed that both groups produced significantly more words at T1 than at baseline, $F_{1,52} = 33.55, p < .001, \eta\text{p}^2 = .392$, after a year of kindergarten (and aging), but there was no effect of Group, $F_{1,52} = .699, p = .407, \eta\text{p}^2 = .013$, and no Time × Group interaction, $F_{1,52} = .149, p = .701, \eta\text{p}^2 = .003$. This shows that both groups improved similarly on verbal fluency.

Raven's Coloured Progressive Matrices

Analyses were performed on both raw Raven scores and Raven percentile scores. Analyses on raw scores are reported in Appendix A. Percentile score analyses did not yield an overall effect of Group, $F_{1,52} = 1.93, p = .171, \eta\text{p}^2 = .036$, but did reveal a significant effect of Time, $F_{1,52} = 29.07, p < .001, \eta\text{p}^2 = .359$, on age-controlled Raven intelligence, qualified by the critical significant Time × Group interaction, $F_{1,52} = 7.69, p = .008, \eta\text{p}^2 = .129$. Intelligence of bilingual school children (T0 = 50.61, *SD* = 24.6; T1 = 76.04, *SD* = 20.6) improved significantly more than intelligence of monolinguals (T0 = 50.39, *SD* = 27.9; T1 = 58.54, *SD* = 30.2). Planned comparisons showed that the monolingual children did not improve significantly, $F_{1,52} = 3.43$, $p = .070$, whereas the bilingual children did, $F_{1,52} = 33.34, p < .001$ (Figure 1).

Simon Task

For the Simon task, Congruency was added as a within-subject (task) factor. RT analyses showed main effects of Time, $F_{1,52} = 35.67$, $p < .001$, $\eta\text{p}^2 = .407$, and Congruency, $F_{1,52} = 101.99$, $p < .001$, $\eta\text{p}^2 = .662$, showing that both groups improved over time and reacted faster to congruent trials, which validates the paradigm. There was, however, no effect of Group, $F_{1,52} = 1.17$, $p = .284$, $\eta\text{p}^2 = .022$, and no significant Time × Congruency interaction, $F_{1,52} = 2.53$, $p = .139$, $\eta\text{p}^2 = .042$. Neither were there any interaction effects with Group (for all interaction effects $p > .591$). Accuracy analyses yielded a main effect of Time, $F_{1,52} = 9.56$, $p = .003$, $\eta\text{p}^2 = .155$, and Congruency, $F_{1,52} = 35.18$, $p < .001$, $\eta\text{p}^2 = .404$, showing that both groups actually became a little less accurate over time, probably due to the fact they reacted much faster. There was no main effect Group, $F_{1,52} = .095$, $p = .760$, $\eta\text{p}^2 = .002$, and no

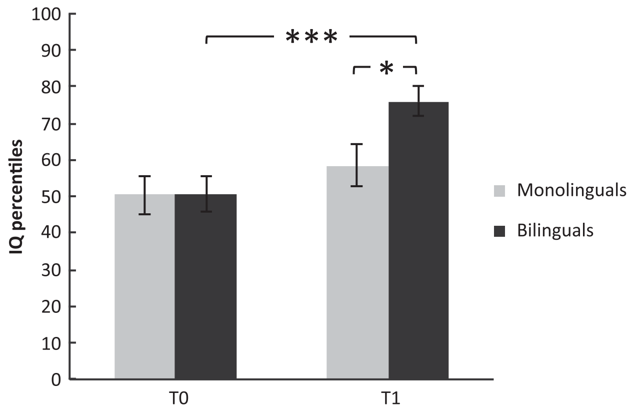

Figure 1 IQ percentiles for monolinguals and bilinguals at T0 and T1, derived from Raven's Coloured Matrices. Error bars reflect $\pm$ 1 standard error of the mean. All group and planned comparisons were performed, but only significant comparisons are indicated by horizontal bars. *$p < .05$, **$p < .01$, ***$p < .001$.

Time $\times$ Congruency interaction, $F_{1,52} = 1.36, p = .249, \eta p^2 = .025$. Interaction effects with Group were also not found (for all interaction effects $p > .387$; Figure 2). Intraclass correlation (ICC) analyses determined the reliability of RTs, ICC(C,k) = 0.651, $F_{1,52} = 2.86, p = .096$, 95% CI = [−0.85, 0.99].

Discussion

For almost a century, research has tried to determine the effects of bilingualism on general cognitive development. The earlier studies focused on the broad concept of intelligence and ended in disagreement following an evolution from a negative to a positive view on bilingualism. Recent literature on the more specific concept of cognitive control has also yielded inconsistent results, and showed the opposite evolution from positive effects to null results. For all of these studies however, bilingual effects were assessed by cross-group comparisons, which leaves room for confounding variables.

This is the first longitudinal study assessing the domain-general cognitive effect, for both cognitive control and general intelligence, of becoming a bilingual *within* participants, starting from two groups of IQ-matched, young, monolingual children who did not have any previous exposure to another language. There have been other longitudinal studies investigating the effects of L2 training, but these included participants that already had L2 knowledge prior to the study (e.g., Macnamara & Conway, 2014; Mårtensson & Lövdén, 2011). Another recent study by Nicolay and Poncelet (2015) looked at the effects of L2

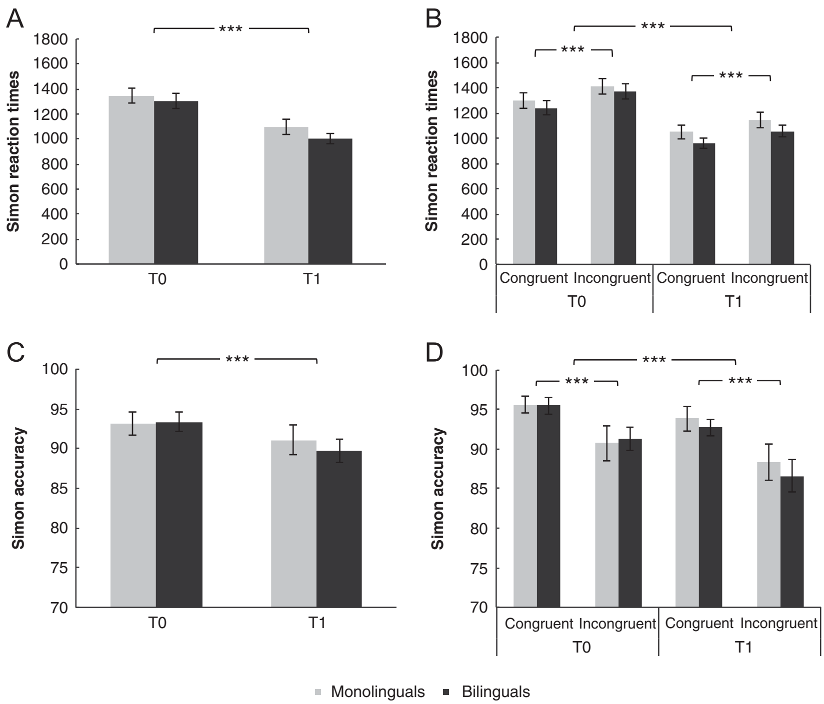

Figure 2 Reaction time and accuracy performance on the Simon task for monolinguals and bilinguals at T0 and T1. Overall reaction time performance. (a) Reaction time performance broken for congruency. (b) Overall accuracy performance. (c) Accuracy performance broken for congruency. (d) Error bars reflect ± 1 standard error of the mean. All planned comparisons and interaction analyses were performed, but only significant ones are indicated by horizontal bars. Top horizontal bars represent interactions. *$p < .05$, **$p < .01$, ***$p < .001$.

immersion in children over a period of 3 years, and reported attentional advantages for immersion children, but no intelligence advantage. However, here the immersion group was not matched at onset with the monolingual controls for this measure. Only raw Raven scores were mentioned, not age-adjusted, which did not differ significantly at the .05 level ($t_{99} = 1.93$, $p = .068$), but which were still substantially divergent. Furthermore, converting these average scores in age-adjusted percentiles showed that the immersion group scored 10 points higher (75th percentile) than the monolingual group (65th percentile) at the onset of the study. Apart from the fact that the groups did not match, these were exceptionally high scores to begin with, which also raises the question whether this study did not include a selective sample of highly intelligent children.

This already implied that especially the immersion group could not cognitively advance much more. In fact, the study even showed an odd pattern of losing 10 to 15 centile points in 3 years (immersion: 60th percentile, monolingual: 55th percentile). Hence, it is difficult to conclude whether L2 immersion had any other cognitive effects, apart from attention.

The results of the current study are somewhat different than those of Nicolay and Poncelet (2015), perhaps because it did match both groups at baseline on intelligence and cognitive functioning, as well as verbal fluency. First, it established that L2 exposure does not reduce L1 fluency, although subtler effects on production (e.g., lexical access speed) may exist (e.g., Gollan, Montoya, & Werner, 2002) for which the task that was currently administered as a matching variable is not well suited. Second, this study demonstrated a bilingual advantage for nonverbal intelligence, a measure on which our groups were initially matched. After 1 year of schooling, we observed gains in age-controlled intelligence scores of 17 percentile points for children attending bilingual kindergarten, whereas monolingual children only improved numerically. The cognitive control advantage was absent in this study, so the most straightforward interpretation is then that the bilingual cognitive control advantage just does not emerge if one adopts a longitudinal design with onset-matched cognitive control and intelligence skills. Another alternative explanation may be due to the low reliability of RTs and accuracy in control tasks as measures of individual cognitive ability differences (see also Miyake & Friedman, 2012; Paap & Greenberg, 2013). The intelligence test did not have this problem, as it has been standardized. Another possibility is that our group of bilinguals did have sufficient L2 exposure, but that the cognitive control advantage only emerges after particular bilingual experience, such as frequent L2 production or language switching (e.g., Prior & Gollan, 2011; Soveri, Rodriguez-Fornell, & Laine, 2011; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016).

Although the current research was partly focused on the effect of L2 immersion on cognitive control, the relation between bilingualism and cognitive development extends much further than this. There is, for instance, a vast amount of literature reporting how L2 exposure also improves spatial reasoning and perspective taking (Greenberg et al., 2013), abstract thinking (Cummins, 1978; Leopold, 1961; Yelland, Pollard, & Mercuri, 1993), abstract and symbolic representation skills (Adesope, Lavin, Thompson, & Ungerleider, 2010), and analytical ability (Marinova-Todd, 2012). These studies show that bilingual benefits are not limited to cognitive control, but instead suggest that the fundamental experience of being able to refer to the same object in two different ways (i.e., the name in L1 and the L2 translation equivalent) could very well accelerate the development of abstract reasoning in young children, prompting them to perform better on reasoning tests such as the Raven. Given the present findings, the recent cognitive control literature in the field of bilingualism could perhaps benefit from adopting this broader cognitive perspective again.

This study assessed bilingual cognitive advantages longitudinally, monitoring cognitive development within participants. As such, this is perhaps the most convincing demonstration of a bilingual advantage and a considerable extension of the literature that compares cognitive control measures across groups, sometimes not even taking intelligence measurements into account. We must, however, acknowledge that children were not randomly assigned by the authors to one of the kindergarten conditions. Such a study might never occur, as random educational assignment is unlikely to be acceptable for parents after full disclosure of the study's aim (e.g., who would accept the monolingual condition after disclosing the bilingual advantage hypothesis?) and perhaps not even ethically acceptable. Parents chose where their child would initiate kindergarten, and perhaps also whether or not they would enrol them in the immersion program, 2 years prior to our study. Nevertheless, our groups were perfectly matched at T0 (all ps $> .30$) on all relevant variables (verbal fluency, cognitive control, intelligence) after already having completed 2 years in their respective schools. It therefore seems very unlikely that any difference between parents would specifically benefit cognitive development in the third kindergarten year. Furthermore, the children were matched for parents' SES, using the method of the PISA to avoid a bias in children's backgrounds. SES matching is common (also accepted in sociological work focusing specifically on nurture effects) and often the only practical way to control children's background. In this respect, it is important to note that we also did *not* observe a bilingual (dis)advantage for verbal fluency or cognitive control, which suggest otherwise similar cognitive development.

To conclude, we have obtained evidence that the mental exercise and processes that are associated with becoming bilingual may have positive, long-term effects on general cognitive abilities, even outside the linguistic domain. One limitation is that we cannot conclusively state that these benefits on fluid intelligence hold as children become older and enter the formal education system. In order to ascertain whether or not this is the case, longer follow-up studies will be necessary. Nonetheless, even without the assurance that this bilingual advantage remains after a certain period of time, our findings are still extremely relevant for policymakers in education.

Final revised version accepted 19 February 2016

Note

1 We still analyzed the evolution of ANT performance between T0 and T1. Similar to the Simon task, no Group effect or Time × Group interaction was obtained (RT: $F_{1,52} = 2.595, p = .114$; ACC: $F_{1,52} = .761, p = .387$). ICC analysis determined RT reliability, ICC(C,k) = 0.280, $F_{1,52} = 1.39, p = .243$, 95% CI = [−2.82, 0.99].

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. doi: 10.3102/0034654310368803

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., et al. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*, 398. doi: 10.3389/fpsyg.2014.00398

Ben-Zeev, S. (1977). The influence of bilingualism on cognitive strategy and cognitive development. *Child Development*, *48*, 1009–1018. doi: 10.2307/1128353

Bialystok, E. & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*, 67–73. doi: 10.1016/j.cognition.2011.08.003

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symtoms of dementia. *Neuropsychologia*, *45*, 459–464. doi: 10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 859–873. doi: 10.1037/0278-7393.34.4.859

Bialystok, E., Klein, R., Craik, F. I. M., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, *19*, 290–303. doi: 10.1037/0882-7974.19.2.290

Bialystok, E., Martin, M. M., & Viswanathan, M. (2005). Bilingualism across the lifespan: The rise and fall of inhibitory control. *International Journal of Bilingualism*, *9*, 103–119. doi: 10.1177/13670069050090010701

Brunner, E. D. (1929). *Immigrant farmers and their children*. New York: Doubleday, Doran & Co.

Carlson, S. M., & Meltzoff, A. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*, 282–298. doi: 10.1111/j.1467-7687.2008.00675.x

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *26*, 1283–1296. doi: 10.1037//0278-7393.26.5.1283

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59–86. doi: 10.1016/j.cognition.2006.12.013

Cummins, J. (1978). Bilingualism and the development of metalinguistic awareness. *Journal of Cross-Cultural Psychology*, *9*, 131–149. doi: 10.1177/002202217892001

Daley, T. C., Whaley, S. E., Sigman, M. D., Espinosa, M. P., & Neumann, C. (2003). IQ on the rise: The Flynn Effect in rural Kenyan children. *Psychological Science*, *14*, 215–219. doi: 10.1111/1467-9280.02434

Darcy, N. T. (1946). The effect of bilingualism upon the measurement of the intelligence of children of preschool age. *Journal of Educational Psychology*, *37*, 21–44. doi: 10.1037/h0056165

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*, 496–518. doi: 10.1006/jmla.1999.2654

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., et al. (2014). The inhibitory advantage in bilingual children revisited. *Experimental Psychology*, *61*, 234–251. doi: 10.1027/1618-3169/a000243

Gollan, T. H., Montoya, R. I., & Werner, G. (2002). Semantic and letter fluency in Spanish-English bilinguals. *Neuropsychology*, *16*, 562–576. doi: 10.1037/0894-4105.16.4.562

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*, 67–82. doi: 10.1017/s1366728998000133

Greenberg, A., Bellana, B., & Bialystok, E. (2013). Perspective-taking ability in bilingual children: Extending advantages in executive control to spatial reasoning. *Cognitive Development*, *28* (1), 41–50.

Hakuta, K., & Diaz, R. M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. In K. E. Nelson (Ed.), *Children's language* (Vol. *5*, pp. 319–344). Hillsdale, NJ: Erlbaum.

Hill, H. S. (1935). *The effects of bilingualism on the measured intelligence of elementary school children of Italian parentage*. Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Kousaie, S., & Phillips, N. A. (2012). Ageing and bilingualism: Absence of a "bilingual advantage" in Stroop interference in a nonimmigrant sample. *The Quarterly Journal of Experimental Psychology*, *65*, 356–369. doi: 10.1080/17470218.2011.604788

Kovács, Á. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences*, *106*, 6556–6560. doi: 10.1073/pnas.0811323106

Kovács, Á. M., & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*, 611–612. doi: 10.1126/science.1173947

Leopold, W. F. (1961). Patterning in children's language learning. In S. Saporta (Ed.), *Psycholinguistics*. New York: Holt, Rinehart, & Winston.

Macnamara, B. N., & Conway, A. R. A. (2014). Novel evidence in support of the bilingual advantage: Influences of task demands and experience on cognitive control and working memory. *Psychonomic Bulletin & Review*, *21*, 520–525. doi: 10.3758/s13423-013-0524-y

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*, 976–980. doi: 10.1126/science.1238041

Marinova-Todd, S. H. (2012). "*Corplum* is a core from a plum": The advantages of bilingual children in the analysis of word meanings from verbal context. *Bilingualism: Language and Cognition*, *15*, 117–127. doi: 10.1017/S136672891000043X

Martin, C. D., Dering, B., Thomas, E. M., & Thierry, G. (2009). Brain potentials reveal semantic priming in both the "active" and the "non-attended" language of early bilinguals. *NeuroImage*, *47*, 326–333. doi: 10.1016/j.neuroimage.2009.04.025

McCarthy, D. A. (1930). *The language development of the pre-school child*. Minneapolis: University of Minnesota Press.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(8), 8–14. doi: 10.1177/0963721411429458

Morton, J. B., & Harper, S. N. (2007). What did Simon say? Revisiting the bilingual advantage. *Developmental Science*, *10*, 719–726. doi: 10.1111/j.1467-7687.2007.00623.x

Mårtensson, J., & Lövdén, M. (2011). Do intensive studies of a foreign language improve associative memory performance? *Frontiers in Psychology*, *2*, 12. doi: 10.3389/fpsyg.2011.00012

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*, 789–795. doi: 10.1017/s1366728914000868

Organisation for Economic Co-Operation and Development. (2014). PISA in focus. Retrieved from http://www.oecd.org/pisa/pisaproducts/pisainfocus/PISA-in-Focus-N36-(eng)-FINAL.pdf

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*, 232–258. doi: 10.1016/j.cogpsych.2012.12.002

Peal, E., & Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied*, *76*, 1–23. doi: 10.1037/h0093840

Portan Belgian Government. (2012). L'enseignement maternel. Retrieved from http://www.belgium.be/fr/formation/enseignement/maternel_et_primaire/

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. doi: 10.1017/s1355617711000580

Raven, J. C., Court, J. H., & Raven, J. (1998). *Progressive coloured matrices*. Oxford, UK: Oxford Psychologists Press.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. San Antonio, TX: Harcourt Assessment.

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Pappert Lercari, L., et al. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*, 1029–1040. doi: 10.1016/j.neuropsychologia.2003.12.012

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, *51*, 300–304. doi: 10.1037/h0020586

Soveri, A., Rodriguez-Fornells, A., & Laine, M. (2011). Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. *Frontiers in Psychology*, *2*(183), 1–8. doi: 10.3389/fpsyg.2011.00183

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*, 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 sentence context. *Psychological Science*, *20*, 923–927. doi: 10.1111/j.1467-9280.2009.02389.x

Verreyt, N., Woumans, E., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*(1), 181–190. doi: 10.1017/S1366728914000352

Yelland, G. W., Pollard, J., & Mercuri, A. (1993). The metalinguistic benefits of limited contact with a second language. *Applied Psycholinguistics*, *14*, 423–444. doi: 10.1017/s0142716400010687

Appendix A: Raven's Coloured Progressive Matrices

Analyses were also performed on raw Raven scores. These analyses revealed a main effect of Time, $F_{1,52} = 105.68$, $p < .001$, $\eta p^2 = .670$, but not of Group, $F_{1,52} < 1.0$, *ns*. Crucially, the Time × Group interaction was significant, $F_{1,52} = 4.49$, $p = .039$, $\eta p^2 = .080$, with the scores of bilingual children (T0 = 16.2, *SD* = 3.1; T1 = 22.0, *SD* = 3.6) improving significantly more over time than those of monolingual children (T0 = 16.4, *SD* = 3.9; T1 = 20.2, *SD* = 4.3). These findings are similar to those from the percentile score analyses.