IN PRESS/ Bilingualism : Language and Cognition

Running head: Executive control in immersion education

Executive control performance and foreign-language proficiency associated with immersion education in French-speaking Belgium*

Morgane Simonis[1]
Lize Van der Linden[1]
Benoit Galand[1]
Philippe Hiligsmann[2,]
Arnaud Szmalec[1,3,4]

[1]Psychological Sciences Research Institute, Université catholique de Louvain, Belgium
[2]Institut Langage et Communication, Université catholique de Louvain, Belgium
[3]Institute of Neuroscience, Université catholique de Louvain, Belgium
[4]Department of Experimental Psychology, Ghent University, Belgium

Supplementary materials
For supplementary material accompanying this paper, visit

* Acknowledgements:
This work was supported by a Concerted Research Action grant (ARC 14/19-061) from the Université catholique de Louvain (UCL) and the Université de Namur (UNamur), awarded to Philippe Hiligsmann (spokesman; UCL), Benoît Galand (UCL), Laurence Mettewie (UNamur), Fanny Meunier (UCL), Arnaud Szmalec (UCL) and Kristel Van Goethem (UCL). We thank Amélie Bulon, Audrey De Smet, Isa Hendrikx and Luk Van Mensel for their assistance in the data collection.

Address for correspondence:
Morgane Simonis
Psychological Sciences Research Institute
Université catholique de Louvain
Place Cardinal Mercier, 10
1348 Louvain-la-Neuve, Belgium
morgane.simonis@uclouvain.be

Abstract

A large sample study *(n* = 513) was conducted to investigate executive control performance in pupils following an immersion education program. We recruited 10-year-old children (n=128) and 16-year-old adolescents (n=127) who were enrolled in English or Dutch immersion education in French-speaking Belgium for at least 4 school years. They were compared to non-immersed children (n=102) and adolescents (n=156) on a number of executive control tasks assessing inhibitory control, monitoring, switching and attentional abilities. Several control variables such as receptive vocabulary, nonverbal intelligence, socioeconomic status and other potentially relevant background variables were also considered. Our results show significant gains in foreign-language proficiency for the immersed compared to the non-immersed participants. These gains were however not associated with any measurable benefits on executive control. Our findings make a unique contribution to understanding how language and cognition develop through formal education methods that promote bilingualism.

Keywords: bilingualism, foreign-language acquisition, immersion education, executive control

Introduction

Throughout the last decades, many studies concluded that using two or more languages in daily life is beneficial for cognitive functioning. This positive impact of bilingualism was demonstrated especially at the level of nonverbal executive control (e.g., Bialystok, Craik, & Freedman, 2007; Bialystok & Martin, 2004; Bialystok, Martin, & Viswanathan, 2005; Costa, Hernández, & Sebastián-Gallés, 2008; Martin-Rhee & Bialystok, 2008). Executive control is an umbrella term for a conglomerate of higher-order cognitive processes that are responsible for goal-directed behaviour. Throughout the years, several different executive control processes were put forward (e.g., Baddeley, 1996; Diamond, 2013; Miller, 2000; Miyake & Friedman, 2012; Smith & Jonides, 1999). Executive control in these models most often includes, amongst others, inhibitory control, working memory, attention, mental switching, monitoring, planning, updating, and problem solving (Chan, Shum, Toulopoulou, & Chen, 2008; Wang, Chan, & Shum, 2014).

Bilingualism might improve several executive control processes. First, both languages of bilinguals are always simultaneously active, regardless of their language proficiency (Blumenfeld & Marian, 2013; Dimitropoulou, Duñabeitia, & Carreiras, 2011; Duyck & Warlop, 2009). For bilinguals, communication in a particular language therefore requires the inhibition of the non-target language (Green, 1998). This continuous language control demand might train inhibitory control (e.g., Barac & Bialystok, 2012; Bialystok et al., 2004, 2005; Carlson & Meltzoff, 2008). Second, bilingualism might enhance overall monitoring skills (Bialystok, 2015; Costa et al., 2009; see Bialystok, Craik, & Luk, 2012, and Hilchey & Klein, 2011, for reviews). Bilinguals need to continuously monitor their known languages and attend to cues informing them which language to use. This is believed to improve bilinguals' overall performance on executive control tasks (Bialystok et al., 2005; Costa et al., 2008). Third, bilinguals often have to switch back and forth between their languages, depending on the circumstances. This is assumed to train mental switching (Bialystok & Martin, 2004; Prior & Macwhinney, 2010). Fourth, recent studies suggest that, apart from inhibitory control, monitoring, and switching, bilingualism might improve top-down attention modulation abilities (Grundy, Chung-Fat-Yim, C Friesen, Mak, & Bialystok, 2017; Grundy & Keyvani Chahi, 2017). While monitoring involves the adjustment to demands associated with a particular task or situation, top-down attention modulation rather reflects the ability to disengage attention from irrelevant information to focus on relevant information. Seemingly, bilinguals require a constant engagement and disengagement of attention from the non-target language to focus on the target language.

Finally, variations in the characteristics of bilingual language use might also engage and hence, train different aspects of language control. Green and Abutalebi (2013) suggested that different control processes are engaged as a function of bilinguals' particular interactional contexts (i.e., the adaptive control hypothesis). For example, when speaking with monolinguals, bilinguals must sustain attention to the current language while monitoring conflict and suppressing interference from the other. Depending on the linguistic profile of the interlocutor, bilinguals may also have to switch between languages, or code-switch, meaning that they alternate between their languages within the same conversation or utterance. Hence, communicating with monolinguals may primarily train bilinguals' inhibitory control and monitoring abilities, whereas code-switching with other bilinguals is more likely to improve mental switching (Green & Abutalebi, 2013; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016).

Although the bilingual executive control advantage has received wide empirical support, a number of more recent studies contradict its existence (Paap & Greenberg, 2013; Van Der Linden, Van de Putte, Woumans, Duyck, & Szmalec, 2018; see Lehtonen et al., 2018 for review). Therefore, the extent to which cognitive benefits of speaking multiple languages are restricted to specific executive control processes or to specific types of bilingualism, remains important but open questions. Given that half of the world's population is nowadays bilingual (Grosjean, 2010), understanding how this phenomenon influences cognition remains important.

In most previous studies, executive control advantages were examined in bilinguals that acquired a second language as a necessity of life (e.g., raised in multilingual families or after immigration) (e.g., Bialystok & Martin, 2004; Bialystok et al., 2005; Costa et al., 2008; Poulin-Dubois et al., 2011). More recently, some researchers also began to focus on particular educational methods that promote bilingualism. One type of foreign-language[1] education, which we focus on here, is immersion or Content and Language-Integrated Learning (CLIL). CLIL is a didactic method in which certain school subjects (e.g., geography, history, science, or mathematics) are taught in a different language than the main school language.

Only a handful of small-scale studies thus far examined the effects of immersion education on executive control. Carlson and Meltzoff (2008) investigated English children attending Spanish or Japanese immersion education for a period of six months. The immersed children did not outperform their monolingual peers on a wide range of executive control processes, including inhibitory control and mental switching.

Importantly, simultaneous bilingual children outperformed both the immersed and the monolingual groups. These findings suggest that the level of bilingualism attained after six months of immersion education may not be sufficient to obtain detectable executive control advantages. In a similar vein, Poarch and van Hell (2012) observed no executive control differences between monolinguals and German children immersed in English for 1.3 years, as examined with a series of inhibitory control tasks. However, they also observed that simultaneous bilingual children outperformed both the immersed and monolingual children on inhibitory control and attentional abilities. Like Carlson and Meltzoff (2008), Kaushanskaya, Gross and Buac (2014) found no advantage in mental switching for 7-year-old English children immersed in Spanish for two years compared to monolinguals. Nevertheless, several studies found a positive relation between immersion education duration and executive control performance, suggesting that immersion education might yield better executive control (e.g., Bialystok, Peets, & Moreno, 2014; Bialystok & Barac, 2012; Carlson & Meltzoff, 2008).

There are also a few studies that examined the cognitive effects of immersion education in French-speaking Belgium. In these studies, an executive control advantage was found after three years of immersion education in 8-year-old children immersed in English. An advantage of immersion education was found in attentional abilities, but not in inhibitory control (Nicolay & Poncelet, 2013, 2015). A recent longitudinal study of Woumans, Surmont, Struys and Duyck (2016) comparing 5-year-old French-speaking children immersed in Dutch with matched monolinguals, showed that one year of immersion education does not improve inhibitory control. However, the immersed children in Woumans et al. (2016) outperformed the monolingual group on nonverbal intelligence, suggesting an advantage in cognitive functioning. Altogether, the evidence regarding an executive control advantage emerging from immersion education is thus inconclusive. Prior studies also seem to suggest that a certain level or use of foreign-language proficiency is necessary for executive control advantages to emerge in a context of immersion education.

From a theoretical point of view, there are reasonable grounds to assume that immersion education might improve executive control, because this type of education is assumed to foster bilingualism. Immersion education offers a context in which children have more exposure to and proficiency in the foreign language than non-immersed children who learn this language through traditional language courses (Dalton-Puffer, 2011). In line with Grosjean (2010), immersed children are bilinguals because they use both the main school language and the foreign language in daily life (i.e., at school). Amongst the executive control demands inherent to bilingualism, a number of executive control processes may also be trained by immersion education. First, immersed children might train inhibitory control by controlling the non-target language, just like typical bilinguals (Green & Abutalebi, 2013). Furthermore, in immersion education, some interlocutors always have to be addressed in a particular language (e.g., immersion teacher), whereas others (e.g., classmates) may be addressed in different languages. Immersion schools are therefore dual-language environments in which children need to monitor and sustain attention to the target language, attend to cues informing which language to use, select the appropriate language, suppress non-target language interference, and switch efficiently between languages.

Of relevance for the current study, there might be important differences between immersed children and bilingual children enrolled in non-immersion education at the level of language control demands at school, the latter being the context to which children are more frequently exposed to. For non-immersed children, school is typically a single-language context, where all courses are given in the same language (their first or second language). Non-immersed children, also those raised in a bilingual home environment, can therefore use a global strategy of non-target language control (e.g., whole-language inhibition) at school, because everyone has to be addressed in the same language here. In contrast, immersed children require a more local strategy of language control (e.g., word-level inhibition), as they have to switch frequently between the main school language and the foreign language (Hofweber, Marinis, & Treffers-Daller, 2016; Van Assche, Duyck, & Hartsuiker, 2012). Therefore, the non-target language is likely to interfere with the target language at a different level (e.g., language- versus word-level) for immersed compared with non-immersed bilingual children, at least at school. As a result, executive control advantages might be qualitatively different for immersed and non-immersed bilingual children.

In the current study, we assessed whether specifically immersion education (through CLIL) is beneficial for executive control at the level of inhibitory control, monitoring, switching, and attentional abilities, above and beyond the question of whether or not typical bilingualism has an effect on these processes. Hence, we compared executive control performance between immersed and non-immersed children, beyond and above informal bilingual usage at home. Therefore, the study has unique contribution to the literature on cognitive consequences of becoming bilingual through an immersion education experience. Furthermore, we also tried to address a number of limitations in previous studies that might explain the conflicting results in the literature. First, the majority of studies on the cognitive benefits of immersion education used small-sized samples, which limits the reliability of the conclusions. In this study, we compared immersed and non-immersed participants in a large sample of over 500 participants. Second, we aimed at investigating the effects of immersion education at different developmental stages. Therefore, we recruited fifth grade primary children (10 years old) and eleventh grade secondary adolescents (16 years old) who were enrolled in immersion education for approximately the same duration. Third, previous studies investigated executive control especially during the early months or years of immersion education. As the development of executive control may depend on the duration of the experience with multiple languages, we examined whether a period of four to five years of immersion education is beneficial for executive control. Fourth, we investigated the potential beneficial effects of immersion education in different languages (Dutch and English). This should allow us to examine whether cognitive benefits of foreign-language learning generalize across languages, as was found in Carlson et al. (2008). Finally, one major difference between the present study and prior work on executive control in immersion education is that we brought several executive control processes together in one study. Indeed, of the relatively few studies examining the executive control abilities of immersed children, the majority focused on only one or two cognitive control processes. We aimed to investigate inhibitory control, monitoring, switching, and attentional abilities.

To assess the different executive control processes, we used the most extensively used tasks in the bilingualism literature that revealed bilingual advantages. To measure inhibitory control, we used two different tasks. The first task was the Simon task (Simon & Wolf, 1963), which typically measures prepotent response inhibition through the Simon effect (Simon & Rudell, 1967; see the Method section for details). The second task was the Attention Network Task (ANT; Fan, McCandliss, Sommer, Raz, &

Posner, 2002), which is more a measure of interference suppression, operationalized through a stimulus-response incongruency procedure. Although we had no strong a priori rationale to anticipate dissociations between both tasks, we opted for including both measures of inhibition in our study. This is because prepotent response inhibition and interference suppression were argued to be different types of inhibition (Friedman & Miyake, 2004), which might explain the inconsistencies in the bilingualism literature on inhibitory control. To assess monitoring abilities, we used two measures. First, we compared overall reaction times (RTs) between immersed and non-immersed children on both the Simon task and the ANT. Second, we assessed monitoring through the mixing cost using the Dimensional Change Card Sort (DCCS) task (Frye, Zelazo, & Palfai, 1995). The switching cost of the DCCS was used as a measure of mental switching.

In addition to inhibitory control, monitoring and switching abilities, which constitute the main focus of the present study, we also examined attentional abilities. A particularity of attentional abilities is that they are hard to separate from other executive control processes. For instance, a classic Eriksen flanker task measuring inhibitory control involves interference suppression, but also avoiding to attend to misleading information. Likewise, overall RTs on the Simon task and the ANT measuring monitoring abilities are dependent upon how well participants can attend to incoming information. Finally, switching between different tasks in the DCCS requires participants to shift their attention to the relevant characteristics of the stimuli (i.e., form or colour). Thus, attentional abilities are cognitive processes that are assumed to be involved in various executive control tasks, although they are at the same time often considered as executive control processes themselves (Sorge, Toplak, & Bialystok, 2017). Recent work suggested that attentional abilities, rather than inhibitory control, monitoring and switching, might be enhanced by bilingualism (Bialystok, 2015, 2017). In the current study, we therefore also assessed alerting and orienting in the ANT and top-down modulation of attention in the Simon task. Alerting refers to the ability to produce and maintain a state of readiness in order to process non-specific impending inputs and orienting refers to the ability to select the most relevant information from various sensory inputs (Fan et al., 2002; Posner & Petersen, 1990). The ANT is a combination of the classic Eriksen flanker task (Eriksen & Eriksen, 1974), measuring interference suppression, and the cueing task (Posner, 1980). The cueing demands of the ANT allow measuring how people maintain a state of alert and select relevant information from sensory input (alerting and orienting of attention). As noted earlier, Nicolay and Poncelet (2013, 2015) found an advantage for alerting after three years of immersion education. They did, however, not consider orienting abilities. Poarch and van Hell (2012) showed that at least a short period of immersion education does not improve orienting. Because our participants were immersed for a longer duration (four to five years), enhanced orienting skills for the immersed children, if they exist, may be more readily observable in our study. Finally, top-down attention modulation can be measured through the Simon task and recent evidence shows that bilinguals outperform monolinguals here (e.g., Grundy et al., 2017; Grundy & Keyvani Chahi, 2017). Altogether, we thus anticipated immersed children to outperform non-immersed children on the Simon task (inhibitory control, monitoring, attentional abilities), the ANT (inhibitory control, monitoring, attentional abilities), and the DCCS task (switching, monitoring).

In summary, we conducted a large-sample study with primary and secondary education non-immersed and immersed participants that were enrolled in CLIL for at least four years, to examine whether or not immersion education is beneficial for executive control. We compared the performance of the groups on three tasks assessing inhibitory control, monitoring, switching, and attentional abilities (alerting, orienting, and top-down attention modulation). Overall, we anticipated executive control advantages for the immersed groups over the non-immersed groups. We also predicted the executive control advantages to be more pronounced in primary than in secondary education for several reasons. First, primary immersed children in this study were enrolled in immersion education for a longer period than secondary immersed adolescents and the duration of immersion education has been found to be positively correlated with executive control (Bialystok & Barac, 2012). Second, immersion education in French-speaking Belgium involves a higher proportion of weekly immersion classroom hours in primary than in secondary education, which might lead to more executive control training. Third, immersed primary children were compared to non-immersed primary children who have not yet received foreign-language courses. In secondary education, on the other hand, the non-immersed adolescents all received traditional foreign-language courses for the same duration as the immersed adolescents, although the latter were exposed to the foreign language more frequently. Finally, given that the bilingual executive control advantage is believed to be observable specifically when executive control processes are still developing (Bialystok et al., 2005) the immersion advantage is more likely to emerge in primary children because most executive control processes are not mature until adolescence (Anderson, 2002; Anderson, Anderson, Northam, Jacobs, & Catroppa, 2001; Best & Miller, 2010; Diamond, 2013).

Method

Participants

Participants (*n*=813) from fifth grade primary (about 10 years old) and eleventh grade secondary (about 16 years old) education were recruited from twelve primary and nine secondary schools in Belgium. Belgium has four official linguistic regions (Dutch-speaking Flanders, French-speaking Wallonia, French-Dutch bilingual Brussels and German-speaking East cantons). Participants were recruited in the French-speaking region, which provides foreign language (Dutch/English) education through CLIL since 1998 (see Hiligsmann et al., 2017 for an overview). Immersion pupils represent approximately 4% of the primary and secondary total pupil population in Wallonia (ETNIC, February 2018). CLIL is available from the third year of kindergarten (about 5 years old), but children are also allowed to enter CLIL at a later age, namely in the seventh grade (about 12 years old). In the French-speaking schools that do not offer the CLIL program, Dutch or English are taught in traditional foreign language classes. These foreign languages are introduced only at the beginning of the fifth grade in primary (about 10 years old), with a frequency of one hour per week. Prior to this foreign-language initiation at school, pupils in Wallonia usually have no significant exposure to foreign languages. Thus, apart from simultaneous bilinguals, children in Wallonia are generally monolinguals when starting traditional foreign language courses or entering immersion education.

In the current study, primary children were in immersion since their final kindergarten year. Thus, they already completed five years of immersion education at the time of testing. Primary non-immersed children started traditional foreign-language introduction less than two months prior to testing, for one hour per week. Secondary adolescents were in immersion since their seventh grade. Thus, they already completed four years of immersion education at the time of testing. The non-immersed adolescents received traditional courses of a foreign language (Dutch/English) for 4 hours a week during the same period as the immersed adolescents. For immersion classes, depending on the school program, the mean proportion of school subjects taught in the foreign language was 50% (range 41-60%) in primary education and 27% (range 18-32%) in secondary education. The other subjects were taught in French.

Participants completed a questionnaire about variables such as age, gender and bilingualism. As to bilingualism, in terms of other languages than French outside the school context was measured on a 3-point Likert scale (1=Never; 2=Sometimes (e.g., with grandparents/friends); 3=Mostly (e.g., at home)). Parents also completed a questionnaire to identify possible developmental disorders. Based on the questionnaire, 17 participants with dyslexia (9 immersed) were excluded. All other participants had no learning, language, hearing, uncorrected visual, or neurological problems. The parental questionnaire also assessed the socioeconomic status (SES) of the family, as SES may have an influence on executive control abilities (Calvo & Bialystok, 2014). The education level of the mother, measured on a 3-point Likert scale, was used as a proxy for SES (1=primary/secondary education; 2=higher education; 3=university degree). Due to non-responders on the SES question, 116 participants were excluded from our sample. Finally, 167 immersed children and adolescents in our sample had not entered immersion education in third kindergarten or in seventh grade, or they had repeated a grade. They were discarded from the analyses to further increase the homogeneity of our sample.

The final sample included 513 participants (255 immersed and 258 non-immersed): 128 immersed and 102 non-immersed fifth grade children and 127 immersed and 156 non-immersed eleventh grade adolescents. Of these participants, 42% immersed children, 52% non-immersed children, 23% immersed adolescents and 35% non-immersed adolescents were active bilinguals that at least sometimes used a second language outside the school context[2]. Each pupil participated voluntarily and parental consent was obtained. The procedure was approved by the Ethics Committee of the Psychological Sciences Research Institute at the Université catholique de Louvain.

Materials and procedures

Participants were tested in groups (nine to 24 participants per session with one to three supervising experimenters). The tasks were computerized using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) and performed on azerty keyboards.

Background measures

Nonverbal intelligence

Nonverbal intelligence was measured with the Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1998).

French receptive vocabulary knowledge

French receptive vocabulary was measured using the Echelle de Vocabulaire en Images Peabody (EVIP; Dunn, Thériault-Whalen, & Dunn, 1993), a French adaptation of the Peabody Picture Vocabulary test (PPVT; Dunn & Dunn, 1981).

Foreign (Dutch and English) receptive vocabulary knowledge

Dutch and English receptive vocabulary was measured using the PPVT. More precisely, the PPVT-III-NL (Dunn & Dunn, 2005) and PPVT-IV (Dunn, Dunn, & Pearson

Assessments, 2007) were used for Dutch and English, respectively.

Executive measures

Simon task

In the Simon task, adapted from Simon and Rudell (1967), participants saw coloured squares on the left or right side of the screen. They were asked to indicate as quickly and accurately as possible whether the square was blue or red by pressing the left (a) or right (b) key on the keyboard, respectively. Position and colour elicited either the same (congruent trials) or different responses (incongruent trials). Congruent trials are usually processed faster and more accurately than incongruent trials. The size of this congruency effect, i.e. the so-called Simon effect (Simon & Rudell, 1967), reflects the ability to inhibit prepotent responses emerging from the location of the stimulus (i.e., inhibitory control).

The Simon effect also depends on the (in)congruency of the previous trial due to top-down attention modulation. As such, the Simon effect is reduced after an incongruent trial, which is known as the Gratton effect (Gratton et al., 1992). According to the conflict-monitoring hypothesis (CMH; Botvinick, Braver, Barch, Carter, & Cohen, 2001), when interference is detected (e.g., on an incongruent trial), the executive control-loop prioritizes the controlled processing route to override the erroneous prepotent response elicited by the automatic route. Therefore, on incongruent trials in the Simon task, controlled processing is biased in a top-down fashion. Subsequent incongruent trials will therefore produce less interference, reducing the Simon effect. We used the Gratton effect as a measure of the top-down attention modulation. Furthermore, overall RTs were taken as a measure of monitoring.

Each trial began with a centered fixation cross (“+”) for 800 ms, followed by a 250 ms blank interval. Then, a blue or red square appeared on the left or the right side of the screen for 1000 ms or until a response was given. A blank 500 ms inter-trial interval preceded the next trial. Response mapping between the colour and response key was counterbalanced across participants.

To familiarize participants with the response mapping and to provide additional instructions if needed, the task started with a Central task in which the coloured squares appeared on the center of the screen (Woumans, Ceuleers, Van der Linden, Szmalec, & Duyck, 2015). The Central task started with eight practice trials with feedback (exercising until 75% accuracy), followed by 40 trials. Next, the Simon task started with eight practice trials with feedback (exercising until 75% accuracy), followed by three blocks of 40 trials in total. Each block contained an equal amount of randomly presented congruent and incongruent trials.

Attention Network Task

On most trials in the ANT, adapted from Fan et al. (2002), participants saw five arrows and were asked to indicate as fast and accurately as possible the direction of the central arrow by pressing a left (a) or a right (p) key. The flanking arrows pointed either in the same (congruent trial) or opposite direction (incongruent trial) than the central arrow. There were also neutral trials, where only the central arrow was presented. Typically, performance is worse on incongruent than on congruent and neutral trials, because of the interference induced by the irrelevant flankers. The difference in performance on incongruent and congruent trials is known as the congruency effect. It reflects the ability to suppress interference of irrelevant information (inhibitory control).

Overall RTs are taken as a measure of monitoring abilities.

In addition, every trial in the ANT was preceded by one of four visual cues (see Figure 1): no cue, double cue (an asterisk above and below the fixation cross), central cue (an asterisk at the location of the fixation cross), and spatial cue (an asterisk at the location of the upcoming target stimulus, above or below the fixation cross). These cues allow investigating alerting and orienting abilities. The alerting effect is reflected by faster RTs when the stimulus is preceded by a double cue than when there is no cue. The orienting effect is examined by comparing performance on spatial cue trials, which indicates the location of the upcoming stimulus, and performance on central cue trials, which do not prime the location. Typically, RTs are lower on spatial cue than on central cue trials.

Each trial began with a centered fixation cross (“+”) for a randomly variable duration between 400 to 1600 ms. Then, a cue was presented for 100 ms, followed by a fixation cross for 400 ms. Subsequently, the target was presented for 1700 ms or until a response was given. The duration of the inter-trial interval, involving the presentation of a fixation cross, was variable depending on the duration of the first fixation cross and participants’ RT so that each trial lasted 4000 ms in total (see Figure 1).

The task started with a practice phase of six neutral trials with feedback, followed by 24 randomized congruent and incongruent practice trials (without cue) with feedback. The actual experiment consisted of three blocks of 48 trials, with each condition represented equally in a random order (three trial types: neutral, congruent, incongruent; four cue types: no, double, central, spatial).

<Insert Figure 1 about here>

Dimensional Change Card Sort task

In the DCCS task, adapted from Zelazo (2006) and Bialystok and Martin (2004), participants were asked to sort coloured geometric shapes depending on their colour or shape. A sorting cue was presented on the top of the screen to indicate the sorting rule with either a large rectangular colour gradient (the cue for colour) or four different empty geometric shapes (the cue for shape). Two buckets were located on the right and left bottom corner of the screen. The left bucket contained a red square and the right bucket contained a blue circle (see Figure 2). Depending on the sorting rule, participants had to sort the presented blue square or the red circle in the appropriate bucket by pressing the left (a) or the right (p) key as fast and accurately as possible (e.g., after a colour cue, a blue square goes in the right bucket (p); after a shape cue, it goes in the left bucket (a)).

Two measures of the DCCS are important here. The first measure, the switching cost, is the difference between switch and non-switch trials in the mixed-task. It reflects the difficulty to switch between sorting rules and is a measure of mental switching. The second measure, the mixing cost, is the difference between performance on single-task trials and non-switch trials from the mixed-task. It measures monitoring and reflects a more global sustained control mechanism that enables to maintain the two competing sorting rules which are necessary to make the correct responses (Braver, Reynolds, & Donaldson, 2003).

Each trial began with a centered fixation cross ("+") for 200 ms, followed by a 500 ms blank interval. The cue then appeared at the top of the screen for 250 ms and remained visible during the stimulus presentation at the center of the screen. The stimulus was presented for 4000 ms or until a response was given. There was an 850 ms blank inter-trial interval.

Participants performed two single-task blocks at the beginning of the task. During the first block, the pre-switch task, they needed to sort the stimuli either by colour or by shape (counterbalanced across participants). They were then asked to perform the second block, the post-switch task, where they needed to sort the stimuli by the other rule. For pre- and post-switch tasks, four practice trials with feedback were included (exercising until 75% accuracy), followed by 10 single-task trials. In the second part, the mixed-task, participants performed both the colour and the shape task in the same block. The mixed-task started with a practice phase of 12 trials with feedback (exercising until 75% accuracy), followed by 40 trials with an equal number of non-switch (same rule as previous trial) and switch trials (different rule than previous trial) of both the colour and shape tasks, randomly presented with a maximum of three consecutive trials of the same rule.

<Insert Figure 2 about here>

Results

Analyses were conducted for the two education levels (primary and secondary) separately. Dutch and English foreign-language learners were treated as a single group since preliminary analyses have shown no effect of Foreign language (Dutch or English) and no interaction between Foreign language and Group (immersion or non-immersion) on the executive control measures (all $\chi^2 < 1$).

Bayes factors (BF_{10}; Rouder, Speckman, Sun, Morey, & Iverson, 2009) were recently proposed as a more informative and reliable approach than *p*-values. They allow for an unbiased estimation of the effect of interest relative to the null model (Wagenmakers, 2007), which can explain why sometimes discrepancies occur between the two approaches. Results were interpreted based on BF_{10} but *p*-values are also reported for the interested reader. Bayesian analysis compares the fit of the data under the null hypothesis (immersed and non-immersed participants perform similarly) compared to the alternative hypothesis (immersed and non-immersed participants perform differently). BF_{10} varies between 0 and ∞. Values greater than 1 indicate increasing evidence for the alternative hypothesis over the null hypothesis and values less than 1 the reverse. As such, BF_{10} makes it possible to directly compare the relative strength of evidence for null and alternative hypotheses, which is not possible with *p*-values. We relied on the guidelines proposed by Jeffreys (1961) for interpreting BF_{10} (see Table S1 for details). These labels are merely used to facilitate interpretation and do not introduce cut-off values.

Background measures

Demographic data and *t*-tests or chi-square tests on the different background measures comparing our two groups (immersion and non-immersion) for the two education levels (primary and secondary) are shown in Table 1. $BF_{10}s$ were computed using JASP (JASP Team, 2017) with a default Cauchy prior width of r = .707.
The results indicate that, concerning the proportion of bilinguals, there was no difference between the immersed and non-immersed groups, neither in primary nor in secondary education (both substantial evidence). There was a higher SES for the immersed group than for the non-immersed group for both education levels (both

decisive evidence).

Analyses on raw scores of the Raven indicated higher intelligence for the immersed than for the non-immersed group both in primary (anecdotal evidence) and in secondary education (substantial evidence). As can be expected, SES and nonverbal intelligence were positively correlated in primary ($r(228) = .26, p < .001, BF_{10} < 150$) and secondary education ($r(281) = .21, p < .001, BF_{10} = 43.92$). When SES was introduced as a covariate in the analysis, the evidence in favour of nonverbal intelligence differences between the groups disappeared for both education levels ($BF_{10} < 3$ for both tests).

Analyses on raw French receptive vocabulary (EVIP) scores revealed no group difference in primary education (substantial evidence) and a higher score for the immersed than for the non-immersed group in secondary education (anecdotal evidence). The EVIP scores were within the normal range for all participants. For foreign-language receptive vocabulary, analyses were conducted on z-scores derived from raw scores of the PPVT-NL-III (Dutch) and the PPVT-IV (English), for each education level separately. Raw scores and BF_{10} for each foreign language are also reported in Table 1. We observed better foreign-language receptive vocabulary for the immersed over the non-immersed groups, for both education levels (decisive evidence). Note that these tests are not yet validated for foreign-language learners, which might explain the rather low performance of our participants. As such, after five years of immersion education, the 10-year-old primary immersed children obtained a mean score equivalent to 5.4-year-old native Dutch speakers (*SD* = 1.5 years) and to 4-year-old native English speakers (*SD* = 2 years). The primary non-immersed children attained a score equivalent to native Dutch speakers younger than 2.3 years old and to native

English speakers younger than 2.6 years old. These low scores are a consequence of the fact that these non-immersed children started foreign-language courses less than two months before testing. After four years of immersion education, 16-year-old secondary immersed adolescents obtained a score equivalent to 11-year-old native Dutch speakers (*SD* = 1.7 years) and to 9.3-year-old native English speakers (*SD* = 1.7 years). After four years of traditional foreign language courses at a rate of 4 hours per week, secondary non-immersed adolescents had a score equivalent to 7-year-old Dutch native speakers (*SD* = 2.2 years) and to 6.5-year-old native English speakers (*SD* = 2.6 years). Given that participants' French receptive vocabulary was within the normal range, receptive vocabulary of the immersed participants was thus lower in the foreign language than in French, which suggests that they were unbalanced bilinguals. Although far from reaching a native-like level of proficiency, immersed participants nevertheless all had better foreign-language knowledge than the non-immersed participants.

To summarize, as expected, we observed better foreign-language proficiency for the immersed groups than for the non-immersed groups. Nevertheless, we also observed differences in certain background variables. Both immersed children and adolescents had a higher SES than their non-immersed peers. In addition, the immersed adolescents were younger than the non-immersed adolescents and they had better nonverbal intelligence. Furthermore, there was a higher proportion of adolescent girls in non-immersion than in immersion. It is worth mentioning that SES (Calvo & Bialystok, 2014), age (Best & Miller, 2010), nonverbal intelligence (Friedman et al., 2006), and gender (Berthelsen, Hayes, White, & Williams, 2017) all influence executive control performance. We therefore took these group differences, and their possible influence on executive control performance, into account by entering them as covariates in the analyses for all executive control tasks.

<Insert Table 1 about here>

Executive measures

For all tasks, preliminary data treatment was as follows: RTs shorter than 200 ms, outliers and trials including incorrect responses were discarded from RT analyses (e.g., Poarch & van Hell, 2012). Outlier analyses were conducted by calculating participants' mean RT for each trial type and then excluding all responses below or above 2.5 *SD* of the mean. This led to an exclusion of 1.47% RT data for the Simon task, of 2.7% RT data for the ANT, and of 2.8% RT data for the DCCS task. Both RTs and accuracy (ACC) data were analysed by fitting Generalized Linear Mixed-effect Models (GLMMs) with maximum-likelihood estimation on individual trials, using the glmer function from the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015). Models on RT data assumed an Inverse Gaussian distribution, and a linear relationship between the predictors and RT to accomodate to the shape of the skewed RT data (Lo & Andrews, 2015). Planned comparisons were performed using the multcomp package (Bretz, Hothorn, & Westfall, 2010) with Bonferroni corrections. For main and interaction effects, $BF_{10}s$ were calculated with the Bayesian Information Criteria technique (Wagenmakers, 2007). We used Bayesian *t*-tests (with a default Cauchy prior width of r = .707 for effect size on the alternative hypothesis; Rouder et al., 2009) for $BF_{10}s$ of planned comparisons.

For each analysis, we applied the simplest model, which included the fixed effects and their interactions, as well as random intercepts for participants (see

Appendix S1 for the models used for each analysis for the different tasks). We also included by-participant random slopes when maximum-likelihood comparisons showed that the data justified their inclusion. The variables Age in years, Gender, Bilingualism, SES, Raven and EVIP were included as covariates. For Bilingualism, levels 2 and 3 of the scale-variable were combined in order to compute a factor-variable that controls for any other language use outside the school context. This procedure allowed us not to confound the potential executive control advantages of immersion education with those associated with second languages acquisition outside the school context.

Table 2 summarizes the comparisons between the Groups (immersion and non-immersion) on ACC and RTs for each Task (Simon task, ANT, and DCCS task) and Education level (primary and secondary). The results of the Group comparisons on the effects that were of main interest in our study are shown (i.e., overall RTs, congruency effect and Gratton effect for the Simon task; overall RTs, congruency, alerting effect, and orienting effect for the ANT; and switching cost and mixing cost for the DCCS task). In Table S2, the interested reader can find the remaining main and interaction effects, such as congruency, Gratton, alerting and orienting effects, that are beyond the scope of the inquiry in the current study. In Table S3, Kendall's tau correlations computed using JASP (JASP Team, 2017) between the different background measures (Age in years, Gender, Bilingualism, SES, Raven, EVIP, and foreign-language Receptive Vocabulary) and the executive control measures (RTs) are presented. Correlational analyses revealed significant negative correlations between the Raven and most executive control measures, indicating improvement (i.e., smaller RTs) in executive control with increasing performance on the Raven. Furthermore, the Raven was positively correlated with the EVIP and the foreign-language receptive vocabulary.

Finally, both the EVIP and foreign-language receptive vocabulary were negatively correlated with the different executive control measures. This correlation can be explained by the positive correlation between these two vocabulary measures and the Raven, which was also found in prior studies (Xiang et al., 2012).

<Insert Table 2 about here>

Simon task

Due to technical errors, the Simon task data of two non-immersed participants (one primary) were not retained. An additional seven children (four immersed) and one immersed adolescent were excluded because they had an ACC of less than 50% (chance level) at the Central task. Overall, mean ACC was high in primary (84.00 (0.36)%) and in secondary (93.00 (0.24)%) education. Mean RTs and ACC by Group (immersion and non-immersion), Trial Type (congruent and incongruent) and Previous Trial Type (congruent and incongruent) for each Education level (primary and secondary) are displayed in Table 3.

In primary education, for ACC, we observed a Simon effect (decisive evidence for higher ACC on congruent than on incongruent trials), but there was no overall Group difference (very strong evidence) and no interaction of Group and Trial Type (decisive evidence). For RTs, we observed a Simon effect (decisive evidence). There was no overall Group difference (decisive evidence) and no interaction of Group and Trial Type (decisive evidence). There was an interaction of Trial Type and Previous Trial Type (decisive evidence). Planned comparisons revealed a Gratton effect (decisive evidence for a larger Simon effect after congruent than after incongruent trials). There was no interaction of Group and the Gratton effect (substantial evidence). Thus, there was no evidence for group differences in Simon task performance, neither on ACC nor on RTs.

In secondary education, for ACC, we observed a Simon effect (decisive evidence), but no overall Group difference and no interaction of Group and Trial Type (decisive evidence for both tests). For RTs, we observed a Simon effect (decisive evidence). There was no main effect of Group and no interaction of Group and Trial Type (decisive evidence for both tests). There was an interaction of Trial Type and Previous Trial Type (decisive evidence). Planned comparisons revealed a Gratton effect (decisive evidence). There was no interaction of Group and Gratton effect (anecdotal evidence). Thus, there was no evidence for group differences in Simon task performance, neither on ACC nor on RTs.

<Insert Table 3 about here>

Attention Network Task

Due to technical errors, the ANT data of three children (one immersed) were not retained. As in Poarch & van Hell, 2012, neutral trials were not analysed and only used as a baseline. Overall, ACC was high in both primary (93.00 (0.24)%) and secondary (97.00 (0.15)%) education. Mean RTs and ACC by Group (immersion and non-immersion), Trial Type (congruent and incongruent) and Cue Condition (no, double, central, spatial) for each Education level (primary and secondary) are displayed in Table 4.

In primary education, for ACC, we observed a congruency effect (decisive evidence for higher ACC on congruent than on incongruent trials) and an orienting effect (strong evidence for a difference between central and spatial cue trials). There was no evidence for any other main or interaction effects (substantial to decisive evidence). For RTs, we observed congruency, alerting (no cue – double cue trials) and orienting effects (all decisive evidence). However, there was no overall Group difference (decisive evidence) and no interaction of Group and Trial Type (very strong evidence). There was also no interaction of Group and the orienting effect (substantial evidence) and of Group and the alerting effect (anecdotal evidence). Thus, performance of the two groups did not differ on ACC and on RTs of the ANT.

In secondary education, for both ACC and RTs, conclusions of the analyses were the same as for primary education. That is, the performance of the two groups was similar in terms of congruency, monitoring, alerting, and orienting effects.

<Insert Table 4 about here>

Dimensional Change Card Sort task

Analyses were first conducted on single-task trials (pre- and post-switch) to investigate whether both groups had the same baseline performance. For both education levels, there were no baseline differences between the groups, neither for ACC nor for RTs (all very strong to decisive evidence). Overall, ACC was high in primary (84.00 (0.36)%) and in secondary (93.00 (0.25)%) education. Mean RTs and ACC for Group (immersion and non-immersion) and Condition (pre-switch trials, post-switch trials, switch trials, non-switch trials) for each Education level (primary and secondary) are shown in Table 5.

In primary education, for ACC, we observed a switching cost (decisive evidence for higher ACC on non-switch than on switch trials), but it did not differ across Groups (strong evidence). For RTs, there was also a switching cost (decisive evidence for shorter RTs for non-switch than for switch trials) that did not differ across Groups (strong evidence). Moreover, for ACC, there was no mixing cost (strong evidence) and no Group difference (very strong evidence). For RTs, there was a mixing cost (decisive evidence for shorter RTs for single-task than for non-switch trials), but it did not differ across Groups (strong evidence).

In secondary education, for ACC, we observed a switching cost (decisive evidence), but it did not differ across Groups (very strong evidence). For RTs, there no switching cost (very strong evidence) and it did not differ across Groups (strong evidence). Moreover, for For ACC and RTs, we observed a mixing cost (decisive evidence), but it did not differ across Groups (very strong evidence).

<Insert Table 5 about here>

Discussion

The primary goal of this study was to examine whether immersion education leads to better executive control. Despite the increasing number of schools and pupils enrolled in immersion education, the cognitive effects of foreign-language acquisition through formal education are just starting to be investigated. We collected data from a large sample of 10-year-old children and 16-year-old adolescents, enrolled in immersion education for five and four school years, respectively. Based on a few previous studies that investigated the cognitive benefits of the first years of immersion education (Carlson & Meltzoff, 2008; Nicolay & Poncelet, 2013, 2015; Poarch & van Hell, 2012; Woumans et al., 2016), as well as a study showing that the duration of immersion education is positively correlated with executive control performance (Bialystok & Barac, 2012), we anticipated the immersed groups to outperform the non-immersed groups on inhibitory control, monitoring, switching, and attentional abilities.

These executive control processes were assessed using three widely used tasks to investigate executive control advantages of bilinguals: the Simon task (measuring inhibitory control, monitoring, and attentional abilities), the ANT (measuring inhibitory control, monitoring, and attentional abilities), and the DCCS task (measuring switching and monitoring). First, in the Simon task, our results yielded clear Simon and Gratton effects for all groups. Despite the fact that these established behavioural markers of executive control were observed, our study did not reveal any group differences. That is, there were no differences in inhibitory control and top-down attention modulation between immersed and non-immersed children and adolescents. Second, the results of the ANT showed that all the groups had the predicted behavioural markers such as the congruency, alerting and orienting effects. However, there was no evidence for group differences on these markers, meaning that immersed and non-immersed children and adolescents performed similarly at the level of inhibitory control or attentional abilities. In addition, on both these inhibitory control tasks, there were no overall RT differences between the immersed groups and non-immersed groups, indicating similar monitoring abilities. Finally, the results of the DCCS task also showed that, despite the presence of a switching cost and a mixing cost, there were no differences between the immersed and non-immersed participants. These results suggest that there is no switching or monitoring advantage for immersed over non-immersed individuals. In conclusion, our findings from more than 500 participants did not reveal any differences in executive control abilities between immersed and non-immersed individuals.

Do our results reflect a true absence of executive control advantages in immersion, or are there alternative explanations for these null-findings? First, the current findings are unlikely to suffer a lack of power considering the large sample-size, which was clearly above those commonly used in earlier research, including the studies that revealed executive control advantages through immersion education. Second, one might argue that non-immersed groups, especially adolescents, also had formal education in a foreign language, which could imply that they also reached a certain level of bilingualism. However, immersed participants outperformed the non-immersed children and adolescents on a foreign-language receptive vocabulary task. Therefore, even if all the participants mastered a foreign language to some extent, the conclusion remains that the established superior foreign-language abilities of the immersion groups did not produce executive control advantages. Third, the absence of executive control advantages in the immersed groups are not likely to be ascribed to a general lack of reliability of the tasks, because the expected markers of executive control were observed (congruency and Gratton effects for the Simon task; congruency, alerting and orienting effects for the ANT; switching and mixing cost for the DCCS). Furthermore, these three executive control tasks are well-established in the bilingualism literature. Altogether, we believe that it is safe to conclude that our findings show no advantage in executive control for individuals enrolled in immersion education.

Few studies thus far already examined the effect of immersion education on executive control (e.g., Bialystok & Barac, 2012; Carlson & Meltzoff, 2008; Nicolay & Poncelet, 2013, 2015; Poarch & van Hell, 2012; Woumans et al., 2016). However, these studies seem to have limited their investigation to the early years of immersion education, which might explain the inconsistent results. Carlson and Meltzoff (2008), for instance, reported no positive effect of immersion education after a period of six months on a wide range of executive control measures, including the ANT and the DCCS task. Poarch and van Hell (2012) reported that children after 1.3 years of immersion education showed no advantage over monolinguals on a Simon task and an ANT. In a study with 7-year-old immersed children for two years, Kaushanskaya et al. (2014) found no advantage for immersed children performing a DCCS task. Children were English native-speakers with 90% of the classroom time instructed in Spanish, which is a higher proportion of foreign-language courses, but a lower duration, compared with our participants. Bialystok and Barac (2012), however, observed a positive relationship between the duration of immersion education and executive control. Woumans et al. (2016) showed that after one year of immersion education, there was no advantage on a Simon task for 5-year-old immersed children, although there was an advantage for the immersed group on nonverbal intelligence. Nicolay and Poncelet (2013, 2015) compared executive control abilities of 8-year-old children immersed for three years with those of monolinguals. In their study, alerting, selective attention, divided attention, switching and response inhibition were assessed with the Test for Attentional Performance in Children (KITAP – Zimmermann, Gondan & Fimm, 2002) and interference suppression was assessed with a short version of the ANT. The authors found that, after three years of immersion education, the immersed children outperformed their monolingual peers on all tested executive control processes, except on inhibitory control. A recurrent conclusion from all those studies on immersion education is that a longer duration of immersion may be a prerequisite for the often-postulated bilingual executive control advantage to emerge. Although the immersed participants of the current study attained a reasonable level of foreign-language proficiency and already spent four to five years in immersion education, we did not observe executive control advantages.

Based on the current and previous studies, it seems that the executive control advantages often observed in typical bilingual populations cannot be easily obtained through immersion education. In what follows, we go further into a number of potential explanations for the absence of measurable evidence for an immersion executive control advantage. First, within a classroom with only one teacher and several pupils, the time devoted to foreign-language production might be limited compared to the time pupils comprehend in that language. This is different from more typical bilingualism, where bilinguals learn their second language by speaking and comprehending this language during one-on-one conversations. Indeed, the bilingual executive control advantage might emerge from experience with speaking multiple languages, rather than from being able to comprehend different languages (see Emmorey, Luk, Pyers, & Bialystok, 2008; Prior & Gollan, 2011). Therefore, although immersed children of this study spoke with their immersion teacher and with their peers in the foreign language, it is possible that foreign-language production was not sufficiently trained for the executive control advantage to develop. Further studies may include a measure of expressive vocabulary in order to elucidate this possibility. Second, Verreyt et al. (2016) showed that frequent language switching (and especially code-switching), rather than high foreign-language proficiency, might be necessary for an executive control advantage to emerge. Although immersion education implies switching frequently between languages, code-switching may be too infrequent to obtain executive control advantages. Finally, another potential explanation for the null-results obtained in this study is that, in the specific context of immersion education, the executive control advantage might be transitory. As suggested by Nicolay and colleagues (2013, 2015), during the first phases of foreign-language learning, specific executive control processes may be more strongly solicited in earlier stages of foreign-language acquisition due to lack of automaticity in language use than in later stages. The Controlled Dose hypothesis (Paap, in press) proposes a similar shift in engagement of executive control for more typical bilinguals. If they exist, the immersion and bilingual advantage might only be present during a particular period of foreign-language acquisition, when individuals are still learning how to control their different languages. Analogous to losing muscles after stopping fitness, improved executive control of bilinguals might not persist indelibly when this mechanism is no longer recruited for language control. This hypothesis offers an explanation for why accumulated experience leads to improved executive control for young bilingual children (Bialystok & Barac, 2012), but also why the bilingual advantage seems to disappear in highly-proficient bilingual adolescents (Bialystok, 2005). Although the immersed participants of this study could be considered unbalanced bilinguals, they might already be experts in language control because they received at least four years of formal education in their two languages. In the same line, Hansen et al. (2016) found an advantage in working memory updating for younger immersed children (grade 2 and 3), but not for older ones (grade 5 and 8). The Controlled Dose hypothesis points to the importance of future work that investigates the longitudinal effects of immersion education on executive control.

We would also like to mention a number of limitations of the current study. First, the immersed groups naturally reflect the characteristics of CLIL in Belgium and were as such not matched with the non-immersed groups on certain background variables that are known to influence executive control (SES for both education levels; age, nonverbal intelligence, and gender for secondary education). The use of multiple covariates, as well as entering covariates in analyses to control for (unwanted) group differences on these variables has been criticized in the bilingualism literature (Paap & Greenberg, 2013; Paap, Johnson, & Sawi, 2014). It is worth mentioning however that the differences in the background measures (except for age) should in theory lead to advantages for the immersed over the non-immersed groups. Indeed, higher SES (Calvo & Bialystok, 2014), better non-verbal intelligence (Friedman et al., 2006) and more boys than girls (Berthelsen et al., 2017) are all linked to better executive control performance. Thus, although there are some marked group differences in background measures, they are unlikely to be responsible for pushing a potential executive control advantage, as we found none. Nevertheless, if future studies, from different countries and involving different social settings, succeed in recruiting samples that are matched on these background variables, we will be able to draw conclusions with relatively more certainty. Relatedly, the group differences in SES are likely the consequence of a self-selection bias in the sense that, in Belgium, although a priori open to anyone, immersion education is known to be particularly attractive to a socially more privileged public. Whereas Woumans et al. (2016) observed a clear advantage in nonverbal intelligence for immersed children, we found a similar advantage, which in our study however disappeared after controlling for differences in SES. This points towards a need for longitudinal studies on immersion education, which are less sensitive to baseline differences in potentially confounding background variables.

Another limitation of this study is that overall ACC for the executive control tasks were almost at ceiling, especially for the adolescent groups. Nevertheless, we observed the established behavioural markers of executive control on RTs, suggesting that our tasks were reliable. Given that Bialystok (2015) stated that more effortful tasks are more likely to yield a bilingual advantage, it is possible that the tasks were not sufficiently sensitive to pick up small group differences in executive control. In this context, previous research has also highlighted the importance of the congruent-incongruent trial split in conflict resolution tasks (e.g., in the ANT; Costa et al., 2009; Hofweber et al., 2016). In line with previous studies on immersion education (Carlson and Meltzoff, 2008; Nicolay and Poncelet, 2013), we used a high-monitoring, and therefore effortful, 50:50 split between congruent and incongruent trials, but we cannot exclude that a different split may yield different results. We further acknowledge that there is a large variability in the RT data. Although common in children (seeYang & Yang, 2016), this variability may have contributed to the lack of significant differences between the immersed and non-immersed participants.

Finally, given the scale of our study in logistical terms, we focused on a well-chosen, but reduced number of executive control processes that were found to be influenced by using multiple languages in daily life: inhibitory control, monitoring, switching, and attentional abilities. Another executive control process, which we did not measure, namely working memory, was recently hypothesized to be modulated by bilingualism (Bialystok, 2017; Yang, 2017). Interestingly though, prior research has observed a link between Simon task performance and working memory capacity (Kane & Engle, 2003). Since we have not found an immersion advantage in the Simon task, this might also lead us to tentatively expect no working memory advantage for immersed over non-immersed pupils. This interpretation needs however to be interpreted with caution, because some researchers do not agree with the hypothesis that performance on the Simon task is related to working memory abilities (Keye, Wilhelm, Oberauer, & Stürmer, 2013). The effects of immersion education on working memory should be investigated in future studies to obtain a more comprehensive view on the broader cognitive implications of formal foreign-language education.

Conclusion

The current study makes a unique contribution to an ongoing debate about whether becoming bilingual through a formal education experience improves executive control abilities or not. This debate is not fully independent of the broader discussion about the existence of a bilingual executive control advantage. Although immersion education is an instructional method that creates the possibility to become bilingual, the current large-scale study has found no measurable evidence that it also improves executive control. It is however also important to keep in mind that immersion education is firstly aimed to enhance proficiency in multiple languages and that this core objective has (to some extent) also been reached in our immersed participants.

Supplementary material

Table S1

Table S2

Table S3

Appendix S1

References

Anderson, P. (2002). Assessment and Development of Executive Function (EF) During Childhood. *Child Neuropsychology*, *8*(2), 71.

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Catroppa, C. (2001). Development of Executive Functions Through Late Childhood and Adolescence in an Australian Sample. *Developmental Neuropsychology*, *20*(1), 385–406.

Baddeley, A. (1996). Exploring the Central Executive. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 5–28. https://doi.org/10.1080/713755608

Barac, R., & Bialystok, E. (2012). Bilingual Effects on Cognitive and Linguistic Development: Role of Language, Cultural Background, and Education. *Child Development*, *83*(2), 413–422. https://doi.org/10.1111/j.1467-8624.2011.01707.x

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67(1)*, 1–48. https://doi.org/10.18637/jss.v067.i01

Berthelsen, D., Hayes, N., White, S. L. J., & Williams, K. E. (2017). Executive Function in Adolescence: Associations with Child and Family Risk Factors and Self-Regulation in Early Childhood. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00903

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bialystok, E. (2015). Bilingualism and the Development of Executive Function: The Role of Attention. *Child Development Perspectives*, *9*(2), 117–121. https://doi.org/10.1111/cdep.12116

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

Bialystok, E., & Barac, R. (2012). Emerging bilingualism: Dissociating advantages for metalinguistic awareness and executive control. *Cognition*, *122*(1), 67–73. https://doi.org/10.1016/j.cognition.2011.08.003

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45*(2), 459–464. https://doi.org/10.1016/j.neuropsychologia.2006.10.009

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, Aging, and Cognitive Control: Evidence From the Simon Task. *Psychology and Aging*, *19*(2), 290–303. https://doi.org/10.1037/0882-7974.19.2.290

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Developmental Science*, *7*(3), 325–339. https://doi.org/10.1111/j.1467-7687.2004.00351.x

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. https://doi.org/10.1177/13670069050090010701

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

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). The benefits of being bilingual: Working memory in bilingual Turkish–Dutch children. *Journal of Experimental Child Psychology*, *128*(Supplement C), 105–119. https://doi.org/10.1016/j.jecp.2014.06.007

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology (Hove, England)*, *25*(5). https://doi.org/10.1080/20445911.2013.812093

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural Mechanisms of Transient and Sustained Cognitive Control during Task Switching. *Neuron*, *39*(4), 713–726. https://doi.org/10.1016/S0896-6273(03)00466-5

Bretz, F., Hothorn, T., & Westfall, P. (2010). *Multiple Comparisons Using R*. CRC Press.

Calvo, A., & Bialystok, E. (2014). Independent Effects of Bilingualism and Socioeconomic Status on Language Ability and Executive Functioning. *Cognition*, *130*(3), 278–288. https://doi.org/10.1016/j.cognition.2013.11.015

Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, *11*(2), 282–298. https://doi.org/10.1111/j.1467-7687.2008.00675.x

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology: The Official Journal of the National*

Academy of Neuropsychologists, *23*(2), 201–216. https://doi.org/10.1016/j.acn.2007.08.010

Costa, A., Hernández, M., Costa-Faidella, & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: now you see it, now you don’t. *Cognition*, *113*, 135–149. https://doi.org/10.1016/j.cognition.2009.08.001

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*(1), 59–86. https://doi.org/10.1016/j.cognition.2006.12.013

Dalton-Puffer, C. (2011). Content-and-Language Integrated Learning: From Practice to Principles? *Annual Review of Applied Linguistics*, *31*, 182–204. https://doi.org/10.1017/S0267190511000092

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Masked translation priming effects with low proficient bilinguals. *Memory & Cognition*, *39*(2), 260–275. https://doi.org/10.3758/s13421-010-0004-9

Dunn, L. M, & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test - Revised.* Circle Pines, NM: American Guidance Service.

Dunn, L. M, & Dunn, L. M. (2005). *PPVT-III-NL: Peabody Picture Vocabulary Test-III-NL*. Pearson Assessment and Information B.V.

Dunn, Lloyd M, Dunn, D. M., & Pearson Assessments. (2007). *PPVT-4: Peabody picture vocabulary test*. Minneapolis, MN.: Pearson Assessments.

Dunn, Lloyd M, Thériault-Whalen, C. M., & Dunn, L. M. (1993). *EVIP: Échelle de vocabulaire en images Peabody*. Toronto: Psycan.

Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: new evidence from Dutch-French bilinguals. *Experimental Psychology*, *56*(3), 173–179. https://doi.org/10.1027/1618-3169.56.3.173

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The Source of Enhanced Cognitive Control in Bilinguals. *Psychological Science*, *19*(12), 1201–1206. https://doi.org/10.1111/j.1467-9280.2008.02224.x

Engel de Abreu, P. M. J. (2011). Working memory in multilingual children: is there a bilingual effect? *Memory (Hove, England)*, *19*(5), 529–537. https://doi.org/10.1080/09658211.2011.590504

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the Efficiency and Independence of Attentional Networks. *Journal of Cognitive Neuroscience*, *14*(3), 340–347. https://doi.org/10.1162/089892902317361886

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of Experimental Psychology. General*, *133*(1), 101–135. https://doi.org/10.1037/0096-3445.133.1.101

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., Defries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179. https://doi.org/10.1111/j.1467-9280.2006.01681.x

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development*, *10*(4), 483–527. https://doi.org/10.1016/0885-2014(95)90024-1

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Guasch, N. V., … Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*(FEB). https://doi.org/10.3389/fpsyg.2014.00011

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121, 480-506.*

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

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, *25*(5), 515–530. https://doi.org/10.1080/20445911.2013.796377

Grosjean, F. (2010). Bilingualism, biculturalism, and deafness. *International Journal of Bilingual Education and Bilingualism*, *13*(2), 133–145. https://doi.org/10.1080/13670050903474051

Grundy, J., Chung-Fat-Yim, A., C Friesen, D., Mak, L., & Bialystok, E. (2017). *Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults* (Vol. 163). https://doi.org/10.1016/j.cognition.2017.02.010

Grundy, J. G., & Keyvani Chahi, A. (2017). Post-conflict slowing effects in monolingual and bilingual children. *Developmental Science*, *20*(1), n/a-n/a. https://doi.org/10.1111/desc.12488

Hansen, L. B., Macizo, P., Duñabeitia, J. A., Saldaña, D., Carreiras, M., Fuentes, L. J., & Bajo, M. T. (2016). Emergent Bilingualism and Working Memory Development in School Aged Children. *Language Learning*, *66*(S2), 51–75.

https://doi.org/10.1111/lang.12170

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, *18*(4), 625–658. https://doi.org/10.3758/s13423-011-0116-7

Hiligsmann, P., Van Mensel, L., Galand, B., Mettewie, L., Meunier, F., Szmalec, A., … Simonis, M. (2017). Assessing Content and Language Integrated Learning (CLIL) in French-speaking Belgium: linguistic, cognitive, and educational perspectives. *Les Cahiers de Recherche Du Girsef*, (109).

Hofweber, J., Marinis, T., & Treffers-Daller, J. (2016). Effects of dense code-switching on executive control. *Linguistic Approaches to Bilingualism*, *6*(5), 648–668. https://doi.org/10.1075/lab.15052.hof

Jeffreys, H. (1961). Theory of probability (3 ed.). Oxford: Oxford University Press, Clarendon Press.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology. General*, *132*(1), 47–70.

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. https://doi.org/10.1111/desc.12142

Keye, D., Wilhelm, O., Oberauer, K., & Stürmer, B. (2013). Individual differences in response conflict adaptations. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00947

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01171

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*(2), 520–525. https://doi.org/10.3758/s13423-013-0524-y

Martin-Rhee, M. M., & Bialystok, E. (2008). The development of two types of inhibitory control in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *11*(01). https://doi.org/10.1017/S1366728907003227

Mezzacappa, E. (2004). Alerting, Orienting, and Executive Attention: Developmental Properties and Sociodemographic Correlates in an Epidemiological Sample of Young, Urban Children. *Child Development*, *75*(5), 1373–1386. https://doi.org/10.1111/j.1467-8624.2004.00746.x

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews. Neuroscience*, *1*(1), 59–65. https://doi.org/10.1038/35036228

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*(1), 8–14. https://doi.org/10.1177/0963721411429458

Nicolay, A.-C., & Poncelet, M. (2013). Cognitive advantage in children enrolled in a second-language immersion elementary school program for three years. *Bilingualism: Language and Cognition*, *16*(03), 597–607. https://doi.org/10.1017/S1366728912000375

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*, 1–7. https://doi.org/10.1017/S1366728914000868

Paap, K. R. (in press). Bilingualism in cognitive science. In *A. De Houwer & L. Ortega (Eds.), Handbook of bilingualism.* Cambridge: Cambridge University Press.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258. https://doi.org/10.1016/j.cogpsych.2012.12.002

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, *26*(6), 615–639. https://doi.org/10.1080/20445911.2014.944914

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, *113*(4), 535–551. https://doi.org/10.1016/j.jecp.2012.06.013

Posner, M. I., & Petersen, S. E. (1990). The Attention System of the Human Brain. *Annual Review of Neuroscience*, *13*(1), 25–42. https://doi.org/10.1146/annurev.ne.13.030190.000325

Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, *108*(3), 567–579. https://doi.org/10.1016/j.jecp.2010.10.009

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*(04), 682–691.

https://doi.org/10.1017/S1355617711000580

Prior, A., & Macwhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*(02), 253. https://doi.org/10.1017/S1366728909990526

Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, *27*(1), 1–11. https://doi.org/10.1080/20445911.2014.976226

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

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

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.*

Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, *6*(1), 99–105. https://doi.org/10.1080/00140136308930679

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science (New York, N.Y.)*, *283*(5408), 1657–1661.

Sorge, G. B., Toplak, M. E., & Bialystok, E. (2017). Interactions between Levels of Attention Ability and Levels of Bilingualism in Children's Executive Functioning. *Developmental Science*, *20*(1). https://doi.org/10.1111/desc.12408

Van Assche, E., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual Word Recognition in a Sentence Context. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00174

Van Der Linden, L., Van de Putte, E., Woumans, E., Duyck, W., & Szmalec, A. (2018). Does extreme language control training improve cognitive control? A comparison of professional interpreters, L2 teachers and monolinguals. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01998

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*(01), 181–190.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems ofp values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Woumans, E., Ceuleers, E., Van der Linden, L., Szmalec, A., & Duyck, W. (2015). Verbal and nonverbal cognitive control in bilinguals and interpreters. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *41*(5), 1579–1586. https://doi.org/10.1037/xlm0000107

Woumans, E., Surmont, J., Struys, E., & Duyck, W. (2016). The Longitudinal Effect of Bilingual Immersion Schooling on Cognitive Control and Intelligence. *Language Learning*, *66*(S2), 76–91. https://doi.org/10.1111/lang.12171

Xiang, H., Dediu, D., Leah, R., van Oort, E., Norris, D. G., & Hagoort, P. (2012). The Structural Connectivity Underpinning Language Aptitude, Working Memory, and IQ in the Perisylvian Language Network. *Language Learning*, *62*(s2), 110–130. https://doi.org/10.1111/j.1467-9922.2012.00708.x

Yang, S., & Yang, H. (2016). Bilingual effects on deployment of the attention system in linguistically and culturally homogeneous children and adults. *Journal of Experimental Child Psychology*, *146*, 121–136. https://doi.org/10.1016/j.jecp.2016.01.011

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

[1] The term “foreign language” is used because the language of immersion was not always the second language of our participants, given that certain participants were raised in a bilingual home environment in which the home languages are different from the foreign language learned at school.

[2] The choice not to exclude bilinguals from the sample was motivated by statistical analyses showing that their exclusion did not alter the results.

Table 1. *Descriptive statistics and mean comparisons for background information in the immersion and non-immersion groups.* BF_{10} = *Bayes factor in favour of the alternative hypothesis.*

	Immersion	**Non-immersion**		
Primary	**Mean (SD)**	**Mean (SD)**	***Test***	BF_{10}
N	128	102		
Age in years	10.38 (0.40)	10.48 (0.55)	$t(166.35) = -1.58$	0.53
Gender F/M	67/61	56/46	$\chi^2 < 1$	0.17
Bilingualism 1/2/3	75/45/8	51/39/12	$\chi^2(2) = 2.89$	0.20
SES 1/2/3	23/44/61	49/30/23	$\chi^2(2) = 26.63$***	$> 100^{+++}$
Raven (max=60)	29.97 (8.12)	27.55 (6.85)	$t(228) = 2.39$*	2.10
EVIP (max=170)	102.10 (19.16)	101.95 (21.92)	$t < 1$	0.14
Dutch/English Receptive Vocabulary	0.59 (0.77)	-0.73 (0.71)	$t(223.10) = 13.24$***	$> 100^{+++}$
PPVT-NL-III (max=204) (80/52)	74.32 (20.59)	30.13 (20.76)	$t(130) = 12.01$***	$> 100^{+++}$
PPVT-IV (max=228) (48/50)	66.62 (33.82)	23.74 (26.25)	$t(88.64) = 6.99$***	$> 100^{+++}$

Secondary	**Mean (SD)**	**Mean (SD)**	***Test***	BF_{10}
N	127	156		
Age in years	16.37 (0.46)	16.64 (0.61)	$t(276.70) = -4.25$***	$> 100^{+++}$
Gender F/M	59/68	101/55	$\chi^2(1) = 9.52$**	17.09^{+}
Bilingualism1/2/3	98/23/6	110/39/7	$\chi^2(2) = 1.94$	0.15
SES 1/2/3	19/49/59	50/68/38	$\chi^2(2) = 18.78$***	$> 100^{+++}$
Raven (max=60)	44.75 (6.71)	42.22 (7.98)	$t(281) = 2.84$**	5.96
EVIP (max=170)	141.44 (12.20)	138.76 (12.28)	$t(281) = 1.82$	0.64
Dutch/English Receptive Vocabulary	0.60 (0.67)	-0.49 (0.95)	$t(274) = 10.86$***	$> 100^{+++}$
PPVT-NL-III (max=204) (73/90)	126.60 (18.84)	93.21 (29.61)	$t(152.90) = 8.73$***	$> 100^{+++}$
PPVT-IV (max=228) (54/65)	147.35 (27.93)	106.75 (35.09)	$t(116.80) = 7.02$***	$> 100^{+++}$

Note. "Alpha"; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$ and "BF_{10}"; $^{---}BF_{10} < 0.01$ (decisive evidence for H0); $^{--}BF_{10} < 0.03$ (very strong evidence for H0); $^{-}BF_{10} < 0.10$ (strong evidence for H0); $^{+}BF_{10} > 10$ (strong evidence for H1), $^{++}BF_{10} > 30$ (very strong evidence for H1), $^{+++}BF_{10} > 100$ (decisive evidence for H1)

Table 2. *Group comparisons (immersion, non-immersion) on RTs (ms) and ACC (1=100% accuracy) for the effects of interest as a function of task and education level. ANT = Attention Network task; DCCS = Dimensional Change Card Sort.* BF_{10} *= Bayes factor in favour of the alternative hypothesis.*

			Primary		**Secondary**	
	Measure		**Test**	BF_{10}	**Test**	BF_{10}
Simon task	RT	Overall	$\chi^2 < 1$	$1.26^{\text{e-14}===}$	$\chi^2(1) = 1.41$	$0.01^{==}$
		Congruency effect	$\chi^2 < 1$	$8.53^{\text{e-17}===}$	$\chi^2(2) = 6.50$*	$0.00^{===}$
		Gratton effect	$z = 0.66$	0.17	$z = 2.04$*	0.94
	ACC	Overall	$\chi^2(1) = 3.06$	$0.05^{=}$	$\chi^2 < 1$	$0.00^{===}$
		Congruency effect	$\chi^2(2) = 2.11$	$0.00^{===}$	$\chi^2(2) = 1.60$	$7.48^{\text{e-05}===}$
ANT	RT	Overall	$\chi^2 < 1$	$0.01^{==}$	$\chi^2 < 1$	$0.00^{===}$
		Congruency effect	$\chi^2(4) = 3.44$	$1.52^{\text{e-08}===}$	$\chi^2(4) = 1.29$	$3.39^{\text{ e-09}===}$
		Alerting effect	$t(17179) = 2.71$*	0.95	$t < 1$	0.14
		Orientation effect	$t < 1$	0.22	$t < 1$	0.13
	ACC	Overall	$\chi^2(1) = 3.69$	$0.04^{=}$	$\chi^2(1) = 2.31$	$0.02^{==}$
		Congruency effect	$\chi^2(4) = 2.39$	$8.78^{\text{e-09}===}$	$\chi^2(4) = 3.99$	$1.07^{\text{e-08}===}$
		Alerting effect	$z = 1.27$	0.23	$z = 0.19$	0.13
		Orientation effect	$z < 1$	$0.00^{===}$	$z = 0.09$	0.13
DCCS task	RT	Overall Single Tasks	$\chi^2 < 1$	$0.01^{==}$	$\chi^2 < 1$	$0.01^{==}$
		Overall Mixed Task	$\chi^2(1) = 2.78$	$0.04^{=}$	$\chi^2 < 1$	$0.00^{===}$
		Switching Cost	$\chi^2 < 1$	$0.00^{===}$	$\chi^2(1) = 1.23$	$0.03^{=}$
		Mixing Cost	$\chi^2(1) = 2.60$	$0.04^{=}$	$\chi^2 < 1$	$0.01^{==}$
	ACC	Overall Single Tasks	$\chi^2 < 1$	$0.01^{==}$	$\chi^2 < 1$	$0.01^{==}$
		Overall Mixed Task	$\chi^2 < 1$	$0.01^{==}$	$\chi^2 (1) = 3.66$	$0.03^{=}$
		Switching Cost	$\chi^2(1) = 2.83$	$0.04^{=}$	$\chi^2 < 1$	$0.01^{==}$

Mixing Cost	$\chi^2(1) = 1.16$	$0.02^{==}$	$\chi^2(1) = 1.63$	$0.02^{==}$

Note. "Alpha"; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$ and "BF_{10}"; $^{===}$ $BF_{10} < 0.01$ (decisive evidence for H0); $^{==}$ $BF_{10} < 0.03$ (very strong evidence for H0); $^{=}$ $BF_{10} < 0.10$ (strong evidence for H0); $^{+}$ $BF_{10} > 10$ (strong evidence for H1), $^{++}$ $BF_{10} > 30$ (very strong evidence for H1), $^{+++}$ $BF_{10} > 100$ (decisive evidence for H1)

Table 3. *Means and standard deviations for RTs (ms) and ACC (1=100% accuracy) for the Simon task, as a function of Group and education level.*

	Immersion		**Non-immersion**	
Primary	**Previous Trial Type**			
	Congruent	Incongruent	Congruent	Incongruent
RT Trial type				
Congruent	530.03 (139.17)	590.63 (144.01)	527.33 (140.32)	596.76 (153.14)
Incongruent	622.98 (128.73)	586.13 (136.18)	625.82 (134.54)	584.38 (138.74)
Congruency effect	97.76 (52.74)	-4.10 (52.48)	95.78 (48.68)	-13.33 (58.61)
ACC Trial type				
Congruent	0.93 (0.24)	0.86 (0.34)	0.92 (0.27)	0.83 (0.37)
Incongruent	0.76 (0.43)	0.89 (0.31)	0.74 (0.43)	0.86 (0.34)
Congruency effect	-0.19 (0.13)	0.01 (0.14)	-0.18 (0.14)	0.02 (0.14)
Secondary	**Previous Trial Type**			
	Congruent	Incongruent	Congruent	Incongruent
RT Trial type				
Congruent	410.85 (95.31)	471.27 (117.32)	413.15 (103.58)	463.91 (115.66)
Incongruent	482.34 (92.90)	446.15 (98.16)	479.69 (95.77)	445.92 (98.04)
Congruency effect	71.18 (38.31)	-23.86 (43.30)	66.46 (37.02)	-17.94 (38.68)
ACC Trial type				
Congruent	0.98 (0.12)	0.92 (0.26)	0.98 (0.14)	0.92 (0.27)
Incongruent	0.89 (0.30)	0.96 (0.19)	0.89 (0.31)	0.96 (0.18)
Congruency effect	-0.09 (0.07)	0.03 (0.08)	-0.09 (0.08)	0.04 (0.07)

Table 4. *Means and standard deviations for RTs (ms) and ACC (1=100% accuracy) for the ANT, as a function of Group and education level.*

Primary	**Immersion**			**Non-immersion**		
	Trial Type					
	Congruent	Incongruent	Congruency effect	Congruent	Incongruent	Congruency effect
RT Cue Condition						
No	699.19 (176.71)	865.74 (231.65)	175.07 (138.36)	701.93 (175.43)	871.60 (223.62)	183.10 (131.91)
Double	640.73 (166.47)	823.48 (207.14)	186.85 (106.89)	661.59 (178.12)	841.48 (210.38)	183.63 (131.00)
Central	661.32 (179.94)	842.50 (214.30)	183.70 (120.59)	672.69 (173.81)	848.92 (210.03)	185.38 (108.88)
Spatial	621.09 (187.44)	771.65 (221.12)	162.56 (115.91)	629.85 (164.90)	797.34 (231.11)	167.16 (124.52)
Alerting effect	62.02 (68.59)	50.99 (122.53)	NA	33.09 (90.02)	35.44 (109.48)	NA
Orienting effect	42.99 (61.75)	69.19 (101.40)	NA	45.81 (72.53)	52.61 (92.59)	NA
ACC Cue Condition						
No	0.97 (0.15)	0.90 (0.30)	-0.09 (0.21)	0.98 (0.14)	0.89 (0.31)	-0.010 (0.20)
Double	0.98 (0.11)	0.88 (0.32)	-0.12 (0.22)	0.99 (0.12)	0.90 (0.29)	-0.09 (0.20)
Central	0.97 (0.15)	0.88 (0.32)	-0.12 (0.22)	0.98 (0.12)	0.90 (0.29)	-0.09 (0.20)
Spatial	0.98 (0.13)	0.90 (0.30)	-0.09 (0.20)	0.98 (0.12)	0.92 (0.27)	-0.08 (0.19)
Alerting effect	-0.01 (0.06)	0.01 (0.11)	NA	0.00 (0.07)	-0.01 (0.11)	NA
Orienting effect	0.00 (0.05)	-0.02 (0.11)	NA	0.00 (0.06)	-0.02 (0.12)	NA

Table 4. Continued.

Secondary	**Immersion**			**Non-immersion**		
	Trial Type					
	Congruent	Incongruent	Congruency effect	Congruent	Incongruent	Congruency effect
RT Cue Condition						
No	516.01 (103.73)	614.03 (124.58)	97.86 (62.09)	517.44 (112.37)	622.56 (146.53)	105.33 (61.91)
Double	481.47 (92.76)	600.29 (121.18)	118.78 (52.04)	484.34 (111.38)	607.87 (135.15)	123.12 (63.27)
Central	484.82 (91.71)	609.78 (117.48)	124.28 (52.79)	488.26 (108.95)	622.16 (138.77)	132.18 (59.75)
Spatial	453.86 (89.01)	554.77 (117.00)	101.50 (55.22)	460.66 (106.40)	564.55 (126.62)	104.04 (49.95)
Alerting effect	34.72 (48.51)	13.80 (41.62)	NA	32.93 (40.82)	15.14 (48.81)	NA
Orienting effect	30.61 (33.95)	53.39 (45.32)	NA	28.61 (36.62)	56.74 (46.97)	NA
ACC Cue Condition						
No	0.99 (0.08)	0.97 (0.17)	-0.02 (0.06)	0.99 (0.09)	0.95 (0.22)	-0.04 (0.11)
Double	0.99 (0.07)	0.97 (0.17)	-0.02 (0.06)	0.99 (0.07)	0.95 (0.22)	-0.04 (0.10)
Central	0.99 (0.07)	0.94 (0.23)	-0.05 (0.08)	0.99 (0.06)	0.93 (0.26)	-0.07 (0.13)
Spatial	0.99 (0.03)	0.96 (0.18)	-0.03 (0.06)	0.99 (0.05)	0.96 (0.20)	-0.04 (0.09)
Alerting effect	0.00 (0.03)	0.00 (0.02)	NA	0.00 (0.03)	0.00 (0.09)	NA
Orienting effect	0.00 (0.06)	-0.02 (0.09)	NA	0.00 (0.02)	-0.02 (0.10)	NA

Table 5. *Means and standard deviations for RTs (ms) and ACC (1=100% accuracy) for the DCCS, as a function of Group and education level.*

	Primary		**Secondary**	
	Immersion	**Non-immersion**	**Immersion**	**Non-immersion**
RT				
Pre-switch	716.15 (403.64)	656.72 (336.34)	454.32 (198.17)	453.56 (167.50)
Post-switch	728.50 (331.49)	721.27 (361.21)	451.72 (171.99)	455.72 (198.22)
Mixed Task non-switch	1380.33 (637.31)	1445.64 (658.73)	899.53 (436.20)	893.32 (448.72)
Mixed Task switch	1444.82 (578.47)	1543.66 (645.95)	906.38 (418.23)	922.95 (439.34)
Switching Cost	72.41 (220.68)	102.76 (247.70)	5.09 (148.15)	34.62 (147.61)
Mixing Cost	720.66 (309.48)	747.24 (325.88)	444.91 (237.45)	436.86 (209.76)
ACC				
Pre-switch	0.93 (0.26)	0.93 (0.24)	0.98 (0.14)	0.97 (0.16)
Post-switch	0.91 (0.28)	0.90 (0.30)	0.97 (0.17)	0.97 (0.17)
Mixed phase non-switch	0.91 (0.28)	0.89 (0.30)	0.96 (0.19)	0.94 (0.22)
Mixed phase switch	0.77 (0.41)	0.78 (0.41)	0.92 (0.26)	0.91 (0.28)
Switching Cost	-0.14 (0.14)	-0.13 (0.15)	-0.04 (0.089)	-0.04 (0.09)
Mixing Cost	0.00 (0.11)	-0.02 (0.12)	-0.01 (0.06)	-0.02 (0.08)