Relationship between language switching experience and executive functions in bilinguals : an Internet-based study

Jylkka, Jussi

2017

unspecified
publishedVersion

Relationship between language switching experience and executive functions in bilinguals: an Internet-based study

Jussi Jylkkä, Anna Soveri, Jenny Wahlström, Minna Lehtonen, Antoni Rodríguez-Fornells & Matti Laine

Published online: 03 Feb 2017.

Submit your article to this journal

Article views: 119

View related articles

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

Relationship between language switching experience and executive functions in bilinguals: an Internet-based study*

Jussi Jylkkä[a], Anna Soveri[a], Jenny Wahlström[a], Minna Lehtonen[a,b], Antoni Rodríguez-Fornells[c] and Matti Laine[a]

[a]Department of Psychology, Abo Akademi University, Åbo, Finland; [b]Cognitive Brain Research Unit, Cognitive Science, Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland; [c]Psicologia Bàsica, Universitat de Barcelona, Barcelona, Spain

ABSTRACT
We examined the relationship between self-reported everyday language switching experience and the performance of early bilinguals in tasks measuring different executive functions. Our participants were Finnish–Swedish early bilinguals, aged 16–41 years ($N = 66$, Experiment 1) and 18–69 years ($N = 111$, Experiment 2). An earlier study using a sample from a similar population discovered a negative relationship between self-reported language switching and a mixing cost in error rates in a number–letter task. This finding was not replicated. Instead, we found that a higher rate of reported contextual language switching predicted larger switching cost reaction times in the number–letter task, and that a higher rate of reported unintended language switches predicted larger error rates in a spatial n-back task. We conclude that these results likely reflect individual differences in executive skills, and do not provide evidence for the hypothesis that language switching trains executive functions.

Introduction

Executive functions (EF) the weaker language consist of a range of cognitive abilities, central to coordinating purposeful goal-oriented and socially responsible behaviour, mainly subserved by the frontal lobes and fronto-subcortical circuits. In an influential attempt to clarify the mental architecture of executive functions, Miyake, Friedman, Emerson, Witzki, and Howerter (2000) analysed subjects' performance on executive tasks using a latent variable analysis. They differentiated between three central executive functions: inhibition of unwanted responses, shifting between tasks and mental sets, and updating and monitoring of working memory.

Language switching in bilinguals is proposed to engage executive processes, such as inhibiting the context-irrelevant language, switching between the two languages, and monitoring the activation levels of the two languages (see, e.g. Linck, Schwieter, & Sunderman, 2012; Rodriguez-Fornells, de Diego Balaguer, & Münte, 2006). It has been hypothesised that the training in these executive processes

that bilinguals have gotten via frequent language switching could be reflected in better performance in tests measuring executive functions (see e.g. Bialystok, 2010). Indeed, there are studies indicating that compared to monolinguals, early bilinguals show an advantage on executive functions, thought to stem from life-long bilingual language use (e.g. Bak, Nissan, Allerhand, & Deary, 2014; Bialystok, 1999; Bialystok, Craik, Klein, & Viswanathan, 2004; Bialystok, Craik, & Luk, 2008; Bialystok, Craik, & Ryan, 2006; Bialystok & Majumder, 1998; Bialystok & Martin, 2004; Bialystok & Viswanathan, 2009; Calvo & Bialystok, 2014; Costa, Hernández, & Sebastián-Gallés, 2008; Kapa & Colombo, 2013; Morales, Calvo, & Bialystok, 2013; Soveri, Laine, Hämäläinen, & Hugdahl, 2011; Soveri, Rodriguez-Fornells, & Laine, 2011). On the other hand, many of the more recent studies suggest that the performance of bilinguals is equal to, or even weaker than, that of monolinguals (e.g. Ansaldo, Ghazi-Saidi, & Androver-Roig, 2015; Antón et al., 2014; von Bastian, Souza, & Gade, 2016; De Bruin, Treccani, & Della Sala, 2015;

Gathercole et al., 2014; Hilchey & Klein, 2011; Kousaie & Phillips, 2012; Kousaie, Sheppard, Lemieux, Monetta, & Taler, 2014; Mor, Yitzhaki-Amsalem, & Prior, 2014; Paap & Greenberg, 2013; Paap & Sawi, 2014; Ratiu & Azuma, 2015).

Whether bilingualism enhances EFs is, thus, still an open question. Here we do not directly address the bilingual advantage hypothesis, which would require comparing bilinguals and monolinguals. Instead, we focus on one of the key assumptions underlying the advantage hypothesis, namely that bilingual language switching trains EFs. There is evidence that high-frequency language switchers perform better in executive tasks than low-frequency switchers (Prior & Gollan, 2011; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016; but see also Johnson, Sawi, & Paap, 2015).

Code switching, that is, alternating between or mixing two languages within a single discourse, is considered to be of particular importance to training executive functions, as switching requires inhibiting the non-target language and monitoring language use. In addition to intra-individual factors such as language proficiency, the extent of code switching depends on social factors, such as social roles (socio-economic status, educational background, and relationships between the participants), situational factors (discourse topic and language suitability in specific contexts), language attitudes (social dominance, group membership, security, and mixing options in a given social context; see Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman, & Münte, 2012). We will largely ignore these sociolinguistic issues in this paper and instead focus on the relationship between self-reported language switching and performance on tasks measuring executive functions.

Our study is a replication of Soveri, Rodriguez-Fornells et al. (2011), who examined whether individual differences in self-reported everyday language switching behaviour predicted performance in executive tasks. Soveri et al. stipulated that the more a subject switches between two or more languages in daily life, the better his/her executive task performance is. The authors tested 38 Swedish–Finnish early bilinguals between 30 and 75 years of age. They used a Swedish translation of the Bilingual Switching Questionnaire (BSWQ) devised by Rodriguez-Fornells et al. (2012) to provide subjective estimates on four different types of everyday language switching. These were as follows: (1) tendencies to switch from Swedish to Finnish (e.g. "When I cannot recall a word in Swedish, I tend to immediately produce it in Finnish"), (2) tendencies to switch from Finnish to Swedish (e.g. "When I cannot recall a word in Finnish, I tend to immediately produce it in Swedish"), (3) contextual switches (e.g. "There are situations in which I always switch between the two languages"), and (4) unintended switches (e.g. "It is difficult for me to control the language switches I introduce during a conversation (e.g. from Swedish to Finnish)"). The questionnaire has been validated with a sample of 566 Spanish–Catalan bilinguals (Rodriguez-Fornells et al., 2012).

Soveri, Rodriguez-Fornells et al. (2011) used the Simon task and the Flanker task to assess inhibition, a spatial n-back task to assess working memory updating, and a number–letter task to assess set shifting and monitoring. In the Simon and Flanker tasks the subject needs to suppress conflicting information in responding to a stimulus. In the spatial n-back the subject needs to judge whether a stimulus on the computer screen is in the same location as a stimulus presented one or two trials earlier. In the number–letter task the subject is presented with a number–letter pair and the task is to categorise either the number as even or odd or the letter as a vowel or consonant depending on the location of the stimulus. The choice of these measures was based on the above-mentioned model by Miyake et al. (2000). Soveri, Rodriguez-Fornells et al. (2011) used multiple regression analyses to examine the relationship between executive function performance as the dependent variable and the BSWQ variables and selected background variables as predictors. Most consistent effects were found in the number–letter task, where a higher rate of everyday language switches was related to a smaller mixing cost in errors. The mixing cost is the difference in performance between single-task blocks and repetition trials in a mixed-task block, assumed to reflect monitoring and sustained attentional control (Braver, Reynolds, & Donaldson, 2003).

The connection between language switching and EF has been examined in at least two studies following Soveri, Rodriguez-Fornells et al. (2011). Verreyt et al. (2016) found that bilingual subjects who reported switching at a high frequency outperformed both low-frequency switching bilinguals and monolinguals in a Simon and Flanker task (N = 65; approximately 20 per language group). Frequency of switching was assessed with a single question on a scale from zero to seven. On the

other hand, Johnson et al. (2015) analysed a database of 168 bilinguals and found that high-frequency switchers performed equally to low-frequency switchers on a Simon, Flanker, and a colour–shape switching task. Again, language switching frequency was assessed only with a single question on a five-point scale. The BSWQ, a validated survey consisting of altogether 12 questions, could provide more accurate estimates of switching frequency. Moreover, the BSWQ gives separate estimates for intended, unintended, and contextual switches, which presumably correlate differently with EFs.

Rodriguez-Fornells et al. (2012) also investigated the relationship between the BSWQ variables and the inhibition measures from the Flanker, Stroop, and a stop-signal task in a very large sample ($N =$ 566), and discovered some evidence for a connection between language switching and executive functions. They found a weak positive correlation between the rate of unintended switches and reaction time in the stop-signal task, which they took to indicate that high cognitive control (indicated by a low number of reported unintended switches) predicted better performance (i.e. shorter reaction times) in the task. They also discovered a moderate negative correlation between tendencies to switch from) the weaker language L2 to the dominant language L1 and the Stroop incongruence effect in reaction times and error rates. No correlations were found in the Flanker task, or between any of the other language switching measures and the executive tasks. Rodriguez-Fornells et al. (2012) emphasise that the correlations are relatively small, and might not turn out significant in a smaller sample.

In the present study, we conducted two experiments that aimed to replicate the Soveri, Rodriguez-Fornells et al. (2011) experiment using larger samples obtained via Internet-based testing. The first experiment recruited young university students. The age of the participants in this sample differed from the Soveri, Rodriguez-Fornells et al. (2011) study, which employed older adults (mean age 52.8 years). However, if the effect in Soveri, Rodriguez-Fornells et al. (2011) was robust, it should be observable also in a different age group. The second experiment, in turn, aimed to draw a sample from the general Swedish–Finnish population and from a wider age range, coming closer to the sample of the Soveri study.

A secondary aim was to see if the basic executive cost effects could be reliably obtained in the Simon,

Flanker, n-back, and number–letter tasks using Internet-based testing. Internet testing has the advantage of recruiting subjects from a wide range of age and demographic background. The available evidence indicates that the quality of data does not suffer when moving from laboratory to Internet-based testing. In their extensive evaluation, Crump, McDonnell, and Gureckis (2013) administered a diverse body of cognitive tasks including Stroop, task switching, Flanker, and Simon tasks to participants online and attained results very similar to those of laboratory experiments, with sample sizes varying from 40 to 58 per test. In the same vein, Linnman, Carlbring, Åhman, Andersson, and Andersson (2006) compared the performance of participants in a web-based Stroop task conducted in a laboratory setting ($n = 40$) and at home ($n = 28$), and obtained similar results.

Experiment 1

Materials and methods

Participants

The participants ($N = 66$, 52 females) were Swedish–Finnish bilingual students from the Swedish university and the Swedish polytechnic in Finland, recruited through e-mail lists. The sample size was not pre-determined; instead, our aim was to gather as many participants as possible through student e-mail lists. The following criteria were used for the exclusion of participants: any neurological illness or intake of psychotropic drugs, dyslexia, or alcohol intake of over six portions over the last 24 hours. The participants were Swedish–Finnish early bilinguals between ages 16 and 41 ($M = 20.3$) and had learned their L2 before the age of 6. They estimated having learned Swedish at 0.70 years since birth on average ($SD = 1.6$) and Finnish at 0.23 years since birth on average ($SD = 0.74$). It is thus probable that many of the subjects estimated when they started learning a language; the mode answer to these questions was 0. Being university and polytechnic students, the subjects were on average highly educated (years of education $M = 15.7$, $SD = 2.4$). The participants' language skills in Swedish, Finnish, and any two other languages were assessed with self-ratings on a scale from 0 (no skills at all) to 6 (corresponding to the skills of a native speaker) in four domains: speaking, speech comprehension, reading, and writing. The self-ratings did not differ between Swedish and Finnish in reading, writing

Table 1. Demographics and BSWQ scores.

	M	SD	Range
Age (years)	20.3	3.8	16–41
Years of education	15.7	2.4	10–23
Everyday use of L1 and L2 (%)	71.1	22.5	10–100
Overall use of any languages (0–100)	67	17	13–91
Age of L2 acquisition (years)	2.4	1.9	1–6
BSWQ: language switching (6–30 pts)	15.8	3.1	9–26
BSWQ: contextual switches (3–15 pts)	7.9	3.2	3–15
BSWQ: unintended switches (3–15 pts)	5.0	2.2	3–13

Notes: This is calculated by subtracting the percentage of the less frequently used language from the percentage of the more often used language, and then inverting the score. High score indicates balanced everyday use of both languages. This is calculated as the inverted sum of cross products of three reported language use percentages. High score indicates balanced use of any four languages.

or speech comprehension ($|Z|$s < 0.18), but the participants reported having slightly better Swedish-speaking skills ($M = 5.71$) than Finnish ($M = 5.42$; $Z = −2.24$, $p = .025$). Demographic information and self-reported language skills are given in more detail in Tables 1 and 2, respectively.

Tasks and questionnaires

The executive functions tests were the same as those used in Soveri, Rodriguez-Fornells et al. (2011). The Simon task (Simon & Rudell, 1967) is a measure of inhibition and conflict resolution. In the task, a blue or red square appears on either the left or right side of the screen. The participant has to press the left button each time a blue square appears and the right button when a red square appears, irrespectively of the location of the square (in our version of the task, the stimulus-to-response mapping was not counterbalanced). On congruent trials, the square is on the same side as the correct response key (e.g. a blue box on the left side) and on incongruent trials the square is on the opposite side. On incongruent trials, the participant has to

Table 2. Self-reported language skills in Swedish, Finnish, and English (scale 0–6).

Language	M	SD
Finnish		
Speaking	5.42	0.81
Reading	5.68	0.68
Writing	5.30	0.88
Speech comprehension	5.80	0.48
Swedish		
Speaking	5.71	0.60
Reading	5.80	0.40
Writing	5.53	0.77
Speech comprehension	5.80	0.50
English		
Speaking	4.27	1.04
Reading	4.84	0.90
Writing	4.33	1.02
Speech comprehension	4.84	0.79

suppress the conflicting spatial information. The Simon effect is calculated by subtracting the average reaction time or error rate on the congruent trials from the average reaction time or error rate on the incongruent trials. Some studies have shown a bilingual advantage in this or a similar task (Bialystok et al., 2004, 2008; Martin-Rhee & Bialystok, 2008; Morales et al., 2013), but even more studies have yielded null results (Ansaldo et al., 2015; von Bastian et al., 2016; De Bruin et al., 2015; Gathercole et al., 2014; Kousaie & Phillips, 2012; Kousaie et al., 2014; Mor et al., 2014; Morton & Harper, 2007; Namazi & Thordardottir, 2010; Paap & Greenberg, 2013; Paap & Sawi, 2014).

In the present version of the test, we used 100 trials, half of which were congruent and half incongruent, separately randomised for each subject. The trials were divided into four blocks, with 5 s breaks in-between. Before the actual test took place, all participants performed a practice sequence. Each experimental trial began with an 800 ms fixation cross, followed by a 250 ms blank interval. After this, a red or blue box appeared and remained on the screen for 1,000 ms, unless a response was given. Finally, the screen was blank for 500 ms.

The Flanker task (adapted from Eriksen & Eriksen, 1974) is another measure of inhibition. In this task, the subject is presented with an array of five horizontal arrows and has to determine the direction of the middle arrow by a left or right key press. On congruent trials, the arrow in the middle points towards the same direction as the other four arrows (the flankers); on incongruent trials, the middle arrow points towards the opposite direction. The Flanker effect is calculated by subtracting the mean reaction time or error rate of the congruent trials from the mean reaction time or error rate of the incongruent trials. Some studies have previously found a bilingual advantage on the Flanker or similar task (e.g. Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Costa et al., 2008; Kapa & Colombo, 2013), but many have also yielded null results (e.g. Antón et al., 2014; von Bastian et al., 2016; Paap & Greenberg, 2013; Paap & Sawi, 2014; Rodriguez-Fornells et al., 2012).

Our version of the test consisted of 100 trials, half congruent and half incongruent, separately randomised for each subject. The trials were divided into four blocks with 5 s breaks in-between. Before the actual test took place, a practice sequence was presented to each participant. Each trial began with an 800 ms fixation cross, which was immediately

followed by a row of five arrows remaining on the screen for 800 ms unless a response was given. Finally, the screen was blank for 500 ms.

The spatial n-back task (adapted from Carlsson et al., 1998) is a measure of working memory updating and monitoring. In the task, a white square is presented in one of eight possible locations on the screen, and the participant is to determine whether the location of the presented square is or is not the same as for the previous square (1-back), or whether the location is the same as the one before the previous square (2-back). The task is divided into 1-back and 2-back sequences. The n-back effect is calculated by subtracting the average reaction time or error rate of the 1-back trials from the average reaction time or error rate of the 2-back trials. Studies comparing monolinguals and bilinguals on working memory tasks have yielded inconsistent results (e.g. Bialystok et al., 2004; Ratiu & Azuma, 2015).

In the current version of the task, there were 80 1-back trials and 80 2-back trials. The trials were divided into two blocks of 80 trials with a 15 s break in-between. Each block consisted of four sequences of 20 trials: two sequences with 1-back trials and two sequences with 2-back trials. Each of the four sequences consisted of 6 targets and 14 non-targets. The order of the sequences was 1-back, 2-back, 2-back, 1-back within the first block, and 2-back, 1-back, 1-back, 2-back within the second block. The presentation order of the trials was pseudorandomised within each sequence. Prior to the actual task, the participant completed a practice sequence. In the beginning of each sequence, the number "1" or "2" appeared on the screen indicating whether the next sequence would be a 1-back or a 2-back sequence. The number remained on the screen for 5,000 ms. Each trial started with a fixation cross in the middle of the screen and a square in one of eight possible locations. The square remained on the screen for 100 ms. A new trial began after a 3,000 ms blank interval irrespective of whether a response was given.

The number–letter task (adapted from Rogers & Monsell, 1995) is a measure of shifting abilities. In this task, a number–letter pair (e.g. "3A") appears in one of two boxes, and the subject has to decide either whether the number is even or odd, or whether the letter is a vowel or a consonant, depending on which box the pair appears in. Each time the number–letter pair appears in the upper box, the subject has to determine the number, and each time the pair appears in the lower box the subject has to determine the letter. The responses are given with two keys, one for vowels or even numbers, and another for consonants or odd numbers. The task consists of three blocks: two single-task blocks (number only and letter only) and one mixed-task block. In the mixed-task block, a trial is either a repetition trial, where the task of the subject is the same as on the previous trial, or a switch trial, where the task shifts. Two executive measures are calculated from the number–letter task. The switching cost is calculated by subtracting the average reaction time or error rate of the repetition trials in the mixed-task block from the average reaction time or error rate of the switch trials in the mixed-task block. The mixing cost is calculated by subtracting the average reaction time or error rate of the single-block trials from the average reaction time or error rate of the repetition trials in the mixed-task block. Whereas the switching cost is typically taken to measure simple set shifting abilities, the mixing cost is generally considered to reflect a monitoring cost. On repetition trials, unlike single-task trials, the task does not change, but the subject is aware that it may change at any moment. The single-task blocks consisted of 32 trials each (block 1: number; block 2: letter). The mixed-task block consisted of 32 switching trials and 48 repetition trials. Of the repetition trials, 24 were number trials and 24 were letter trials. The number–letter pairs appeared in the two squares randomly. Each block was preceded by a practice sequence. The results from earlier studies comparing bilinguals and monolinguals on set shifting tasks have provided inconsistent results (e.g. Paap, Johnson, & Sawi, 2014; Yow & Li, 2015).

Although the Simon, Flanker, n-back, and number–letter tasks are each designed to tap a specific executive function, it is probable that the tests partly overlap. To take two examples, in the switch trials of the number–letter task, the subject has to inhibit the earlier response pattern. In the number–letter task, the subject has to maintain in working memory as to which task to perform. This overlap is natural, since the executive functions themselves are strongly interrelated (Miyake et al., 2000).

The participants filled in forms regarding their background, including questions about their date of birth, education, occupation, vision, hearing, possible reading difficulties, possible neurological or

psychiatric illnesses, medication, subjective level of alertness, and possible alcohol intake during the last 24 hours prior to testing. The participants also filled in questionnaires concerning their language background and language skills: age of acquisition of Swedish and Finnish, percentage of average use of Swedish, Finnish, and two other possible languages during the last three years (adding up to 100%), and subjective estimations of written and spoken language skills in Swedish, Finnish, and in any other two languages.

In Soveri, Rodriguez-Fornells et al. (2011), self-reports on the percentages of language use were employed to assess the extent of the everyday use of Swedish and Finnish by subtracting the less frequently used language from the more frequently used language. Here (unlike in Soveri et al.), the scale was inverted so that a larger number represented a more balanced use of both languages. The subtraction method omits the use of languages besides Swedish and Finnish, and the majority of our participants reported using at least one additional language, most often English (85% of the participants; see Table 2 for self-reported skills in English). In our questionnaire, the subject could list up to two languages used in addition to Swedish and Finnish. Our assumption was that a very balanced use of several languages would be associated with more switching than a very unbalanced use of the languages. Thus as a measure of the degree of switching between any of the languages, a sum of cross products of the language use (in proportion) was employed. For four languages, this yields $L1 * L2 + L1 * L3 + L1 * L4 + L2 * L3 + L2 * L4 + L3 * L4$, where the proportions are coded as coefficients between 0 and 1. Finally, for the ease of interpretation, the whole formula is multiplied by $266.666 \ldots$, so that a situation where a subject uses four languages to an equal degree (that is, each 25% of the time) results in a value of 100, and a situation where he or she only uses one language results in the value of 0. The function is near-to-linear with respect to the percentages of language use.

The Internet testing platform

Participants performed the tests on a computer and Internet browser of their own choice. The testing platform was programmed with Javascript (TM). The Java applet was automatically installed on the participant's computer and was run locally on the participant's computer after installation. Thus, the results are not affected by the speed of the participants' Internet connection. After all the tests were completed, the results were sent to a server, from where they could be downloaded by the experimenter. The responses in all the tests were given on keyboard.

The participants were given written instructions to perform the tests in a single session and in surroundings with minimal distractions. It was possible, however, to pause between tests and continue on the same computer later.

The BWSQ and other self-reports of language switching

The participants completed a Swedish translation of the BWSQ (Rodriguez-Fornells et al., 2012). Following Soveri, Rodriguez-Fornells et al. (2011), subscales (1) and (2) were combined into one, titled Language switches, by adding them up.

As most of the participants could be expected to also know other languages apart from Swedish and Finnish, in addition to the BSWQ we asked about their switching behaviour between any languages through six additional questions concerning the total amount of language switches which were of the following type: "[during time interval i] I switched languages x times" or "[during time interval i] I made many short language switches". Both question types had three variants with different time intervals: switches made yesterday, the average amount of switches made during a day, and the switches made between the ages zero and six. Principal Components Analysis on the response data of Experiment 1 to these six additional questions showed that the questions concerning switches yesterday and switches on the average formed a factor (loadings greater than .7; unrotated factor solution was extracted on the basis of eigenvalues >1), and the questions about switching between the ages zero and six formed another factor (loadings greater than .7). These two factors were titled Recent switching and Childhood switching, respectively, and were calculated as means of the individual question scores.

Statistical analyses

The Java applet only included reaction times from trials that were answered correctly. Outlier removal was done in two stages. First, the Java applet automatically removed trial level outliers: reaction time to an item was deleted as an outlier if it deviated more than ± three standard deviations

from the subject's average score in that task. This removal was done only for the reaction times. Second, outlier removal was also done on the between-subjects level: the mean reaction time of a participant was removed if it deviated more than ± three standard deviations from the average reaction time of all the subjects in that task, and error rate was deleted if it was higher than three standard deviations above the group average (error rate deletion was done only in this direction, because 0% error rate is not likely to be due to chance). The between-subjects outlier removal was done to rule out the possibility that the subjects responded contrary to instructions, either aiming to maximise speed over accuracy or accuracy over speed. The risk that subjects would use either of these strategies is greater in Internet testing than in laboratory testing, as the test leader is not observing the subjects. Moreover, the reaction times or error rates of a subject could be exceptionally high overall due to, say, distractions in the subject's environment or technical problems, and would in that case not be deleted at the trial level. The overall between-subjects deletion rate was 1% for reaction times and 3% for error rates in our two experiments.

Multiple regression analyses were conducted separately for each of the cost measures in the executive tasks. To enable direct replication, the first two regression models followed those presented in Soveri, Rodriguez-Fornells et al. (2011). The first model included the following background factors: (i) participant's age, (ii) age of L2 acquisition, and (iii) everyday use of both languages (i.e. percentage of everyday use of Finnish and Swedish). The second model included the three BSWQ measures, namely (i) Language switches, (ii) Contextual switches, and (iii) Unintended switches.

In addition to these two models that replicated Soveri et al.'s analysis, we wanted to examine the connection between the switching measures derived from the additional questions regarding language switching, that is, the assumed switching between any of the languages (not only Swedish and Finnish), and the executive tasks. Our third regression model thus included the predictors Recent switching and Childhood switching, in addition to the overall measure of switching between any languages (the sum of cross products of languages used). The predictors were inserted to the analyses simultaneously in each model.

Results

The reaction time variables were log-transformed before calculating the cost measures. The cost measure reaction times that were computed from the transformed values were normally distributed (Kolmogorov–Smirnov $ps > .05$). None of the error rate variables in the executive tasks were normally distributed (Kolmogorov–Smirnov $ps < .05$). No significant collinearity was found between the predictors in the three regression models (tolerances > .6).

Performance in the executive tasks is reported in Table 3 as non-transformed values. We expected that the mean reaction times and error rates would be higher for incongruent trials than for congruent trials in the Simon and Flanker tasks. As to the n-back task, we anticipated that the mean reaction times and error rates would be higher in the 2-back trials than in the 1-back trials. In the number–letter task, we expected the reaction times and error rates to be higher in the switching trials versus repetition trials, and in the repetition trials of the mixed-task block versus the single-task blocks. The reaction time cost effects were in the expected direction in all four tasks. Error rate cost effects were in the expected direction only in the Simon and Flanker tasks. Reaction time cost effects were all significant ($|t|s > 5$). The error rate cost effects were significant in all tasks ($|Z|s > 3$) except for the n-back effect and the number–letter mixing cost.

Data from the spatial n-back task was deleted from 26 subjects due to an error in the experiment programme, which caused that the target to non-

Table 3. Performance in the executive tasks (non-transformed reaction times).

	RT in ms		Errors in %	
	M	SD	M	SD
Simon task				
Congruent	463	72	2.6	2.9
Incongruent	489	73	4.4	4.4
Simon effect	26	38	1.7	4.2
Flanker task				
Congruent	448	49	1.9	2.5
Incongruent	508	49	6.7	5.2
Flanker effect	62	23	4.8	4.9
N-back task				
1-back	647	109	11.6	9.5
2-back	777	207	9.4	8.5
N-back effect	130	166	−1.7	11.6
Number–letter task				
Single-task trials	642	90	2.8	2.9
Repetition trials	765	133	2.3	3.7
Switching trials	1,124	223	5.1	6.3
Switching cost	359	164	2.8	5.1
Mixing cost	123	96	−0.4	4.5

Table 4. Summary of the multiple regression analyses with background variables as predictors of processing cost in log-transformed RTs.

				Number–letter task	
Variable	Flanker effect B	Simon effect B	N-back effect B	Switching cost B	Mixing cost B
Constant	0.10	−0.01	−0.10	0.28	0.23
Age	0.00	0.00	0.00	0.01	−0.00
Age of L2 acquisition	0.01	0.01	0.02	0.01	−0.02
Everyday use of both languages	0.00	0.00	0.00	0.00	0.00
R^2	0.09	0.05	0.08	0.05	0.10
F	0.61	0.30	0.25	0.31	0.64

Table 5. Summary of the multiple regression analyses with background variables as predictors of processing cost in errors.

				Number–letter task	
Variable	Flanker effect B	Simon effect B	N-back effect B	Switching cost B	Mixing cost B
Constant	8.5	9.3	−1.3	6.9	10.7
Age	−0.24	−0.18	−0.04	−0.13	−0.36
Age of L2 acquisition	−0.34	−0.45	−2.6	0.02	−0.77
Everyday use of both languages	0.03	−0.04	0.10	−0.02	−0.03
R^2	0.07	0.09	0.23	0.02	0.18
F	0.44	0.62	0.78	0.12	1.3

target ratio was not accurate. In all other tasks, data from all the 66 subjects was included.

None of the multiple regression models were significant for any of the reaction time cost measures, or for any of the error rate cost measures (Tables 4–7).[1]

Discussion

As expected, all four tasks showed significant executive load effects in reaction times, which speaks for the validity of using Internet-based testing with these tasks. The cost effects in error rates were significant except for the n-back effect and the number–letter mixing cost. The absolute reaction times were all slightly faster than in the Soveri, Rodriguez-Fornells et al. (2011) study, which is probably due to the clearly younger mean age of the

participants in the present experiment (M = 20.3), compared to that in Soveri et al. (M = 52.8).

None of the multiple regression models significantly predicted the subjects' performance in the executive tasks, which suggests that in this sample, greater language switching experience was not associated with better executive functioning. One of the Soveri, Rodriguez-Fornells et al. (2011) findings was the negative association between Language switches and the number–letter mixing cost in error rates. One possible reason for failing to replicate this result is that in our sample the number–letter mixing cost in error rates was not significant. As to the executive cost measures that were significant, one possible reason for the null results is that our sample consisted of young university students (mean age 20.3 years). Previous research indicates that executive functions peak between the ages 20 and 29 (De Luca & Leventer, 2008), which possibly makes it more difficult to observe any practice effects.

Also the programming error which led to the deletion of data from 26 participants in the n-back task weakens the power of Experiment 1 to discover any effects in this task, although the remaining 40 participants is still more numerous than the sample included in the Soveri et al. study. The aim of Experiment 2 was to gather a larger sample and from a wider age range than in Experiment 1.

Experiment 2

Materials and methods

Participants

In contrast to the sample of Experiment 1 that was limited to Swedish–Finnish university students, Experiment 2 drew a sample from the general Swedish–Finnish population in Finland. The participants (N = 111, 85 female) were recruited through Facebook advertising directed at Finnish users who used Facebook in Swedish. This targeting in the Facebook was motivated by the fact that, due to being in a language minority, many Swedish-speaking Finns are also highly proficient in Finnish, that is, bilingual. Our goal was to gather as many participants as possible. Two separate advertisement

---

[1]If we examine the individual predictors, we see in the model with the BSWQ variables as predictors that Contextual switches correlated negatively with the Simon effect reaction time, and Language switches correlated negatively with the N-back effect error rate. In the current type of analysis, however, many pairwise comparisons are conducted, which greatly increases the risk of Type I error. The overall model F is less affected by multiple comparisons, as it gives an overall estimate of the model fit. Here we apply the conservative strategy of examining individual predictors only if the overall model F is significant.

Table 6. Summary of the multiple regression analyses with the three BSWQ variables as predictors of processing cost in log-transformed reaction times.

| | | | | Number–letter task | |
| | Flanker effect | Simon effect | N-back effect | Switching cost | Mixing cost |
Variable	B	B	B	B	B
Constant	0.13	0.04	0.17	0.34	0.19
Language switches	−0.00	0.00	0.00	0.00	0.01
Contextual switches	0.00	−0.01*	−0.01	−0.01	−0.01
Unintended switches	0.00	0.01	0.01	0.01	−0.00
R^2	0.01	0.10	0.02	0.03	0.04
F	0.24	2.3	0.20	0.52	0.85

Table 7. Summary of the multiple regression analyses with the three BSWQ variables as predictors of processing cost in errors.

| | | | | Number–letter task | |
| | Flanker effect | Simon effect | N-back effect | Switching cost | Mixing cost |
Variable	B	B	B	B	B
Constant	4.6	1.9	13	5.5	−1.2
Language switches	0.06	−0.13	−1.5*	−0.16	0.05
Contextual switches	0.11	0.17	1.1	−0.15	0.00
Unintended switches	0.07	0.09	−0.01	0.18	−0.02
R^2	0.00	0.01	0.12	0.02	0.00
F	0.07	0.23	1.6	0.30	0.02

campaigns were used. The participation rate reduced sharply by the end of the second campaign, for which reason we concluded that a third campaign would not result in significantly more participants, possibly because most of the targeted population had already been exposed to the ad. The exclusion criteria were the same as in Experiment 1. The mean age of the participants was 38.8 years ($SD = 14.9$, range 18–69). All participants had learned their L2 prior to the age of 7. They estimated having learned Swedish at 0.73 years of age on average ($SD = 1.6$) and Finnish at 1.1 years of age ($SD = 1.8$). Of the participants, 87 percent reported using English to some extent in their everyday life in addition to Swedish and Finnish. The participants were on average as highly educated as the subjects in Experiment 1 with a mean of 16.1 years of education ($SD = 3.8$). The subjects rated themselves as having slightly better Swedish-speaking ($M = 5.7$) and writing ($M = 5.5$) skills than those of Finnish ($M = 5.4$ and $M = 5.2$, respectively; $|Z|s > 2.3$). Demographic information and self-estimated

language skills are reported in more detail in Tables 8 and 9, respectively.

Principal Components Analysis was conducted on the self-reported language switching questions of the data of Experiment 2, resulting in the same two factors as in Experiment 1, that is, "Recent switching" and "Childhood switching" (factor loadings > .7; unrotated factor solution extracted on the basis of eigenvalues >1).

The tasks and questionnaires were the same as in Experiment 1.

Results

The reaction time variables were log-transformed before calculating the cost measures. All cost

Table 8. Demographics and BSWQ scores.

	M	SD	Range
Age (years)	38.8	14.9	18–69
Years of education	16.5	3.3	11–25
Everyday use of L1 and L2 (%)	69.5	25.2	2–100
Overall use of any languages (0–100)	65.8	17.8	3–93
Age of L2 acquisition (years)	3.0	1.8	1–6
BSWQ: language switching (6–30 pts)	15.3	3.3	7–24
BSWQ: contextual switches (3–15 pts)	7.6	2.3	3–14
BSWQ: unintended switches (3–15 pts)	5.1	2.2	3–13

Table 9. Self-reported language skills in Swedish, Finnish, and English (scale 0–6).

Language	M	SD
Finnish		
Speaking	5.4	0.8
Reading	5.7	0.6
Writing	5.2	1.0
Speech comprehension	5.7	0.6
Swedish		
Speaking	5.7	0.6
Reading	5.8	0.5
Writing	5.5	0.8
Speech comprehension	5.9	0.4
English		
Speaking	4.48	0.80
Reading	4.94	0.78
Writing	4.40	0.97
Speech comprehension	4.93	0.78

Table 10. Performance on the executive tasks (non-transformed reaction times).

	RT in ms		Errors in %	
	M	SD	M	SD
Simon task				
Congruent	498	74	2.4	5.6
Incongruent	524	69	4.7	6.8
Simon effect	28	30	2.3	4.5
Flanker task				
Congruent	501	65	4.1	10.1
Incongruent	559	64	9.4	12.7
Flanker effect	58	24	5.1	6.7
N-back task				
1-back	703	151	23.3	26.5
2-back	842	236	31.9	33.3
N-back effect	142	173	8.6	23.1
Number–letter task				
Single-task trials	697	119	8.6	15.3
Repetition trials	875	202	6.1	12.7
Switching trials	1,240	295	8.5	13.3
Switching cost	359	204	2.4	5.7
Mixing cost	167	133	−2.0	10.7

measures were normally distributed (Kolmogorov–Smirnov $ps > .05$). None of the error rates were normally distributed. No significant collinearity was found between the predictors in the regression models (tolerances > .7).

Performance on the executive tasks as non-transformed values is reported in Table 10. All the reaction time cost measures were significant ($|t|s > 8$). All the error rate cost measures were significant as well ($|Z|s > 2.5$), including an anomalous negative mixing cost in error rates in the number–letter task.

The multiple regression model with the background variables age, age of L2 acquisition, and everyday use of both languages was not significant in any of the cost measures, measured in either reaction times or error rates (Tables 11 and 12).

The multiple regression model with the three BSWQ variables was significant in the number–letter task switching cost measured in reaction

Table 11. Summary of the multiple regression analyses with background variables as predictors of processing cost in transformed reaction times.

				Number–letter task	
	Flanker effect	Simon effect	N-back effect	Switching cost	Mixing cost
Variable	B	B	B	B	B
Constant	0.11**	0.05	0.08	0.38**	0.25*
Age	0.00	0.00	0.00	−0.00	0.00
Age of L2 acquisition	−0.00	−0.00	−0.02	0.02	−0.00
Everyday use of both languages	0.00	0.00	0.00	0.00	−0.00
R^2	0.02	0.01	0.06	0.10	0.03
F	0.31	0.14	1.2	2.1	0.66

times. Contextual switches was a significant individual predictor. The more contextual switches a subject reported, the larger the switching cost in the number–letter task, as measured in reaction times (Table 13). As to the error rates, the model was significant in the *n*-back task: the more unintended switches a subject reported, the higher the error rate in this task (Table 14).

The third regression model with Overall use of any languages, Recent switching, and Childhood switching was not significant on any of the executive tasks in reaction times or error rates.

Discussion

All the reaction time cost effects in the executive tasks were significant and in the hypothesised direction, further confirming the validity of the Internet testing platform. On the other hand, the number–letter mixing cost in error rates was anomalous. The main finding of Soveri, Rodriguez-Fornells et al. (2011), which was precisely about this effect, could not be replicated.

In our Experiment 2, the BSWQ model predicted the switching cost reaction times in the number–letter task, with Contextual switches as a significant individual predictor. Contextual switches predicted reaction times positively, that is, the more contextual switches a subject reported, the higher the switching cost in reaction time in the number–letter task. The BSWQ model also predicted error rates in the *n*-back task, with Unintended switches as a significant individual predictor: more unintended switches were related to a higher processing cost in error rates in the *n*-back task.

What could explain these two statistically significant findings? It is plausible that a high rate of reported unintended switches reflects relatively weaker cognitive control in a subject, which in turn would be observed as a higher error rate in the *n*-back task. This is in line with the reasoning of Rodriguez-Fornells et al. (2012) concerning unintended switches. It is, however, unclear why the lack of cognitive control would only manifest itself in the *n*-back task and not in the other tasks. One possible explanation for this is that the rate of unintended switches correlates negatively with the ability to maintain specific goals. A subject who lacks in capacity of goal maintenance may make many unintended language switches in a conversation, and may also have a tendency of failing to maintain a specific goal in the *n*-back task.

Table 12. Summary of the multiple regression analyses with background variables as predictors of processing cost in errors.

				Number–letter task	
	Flanker effect	Simon effect	N-back effect	Switching cost	Mixing cost
Variable	B	B	B	B	B
Constant	2.2	−0.80	−17	3.5	−3.5
Age	0.01	0.05	0.42*	−0.06	−0.03
Age of L2 acquisition	0.47	−0.18	1.5	0.16	−1.5
Everyday use of both languages	0.02	0.03	0.08	0.01	0.10
R^2	0.02	0.04	0.08	0.03	0.10
F	0.41	0.86	1.8	0.69	2.2

Table 13. Summary of the multiple regression analyses with the three BSWQ variables as predictors of processing cost in transformed reaction times.

				Number–letter task	
	Flanker effect	Simon effect	N-back effect	Switching cost	Mixing cost
Variable	B	B	B	B	B
Constant	0.12**	0.07	0.12	0.23**	0.36
Language switching	−0.00	−0.00	0.01	0.00	−0.01*
Contextual switches	0.00	−0.00	−0.02	0.02*	0.01
Unintended switches	0.00	−0.00	0.01	−0.01	0.00
R^2	0.00	0.00	0.04	0.08	0.04
F	0.12	0.14	1.3	2.9*	1.4

How could the positive association between contextual switches and switching cost reaction time in the number–letter task be explained? At first sight, it might appear that long experience of contextual language switches should facilitate set shifting and inhibition and hence diminish the cost in reaction times, but we found the opposite. This might be explained by assuming that contextual switches are for a large part automatic and context-driven bottom-up processes (see Rodriguez-Fornells et al., 2012). Hence, they might reflect the subjects' sensitivity to external cues and lack of inhibition, which could manifest itself as increased switching cost reaction time in the number–letter task.[2]

General discussion

Our primary goal in the two experiments was to replicate the experiment by Soveri, Rodriguez-Fornells et al. (2011) in larger samples and different age groups. We examined the connections between the subjects' self-reported language switching frequency and their performance on four commonly used executive tasks, with the hypothesis that more reported language switching would be associated with better performance on the tasks. Soveri et al. discovered a negative association

between the Language Switching measure of BSWQ and the mixing cost in the number–letter task, which suggests that everyday bilingual language switching might improve working memory monitoring and set shifting, and in particular the top-down management of competing task sets. In contrast, our Experiment 1, including a sample of mostly young university students, did not reveal any significant associations between reported language switching and the executive cost measures. Experiment 2 which included a larger sample, drawn from the general Swedish–Finnish population and from a wider age range than that of Experiment 1, showed that Contextual switches positively predicted reaction times in the number–letter task, and that Unintended switches positively predicted error rates in the n-back task. Neither of these findings, however, provides evidence for the hypothesis that language switching trains EFs. Instead, they are best explained by an individual differences account. It is possible that those individuals who are low in cognitive control and sensitive to external cues make many unintended and contextual switches, and also perform worse in corresponding executive measures. More generally, intentional language switching, which was hypothesised to facilitate executive functions,

[2]As noted by an anonymous reviewer, these interpretations rest on what the BSWQ factors are supposed to measure. The face validity of the individual questions assumedly measuring unintended switches or contextual switches could be questioned. For instance, item 3 in the BSWQ ("I tend to switch languages during a conversation"), which is assumed to measure contextual switches, may not distinguish between contextually driven switches and intentional switches. On the other hand, the individual questions presumably measuring unintended switches have higher face validity; for example, item 8: "It is difficult for me to control the language switches I introduce during a conversation".

Table 14. Summary of the multiple regression analyses with the three BSWQ variables as predictors of processing cost in errors.

Variable	Flanker effect B	Simon effect B	N-back effect B	Number–letter task	
				Switching cost B	Mixing cost B
Constant	0.70	2.0	14.4	0.69	−9.9
Language switching	0.25	0.16	−0.74	0.29	0.06
Contextual switches	0.24	−0.17	−1.0	−0.35	0.33
Unintended switches	−0.23	−0.16	2.6*	−0.01	0.86
R^2	0.03	0.02	0.07	0.02	0.04
F	1.1	0.56	2.8*	0.81	1.5

did not predict performance on any of the executive tasks in either of the experiments.

The main finding of Soveri, Rodriguez-Fornells et al. (2011) was the negative relationship between language switching and the number–letter mixing cost in errors. In our Experiment 1, the number–letter mixing cost in error rates was not significant, and in Experiment 2 it was anomalous. This is probably one reason why we failed to replicate the Soveri, Rodriguez-Fornells et al. (2011) main finding, and may suggest that the measure itself lacks validity and/or reliability (cf. Paap & Sawi, 2014). In fact, a recent laboratory-based test–retest reliability study of several EF tasks (Soveri, Lehtonen, Karlsson, Lukasik, Antfolk, & Laine, submitted) also found that the number–letter task mixing cost in error rates was negative.

Age is possibly a central reason why our results differed from those of Soveri, Rodriguez-Fornells et al. (2011). This is so particularly with respect to Experiment 1, which employed young university students. There is evidence that executive functions peak between the ages 20 and 29 (De Luca & Leventer, 2008). Because of this, young participants may show less variation in the EF tasks than older participants, and may not show a training effect. Experiment 2, however, employed older participants, whose mean age did not greatly differ from that of the Soveri, Rodriguez-Fornells et al. (2011) sample. Yet we were not able to find evidence that language switching would be linked to EFs.

In Experiment 2 we found a positive correlation between Unintended switches and the n-back effect in error rates, and between Contextual switches and the number–letter switching cost in reaction time. These positive results differ from both Experiment 1 and Soveri, Rodriguez-Fornells et al. (2011). A probable reason for the difference between our two experiments is the age distribution: it is possible that the association between low EF and lapses in linguistic control is only

visible in older adults, not in young adults. On the other hand, the fact that no such association was found in Soveri, Rodriguez-Fornells et al. (2011), where a sample of similar age was used, weakens this conclusion. All in all, the inconsistency in the results of our two experiments and those of Soveri, Rodriguez-Fornells et al. (2011) prevents from making any strong conclusions. If there is a relationship between language switching experience and EFs, this connection may be weak and limited to some specific circumstances and age groups.

A possible weakness in both the present experiments and that of Soveri, Rodriguez-Fornells et al. (2011) is that the BSWQ, or its Swedish translation, might not tap actual language switching experience of the participants well enough. The psychometric qualities of the questionnaire in the Rodríguez-Fornells et al. sample (2012) were good, and the Language switching measure also correlated well with self-reported language proficiency (rs = .21–.47), but the measures correlated less well with tested L1 fluency (number of words produced in a given time that begin with a specific letter; rs = .11–.16). More generally, it is not clear how well language users can estimate the frequency and type of language switches they actually make. It is well known that self-reports about past behaviours are prone to errors and biases (see e.g. Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). Rodriguez-Fornells et al. (2012) note that validating the BSWQ would require comparing it to measures of real-life language switches as well, such as switching behaviour in bilingual conversations or picture naming. Another question is the validity of the Swedish translation of the BSWQ. It is possible that the translation loses some of the reliability and/or validity of the original questionnaire, which would call for an independent validation of the translation.

It could also be that the BSWQ does not measure the type of language switches that are most demanding for EFs. In their *Adaptive Control*

Hypothesis, Green and Abutalebi (2013) distinguish between three different interactional contexts; single language, dual language, and dense code switching. They propose that dense code switching puts minimal demands on EFs, as there the subject utilises whichever language route is most readily available. In contrast, in the dual language context speakers communicate in one language at a time, aiming to keep the languages separate, which demands inhibition of the non-target language. Thus, it is possible that dense code switching does not train EFs, whereas dual language switching does. The BSWQ does not clearly distinguish between these two types of language switching contexts, which may be a reason why it did not correlate with the EFs. The distinction between dense code switching and dual language switching should be taken into account in future studies.

An anonymous reviewer suggested two more possibilities why we failed to replicate the Soveri, Rodriguez-Fornells et al. (2011) results. First, it could be that the samples were too homogenous regarding their switching for significant effects to emerge. However, in both of our experiments the variation in the BSWQ variables was very similar to that in Soveri, Rodriguez-Fornells et al. (2011). In fact, the responses were more varied in our samples in the theoretically most important Intended Language Switches variable than in Soveri, Rodriguez-Fornells et al. (2011). Because of this, it is unlikely that our failure to replicate the Soveri, Rodriguez-Fornells et al. (2011) results is due to a lack of variance in the BSWQ variables. Again, it might be that there is too little variation in language switching in the Swedish–Finnish population in general for any EF advantages to be discovered. To circumvent this potential problem, other, assumedly more varied populations should be tested in future studies.

Second, one could ask whether the failure to replicate the Soveri, Rodriguez-Fornells et al. (2011) study could be due to using Internet testing. Internet testing may elicit a milder cumulative testing effect and fatigue than laboratory testing because subjects are likely to be in the comfort of their homes and can take breaks. This is indeed a possibility that we cannot completely rule out. However, the validity of Internet testing is supported by the fact that all the cost effects observed in the EF tasks in reaction times were significant and in the expected direction.

Setting aside the discussed limitations of our study, the current results do not lend much

support to the hypothesis that language switching improves executive functions in bilinguals. This is in line with a meticulously conducted review of 31 experiments by Hilchey and Klein (2011), which showed that the evidence that bilingualism boosts inhibitory skills is sporadic at best (see also the updated review by Hilchey, Saint-Aubin, & Klein, 2015). Whereas Hilchey and Klein (2011) examined only inhibition, Paap (2014) discusses the methodological shortcomings of studies concerning the facilitating effect of bilingualism on executive functions more generally. He argues that bilingual advantages in inhibitory control, switching, and monitoring are difficult to replicate. The small *N* in previous studies may easily result in false positive results, and a confirmation bias to report positive findings may result in negative findings not being reported as often (see also Paap, Johnson, & Sawi, 2015, 2016). In their meta-analysis, De Bruin et al. (2015) found that bilinguals outperformed monolinguals with a .30 standardised mean difference, but also discovered evidence for a publication bias.

It is noteworthy that our study did not address the question whether bilinguals outperform monolinguals, but instead focused on one of the assumed mechanisms underlying the bilingual advantage, namely that language switching experience boosts EF functioning in bilinguals. Our results are compatible with the hypothesis that some other factors besides self-reported (possibly quite recent) switching experience could enhance EF in bilinguals.

In addition to the BSWQ variables, we also examined the connection between various background factors and performance on our executive tasks. None of the regression models including these variables predicted executive performance. It is noteworthy that the multiple regression model with age, age of L2 acquisition, and everyday use of both languages did not yield any significant findings. The fact that balanced use of two languages did not correlate with the EF measures is in contrast with a recent study by Yow and Li (2015), which discovered that balanced use of two languages predicted a smaller mixing cost in the number–letter task. On the other hand, our result is in line with a study by Paap et al. (2014), which found no significant effects of L1/L2 ratio on a colour–word task mixing or switching effect, or the Simon or Flanker effects.

Our secondary goal was to see if the basic executive cost effects could be reliably obtained in the

Simon, Flanker, *n*-back, and number–letter tasks using Internet-based testing. All the cost effects were discovered in reaction times, and most were also found in error rates. The absolute reaction times and error rates were similar to those attained in the same tests in a laboratory setting by Soveri, Rodriguez-Fornells et al. (2011), which suggests that the Internet subjects took the tasks as seriously as those in the laboratory. Some cost effects in reaction times were smaller in our experiment than in the laboratory experiment, which is probably due to the younger age of the participants in our study. All in all, our results are in line with those of Crump et al. (2013).

In sum, in our two experiments, we could not replicate Soveri, Rodriguez-Fornells et al.'s (2011) finding concerning the negative association between language switching and the mixing cost in the number–letter task. Instead, we discovered that contextual switches positively predicted reaction times in the number–letter task switching cost, and that unintended switches explained the *n*-back effect in error rates. These two findings can be hypothesised to indicate that the lack of cognitive control, manifested as a high rate of contextual switches and unintended switches, predicts poorer performance in these two tasks. Thus, our results do not provide evidence for the hypothesis that bilinguals' language switching experience improves executive functioning.

References

Ansaldo, A. I., Ghazi-Saidi, L., & Androver-Roig, D. (2015). Interference control in elderly bilinguals: Appearances can be misleading. *Journal of Clinical and Experimental Neuropsychology*, *37*, 455–70.

Antn, E., Duabeitia, J. A., Estvez, A., Hernndez, J. A., & Castillo, A. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology*, *5*, 112.

Bak, T. H., Nissan, J. J., Allerhand, M. M., & Deary, I. J. (2014). Does bilingualism influence cognitive aging? *Annals of Neurology*, *75*, 959–963.

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

Bialystok, E. (2010). Bilingualism. *WIREs Cognitive Science*, *1* (4), 559–572.

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*, 290–303.

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

Bialystok, E., Craik, F. I. M., & Ryan, J. (2006). Executive control in a modified antisaccade task: Effects of aging and bilingualism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1341–1354.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics*, *19*, 69–85.

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

Bialystok, E., & Viswanathan, M. (2009). Components of executive control with advantages for bilingual children in two cultures. *Cognition*, *112*, 494–500.

Braver, T. S., Reynolds, J. R., & Donaldson, D. I. (2003). Neural mechanisms of transient and sustained cognitive control during task switching. *Neuron*, *39*, 713–726.

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

Carlsson, S., Martinkauppi, S., Rama, P., Salli, E., Kovenoja, A., & Ahonen, H. J. (1998). The distribution of cortical activation during visuospatial n-back tasks as revealed by functional magnetic resonance imaging. *Cerebral Cortex*, *8*, 743–752.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition*, *113*, 135–149.

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