Does Bilingualism Confer Cognitive Benefits? A Tale of Three Probes

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Lily He Wang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Wilma Koutstaal

December 2015

Acknowledgements

Of all of my advisor's students, I've taken the longest to arrive at this point. Frankly I wish at times that I had worked harder or pushed myself more in order to avoid being this kind of record-setter. Part of the "problem" may have been that the years past were anything but drab, or a crisis mode from which I was eager to extricate myself.

The person I'm most indebted to is Wilma. You gave me an opportunity to volunteer in the lab when no one else would and later gambled to take me on as an advisee. Aside from a tender heart and gentle spirit, you've modeled for me the passion and consistency required to thrive in an academic environment. I've most appreciated your mentoring style towards me: allowing me the freedom to explore and decide authentically while providing helpful resources and guidance along the way.

I'm grateful for the dissertation-writing process. It was the most rewarding venturing out of my comfort zone to learn about new statistical concepts, analysis methods, and software; the process boosted my confidence in being able to continually learn as the demand arises. If there is one suggestion I would make with regards to the graduate curriculum, it'd be interleaving years of course-taking with years of hands-on projects in a more focused and deliberate manner.

My dear committee, thank you for staying with me. Your incisive and helpful comments at the prospectus meeting have directed the shape and movement of the current research endeavor. Yuhong and Sashank, you've been with me since the beginning, i.e., the prelim. Niels' addition was different and nice, and I tried to keep your world in mind as I worked through various problems.

How could I forget the 15+ (a rough estimate) student assistants that have helped me with data collection, data entry and cleaning, even data analysis?! Some of them are no doubt more memorable than others, for better or worse, but each of their contribution was unique in its own right.

On an utterly different level, I'm eternally grateful to the Triune God for using the Ph.D. experience to solidify my identity first and foremost as a daughter of the Most High King; no stupid questions, no lack of clever ideas, no insignificant findings, no rejected manuscripts or research blunders can ever change that. Thank You for desiring to grow my character more than I do, in gratitude, kindness, integrity, perseverance, and humility.

After all these years, I'm still far more comfortable telling others that I'm in cognitive psych rather than I'm a cognitive psychologist. The dissertation itself may prove to validate my sentiment.

To Shane,
I'm so blessed to have you in my life. I think we both agree that I would've gotten here sooner had we met earlier.

To Mom and Bob,
You're hesitant to ask lest I feel pressured but I know you wonder (and care). Thank you for supporting me in numerous ways, financial included.

To my unborn child,
If you end up arriving safely in our world, know that you're in my thoughts long, long before I'm in yours.

Abstract

Early reported findings pointed to a wide range of cognitive benefits of bilingualism. Recent meta-analytic and experimental results, however, cast serious doubts on whether these reported bilingual advantages in cognitive tasks were real, especially so in young adults. This dissertation uses a multi-measure, multi-method approach to comprehensively evaluate monolingual and bilingual differences among college-aged participants. It examines the core cognitive domains of executive function (Chapter 2), conflict monitoring in cognitive control (Chapter 3), and creativity (Chapter 4). Results showed that monolinguals and bilinguals did not differ on any of five dimensions of executive function, or in conflict monitoring, as assessed by both easier and more difficult tasks and task conditions. As hypothesized, bilinguals, however, outperformed monolinguals on measures of nonverbal creativity, whereas the reverse was true for verbal creativity. Additional analyses examined the possible contributors to this difference, focusing on objectively-assessed measures of English proficiency (listening, speaking, reading, and writing), visual-spatial reasoning ability, and the creativity-related personality characteristic of Openness/Intellect. English proficiency was significantly correlated with visual-spatial reasoning in bilinguals but not in monolinguals. Mediation analyses revealed that the Intellect aspect of the Openness/Intellect subscale (but not the Openness aspect) mediated the relationship between visual-spatial reasoning and English proficiency. These findings suggest that there are complex interrelations among language use and higher-order problem-solving abilities, including enduring personality traits revolving around cognitive exploration. We conclude, that, at least for young adults, there is no uniform overall

cognitive advantage conferred by bilingualism, but facility in two or more languages can beneficially influence measures of nonverbal creativity.

Table of Contents

Acknowledgements …………………………………………………………………….…. i

Dedication ……………………………………………………………………….…. iii

Abstract ……………………………………………………………………….…. iv

List of Tables …………………………………………………………………..… viii

List of Figures ……………………………………………………………….…. x

Chapter 1 Introduction …………………………………………………………..… 1

Chapter 2 The Probe on Executive Functions (EF) ………………………………. 12

2.1 Conceptualizations of EF ……………………………………………… 12

2.2 EF and related constructs ……………………………………………… 14

2.3 EF in the context of bilingualism research ……………………….... 16

2.4 The current study ……………………………………………….... 18

2.4.1 Method …………………………………………………..... 20

2.4.2 Results …………………………………………………..... 32

2.4.3 Discussion ……………………………………………..… 40

Chapter 3 The Probe on Conflict Monitoring …………………………………….. 45

3.1 Introduction ……………………………………………………………. 45

3.2 Conflict monitoring in bilingualism ……………………………..... 47

3.3 Feature integration theory and the sequential modulation task 53

3.4 The standard Simon task ………………………..…………….... 55

3.5 The arrow Simon task ……………………………………...….. 57

3.6 Method ……………………………………………………….…..... 58

3.7 Results …………………………………………………………....... 61

3.7.1 English proficiency ……………………..…………..… 61

3.7.2 Sequential Modulation ……………………..………….…. 61

3.7.3 Standard Simon and Arrow Simon ………….…………... 75

3.8 Discussion …………………………………….. …………... 82

Chapter 4 The Probe on Creative Performance …………………………..…………. 86

4.1 Introduction ………………………………….…………….. 86

4.1.1 Personal characteristics related to creativity …….…….… 88

4.1.2 Features of the creative work ………..…………………..… 90

4.2 Method …………………………………………..…………... 92

4.3 Results ………………………………………….…………….. 93

4.4 Discussion ………………………………………...…..…… 111

Chapter 5 General Discussion and Conclusion ……………………………………... 115

Bibliography ………………………………………………….………………... 122

Appendices ……………………………………………………………...……….... 146

List of Tables

Table 1-1 Language group means and independent-samples *t*-tests on demographic variables and CCF.

Table 2-1 Pairwise correlations of the EF variables for monolinguals.

Table 2-2 Pairwise correlations of the EF variables for bilinguals.

Table 2-3 Significance of difference between correlation coefficients of the language groups.

Table 2-4 Pairwise correlations of the EF variables for all.

Table 2-5 Descriptive statistics for the dependent measures of monolinguals.

Table 2-6 Descriptive statistics for the dependent measures of bilinguals.

Table 2-7 Task-specific *t*-tests between language groups.

Table 3-1 Mean response time (ms) and accuracy with respect to previous congruency and current congruency shown by language group.

Table 3-2 Mean response time (ms) and accuracy with respect to previous congruency and current congruency on complete-change trials shown by language group.

Table 3-3 Mean response time (ms) and accuracy with respect to previous congruency and current congruency on complete-change trials shown by self-reported English proficiency level for bilinguals.

Table 3-4 Mean accuracy with respect to previous congruency and current congruency on all trials and complete-change trials shown by objectively assessed proficiency level for bilinguals.

Table 3-5 Mean response time (ms) and accuracy separated by whether the location and response were a repetition or change from the previous trial.

Table 3-6 Mean response time (ms) and accuracy separated by whether the location and response were a repetition or change from the previous trial for incongruent-incongruent trial sequences.

Table 3-7 Mean response time (ms) separated by whether the location and response were a repetition or change from the previous trial for incongruent-incongruent trials sequences for low vs. high SEP bilinguals.

Table 3-8 Mean response time (ms) and accurancy of congruent trials, incongruent trials, and the Simon effect by language group.

Table 4-1 Language group means, SDs, independent-samples t-tests, and confidence intervals on variables of the person.

Table 4-2 Pearson correlations of CCF, English proficiency, Openness/Intellect, and CSP for monolinguals.

Table 4-3 Pearson correlations of CCF, English proficiency, Openness/Intellect, and CSP for bilinguals.

Table 4-4 Regression coefficients, standard errors, and model summary information for the personality influence parallel multiple mediator model depicted in Figure 4-3.

Table 4-5 Language group means, SDs, and independent-samples t-tests on creative products.

Table 4-6 Pearson correlations of TTCT-V, TTCT-F, CAQ, and RFFT for monolinguals.

Table 4-7 Pearson correlations of TTCT-V, TTCT-F, CAQ, and RFFT for bilinguals.

List of Figures

2.1 A schematic illustration of the Antisaccade task.

2.2 A schematic illustration of the Letter Memory task.

2.3 Letter and number layout of Trail B.

2.4 A schematic illustration of the Flanking Color Patches task.

2.5 A screen capture of the Corsi Block task.

3.1 Mean response time (ms) with respect to previous congruency and current congruency shown by language group.

3.2 Mean response time (ms) of complete-change trials with respect to previous congruency and current congruency shown by language group.

3.3 Mean response time (ms) with respect to whether the location and response were a repetition or change from the previous trial.

3.4 Mean response time (ms) with respect to whether the location and response were a repetition or change from the previous trial for incongruent-incongruent trial sequences.

4.1 Scatterplot matrix of the "person" variables.

4.2 Correlations of monolinguals and bilinguals on person variables with 95% confidence intervals (CIs).

4.3 A conceptual diagram of a multiple mediation analysis with CCF as the antecedent, English proficiency as the consequent, and Openness, Intellect, and CPS as mediators.

4.4 Scatterplot matrix of the "product" variables.

4.5 Correlations of monolinguals and bilinguals on product variables with 95% confidence intervals (CIs).

Chapter 1

Introduction

How does fluency in two or more languages influence broader cognitive performance? Do bilinguals outperform monolinguals on tasks that call upon cognitive control or interference resolution? Do bilinguals' varied linguistic and cultural experiences enhance their creativity compared to monolinguals? Research aimed at addressing these intriguing—and theoretically and practically important—questions, is rapidly intensifying. While the number of publications on, and citations to, the topic of bilingualism remained largely unchanged from 1993 to 2003, there has been a steep and continual growth of relevant research in the past decade (Kroll & Bialystok, 2013).

Early reported findings revealed a number of cognitive advantages of bilinguals compared with their monolingual peers, across different age groups. Bilingual children showed enhanced ability in judging the grammaticality of sentences in the face of semantic interference (Bialystok, 1986; Cromdal, 1999), in ignoring perceptually misleading information (Bialystok, 2001), and in performing Theory of Mind tasks (Goetz, 2003). Bilingual adults exhibited a superior ability to extract novel ideas (Kharkhurin, 2009) and outperform monolinguals in nonverbal (but not verbal) creative behavior (Kharkhurin, 2010). On interference tasks such as the Simon or flanker tasks that require, for example, mapping motor responses to spatially congruent versus incongruent locations, bilingual adults were found to outperform monolingual adults. They showed faster overall response time (RT) while also showing a reduced interference effect, that is, a smaller increase in RT for incongruent relative to congruent trials. There was also preliminary evidence indicating that bilingualism across the lifespan protected individuals against an earlier onset of symptoms of dementia into old age (Bialystok, Craik, & Freedman, 2007; Bialystok, Craik, Klein, & Viswanathan, 2004). These findings were of interest to the scientific community as they were adding to a growing corpus of literature purportedly demonstrating a link between engaging in certain challenging mental activities, such as working memory training, and enhanced cognitive functioning in various populations and age groups (for recent reviews see, for example, Melby-Lervåg & Hulme, 2013; von Bastian & Oberauer, 2014). In turn, this promised to address an enduring question in psychology of whether engaging in cognitive training could lead to *domain-general* cognitive advantages, that is, beneficial transfer effects evidenced by enhanced performance on tasks that are quite different from those that had been directly practiced or trained.

Yet, in a meta-analytic review of 31 experiments using non-verbal interference tasks with bilingual and monolingual individuals, Hilchey and Klein (2011) found any bilingual advantage on the conflict effect, i.e., the *difference* in response time (RT) for incongruent relative to congruent trials, to be sporadic at best. Rather, bilinguals appeared to simply be faster overall, showing a global RT advantage, i.e., an overall mean RT advantage across both congruent and incongruent trials.

More recently, Paap, Johnson, and Sawi (2014) summarized the outcomes of 76 tests from 35 different studies since the meta-analytic review of Hilchey and Klein (2011), comparing language group performance on non-verbal interference effects or switching costs. Of the 56 experiments using non-verbal interference tasks, 9 reported a bilingual advantage in interference and 3 in monitoring, which was operationalized as either the overall mean response time across congruent and incongruent trials, or the average difference in response times between the congruent trials from a mixed block of both congruent and incongruent trials and a block of neutral (control) trials. Three additional experiments demonstrated a bilingual advantage in both interference and monitoring. Thus on a simple "count" measure, only 15/56 experiments, or about 27% of the experiments, reported a bilingual advantage on this type of task. It is worth noting that more than half of the significant findings on interference and monitoring had a sample size of 30 or less per language group. In addition, Hilchey, Saint-Aubin, and Klein (2014) focused on studies of non-linguistic effects using the Simon and flanker tasks that had appeared since their 2011 review; the data from these further studies continued to disconfirm bilinguals' advantage on the interference effect, while the empirical support for an overall RT advantage had also disappeared.

Taken together, these recent results cast serious doubts on whether the bilingual advantages reported earlier were real or mere artefacts of a combination of issues: small *n* experiments, the classic file-drawer problem where the null results are quietly tucked away, and a publication bias with results in support of a bilingual advantage being more likely to be published than null or mixed findings (de Bruin, Treccani, & Della Sala, 2014). As described by Paap (2014), small *n*s dominated the early-published reports of bilingual advantages—a methodological shortcoming that is even more problematic when random assignment cannot be used to compare the populations of interest. As the sample size grows, however, researchers can be increasingly confident of their ability to reject the null when the null is in fact false (Rouder et al., 2009). However, three recent studies conducted on children, with large *n*s of 360, 504, and 650 and an equal number of participants per language group, did not find differences between monolinguals and bilinguals on a range of tasks, specifically the Attention Network Task (ANT), verbal Stroop, numerical Stroop, a card-sorting task, and a Simon task (Antón et al., 2014; Duñabeitia et al., 2013; Gathercole et al., 2014).

The relations between bilingualism and cognitive skills of attention control, working memory, metalinguistic awareness, metacognitive awareness, symbolic reasoning, divergent thinking, and problem solving were meta-analytically examined by Adesope, Lavin, Thompson, and Ungerleider (2010). Based on statistical outcomes, the researchers concluded that publication bias did not pose a serious threat to the meta-analytic results. However, de Bruin et al. (2014) provided the number of actually conducted but unpublished null-effect studies and suggested that the results of a meta-analysis can in fact be affected by such a publication bias. The researchers themselves even candidly admitted to having written up for publication the results of only one of four bilingualism-related experiments because the results of the other three experiments were not significant.

This dissertation was born amidst such conflicting reports for a general bilingual cognitive advantage. Because "cognitive" is undoubtedly vague and broad, we chose to focus on comparing monolingual and bilingual differences in the areas of executive function (EF), conflict monitoring, and creative performance. The rationale for selection of these three areas is presented in the following three chapters, which address each of these topics in turn.

Because a number of computer-based and paper-and-pencil tasks assessing the aforementioned areas were administered to the same participants over two 2.5-hour sessions approximately one week apart, aspects of the method that are shared in common across the chapters, i.e., participant characteristics, experimental procedure, and some of the test materials, are presented next for the sake of clarity; the content unique to each of the chapters appears within the respective chapters. Similarly, a broader discussion that pertains to the study at large, regardless of the topic or content area, appears at the end as a separate chapter.

General Method

Participants

Recruitment took place in the spring semester of two consecutive years at a large American research university. An online questionnaire soliciting the prospective participant's language background and experiences was sent to psychology students and the university community, resulting in a total of 1669 completed questionnaires between the two waves of recruitment. Eighty-five percent (or 1419) of the respondents indicated a willingness to continue in the lab portion of the research and were further screened and categorized into monolinguals, bilinguals (multi-linguals), or language learners based on the following criteria:

Monolinguals – Indicated no exposure or experience with languages other than English;

Bilinguals (or Multi-linguals) – Began speaking their second language (L2) (and third language, L3, if relevant) by the age of 8 years, and self-rated their L2 (and L3) proficiency as ≤ 3 on a 7-point scale ranging from native (1) to very poor (7);

Language learners – Began learning L2 at 14, 15, or 16 years of age, as defined by the average age at which they began speaking, reading, and writing L2.

Forty-nine respondents were classified as monolinguals, 87 as bilinguals, 33 as multi-linguals, and the majority as language learners. Out of these, 46 monolinguals, 73 bilinguals, 30 multi-linguals, and 117 language learners were randomly selected and contacted by email for the lab portion of the research. The participation rates of the four groups were 19.56%, 47.95%, 43.33%, and 24.79%, respectively. Because one bilingual and one language learner did not complete Session 2 of the study, there were a total of 84 participants for data analysis: 9 monolinguals, 34 bilinguals, 13 multi-linguals, and 28 language learners.

Given that all participants were students at the same university, for which there was a 2-year foreign language requirement for admission into the university, even the self-identified monolinguals had previously studied a foreign language. Thus we combined self-identified monolinguals with language learners to form the monolingual group, $n = 37$. The small number of multi-linguals was combined with bilinguals to form the bilingual group ($n = 47$), representing speakers of Spanish, Chinese, French, Hindi, Korean, Russian, Hmong, Japanese, Malay, and Oromo, with the languages listed from the most numerous to the least numerous speakers.

To ensure that factors potentially associated with executive control would be ruled out so that any group difference would be due to bilingualism rather than any confounding variables (Hilchey & Klein, 2011), all participants were asked to complete three measures: (i) A one-page *Demographic Questionnaire*, used to obtain demographic information such as the participant's age, gender, handedness, years of formal education, experience with video games and musical instruments, and the socioeconomic conditions of their pre-teen and teenage years. (ii) The *Brief Symptom Inventory* (BSI, Derogatis & Melisaratos, 1983), a 53-item inventory measuring 9 primary symptom dimensions, including anxiety and depression. Each item is rated on a 5-point scale (0-4), ranging from "not at all" to "extremely." (iii) The *Cattell Culture Fair (CCF,* Scale III, Cattell & Cattell, 1973) test, a non-verbal measure of fluid intelligence and visual-spatial reasoning. The test requires inductive reasoning about perceptual patterns and consists of four timed subtests (series completion, classification, matrices, and conditions).

The mean scores of the two language groups from these three sets of control measures are provided in *Table 1-1*, together with the results of independent samples *t*-tests comparing the language groups on each measure.

Table 1-1. *Language Group Means, Independent-Samples t-tests, and Confidence Intervals (CIs) on Demographic Variables and CCF.*

	ML (n = 37)	BL (n = 47)	*t* / χ	df	*p*	95% CI
Gender (male/female)	13 / 24	13 / 34	.54	1	.462	[0, .12]
Age (years)	20.70	21.47	-1.31	82	.195	[-1.93, .40]
Edu (years)	14.31	15.41	-2.78	70.55	.007**	[-1.90, -.31]
GPA	3.38	3.39	- .01	82	.990	[-.19, .19]
SES_Before	2.32	2.49	-.91	82	.366	[-.53, .20]
SES_Teen	2.35	2.47	-.54	82	.593	[-.55, .32]
Edu_Father	3.92	3.49	1.05	78	.299	[-.39, 1.25]
Edu_Mother	3.81	4.11	-.69	80	.492	[-1.17, .57]
BSI_Anxiety	2.54	4.53	-2.58	82	.012*	[-3.53, -.46]
BSI_Depression	2.76	4.00	-1.50	82	.138	[-2.89, .41]
Video Game	1.52	1.49	.05	82	.957	[-1.35, 1.43]
Music	.99	1.12	-.17	82	.864	[-1.65, 1.39]
CCF	30.08	28.81	1.44	82	.154	[-.49, 3.03]

Note. ML = monolinguals, BL = bilinguals; CI = confidence interval; Edu = years of formal education starting from Grade 1; GPA = self-reported Grade Point Average on a 4-point scale; SES_Before = socioeconomic status before teenage years, ranked from Excellent (1) to Poor (5); SES_Teen = socioeconomic status during the teenage years, ranked from Excellent (1) to Poor (5); Ed_Father = highest level of education attained by the biological father, ranked from PhD/MD/JD (1) to Less than High School (8); Ed_Mother = highest education level attained by the biological mother, ranked from PhD/MD/JD (1) to Less than High School (8); BSI = Brief Symptom Inventory anxiety and depression subscales; Gaming = video game play by number of hours on average per week; Music – musical instrument training by number of hours on average per week; CCF = Cattell Culture Fair. The degrees of freedom (df) differ because a) two

bilinguals did not provide any information on parental education, b) two more bilinguals only provided the education information of the mother, and c) equal variances between language groups are not assumed for Edu.
* $p < .05$. ** $p < .01$.

The language groups were comparable on most of the demographic variables, the cognitive variable of CCF, as well as video game and music training experience. However, bilinguals where shown to have more years of formal education because of the greater number of graduate students in the bilingual sample. Bilinguals also scored higher on the anxiety subscale of the Brief Symptom Inventory (BSI), which contained items such as feeling fearful, feeling nervous or shakiness inside, and feeling tense or keyed up. Participants endorsed the items on a five-point scale from 0 (not at all) to 4 (extremely) based on how they felt in the past two weeks. The language groups were not found to differ along any other dimensions.

Procedure

After participants were screened via an online survey, they were invited to the laboratory for two 2.5-hour sessions that took place approximately one week apart. A number of computer-based and paper-and-pencil tasks on executive function, creativity, and language assessments were intermixed throughout the sessions to provide variety and to reduce fatigue and carry-over effects. Participants were given regular brief breaks between tasks. See Appendix A for a listing of the tasks and the task order for each of the lab sessions.

Materials

English Proficiency

For all participants we conducted an objective assessment of English proficiency using an aggregate measure of performance across the linguistic domains of listening, speaking, reading, and writing. Given that evidence suggests that cognitive advantages of bilingualism stem from speaking the languages on a regular basis (Emmory, Luk, Pyers, & Bialystok, 2008), communicative competence—that is, listening and speaking—was weighted more heavily (as reflected by the total number of test items) in the objective assessment than was reading and writing.

Our current approach to assessing language proficiency is novel in two respects. First, compared with studies using self-ratings of proficiency based on Likert-type scales, we provide an objective measure of English proficiency for both monolinguals and bilinguals in their shared English language, in addition to self-reported proficiency ratings of the relevant languages where applicable. While testing the bilinguals in one language provides an incomplete assessment of their proficiency level (Gollan, Wisseberger, Runnqvist, Montoya, & Cera, 2012), this approach is a viable option for objectively evaluating proficiency in a mixed group of bilinguals, as is the case of our sample, and may be more informative than using self-ratings as a proxy for language proficiency alone. Second, relative to studies that measure proficiency with vocabulary assessments or naming tests, which have been found to indicate greater English dominance than self-report and interview measures in college-aged bilinguals (Gollan et al., 2012; Sheng, Lu, & Gollan, 2013), our composite-measure consists of direct assessments in each of the four primary linguistic domains.

Recent studies incorporating some form of language proficiency have identified high-proficient bilinguals to be better at convergent thinking processes (Hommel, Colzato, Fischer, & Christoffels, 2011); faster on all trial types of a saccadic arrow Stroop task (Singh & Mishra, 2013), and showing a smaller attentional blink effect (Khare, Verma, Kar, Srinivasan, & Brysbaert, 2012) than low-proficient bilinguals.

The approach we used to assess the listening, speaking, reading, and writing abilities of participants is as follows:

Listening – Participants listened to two two-minute audio segments on a headset, and answered comprehension questions following each segment. The first segment was a mini-lecture regarding the fossil record and the second segment was an excerpt from a graduation speech. For each segment, participants were permitted to take notes while listening, and were then given up to 60 seconds to answer five multiple-choice questions. The listening score was the total number of comprehension questions answered correctly (out of 10).

Speaking – Participants verbally responded to two prompts requesting that they, first, relate a happy childhood memory and, second, discuss the usefulness of advertising in society. Fifteen seconds and 30 seconds were given for the preparation of the responses for the two prompts respectively, with participants given the option of preparing their response mentally and/or using a sheet of paper that was provided. Participants then had 45 seconds and 60 seconds, respectively, to deliver their responses while being recorded using a Samson Go microphone. The two responses per participant were each scored by two trained raters using a 6-point rubric, with the scores then summed for a single speaking score (out of 12).

Reading – Four reading passages of varying lengths and difficulty were selected from a Graduate Record Examination (GRE) preparatory book (The Official Guide, 2012). Participants were given 12 minutes to read the passages and to answer a total of 8 comprehension questions that followed. The reading score was the total number of comprehension questions answered correctly (out of 8).

Writing – Participants were given 12 minutes to describe their choice of whether they would like to travel forward or backward in time by typing in a WordPad document on the desktop computer in the testing room. The writing output was scored by two trained raters using a 5-point scale for each of five scoring categories. The scores from the five categories were then averaged for a single writing score (out of 5).

This concludes the general methods shared in common across the three probes. We turn next to a consideration of how bilingualism might influence executive functions.

Chapter 2

The Probe on Executive Functions

In this chapter we seek to contribute to the existing body of empirical research on the broader cognitive effects of bilingualism using an executive function (EF) framework, with the aim of systematically investigating differences *a*) between monolingual and bilingual groups; and *b*) within bilinguals of varying degrees of English proficiency and language mixing frequency. The adoption of an EF framework was motivated by considerations of practical significance and theoretical meaningfulness. As a construct, EF is found to affect mental health, physical well-being, school achievement, job success, marital harmony, and public safety (Diamond, 2013). And theoretically, EF is conceived to be fundamental to any non-automated or goal-directed thought and behavior. We now turn to a brief overview of EF and its existing connections with bilingualism research.

2. 1. Conceptualizations of EF

EF can be defined as a set of related processes primarily linked to the prefrontal cortex of the brain that regulate human thought and behavior in order to achieve a goal (Banich, 2009; Miyake & Friedman, 2012). A focal point of empirical research and theoretical discussions of EF concerns its components, their relations to one another, as well as the cognitive and biological underpinnings of those components.

Beyond the general agreement that damage to the prefrontal lobe impairs EF (Banich, 2009), little consensus exists on the core components or processes making up the EF. Conceptualizations as varied as three (Fuster, 1980), four (Baddeley, 1996; Luria, 1973), five (Shallice & Burgess, 1991; Stuss & Benson, 1986), or six core EF components have been put forth (Fisk & Sharp, 2004; Floyd, Bergeron, & Hamilton, 2004; Huizinga et al., 2006, as cited in Packwood, Hodegetts, & Tremblay, 2011). The difficulty associated with identifying the number and nature of EF components is manifold. First, the so-called task impurity problem has plagued EF research for decades (Burgess, 1997; Phillips, 1997). Task impurity refers to the fact that no task exists that uniquely and exclusively assesses a single EF process; rather multiple target EF and non-EF processes are implicated by any given task, making EF elusive to assessment. Second, EF measures from the neuropsychological tradition tend to have low internal and test-retest reliability. This could, in part, arise as a result of a decreased capacity of the tests to capture target EFs with repeated testing exposure or practice, since the very nature of EF involves the application of reasoning strategies to novel problems. And because EF is best assessed in non-routine and unstructured situations, measuring EF in a laboratory setting becomes particularly challenging. Third, construct validities of commonly accepted EF tests are not well established, as demonstrated by divergent ways of labeling the factors produced by factor analytic techniques (Miyake et al., 2000).

The latent variable approach could alleviate some of the aforementioned problems and was adopted by Miyake et al. (2000) to assess three EF processes commonly identified in the literature: *Inhibition*, *Updating*, and *Shifting*. By selecting multiple tasks purportedly and primarily tapping each of the three processes, the commonality shared by the tasks—that is, presumably the target EF—was extracted and thus constituted a "purer" measure of the target process. These "purer" processes were then examined in relation to each other and used to predict performance on other commonly used EF tasks. Miyake et al. (2000) found the processes of *Inhibition*, *Updating*, and *Shifting* to be separable yet also found that they worked together differentially to accomplish complex cognitive tasks.

This general pattern reflecting both the unity and diversity of EF has been replicated in samples of young adults (Friedman et al., 2006), older adults (Fisk & Sharp, 2004; Hedden & Yoon, 2006), children (Huizinga, Dolan, & van der Molen, 2006; Lehto, Juujarvi, Kooistra, & Pulkkinen, 2003; van der Sluis, de Jong, & van der Leij, 2007), and clinical populations with attention deficit hyperactivity disorder (ADHD, Willcutt et al., 2001). Neuroimaging studies have also indicated unity and diversity of executive functions in terms of brain localization (Collette et al., 2005; Sylvester et al., 2003).

2.2. EF and related constructs

EF has been reported to correlate with fluid intelligence (*g*F), working memory (WM), semantic memory, and processing speed, among others (Obonsawin, Crawford, Page, Chalmers, Cochrane, & Low, 2002; Decker, Hill, & Dean, 2007; Salthouse, 2005; Salthouse & Davis, 2006; Salthouse & Siedlecki, 2007). It is interesting to note that while EF developed out of neuropsychological assessments of patients with brain lesions relative to their healthy counterparts, fluid intelligence (*g*F), WM, and processing speed are psychometrically-based constructs of cognitive ability that emerged from the experimental tradition of cognitive processes. A difference between the research traditions is reflected in the definition of terms. For instance, EF researchers refer to WM as a subcomponent of EF, whereas many WM researchers use the term far more broadly so that it is nearly synonymous with EF (Diamond, 2013). Empirically though, the common variance of EF tasks correlated highly with that of WM tasks at .97 (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010), and the researchers labeled the shared component executive attention, a concept we will return to a bit later.

Given such high correlations between purported EF tasks and tasks of another construct, Salthouse, Atkinson, and Berish (2003) argued that it is essential to demonstrate the construct validity of EF by both convergent validity and divergent validity. In other words, variables assumed to reflect EF or aspects of EF should exhibit moderate loadings on the hypothesized aspects/constructs, while at the same time showing much weaker relations to other constructs. In a sample of 261 adults, the researchers found that variables of inhibition, updating (WM), and time-sharing (dual task performance) loaded moderately on their respective constructs, but the common variance shared by updating (WM) tasks and by time-sharing tasks each correlated highly with fluid intelligence (*g*F). This led Salthouse and colleagues to conclude that there was a lack of discriminant validity of aspects of EF from fluid intelligence (Salthouse et al., 2003; Salthouse & Davis, 2006). Inhibition, however, appeared to correlate less strongly with *g*F than did updating or time-sharing, thus exhibiting comparatively better discriminant validity as a component of EF.

Inhibition also occupies a special position in Miyake and Friedman's (2012) unity/diversity framework, which is expressed as follows: the unity of the framework is reflected in the presence of common EF in all EF components of *Inhibition*, *Updating*, and *Shifting*, whereas diversity is observed in the updating-specific and shifting-specific components after the common EF is accounted for. Because in their re-analyses of previous works, Miyake and Friedman (2012) found that *Inhibition* correlated perfectly with the common EF, there is no inhibition-specific variance after common EF is taken into account.

The executive attention mentioned earlier may in fact be analogous to the common EF in Miyake et al.'s (2012) unity/diversity framework, with updating-specific and shifting-specific variances left uninvestigated in the tasks used by McCabe et al. (2010).

The common variance account of Miyake et al.'s (2012) unity/diversity framework also coheres with the close connections found between WM and inhibition in the literature of the experimental tradition, regardless of which construct is construed as primary and the other derivative (Hanania & Smith, 2010; Munakata et al., 2011) or if both depend on the same limited-capacity system (Engle & Kane, 2004).

2.3. EF in the context of bilingualism research

Current evidence appears mixed with regards to whether bilinguals demonstrate superior EF functioning to monolinguals. We next broadly delineate the landscape of bilingualism EF research under the unity/diversity framework of *Inhibition*, *Updating*, and *Shifting*.

Whether a bilingual advantage in inhibition is observed seems to depend on a variety of experimental parameters: type of inhibition (Colzato et al., 2008), age of the participants (Dunabeitia et al., 2014, but also see Wimmer & Marx, 2014), whether the stimuli to be inhibited are linguistic (Weissberger, Gollan, Bondi, Clark, & Wierenga, 2015), whether inhibition is assessed using behavioral or neuroimaging measures (Rodriguez-Pujadas et al., 2014), and whether inhibition is recruited jointly with monitoring (Morales, Gomez-Ariza, & Baj, 2013).

In contrast to the many nuances present in the inhibition-related findings, evidence relating to updating is less equivocal; bilinguals do not differ from monolinguals on performance of updating or working memory tasks, regardless of age (Bialystok, Craik, & Luk, 2008; De Abreu, 2011; McCrea, Penningroth, & Radakovich, 2015; Ratiu & Azuma, 2015; but also see Morales, Calvo, & Bialystok, 2013).

Shifting may be the most investigated of the three EF processes in bilingualism research. Bilinguals' frequent switch of their two languages is related to the broader theoretical question of general task switching and the degree to which language switching and general task switching are independent, if at all. The common switching measures used by these studies are: a) switch cost, also referred to as local switch cost, assessed as the performance difference between switch and non-switch trials in mixed-task blocks; and b) mixing cost, also referred to as global/general switch cost, assessed as the performance difference between single-task blocks and non-switch trials in mixed-task blocks.

Some studies have identified a bilingual advantage on switch cost (Christoffels, de Haan, Steenbergen, Wildenberg, & Colzato, 2015; Prior & MacWhinney, 2010), on mixing cost (Yow & Li, 2015), and a reduced switch cost pertaining to certain stimulus types or response modes (Hernandez & Kohnert, 2015; Tse & Altarriba, 2015), whereas others have not (Hernández, Martin, Barceló, & Costa, 2013; Kaushanskaya, Gross, & Buac, 2014; Mor, Yitzhaki-Amsalem, and Prior, 2015; Moradzadeh, Blumenthal, & Wiseheart, 2015; Paap & Greenberg, 2013; Tse & Altarriba, 2014).

The inconsistent findings scattered throughout the literature on EF in bilingualism partially result from the varying and imprecise ways of labeling the constructs that are measured by the experimental tasks. Consider the commonly used Simon task as an example; the task consists of trials of color stimuli or geometric shapes appearing on the same or opposite side of the computer screen as the correct response key. When the stimulus appears on the same side as the correct response key, it is designated a congruent trial; otherwise, it forms an incongruent trial. A non-exhaustive query of recent publications reveals investigators variously interpreting the Simon task as allegedly assessing executive function (Gathercole et al., 2014; Kousaie et al., 2014; Paap & Sawi, 2014), executive control (Coderre & van Heuven, 2014; Kirk, Fiala, Scott-Brown, & Kempe, 2014), inhibition or inhibitory control (Goral, Campanelli, & Spiro, 2015; Liu, Rossi, Zhou, & Chen, 2014; Poarch & van Hell, 2012), attentional control (Tse & Altarriba, 2014), selective attention (Bialystok, 2006), conflict resolution (Mohades et al., 2014), and cognitive control (Blumenfeld & Marian, 2014). In addition, the tasks frequently employed by bilingualism researchers often are not tested by EF researchers, which results in a mix of tasks lacking convergent validity (Paap & Greenberg, 2013). Further contributing to this lack of convergent validity is the small number of dependent measures typically used in bilingualism research; most studies employ one or at most two tasks as the dependent variable, with only a handful adopting four tasks or more (Festman, Rodriguez-Fornells, & Münte, 2010; Kousaie et al., 2014; Paap, Johnson, & Sawi, 2014; Soveri, Rodriguez-Fornells, & Laine, 2011; Yow & Li, 2015).

2.4. The current study

To help address these validity issues and to more adequately bridge bilingualism research with the EF literature, we adopted two EF frameworks for the current study; one was based on neuropsychological and neuroimaging research (Shimamura, 2000) and the other was based on the research of individual differences using data reduction techniques (Miyake et al., 2000). The two frameworks complement each other in having emerged from different research traditions, as well as in emphasizing different levels of analysis at which the core processes are implicated. The frameworks together led to the differentiation of five different processes: *Inhibition*, *Updating*, *Shifting*, *Maintaining*, and *Selecting*. Each process was assessed by two experimental tasks (described later) for a total of 10 EF tasks.

Selecting refers to the ability to focus attention on aspects of information processing, such as attending to perceptual stimuli or to memory representations. In neuroimaging research, the anterior cingulate, polar frontal cortex, and medial frontal regions have been shown to activate in selective attention tasks such as Stroop and flankers (Shimamura, 2000). *Maintaining* refers to the ability to keep information active in short-term memory for a period of seconds after it has been selected. The process of maintaining is often assessed by immediate span tasks involving verbal, spatial, or object information. Neuroimaging studies of span tasks show that prefrontal regions recruit the posterior parietal cortex and the inferior temporal cortex for the maintenance of spatial and object information, respectively (Shimamura, 2000). *Updating* refers to the manipulation and reorganization of information in short-term memory. Rather than merely maintaining or storing information, updating involves actively reorganizing short-term memory contents. Neuroimaging studies have shown the dorsolateral prefrontal areas to be activated in updating tasks such as the n-back, verbal fluency, and self-ordered pointing. *Shifting* refers to the ability to switch from one cognitive process or response set to another, usually involving the same stimuli, whereas (cognitive) *Inhibition* entails deliberately inhibiting dominant, automatic, or prepotent responses or mental representations when necessary.

In addition to completing EF tasks intended to assess these five different processes, bilinguals were also assessed for their English proficiency and amount of language mixing. Because higher language proficiency has been found to produce stronger parallel activation of lexicons in bilinguals (Van Hell & Dijkstra, 2002), it is also expected to modulate the demand for the executive system's intervention in order to keep the unintended lexicon from intruding upon a bilingual's speech production. Recent studies incorporating some form of language proficiency assessment have identified high-proficient bilinguals to be better at convergent thinking processes (Hommel, Colzato, Fischer, & Christoffels, 2011); faster on all trial types of a saccadic arrow Stroop task (Singh & Mishra, 2013), and as showing a smaller attentional blink effect (Khare, Verma, Kar, Srinivasan, & Brysbaert, 2013) than low-proficient bilinguals. In a related approach, bilinguals' susceptibility to cross-language interference is indexed by the amount of language mixing that occurs. Bilinguals who do not mix their languages demonstrate better performance on the Tower of Hanoi, the Go/No Go task, the Wisconsin Card Sorting Task, and divided attention tasks than those who do (Festman et al., 2010; Festman & Muente, 2012). In light of these findings, we will examine whether high-proficient and low-proficient bilinguals, as well as high-mixing and low-mixing bilinguals, differ in EF performance, in addition to comparing between language groups.

2.4.1. *Method*

Materials

Language Mixing

The online language-experiences survey that participants initially completed included a question about language mixing. Participants were asked to respond *Yes* or *No* to the question: *When you are speaking, do you mix words or sentences from two or more languages you know?* For those who selected *Yes*, they were presented with a matrix table with frequency of mixing (*Rarely, Occasionally*, *Sometimes*, *Often*, *Always*, *Does not apply*) designated in columns and the interpersonal or relational categories with whom their language mixing might occur (*Spouse, Children, Parents, Siblings, Friends, Coworkers, and Classmates*) designated in rows. Levels of frequency were assigned 1 to 5 from *Rarely* to *Always*. The total mixing score was the sum of scores across all relational categories.

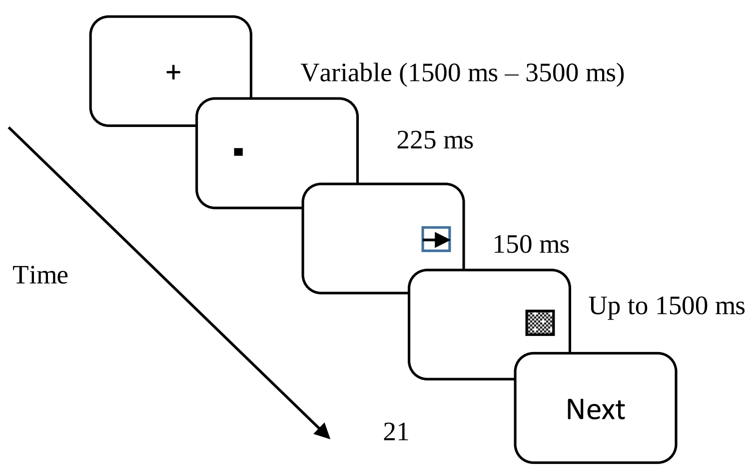
EF Tasks

The ten tasks used to assess the five processes were as follows: Antisaccade and Stop Signal (Inhibition); Tone Monitoring and Letter Memory (Updating); Number Letter and Trails A & B (Shifting); Flankers and Flanking Color Patches (Selecting); and the Spatial Delayed Response Task and Corsi Blocks (Maintaining). These tasks were chosen based on their inclusion in the two aforementioned theoretical frameworks, prevalence of use in the EF literature, and generally acceptable validity and reliability estimates.

Antisaccade (Inhibition)

Antisaccade is an inhibition task administered in the visual modality. In this task a central fixation cross first appeared for a variable duration. Then a cue flashed for 225 ms either to the left or to the right of fixation subtending 0.39 degrees. Next, the target, a 16 mm open black square with an arrow inside, pointing left, right, or up, appeared on the opposite side of the cue for 150 ms. Then the open square was masked by a grey square patch for 1500 ms or until the participant responded, whichever occurred first. Participants were required to indicate the direction of the arrow that was presented in the open square by pressing the left, right, or up key on the keyboard with the index, middle, or ring fingers of their preferred hand. See *Figure 2.1* for a schematic illustration of the task.

Figure 2.1. A Schematic Illustration of the Antisaccade Task.



The task consisted of 22 practice trials and 90 test trials. The fixation duration ranged from 1500 ms to 3500 ms in incremental intervals of 250 ms, resulting in 9 different durations, which were randomized in a fixed order. There were 10 trials for each of the nine fixation durations; the cue appeared equally often on the left and right for each duration. The 90 trials were evenly distributed across the three arrow directions (left, right, up), and the arrow directions were roughly evenly distributed across short (1500 ms, 1750 ms, and 2000 ms), mid (2250 ms, 2500 ms, and 2750 ms), and long (3000 ms, 3250 ms, and 3500 ms) fixation durations. Overall task accuracy served as the dependent variable.

The open black square measured 16 mm in length, and the arrow inside was 14 mm long. The arrow was composed of two sides of an equilateral triangle with a side of 2.5 mm. The cue was a 4 mm solid black square. The grey square patch that masked the open square had a side of 22 mm. The distance from the central fixation to the inner edge of the open square was 110 mm.

Stop Signal (Inhibition)

The second inhibition task included in the study was the Stop Signal task (or STOP-IT), which we obtained from Verbruggen, Logan, and Stevens (2008). In this task, participants responded to a stimulus appearing in the center of the screen by key press. When the stimulus was a square, the participant was required to press the *z* key; when the stimulus was a circle, the participant had to press the */* key. On 25% of the trials, an auditory tone (i.e., the stop signal) was presented after a variable delay (stop-signal delay or SSD). The SSD was initially set at 250 ms and continuously adjusted with a tracking procedure to obtain a probability of stopping of .50. The SSD decreased by 50 ms after unsuccessful stopping and increased by 50 ms after successful stopping.

The task started with a practice block of 32 trials, followed by three experimental blocks of 64 trials per block. After each block, the participant received feedback on their performance in the block just completed. There was an automated 10-second break between blocks. Output was generated by the program. The main dependent variable of interest was the speed at which one was able to refrain from responding at the presentation of the stop signal, or stop signal response time (SSRT), calculated by subtracting mean SSD from the untrimmed mean RT.

Tone Monitoring (Updating)

Tone Monitoring is a task of updating administered in the auditory modality. In this task, participants were presented with four trial blocks, with each block consisting of 25 tones intermixed in a semi-random order. Of the 25 tones in a block, eight were high tones (880 Hz), eight were mid tones (440 Hz), and eight were low tones (220 Hz). The 25th tone was different for each block except that two blocks had a mid tone. Participants were instructed to press the space bar whenever they heard the fourth tone of each pitch. After they pressed the space bar for a particular pitch, the count of that pitch would reset to zero and start over.

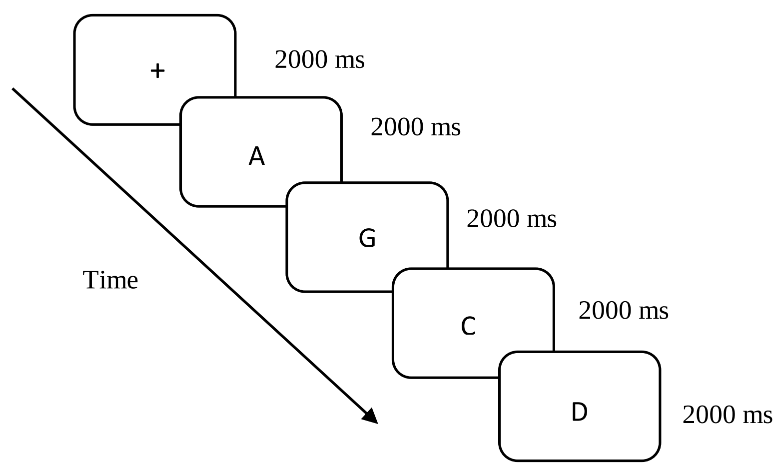
The target tones, that is tones warranting spacebar presses, were located relatively close to each other in two of the four blocks, whereas in the other two blocks, the target tones were distributed farther apart. The tones were 500 ms in duration and the inter-stimulus interval was 2500 ms. Participants were presented with a guided training, as well as a practice block before commencing the test blocks. There was a 9-second automated break between trial blocks.

The dependent variable was the percentage of target tones correctly identified.

Letter Memory (Updating)

The second task of updating was Letter Memory. Capitalized letters in white font were presented pseudo-randomly one at a time on a black background in 2000 ms intervals. The participant was asked to name the letter out loud, as well as the letters presented previously for up to four letters total, in the order they were presented. For instance, if the letters A, G, C, K, and D were presented one after another, the correct response would be A, AG, AGC, AGCK, GCKD. The letter sequence was nine, five, eleven, or seven letters in length, and was presented in the aforementioned order three times, resulting in a total of 12 trials. The task stimuli included 18 letters of the alphabet. Letters F, H, L, M, S, W, and X were excluded because they potentially took longer to name compared to the included letters, and the letter I was excluded given its confusability with the number 1. Each of the 18 included letters appeared five or six times throughout the task. See *Figure 2.2* for a schematic illustration of the Letter Memory task.

Figure 2.2. A Schematic Illustration of the Letter Memory Task.



The letters measured 15 mm tall and 13 mm wide (corresponding to a size of 58 points in the presentation software), and were shown in Arial Unicode Bold font.

Participants were given two practice trials, including a five-letter sequence and a seven-letter sequence.

The participant's responses were recorded to an online server using Blue Microphone's Snowball USB microphone as an mp3 file, while the experimenter checked the participant's responses against a pre-prepared answer sheet.

The dependent variable was the overall percentage of letters correct out of a total possible of 48 letters.

Number Letter (Shifting)

Number Letter was considered a shifting task primarily. The stimulus was a number-letter pair (e.g., G2, 9U) that appeared in one of four quadrants (two quadrants at the top and two at the bottom). In the first block of trials, the number-letter pair appeared in either of the top two quadrants and called for an odd/even number judgment, that is pressing the *v* key for odd and the *m* key for even. In the second block, the stimulus pair appeared in either of the two bottom quadrants and called for a vowel/consonant letter judgment, or pressing *v* for vowels and *m* for consonants. In the third and final block of trials, the stimulus pair appeared consecutively in the four quadrants in a clockwise fashion starting from the upper right position. The participant responded to the number (odd/even decision) when the stimulus appeared in either of the top quadrants and responded to the letter (vowel/consonant decision) when the stimulus appeared in either of the bottom quadrants. The stimulus-response mapping was counterbalanced across participants.

The first and second blocks each had 10 practice trials and 32 test trials. The third block had 12 practice trials and 128 test trials. The inter-trial interval was 2000 ms in all blocks. The number-letter pair was displayed in black font against a white background for

5000 ms, during which participants made their number or letter decision accordingly. In the first two blocks, the number-letter stimulus pairs were presented equally often on the left and right side of the two quadrants in a pseudo-random order. The number and letter positions within a pair (e.g., A6 and 6A) were reversed for half of the trials in each block.

Each of the four quadrants measured 55.5 mm by 68.5 mm. The number-letter pair stood at 6 mm tall and had an approximate width of 10 mm. The font and size of the stimuli were Book Antiqua Bold and 27 points in the presentation software.

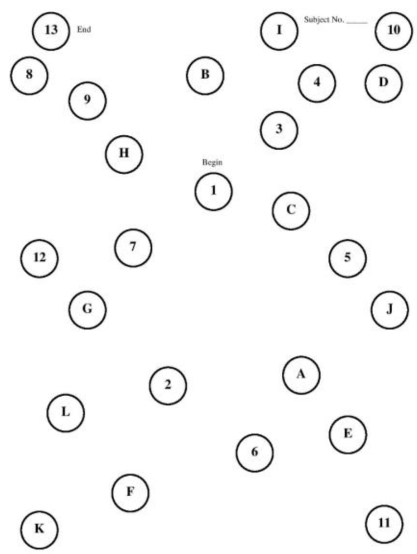
The letters A, E, I, and U were included as vowels; G, K, M, and R were included as consonants. Numbers 2 through 9 were included as numbers. Numbers and letters were chosen to minimize visual confusability (e.g., the letter O was omitted and since the letter I was used, the number 1 was not included). In the first block, numbers 2, 3, 4, and 5 were paired with consonants and numbers 6, 7, 8, and 9 were paired with vowels. In the second block, numbers 2, 3, 4, and 5 were paired with vowels and numbers 6, 7, 8, and 9 were paired with consonants. In the third and final block, all numbers were paired with all letters to yield 64 stimulus pairs. Reversing the number and letter positions in the pair yielded a second set of 64 stimulus pairs for a total of 128, with the first trial in the block excluded from the analysis.

The dependent variable was the difference between the averaged RTs of trials in the third block that required a mental shift of task sets and the averaged RTs of trials from the first two blocks in which no shift was necessary.

Trails A and B (Shifting)

The other shifting task included was Trails A and B. This was a pencil-and-paper task with two parts: Trail A and Trail B. Trail A consisted of the numbers 1-25 semi-arbitrarily displayed on a page. Participants needed to draw lines as quickly as possible to connect the numbers in an ascending order beginning from 1 without lifting their pen from the page. Trail B consisted of 13 numbers (1-13) and 12 letters (A-L) semi-arbitrarily arranged on a page. Participants were instructed to draw lines as quickly as possible to connect the numbers and the letters in an alternating fashion in the order of 1, A, 2, B, 3, C, etc. See *Figure 2.3* for the letter and number layout of Trail B.

Figure 2.3. Letter and Number Layout of Trail B.



Trail A was always administered before Trail B. Before each trail's test session, a short practice session containing one-third of the stimuli was given to the participants to ensure that the participant understood the instructions. Times were recorded in seconds separately for Trail A and Trail B.

The dependent variable was the ratio of Trail B time over Trail A time.

Flankers (Selecting)

Flankers was a selecting task. The stimuli consisted of a horizontal row of five arrows, with the target arrow in red and the four flanking arrows in black. The target arrow could occupy any of the three inside positions. When the target arrow and the flanking arrows pointed in the same direction, it was a congruent trial; an incongruent trial was when the target arrow and flanking arrows pointed in opposite directions. There was also a neutral condition, in which the target arrow appeared in the center and two open diamond shapes flanked the target arrow on each side. The participant responded to the direction of the target arrow by pressing the key *b* for left and the key *n* for right using the index and middle fingers of their preferred hand. The dependent variable was the overall RT.

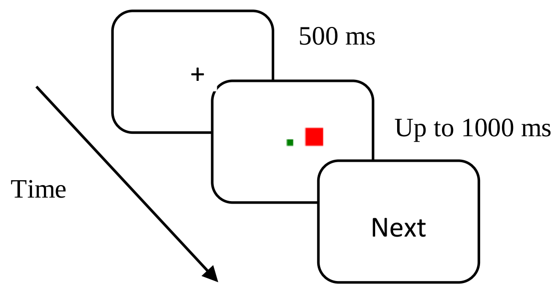
There were 14 practice trials and two test blocks of 91 trials per block, with the second block being a repeat of the first. Forty-two congruent trials, 42 incongruent trials, and 7 neutral trials were semi-randomly intermixed, with no more than four trials of the same type appearing consecutively. Each trial was presented for up to 1500 ms or until the subject responded, and the inter-trial interval (in the form of a central fixation cross) was 1000 ms. There was an automated break lasting nine seconds between blocks.

The stimuli for congruent, incongruent, and neutral trials measured 9 mm tall. The congruent trial measured 37 mm in length, with a distance of 8 mm between arrow tips. The incongruent trial measured 34 mm in length. The flanking diamond shape in the neutral trial was formed by joining together two arrows pointing in opposite directions, resulting in a diamond shape 11 mm in width. The neutral trial measured 58 mm in length, with 1 mm of distance between the diamonds.

Flanking Color Patches (Selecting)

The second selecting task included in the study was Flanking Color Patches. The task consisted of two blocks of trials. In the first block, a small red or green square appeared in the center of the screen with a large square of the same color, a large square of the opposing color, or nothing next to it, thus forming congruent, incongruent, and neutral trials, respectively. The participant was required to press the space bar whenever the small square in the center was green. In the second block, the stimuli and the order of presentation remained the same as the first block but, instead, the participant was asked to press the space bar in response to a small red square appearing in the center. The dependent variable was the overall RT. See *Figure 2.4* for a schematic illustration of the task.

Figure 2.4. A Schematic Illustration of the Flanking Color Patches Task.



There were 10 practice trials and 73 test trials for each block. Thirty-four congruent trials, 31 incongruent trials, and 8 neutral trials were semi-randomly intermixed, with no more than four trials of the same type appearing consecutively. Each trial was presented for up to 1000 ms or until the subject responded, and the inter-trial interval (in the form of a central fixation cross) was 500 ms.

In a block, 37 trials had the small red square in the center and 36 trials had the small green square in the center. Aside from eight neutral trials, the numbers of congruent and incongruent trials were approximately equal between stimulus types. The small square in the center measured 5 mm in length and the large square measured 15 mm in length. The distance between edges of the squares was 10 mm.

Spatial-Delayed Response Task (Maintaining)

The Spatial-Delayed Response Task was a maintaining task that we obtained from Lyons-Warren, Lillie, and Hershey (2004). Participants first focused on a central fixation cross on a computer screen. While fixated, a cue comprised of a solid circle 13 mm in diameter, appeared for 150 ms in one of 32 possible locations towards the outer edge of the screen. Then a variable delay period of five seconds, 15 seconds, or 30 seconds was imposed. During the delay, a series of geometric shapes appeared in the place of the central fixation as distractors. The participant was required to press the spacebar whenever a designated diamond shape appeared. After the delay, the fixation cue returned and the participant pointed to the screen where they remembered seeing the cue, and the experimenter marked the pointed location with the cursor (Lyons-Warren, Lillie, & Hershey, 2004).

There were 10 trials for each of the three delay periods and five cue-present trials, resulting in a total of 35 trials for the task. For cue-present trials, as suggested by its name, the cue was present during the response phase. These trials served to ensure that the participant understood the task instructions and were not pointing randomly on the screen. All trials were presented by the program in a random order. There were six practice trials before the test trials.

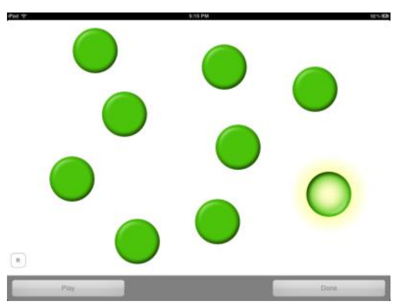
The target detection task on the geometric shapes presented as distractors during the delay indexed how well the participants were paying attention during the delay. Performance on the delay target detection task was measured by overall accuracy (i.e., combined correct rejections, or no response when the stimulus was not a diamond and correct hits, or response if a diamond) and the accuracy of responding to the diamond shape only (i.e., correct hits).

Performance on the primary spatial delayed working memory task was assessed as the mean error of spatial location, or the distance between the recalled and actual target location, calculated in pixels for each participant for each delay period.

Corsi Blocks (Maintaining)

For the second task of maintaining, we administered PathSpan, an iPad version of the Corsi Block tapping test (LeFevre et al., 2010). After registering the participant ID, nine green buttons appeared on a white background in an arbitrary configuration. A trial consisted of the buttons lighting up one at a time for one second and at successive one-second intervals. The participant's task was to tap the buttons in the same order they had earlier lit up. Trial length ranged from two buttons to nine buttons. See *Figure 2.5* for a screen capture of the Corsi Block task administered on the iPad.

Figure 2.5. A Screen Capture of the Corsi Block Task.



The task began at the two-button length, then the trial length iteratively increased by one until reaching the length of nine buttons or exiting the program, whichever occurred sooner. There were three trials for each button length, except for the two-button length, which had four trials. If a participant did not correctly tap all (three) trials of a given button length, the task was stopped. Both the number of buttons correct and the number of trials correct were recorded by the program.

Data output was generated by the program in terms of the number of taps correct per trial and the number of trials correct per participant. The latter measure was used as the dependent variable of the task.

2.4.2. Results

Reliability Estimates

English proficiency – The internal consistency of the English proficiency assessment, as assessed by Cronbach's alpha, was .56.

Analysis Plans

We were most interested in whether the language groups differed in their linear combinations of the EF variables. Secondarily, we aimed to investigate whether bilinguals differing in English proficiency and the amount of language mixing performed differently on the set of dependent variables. The data were analyzed under two approaches: the classic multivariate technique which requires assumptions such as multivariate normality and the variance-covariance matrices being equal between groups to be met prior to conducting the analysis, and a multivariate analysis of variance method based on trimmed means which foregoes such assumptions. Results from both approaches are compared and briefly discussed.

Classic Multivariate Technique

Data for Trails A and B for one monolingual and one bilingual participant were missing due to experimenter error, and were subsequently replaced with the means of their respective groups.

Data Trimming and Outlier Analyses

For tasks of *Selecting* (i.e., flankers and flanking color patches), error trials, trials with an RT below 250 ms, and trials with an RT 3 *SD*s above an individual's mean were eliminated. Similarly for the *Shifting* task of NL, error trials, trials with an RT below 300 ms, and trials with RTs 3 *SD*s above an individual's mean were eliminated. This procedure affected 1.6% of trials for flankers, 0.8% of trials for flanking color patches, and 1.9% of trials for NL.

Overall accuracies of these tasks, although not dependent variables in the study, were also compared between language groups. Both monolinguals and bilinguals achieved 91% or greater on the accuracy of all three tasks, and no significant differences were found between language groups on any of the tasks.

Pertaining to the 10 dependent variables, the influence of extreme scores on RT or accuracy was reduced by replacing observations farther than 3 *SD*s from the mean with values at 3 *SD*s above or below the mean as appropriate for each variable within language group. This procedure affected 2% of the total observations.

Tables 2-1 and *2-2* present pairwise correlations of all dependent variables separately for the two language groups. The directionality of all but the accuracy measures was reversed so that for all measures, higher scores indicated better performance.

Table 2-1. *Pairwise Correlations of the Dependent Variables for Monolinguals (n = 37).*

	Anti	SSRT	TM	LM	NL	Trls	SDR	Corsi	Flkrs	FColors
Anti	-									
SSRT	**.39***	-								
TM	.32*	-.16	-							
LM	.10	.16	**-.11**	-						
NL	.24	-.004	-.01	-.14	-					
Trails	-.02	.08	.08	.20	**-.16**	-				
SDR	.11	.16	-.09	.13	.11	.02	-			
Corsi	.10	-.11	.14	-.02	.20	-.22	**.19**	-		
Flanker	.31[a]	.06	.12	.07	.21	-.15	.10	.15	-	
FColors	.20	.09	.30[a]	-.07	.20	-.14	.18	.19	**.61****	-

Anti – antisaccade. SSRT – stop signal response time. TM – tone monitoring. NL – number letter. Trls – Trails AB. SDR – spatially-delayed response task. Corsi – span task. Flkrs – flankers. FColors – flanking colors.
*significant at the .05 level. **significant at the .01 level. [a] *p*s = .06-.07.

Table 2-2. *Pairwise Correlations of the Dependent Variables for Bilinguals (n = 47).*

	Anti	SSRT	TM	LM	NL	Trls	SDR	Corsi	Flkrs	FColors
Anti	-									
SSRT	**.30***	-								
TM	.44**	-.04	-							
LM	.27[a]	.21	**-.03**	-						
NL	.08	.03	-.15	-.006	-					
Trails	.06	.21	-.02	.17	**.02**	-				
SDR	-.12	-.02	-.03	-.09	.10	.10	-			
Corsi	.21	.40**	.36*	-.04	.008	-.17	**.29***	-		
Flanker	.45**	-.03	.21	.08	.28[a]	-.16	-.21	.04	-	
FColors	.45**	.26[a]	.29*	.05	.12	-.23	.004	.26[a]	**.58****	-

Anti – antisaccade. SSRT – stop signal response time. TM – tone monitoring. NL – number letter. Trls – Trails AB. SDR – spatially-delayed response task. Corsi – span task. Flkrs – flankers. FColors – flanking colors.
*significant at the .05 level. **significant at the .01 level. [a] *p*s = .06-.08.

Visual inspection of *Tables 2-1* and *2-2* reveals the EF variables being generally more correlated with one another in bilinguals than monolinguals. Tasks of *inhibition* (Anti, SSRT) and of *selecting* (Flanker, FColors) were correlated at similar magnitudes in each language group, whereas tasks of *maintaining* (Corsi, SDR) were only significantly correlated in bilinguals. To quantitatively compare the correlation coefficients between language groups, Fisher's r-to-z transformations were first performed, then the z scores compared and analyzed for statistical significance. See *Table 2-3* for a summary.

Table 2-3. *Significance of Difference between Two Correlation Coefficients.*

	Anti	SSRT	TM	LM	NL	Trls	SDR	Corsi	Flkrs	FColors
Anti	-									
SSRT	.66	-								
TM	.56	.57	-							
LM	.46	.82	.71	-						
NL	.46	.89	.54	.57	-					
Trails	.73	.56	.67	.89	.41	-				
SDR	.30	.43	.78	.33	.97	.71	-			
Corsi	.65	.02*	.30	.94	.39	.83	.63	-		
Flanker	.47	.70	.72	.98	.76	.97	.16	.62	-	
FColors	.21	.44	.98	.61	.70	.66	.44	.76	.87	-

*significant at the .05 level.

Because the pairwise correlations between language groups were mostly found not to significantly differ, the data from the language groups were collapsed to form one group. Pairwise correlations of the EF variables were then conducted on all study participants. See *Table 2-4*.

Table 2-4. *Pairwise Correlations of Dependent Variables of EF (N = 84).*

	Anti	SSRT	TM	LM	NL	Trls	SDR	Corsi	Flkrs	FColors
Anti	-									
SSRT	**.33****	-								
TM	.40**	-.07	-							
LM	.20[a]	.18	**-.06**	-						
NL	.14	.02	-.09	-.06	-					
Trails	.05	.18	.04	.17	**-.03**	-				
SDR	-.004	.08	-.02	.004	.11	.12	-			
Corsi	.18	.24*	.29**	-.04	.08	-.15	**.27***	-		
Flanker	.40**	-.004	.17	.08	.26*	-.16	-.10	.07	-	
FColors	.36**	.20[a]	.29**	.01	.14	-.20[a]	.07	.23*	**.59****	-

*significant at the .05 level. **significant at the .01 level. [a] *p*s = .06-.07.

As can be seen in *Table 2-4*, Trails A&B and Letter Memory, compared to the other EF variables, exhibited weak and insignificant correlations with the other dependent variables; thus they were eliminated from subsequent analyses. SDR and Number Letter both showed sparse correlations with the other EF variables as well but warranted different decisions with regards to their inclusion in the data set. Because SDR was found to significantly correlate with Corsi, a task tapping into the same hypothesized EF process as SDR (*maintaining*), in bilinguals and the collapsed group, SDR was retained for further analyses. Number Letter, on the other hand, was eliminated because it exhibited weaker correlations within language group than the two groups combined.

To improve normality of the 7 retained EF variables, accuracy data for Corsi Blocks, Antisaccade, and Tone Monitoring were arcsine transformed and the original RT data for SSRT and Flanking Color Patches were first log transformed then reverse scored, so that for all measures, higher scores indicated better performance. After these transformations, the data showed acceptable skewness and kurtosis (see *Table 2-5* and *Table 2-6*). The reliability estimates for the measures used in this study were derived from computing Cronbach's alpha or the split-half correlations (based on correlations between two subsections or correlations between even-numbered and odd-numbered items), adjusted by the Spearman-Brown prophecy formula. These estimates are also listed in *Table 2-5* and *Table 2-6*.

Table 2-5. *Descriptive Statistics for the Dependent Measures of Monolinguals (n = 37).*

Task	M	SD	Range	Skewness	Kurtosis	Reliability
SDR	-980.56	241.71	-1562 – -446	-.24	.26	.68[c]
Corsi[a]	.79	.15	.37 – 1.08	-.44	.59	.63[d]
Anti[a]	1.07	.15	.73 – 1.42	-.08	.22	.93[c]
SSRT[b]	-2.42	.06	-2.54 – -2.29	-.01	.55	n/a
TM[a]	1.00	.31	.40 – 1.57	.33	-.35	.68[c]
Flkrs	-489.75	42.95	-585.72- -402.96	-.37	.32	.95[c]
FColors[b]	-2.59	.04	-2.69 – -2.51	-.29	-.57	.88[c]

Table 2-6. *Descriptive Statistics for the Dependent Measures of Bilinguals (n = 47).*

Task	M	SD	Range	Skewness	Kurtosis	Reliability
SDR	-1101.09	228.42	-1721.19 – -685	-.63	.18	.55[c]
Corsi[a]	.74	.17	.37 – 1.08	-.14	-.55	.73[d]
Anti[a]	1.04	.17	.64 – 1.42	-.42	.07	.96[c]
SSRT[b]	-2.43	.07	-2.60 – -2.23	.30	1.33	n/a
TM[a]	.94	.33	.27 – 1.57	.22	-.22	.73[c]
Flkrs	-485.73	58.29	-622.83 – -378.22	-.46	-.34	.95[c]
FColors[b]	-2.58	.05	-2.70 – -2.50	-.73	-.43	.91[c]

[a] – arcsine transformation is applied the accuracy data. [b] – log transformation is applied to the RT data. [c] Reliability was calculated by adjusting split-half correlations with the Spearman-Brown prophecy formula. [d] Reliability was calculated using Cronbach's alpha. n/a – only one output was generated by the computer program as an average of the three test blocks.

Then a MANOVA examined the 7 EF tasks as the dependent variable and language group as the independent variable. Using Wilks's statistic, there was not a significant effect of language group on the linear combination of DVs, $\Lambda = .09$, $F\ (7, 76) = 1.07$, $p = .39$.

To examine whether bilinguals differing in English proficiency differed on the performance of DVs, a median split was conducted on English proficiency; one individual with the median score on the proficiency measure was excluded from the analysis. The multivariate test revealed that low-proficient bilinguals performed similarly to high-proficient bilinguals, $\Lambda = .18$, $F\ (7, 38) = 1.16$, $p = .35$.

Similarly, bilinguals differing in the frequency with which they mixed two (or more) languages were compared on the set of DVs. A median split was conducted on mixing frequency; three individuals with the median score on the mixing frequency measure were excluded from the analysis. Using Wilks's statistic, there was not a significant effect of mixing frequency on the linear combination of DVs, $\Lambda = .09$, $F\ (7, 35) = .51$, $p = .824$.

See *Table 2-7* in Appendix B for task-specific *t*-tests conducted between language groups.

Robust Multivariate Method

Arcsine and log transformations were performed on several dependent variables in order to meet the basic assumptions of MANOVA, including achieving multivariate normality within groups and variance-covariance matrices of the groups being equal. A more preferred approach to transforming the data though would be using modern robust methods, which are designed to maintain Type I error rate and statistical power relatively well when classic assumptions are either met or violated. All subsequent analyses were conducted in *R* with functions from the WRS library (Wilcox & Schönbrodt, 2014).

First, a projection-type method was implemented in *R* using the *outproad()* function (Wilcox, 2012) to detect outliers in high-dimension data sets, i.e., more than 9 variables, in each of the language groups. One and two outliers were identified in monolinguals and bilinguals, respectively.

After removing these outliers, Yanagihara and Yuan's (2005, as cited in Wilcox, 2012) heteroscedastic approach to comparing the means of two groups was extended to 20% trimmed means. The analysis was implemented in *R* using the *YYmanova()* function. The language groups were found not to differ in their performance of the 10 EF variables, $t = .90$, $p = 0.54$.

Alternatively, a linear contrast of the independent groups was conducted on 20% marginal trimmed means using a percentile bootstrap method, implemented in *R* using the *linconSpb()* function (Wilcox, 2012), which takes into account the overall structure of the data within each language group. The result converges with the Yanagihara and Yuan method extended to 20% trimmed means; performance of the groups did not significantly differ from each other, $p = 0.63$.

After comparing between language groups, bilinguals were further divided into groups of high and low English proficiency via a median split procedure; one individual at the median score on the proficiency measure was excluded from the analysis. Bilinguals high on English proficiency were found not to differ significantly from those low on English proficiency, $T = 1.00$, $p = 0.48$.

Similarly, a median split procedure was conducted on a quantified measure of language mixing to differentiate bilinguals who mixed their two (or more) languages often from those who did not; three individuals at the median score on the language mixing measure were excluded from the analysis. Bilinguals high on language mixing were found not to differ significantly from those low on language mixing, $T = 1.19$, $p = 0.38$.

In summary, the classic multivariate method and the robust method arrived at the same conclusions with regards to performance on the EF variables between language groups, between high-proficient and low-proficient bilinguals, and between high-mixing and low-mixing bilinguals. The convergence of findings affords us confidence that the approach chosen for analysis (classic or robust) did not differentially influence the pattern of results that was observed.

2.4.3. Discussion

We compared the performance of monolingual and bilingual young adults on 10 EF tasks purportedly tapping into processes of *inhibition*, *shifting*, *updating*, *maintaining*, and *selecting* using both classic and robust methods for multivariate analysis. Results from these methods converged in demonstrating that the two language groups did not show a difference on the combination of EF variables. In addition, we examined bilinguals of high vs. low levels of English proficiency and bilinguals of high vs. low degrees of language mixing on the same dependent variables, and concluded that the assessed central executive functions did not differ for higher vs. lower levels of English proficiency or higher vs. lower degrees of language mixing. We next consider these outcomes in relation to the participant sample we tested, and also the dependent measures and analyses approaches we adopted, before returning to the overarching question of how executive function is influenced by the early acquisition and lifelong use of two or more languages.

On Participants

As the discussion pertaining to the study participants applies to both the current chapter and the following two chapters on language-group comparisons of conflict monitoring and of creative performance, consideration of the possible role of other participant differences between the language groups will be taken up in Chapter 5.

On Dependent Variables

The dependent variables chosen for the study emerged from two theoretical frameworks of EF. The tasks showed a range of reliability estimates, from 0.63-0.95 for monolinguals and from 0.55-0.96 for bilinguals. While some of these reliability values may seem wanting, they are in fact comparable to the reliability estimates of the dependent measures in Miyake et al. (2000). The English proficiency assessment showed a low reliability of 0.56. The inclusion of five executive processes of two tasks per process constituted an extensive test battery unparalleled in the bilingualism literature. At the same time, though, the large number of processes precluded our having more tasks per process, making analyses such as latent variable analysis unviable.

Interestingly, the EF variables were found to more strongly correlate with one another in bilinguals than in monolinguals.

When the EF variables were examined separately, monolinguals were found to outperform bilinguals on SDR and Trails A&B (not reported here but see Appendix B for further details), which were maintaining and switching tasks, respectively. One advantage of administering a battery of EF tasks instead of only two or three is that a general pattern can be deduced from performance on the task set rather than a couple of seemingly random and isolated tasks.

On the two *selecting* tasks of flankers and flanking color patches, the overall RT was used as the dependent variable, which was arguably a crude measure of selecting. A better indicator may have been the difference between mean RT of a pure block of neutral trials (that involve no conflict) and the averaged RT of congruent and incongruent trials in a block that mixes both types of trials together. Because our task setup did not have neutral trials appear in a stand-alone block and the number of neutral trials was small (7) relative to the number of congruent (42) and incongruent (42) trials, no effort was made to separate the trial types in calculating the RT.

Given that our primary interest pertained to performance on the linear combination of all dependent variables, relations of the dependent variables within themselves were not investigated. If such analyses were to be carried out, a confirmatory factor analysis could be first conducted to verify the existence of the five hypothesized processes. If the five-factor model was not supported, an exploratory factor analysis could be performed to identify the number of separable cognitive processes represented in the data.

On Analysis Methods

In the current study, we used both a classic multivariate technique and a modern robust method to compare performance on the dependent variables between the two language groups as well as within bilinguals of varying degrees of English proficiency and language mixing. The two different analysis methods reached the same conclusion: monolinguals and bilinguals were not found to differ in their performance on the EF variables. Furthermore, bilinguals with high vs. low English proficiency did not significantly differ in performance, nor did bilinguals with high vs. low amounts of language mixing.

In lieu of simple group comparisons, individual-differences approaches pertaining to participant characteristics as well as dependent variables can be employed. Specifically, participant characteristics such as second language (L2) proficiency, relative language dominance, and whether language mixing is normative (Titone, Pivneva, Sheikh, Webb, & Whitford, 2015) can be adopted separately or in combination to identify factors that differentiate bilinguals' performance from monolinguals' and from each other. In step with the EF research tradition, latent variable analyses could be performed on the dependent variables adopted in bilingualism research to uncover underlying task structures between and within language groups. The field could also benefit from more widespread use of modern robust methods, which do not assume specific shapes of variable distributions a priori yet also maintain satisfactory levels of Type I error rate and power.

On the Ultimate Question

Our primary interest lies in the question of whether bilinguals have an advantage over monolinguals on EF functioning. Data from the current study indicate the lack of such an advantage; not only did bilinguals fail to outperform monolinguals on the EF processes and tasks, separate task-specific analyses revealed a reversed pattern (not reported here but see Appendix B for further details), such that monolinguals showed superior performance on a maintaining task (spatial delayed response) and a switching task (Trails A&B). However, monolinguals' observed advantage was restricted to the specific tasks and cannot be extended to the respective EF processes.

The general lack of a bilingual advantage is consistent with the summary results of Paap and Greenberg (2013) who found the language groups to perform similarly on all 15 indicators of executive processing but one, on which monolinguals performed better. Similarly, de Bruin, Treccani, and Della Sala (2015) found that a language-group difference only emerged in one of 4 tasks, on which bilinguals performed better.

Our findings are consistent with the most recent meta-analysis reported by Paap, Johnson, and Sawi (2015); significant bilingual advantages found in the literature on the interference effect, switch cost, mixing cost, and monitoring since 2011 did not occur with large-*n* studies but appeared mostly when the average number of participants per group was small (n < 30). This observation is a cause for concern because as the sample size grows, researchers should be increasingly confident of their ability to reject the null when the null is false, unless of course the null is true (Rouder, Speckman, Sun, Morey, & Iverson, 2009).

Some researchers caution against a simple binary approach to comparing language groups (Mishra, 2015; Titone et al., 2015). Instead, the research strategy of comparing different levels of bilingualism on the dimensions of proficiency, switching frequency, etc. is advocated particularly in linguistic contexts where "pure" monolinguals do not exist. On the one hand, a more fine-grained approach of this sort has the potential of yielding rich and detailed information about the effects of various types of bilingual experiences. On the other hand, the lack of a clear-cut group difference speaks to limits on the scope and/or magnitude of the advantage that may be conferred by bilingualism-related experiences. If an advantage can only be realized under very specific and limiting conditions, the practical significance of these experiences would undoubtedly be diminished. Studies probing parental reasons for raising children in a bilingual environment identified numerous factors as important considerations, ranging from the likelihood of better career opportunities for their offspring in a competitive global market, to the development of communication skills with family relatives, to the strengthening of cultural and personal identity (Mosty, Lefever, & Ragnarsdóttir, 2013; Schecter, Sharken-Taboada, & Bayley, 1996). It still remains to be seen whether general cognitive development becomes a more prominent factor as the findings in this area continue to accumulate.

Chapter 3

The Probe on Conflict Monitoring

3.1. Introduction

The meta-analytic review of Hilchey and Klein (2011) identified a global response time (RT) advantage in bilinguals. It was therefore proposed that bilingualism was associated with a *general* processing advantage that was realized in the superior performance of *both* congruent and incongruent trials, rather than enhanced inhibitory control per se, i.e., the *difference* in RT for incongruent relative to congruent trials. They further suggested that investigating such a general processing advantage within the conflict monitoring theory would be both intuitive and potentially fruitful—provided that due attention was devoted to competing theories and relevant task paradigms.

The classic conflict monitoring theory identifies the anterior cingulate cortex (ACC) in the human frontal lobe as a domain-general neurocognitive system for modulating trial-by-trial control. Upon detecting intra-trial conflict, the ACC passes this information on to centers in the brain responsible for cognitive control, causing them to intervene more strongly in processing the subsequent trial (Botvinick, Braver, Barch, Carter, & Cohen, 2001). A later extension of the original theory went beyond intra-trial conflict to implicate the ACC also in encoding information about mental effort, or that the presence of conflict simply reflects the extent to which the task is cognitively demanding (Bush et al., 2000; Botvinick, Cohen, & Carter, 2004).

Conflict monitoring as described shares tenuous connections with the bilingual advantages reported in the literature. While an accumulating number of studies utilizing brain activation techniques make direct or indirect references to the role of the ACC, most of these (as will be reviewed below), were conducted with conflict monitoring neither as a grounding theory nor the focus of the inquiry. Furthermore, the findings were not entirely in accordance with the aforementioned meta-analysis (Hilchey & Klein, 2011), nor did they necessarily agree with one another. Behaviorally, experiments using tasks developed out of the conflict monitoring theory to investigate monolingual and bilingual differences are virtually non-existent.

To address this theoretical and empirical gap, we here present the results of three behavioral tasks from the perspective of conflict monitoring theory, in each of which we contrast the performance of monolingual and bilingual participants. One of these tasks—a four-choice sequential modulation task—was explicitly adopted to allow the empirical disambiguation of the contributions of conflict monitoring from a theoretical alternative that focuses, instead, on the role of episodic binding of stimulus and response features or "feature integration" on a trial-by-trial basis. The other two tasks were variants of the Simon task developed to examine performance under varying levels of difficulty: a standard Simon task, where we include block types with varying degrees of difficulty (as indexed by the overall RT and Simon effects of different magnitudes) and a version of Simon using arrows as the stimuli. The arrow Simon task we used had the same block and trial setup as the standard Simon task, but is more perceptually demanding overall. For each task, we consider performance both when language groups are assigned categorically and when the bilinguals are assigned to high-proficient and low-proficient groups on the basis of self-report and objective measures of language proficiency (listening, speaking, reading, and writing).

We elaborate on the rationale and empirical approach we adopted for each of the three tasks and the questions they were designed to address below. But first we provide a brief overview of the limited findings, to date, examining conflict monitoring in bilingualism.

3.2. Conflict monitoring in bilingualism

To our best knowledge, the only behavioral experiment to date bearing a connection to conflict monitoring in bilingualism was conducted by Costa, Hernández, Costa-Faidella, and Sebastián-Gallés (2009). The researchers had speculated in their earlier work (Costa, Hernández, & Sebastián-Gallés, 2008) whether bilinguals' faster performance on congruent trials was due to their superior conflict monitoring processes compared to monolinguals. In a follow-up experiment using a cued flanker task from the Attention Network Task (ANT), the proportion of congruent versus incongruent trials was systematically varied to create the so-called high-monitoring and low-monitoring contexts. Both the condition with more congruent trials and the condition with more incongruent trials were considered lower in monitoring demand than the condition with an equal number of congruent and incongruent trials, as the latter required a continual monitoring of every trial in order to respond appropriately. Results supported the presence of a trial-type frequency effect (Botvinick et al., 2001), which is the dependence of the magnitude of interference effect on the frequency of incongruent trials, with less interference found under conditions of more frequent incongruent trials. The trial-type frequency effect had been reported in a number of experiments using the Stroop task and was an empirical phenomenon for which the conflict monitoring theory had accounted. In addition, bilinguals were found to outperform monolinguals on the overall RT only in the high-monitoring condition.

A number of studies have used neuroimaging techniques to examine aspects of conflict monitoring in bilingualism. Using magneto-encephalography (MEG), Bialystok et al. (2005) examined the neural correlates of Simon task performance in two bilingual groups (Canton-English and French-English) and one monolingual group. Generally, fast responding in bilinguals was associated with greater activation in cingulate, superior frontal, and inferior frontal regions whereas faster responding in monolinguals was associated with middle frontal activity; activations in both groups were lateralized to the left hemisphere. Slower responding was signaled by more activation in the visual cortex in bilinguals and more activation in the motor cortex in monolinguals. The notable result here was the recruitment of left hemisphere regions in the inferior frontal cortex that bordered on language centers for faster responding by the bilinguals, including the cingulate areas. Behaviorally, only one of the two groups of bilinguals (the Canton-English bilinguals) performed the task faster than the English-speaking monolinguals, with the other group (French-English bilinguals) performing similarly to the monolinguals.

Using functional magnetic resonance imaging (fMRI), Luk, Anderson, Craik, Grady, and Bialystok (2010) investigated neural correlates of the response inhibition and interference suppression aspects of cognitive control in monolingual and bilingual young adults. Brain activations of congruent, incongruent, and no-go trials in a modified flanker task were compared. Results showed that the same brain network was recruited by the language groups for congruent and no-go trials, but different networks were recruited for incongruent trials. Specifically, the network recruited by bilinguals included similarly activated regions as the language control network—left prefrontal cortex, left ACC, left caudate nucleus, and bilateral supramarginal gyri—for all trial types, whereas increased activity in the left temporal pole and left superior parietal lobe was associated with incongruent trials for monolinguals. Overall RTs between language groups did not differ. Central to our interest here is the recruitment of left ACC by bilinguals, and not monolinguals, for incongruent trials.

Also using fMRI, Garbin et al. (2010) investigated executive control in monolinguals and bilinguals with a non-verbal task-switching paradigm. Behavioral results showed that monolinguals exhibited a larger switching cost—that is, a larger difference between switch and non-switch trials—in terms of both RT and accuracy, with the overall RT and accuracy not differing between language groups. Relative to non-switch trials, monolinguals had greater neural activity in right inferior frontal gyrus, ACC, and left inferior parietal lobe for switch trials, whereas bilinguals showed greater neural activity in left inferior frontal gyrus for switch trials, which was also involved in bilingual language control. The differential patterns of activation on switch trials were taken to indicate that bilinguals' language control and non-linguistic cognitive control shared certain cortical areas.

Complementary to the high spatial resolution provided by fMRI, event-related potentials (ERPs) are extracted from the ongoing electroencephalograph and have excellent temporal resolution on the order of milliseconds. The ERP components of N2 and error-related negativity (ERN), both of which are thought to correlate with ACC activity measured by fMRI, were assessed in monolingual and bilingual young adults as they performed the Stroop, Simon, and flanker tasks (Kousaie & Phillips, 2012a). The N2 component peaks approximately 200-350 ms following a stimulus; it is thought to reflect the ACC activation prior to the response on *correct* conflict trials. The ERN, on the other hand, is a negative-going ERP that peaks approximately 100-150 ms after an error and has been suggested to reflect the ACC activation immediately following the response on *incorrect* conflict trials. The researchers predicted larger amplitudes of N2 and ERN for bilinguals to reflect their greater activity in the ACC, but the results showed both components to be in fact larger for monolinguals on the Stroop task and not to differ between language groups on Simon or flankers. Analyses of accuracy and RT data revealed no language group differences behaviorally on any of the tasks.

Perhaps the study most directly motivated by the conflict monitoring theory was conducted by Abutalebi and colleagues (2012). These researchers used a combined functional and structural neuroimaging design, comparing monolingual and bilingual performance on a flanker task and a linguistic switching task in the same experiment. The switching task involved within-language production of nouns or verbs for monolinguals and between-language picture naming for bilinguals. Results showed both language control and general cognitive control as assessed by the conflict effect on a flanker task engaged the dorsal ACC. Bilinguals exhibited less ACC activity than monolinguals across both sessions of the study while also showing a reduced conflict effect functionally and behaviorally in the second session. No language-group difference in the overall RT, however, was found.

Most of the behavioral outcomes of the aforementioned studies did not reveal a language-group difference in the overall RT; the ones that did involved either a selected condition (i.e., the high-monitoring condition) or a subgroup of bilinguals (i.e., Cantonese-English rather than French-English). One study focusing on the interference effect and another study on the switching cost demonstrated a bilingual advantage. These findings do not cohere well with the meta-analytical conclusion of a global RT advantage for bilinguals, although a reduced conflict effect was occasionally observed. In terms of neural activity, the reported brain regions in bilinguals that are common to language control and cognitive control processes differed between Abutalebi et al. (2012) and Garbin et al. (2010), i.e., dorsal ACC and left IFG, respectively. The former study empirically compared performance of language switching in bilinguals with their demonstrated conflict effect on a flanker task, whereas the latter study compared regions of previously identified language control centers in bilinguals with their demonstrated switch cost on a non-verbal switching task. In addition, bilinguals' unique recruitment of the ACC for incongruent trials on a flanker task (Luk et al., 2010) does not appear consistent with their smaller amplitude of the N2 component for correct incongruent trials on a Stroop task (Kousaie & Phillips, 2012a).

Now turning to the extension of the classic conflict monitoring theory that posits the ACC as assuming a role in indexing cognitive demand, empirical evidence seems to point to a possible connection between a bilingual processing advantage and this extension of the conflict monitoring theory. A key result of the aforementioned Costa et al. (2009) study was that bilinguals outperformed monolinguals in overall RT only in the high-monitoring condition, which was also the most cognitively demanding condition. These results broadly align with outcomes found by Bialystok (2006) contrasting two Simon tasks with colored squares and arrows as stimuli. Both Simon tasks included a low-switch condition and a high-switch condition, with the difference between the conditions being the number of times the required response switched. The arrow task and the high-switch condition proved to be the more difficult of the two tasks and the two conditions. The finding most pertinent to our current discussion was that on the most demanding condition of all, i.e., the high-switch condition of the arrow task, bilinguals responded more rapidly than monolinguals whereas the groups performed equivalently on the low-switch condition of the same task. In other work, this same research group used a modified version of the squares Simon task in which the stimuli were presented centrally instead of to the side but cognitive load was increased by changing the number of stimulus-response mappings from two-to-two to four-to-two. In this particular task, bilinguals were shown to perform faster than monolinguals in the first seven of ten testing blocks (Blocks 1–7) whereas the language groups' performance converged in the final three blocks (Bialystok et al., 2004).

A bilingual advantage has also been reported in conditions of a saccadic eye-movement task consisting of both single trial-type and mixed trial-type blocks. Task trials began with a schematic drawing of a human face containing blank eyes, which would turn green or red for 500 ms, indicating the type of trial to perform, i.e., prosaccade or antisaccade, respectively. Then an asterisk would flash briefly either to the left or right of the face and participants were required to press the *shift* key on the same side of the asterisk for a prosaccadic trial and the *shift* key on the opposite side for an antisaccadic trial. In addition, the schematically depicted eyes could look straight ahead or toward the left or right, thus making the gaze direction relevant (green eyes looking towards the asterisk or red eyes looking away from the asterisk) or irrelevant (green eyes looking away from the asterisk or red eyes looking towards the asterisk). The most difficult condition in the entire task consisted of the irrelevant eye gaze manipulation on saccadic trials in a mixed block. Bilinguals at both the young adult and older adult levels were found to be less affected by the misleading gaze direction than were monolinguals (Bialystok, Craik, & Ryan, 2006).

Taken together, these studies provide comparatively stronger convergent evidence for a possible connection between a bilingual processing advantage and the extended version of conflict monitoring theory, in which the ACC assumes a role in indexing cognitive demand.

3.3. Feature integration theory and the sequential modulation task

Our discussion of conflict monitoring theory has thus far focused on how this theory accounts for participants' performance on a given trial, particularly the effects of stimulus-response congruency and broader effects of task difficulty. However, the conflict monitoring theory also provides a cogent account for the empirically validated phenomenon of first-order sequencing effects, or the effects of trial-to-trial (rather than only within-trial) congruency or incongruency. Originally reported in a traditional interference task of the letter flanker task (Gratton, Coles, & Donchin, 1992), the most crucial finding from the study was that performance on the current trial was affected by the congruency (or incongruency) of the previous trial. Specifically, an incongruent trial was typically performed faster, and more accurately, if it followed *another incongruent* trial than if it followed a *congruent* trial. That is, there was an interaction of the previous trial type (congruent or incongruent) and the current trial type (congruent or incongruent). The conflict monitoring theory explains the interaction effect in this way: the detection of conflict by the ACC at the presentation of an incongruent trial triggers an up-regulation in the cognitive control centers, thus facilitating performance on the following trial, whether it is congruent or incongruent. When presented with a congruent trial, however, the ACC is not activated nor are the centers of cognitive control recruited, thus incurring a RT cost when the following trial is incongruent.

An alternative explanation of these first-order sequencing effects is provided by feature integration theory (Hommel, 1998, 2004). This theory focuses on how particular features of a stimulus and response on one trial may be integrated or bound together in an episodic memory record, and the potential carry-over effects of such integration to subsequent trials. Feature integration theory evolved out of earlier accounts of the spontaneous binding of object features in perception (Kahneman, Treisman, & Gibbs, 1992) to a broader account involving the binding or integration of *both* stimulus features and responses—thus the theory is sometimes designated with the more inclusive term “event files” (Hommel, 1998). On the event file account, once a binding occurs, the activation of one feature activates the rest of the features. Perceiving or acting upon a new combination of features that are partially overlapping with the existing set would incur a RT cost, because the existing features would first need to be unbound and then bind with the new features. The feature integration account explains the interaction effect of congruency for a current trial and the subsequent trial (found in traditional Flanker or Simon interference tasks) in this way: following a congruent trial, a congruent trial is either a complete repetition or a complete alternation of the previous trial (which leads to faster responses), whereas an incongruent trial is always a partial repetition/alternation of the previous trial (which leads to slower responses), thus increasing the RT difference between trial types on the subsequent trial. Following an incongruent trial, however, a congruent trial is a partial repetition/alternation from the previous trial and an incongruent trial entails a complete repetition/alternation of the previous trial, thus reducing the RT difference between trial types.

Traditional interference tasks, which are commonly used in research investigating broader cognitive effects of bilingualism, cannot arbitrate between these two theoretical accounts (Egner, 2007). In a two-value (two stimulus forms, two stimulus locations, and two responses) forced-choice task such as the Simon task and the flanker task, the effects of congruency and of feature overlap are perfectly confounded (Wuhr, 2005). Specifically, the second trial in a congruent-congruent (CC) or incongruent-incongruent (II) sequence is either a complete-repetition (same stimulus features, same response) or a complete-alternation (different stimulus features, different response) of the first trial. And the second trial in an incongruent-congruent (IC) or congruent-incongruent (CI) sequence involves only partial repetitions (some stimulus features repeat or the response repeats). So it remains unclear if a bilingual advantage, if observed on a traditional interference task, is fully attributable to conflict monitoring or to differences in the formation (or dissipation) of ongoing episodic memory representations of stimulus and response features that are dynamically bound together.

To directly contrast the predictive adequacy of the conflict monitoring account with the feature integration or event file theory in the context of bilingualism research, we adopted a four-choice sequential modulation task developed by Akçay and Hazeltine (2007) to better disambiguate the effects of congruency and feature overlaps. We separately examine evidence for conflict monitoring and for feature integration in college students who are monolingual versus individuals who are fluent in two or more languages.

3.4. The standard Simon task

The standard Simon task evolved out of an auditory task involving low-pitched or high-pitched tones presented to either the left ear or the right ear, in which participants indicated if the tone was high or low in pitch using left or right key presses (Simon & Small, 1969). Responses were faster when tones were presented on the side consistent with the designated stimulus-response (S-R) mapping; for example, responses were faster when low-pitched tones corresponded with the left key press and were presented to the left ear, than contrary to it (e.g., if low-pitched tones corresponded with the left key press and were presented to the right ear). The visual version of the task uses stimuli that differ on a variety of relevant stimulus dimensions such as color, letters, and geometric shapes (Lu & Proctor, 1995). For instance, a patch of blue may be associated with a left key press and a patch of red with a right key press. When the stimulus appears on the same side of the computer screen as the correct response key, the stimulus location and response position converge and the trial is considered congruent. When the stimulus appears on the opposite side of the correct response, the trial is incongruent. Executing a response on incongruent trials usually takes 20–30 milliseconds longer than congruent trials in a standard Simon task (Lu & Proctor, 1995), as a temporal cost is incurred for resolving the positional conflict between the stimulus and the correct response (Martin-Rhee & Bialystok, 2008). This difference in RT between congruent and incongruent trials is called the Simon effect.

The Simon effect serves as an index of the degree of conflict. In a parametric manipulation of congruency proportions in the Simon task (Stürmer, Leuthold, Soetens, Schröter, & Sommer, 2002), Simon effects from the 80% congruent (20% incongruent), 50% congruent (50% incongruent), and 20% congruent (80% incongruent) conditions were observed that were ranked in an order of significantly reduced magnitudes. So it appeared that the magnitude of the conflict positively corresponded with the proportion of congruent trials. Similarly, in a cued flanker task, the conflict effect was found to be larger in the 75% congruent condition than in the 50% congruent condition (Costa et al., 2009). The metric of overall RT is usually used as an index of condition or task difficulty; more challenging tasks generally take longer to complete than easier ones. No difference in RT was found to exist between the 50% congruent and 75% congruent conditions of the aforementioned cued flanker task, although the researchers had conceived the 50% condition to be higher in monitoring demand than the 75% condition (Costa et al., 2009). For variations of the 50% congruent condition, the condition with fewer inter-trial response switches was performed faster than the condition with more inter-trial response switches, while the error rates between the switch conditions did not differ (Bialystok, 2006). Taken together, when the magnitude of the Simon effect is used to index task difficulty, the condition with the largest proportion of congruent trials presents the strongest conflict. If, however, the overall RT is used to index task difficulty, the high-switch condition of the 50% congruent condition would likely demand the greatest mental effort.

In accordance with these findings, we constructed a standard Simon task with different degrees of conflict and overall RTs across four experimental conditions: 20% congruent, 80% congruent, 50% congruent low-switch, and 50% congruent high-switch. If a bilingual advantage is most evident in more cognitively demanding conditions, bilinguals, compared to their monolingual peers, would be expected to show a smaller Simon effect in the 80% congruent condition and a reduced overall RT in the high-switch condition of the 50% congruent condition.

3.5. The Arrow Simon task

Stimuli in the arrow Simon task consist of left- or right-pointing arrows, where the arrow position on the screen either matches or mismatches the spatial position of the response keys, thereby creating congruent and incongruent trials respectively. The arrow Simon notably differs from the standard Simon task in that the stimuli in arrow Simon possess inherent S-R mappings (i.e., arrows themselves symbolically indicate directions), whereas the stimuli for the standard task are typically letters (e.g., X and O) or different colored squares, with no associated inherent directionality. In particular, arrow Simon has been found to elicit greater perceptual conflict and longer overall RTs (Bialystok, 2006). The arrow Simon task included in our study is an exact replication of the standard Simon task described earlier, except with arrows as stimuli.

The inclusion of both standard Simon and arrow Simon allows a direct comparison of conditions with varying degrees of difficulty within task as well as between tasks.

In sum, we aim to evaluate the following questions: (1) Do bilinguals (relative to non-bilinguals) demonstrate an RT or accuracy advantage according to the classic account of conflict monitoring theory or feature integration? (2) Do bilinguals (relative to non-bilinguals) show an RT or accuracy advantage in tasks/conditions that are especially cognitively demanding? (3) Do high-proficient and low-proficient bilinguals show systematic differences in conflict monitoring or feature integration depending on how English proficiency is measured?

3.6. *Method*

Stimulus Materials

Sequential Modulation

This four-choice task was adopted from Akçay and Hazeltine (2007), with a few modifications. The primary stimuli were four colored squares (red, green, blue, and yellow), measuring 30 mm in length and presented on a white screen one at a time. Each color-square could appear within any of four open black squares (boxes) horizontally displayed on screen, with 45 mm between each box. Four response keys (d, f, j, and k) were mapped onto the four colors, respectively, and participants were instructed to respond using the index and middle fingers of their right and left hands, with their fingers lightly resting on the appropriate keys in the “typing” position. When the response position aligned with the stimulus position, it was a congruent trial; otherwise it formed an incongruent trial. Twelve blocks of test trials with 97 trials per block were preceded by 32 practice trials. The congruent and incongruent trials were pseudo-randomized in each block such that no more than six trials of the same type appeared consecutively. The last trial of each block also served as the first trial of the following block, which was discarded from data processing and analysis.

For each trial, the black open squares were displayed for 500 ms, followed by the presentation of a color-square, which remained until the subject responded, or up to 2000 ms. An automated 9-second break separated the blocks.

Standard Simon

In the Standard Simon task, a blue square or a red square was presented on either the left side, or the right side, of the computer screen. The stimulus square measured 16 mm in length, and the distance from the central fixation to the inner edge of the square was 165 mm. Participants responded to the blue square by pressing the *z* key and to the red square by pressing the */* key on a standard keyboard. When the stimulus location and response position matched, it was a congruent trial; otherwise it was an incongruent trial. The stimulus-response mapping was counterbalanced across participants.

Eight practice trials preceded the presentation of 16 test blocks, which consisted of 8 congruent-and-incongruent-trial intermixed blocks and 8 congruent-or-incongruent-trial-only mini-blocks. There were four congruency-proportion conditions for the mixed blocks: equal number of congruent and incongruent trials with a high-switch frequency, defined as having trial congruency and response position switch more than 25 times in a 40-trial block; equal number of congruent and incongruent trials with a low-switch frequency, defined as having trial congruency and response position switch fewer than 15 times in a 40-trial block; more congruent trials (80%); and more incongruent trials (80%). Similarly, there were four congruency-response-mappings for the single-trial-type mini-blocks: congruent-left, congruent-right, incongruent-left, and incongruent-right. The four congruency-proportion conditions and the four congruency-response-mapping conditions were each repeated once to yield a total of 16 blocks. The mixed blocks and the single-trial-type mini-blocks alternated in the order of presentation beginning with a mixed block, and the congruency-proportion conditions and the congruency-response-mappings were semi-randomized within their block types and presented in a fixed order.

Mixed blocks had 40 trials per block and single-trial-type mini-blocks had 10 trials per block. Each trial began with a central fixation, displayed for 800 ms, followed by the presentation of a blank screen for 250 ms. Then the stimulus appeared until the participant responded, or up to 1500 ms. The inter-trial interval displayed a central fixation for 500 ms. The blocks were separated by an automated 9-second break.

Arrow Simon

In the arrow Simon task, a left- or a right-pointing arrow was presented on either the left side, or the right side, of the middle of the screen. The arrow, shown in red, was 31 mm in length. The arrowhead was formed by two sides of an equilateral triangle with a side of 11 mm. The distance from the central fixation to the inner edge of the arrow stimulus was 15.8 mm horizontally. Participants responded to the left-pointing arrow by pressing the *z* key and to the right-pointing arrow by pressing the */* key on a standard keyboard.

When the arrow direction and response position matched, it was a congruent trial; otherwise it was an incongruent trial. Trial and block setup were the same as for the standard Simon task.

3.7. Results

3.7.1 English Proficiency

The language groups differed on objectively measured English proficiency (OEP), with monolinguals (*M* = 23.27, *SD* = 3.97) scoring higher than bilinguals (*M* = 21.46, *SD* = 3.68), t (82) = 2.16, $p = .04$. In contrast, the groups did not differ on self-reported English proficiency (SEP), with monolinguals (*M* = 8.99, *SD* = 1.18) and bilinguals (*M* = 8.88, *SD* = 1.19) reporting similar levels of proficiency, t (82) = .43, $p = .67$. These data reveal a clear difference between using objective English proficiency and self-report English proficiency to determine the comparability of English ability between language groups.

The OEP correlated with SEP at r (37) = .29, $p = .08$ for monolinguals and at r (47) = .57, $p < .001$ for bilinguals. The difference between these correlations approached significance, Z = -1.53, $p = .06$. Bilinguals tended to be better than their monolingual peers at assessing their language proficiency.

3.7.2 Sequential Modulation

Individuals with an overall task accuracy three *SDs* from their respective language group means or more extreme were excluded. This led to the exclusion of data for 6 monolinguals and 3 bilinguals, resulting in 31 monolinguals and 44 bilinguals for all subsequent analyses on the sequential modulation task unless otherwise noted. The means and standard deviations of task accuracy post-exclusion were .914 (.04) and .897 (.05) for monolinguals and bilinguals, respectively. This difference was not significant, t (73) = 1.71, $p = .09$.

For the RT analysis, error trials, trials immediately following an error trial, and correct trials less than 250 ms or three or more *SD*s above a given individual's RT mean were excluded. This led to the elimination of 1402 correct trials, or 1.4% of all trials.

Congruency effects and feature overlap effects are examined in turn, first including all of the data, then on subsets of the data that allowed analytical examination from the perspectives of the conflict monitoring versus feature integration accounts.

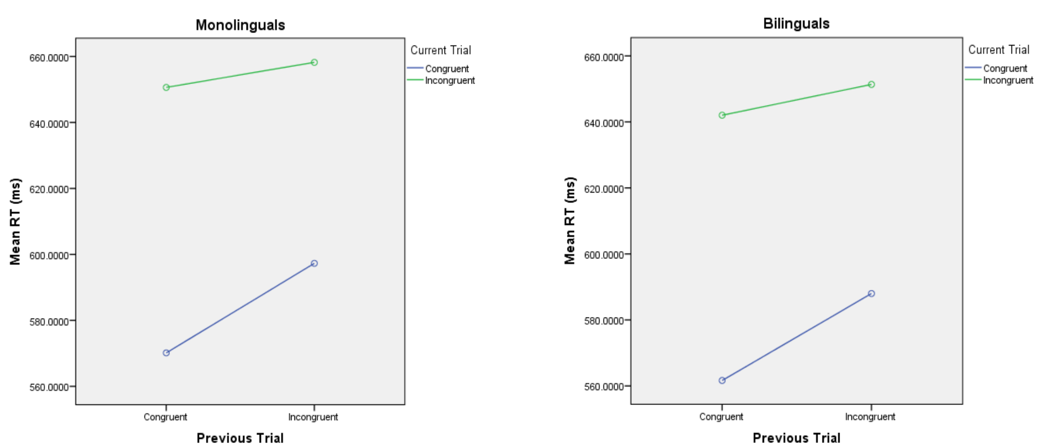
Congruency Effects

A 2 x 2 x 2 mixed-factor ANOVA was conducted on RT, including the factors of *previous trial* congruency (congruent/incongruent), *current trial* congruency (congruent/incongruent), and language group (monolingual/bilingual). Mean RTs and accuracies with respect to previous trial congruency and current trial congruency are shown in *Table 3-1*. The main effects of previous congruency and current congruency were both significant, F (1, 73) = 127.86, $p < .001$ and F (1, 73) = 960.10, $p < .001$; and so was the interaction, F (1, 73) = 33.91, $p < .001$. The three-way interaction involving language group was not significant, F (1, 73) = .26, $p = .61$. See *Figure 3.1* for a visual illustration of this relationship exhibited by the RTs.

Table 3-1. *Mean Response Time (ms) and Accuracy with Respect to Previous Congruency and Current Congruency Shown by Language Group (nML = 31, nBL = 44) in the Sequential Modulation Task.*

Previous	**Current**	**ML RT**	**BL RT**	**ML ACC**	**BL ACC**
Congruent	Congruent	570.14 (14.75)	563.94 (12.38)	.96 (.005)	.95 (.004)
	Incongruent	650.62 (14.44)	643.86 (12.12)	.88 (.011)	.85 (.009)
	Conf. Effect	80.48	79.92		
Incongruent	Congruent	597.29 (15.21)	591.01 (12.76)	.94 (.007)	.93 (.005)
	Incongruent	658.22 (15.78)	654.53 (13.24)	.88 (.011)	.86 (.010)
	Conf. Effect	60.93	63.52		

Figure 3.1. Mean Response Time (ms) with Respect to Previous Congruency and Current Congruency Shown by Language Group.



We first examine the magnitude of conflict effects compared with those reported in the previous study using the Sequential Modulation Task, focusing on the results of the monolinguals when comparing with prior research for the purpose of consistency, unless otherwise noted. The results of Akçay and Hazeltine (2007) were largely replicated by our findings. The conflict effect following an incongruent trial was comparable to theirs (60 ms), but the conflict effect following a congruent trial was about 20 ms smaller. This was primarily due to our (monolingual) participants performing slower on congruent trials (that followed a congruent trial), thus reducing the RT difference between trial types after a congruent trial. The conflict effect following an incongruent trial (60.93 ms) was significantly smaller than following a congruent trial (80.48 ms), which is consistent with prior research on sequencing effects (Gratton et al., 1992). The language groups did not differ on the congruency effect involving all trials.

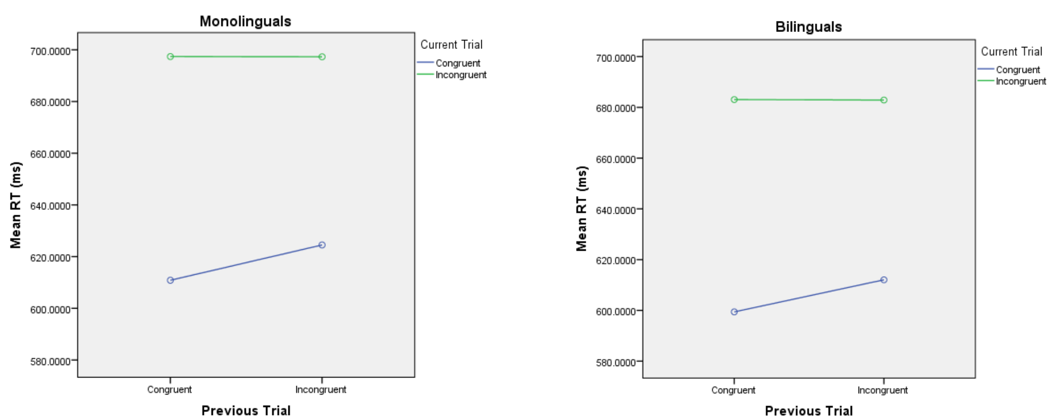
To test the conflict monitoring hypothesis specifically, the same three-way mixed-factor ANOVA was conducted on stimulus-location-change and response-position-change (complete-change) trials that were devoid of stimulus or response feature repetitions. If sequential modulations were obtained on trials with no overlapping features, the conflict monitoring hypothesis would be supported. We also excluded negative priming trials as Akçay and Hazeltine (2007) had done; negative priming trials were trials on which the stimulus location corresponded with the response position of the previous trial or on which the response position corresponded with the stimulus location of the previous trial.

Mean RTs and accuracy with respect to previous congruency and current congruency for complete-change trials are shown separately by language group in *Table 3-2*. Results showed a main effect of previous congruency, F (1, 73) = 4.98, p = .03, a main effect of current congruency, F (1, 73) = 717.81, $p < .001$, and a significant interaction, F (1, 73) = 4.25, p = .04. The three-way interaction of previous congruency, current congruency, and language group was not significant, F (1, 73) = .06, p = .82. See *Figure 3.1* for a visual illustration of this relationship exhibited by the RTs.

Table 3-2. *Mean Response Time (ms) and Accuracy with Respect to Previous Congruency and Current Congruency on Complete-Change Trials Shown by Language Group (nML = 31, nBL = 44) in the Sequential Modulation Task,*

Previous	**Current**	**ML RT**	**BL RT**	**ML ACC**	**BL ACC**
Congruent	Congruent	610.99 (17.67)	601.42 (14.84)	.95 (.006)	.94 (.005)
	Incongruent	697.41 (17.65)	683.07 (14.81)	.86 (.013)	.83 (.011)
	Conf. Effect	86.42	81.65		
Incongruent	Congruent	624.50 (17.14)	612.05 (14.39)	.94 (.007)	.93 (.006)
	Incongruent	697.31 (18.55)	682.88 (15.57)	.88 (.012)	.87 (.010)
	Conf. Effect	72.81	70.83		

Figure 3.2. Mean Response Time (ms) of Complete-Change Trials with Respect to Previous Congruency and Current Congruency Shown by Language Group.



The main effects and congruency interaction again replicated those of Akçay and Hazeltine (2007). The conflict effect following a congruent trial was 86.42 ms and the conflict effect following an incongruent trial was 72.81 ms. The magnitude difference of 13.61 ms was slightly less than the 17 ms reported by Akçay and Hazeltine (2007). While bilinguals showed smaller RTs and smaller conflict effects in all conditions numerically, these differences did not reach statistical significance. Moreover, bilinguals had numerically (slightly) lower accuracy rates in all conditions, which may be an indication of speed-accuracy tradeoff at play. The language groups did not differ in the congruency effect involving complete-change trials.

When the congruency effect was examined devoid of feature overlaps, sequential modulation remained, although in an attenuated form as evidenced by the smaller magnitude of interaction, i.e., F (1, 73) = 4.25, p = .04, than that of the mixed-factor ANOVA including all trials, i.e., F (1, 73) = 33.91, p < .001. The presence of sequencing effects on complete-change trials provides support for the conflict monitoring hypothesis. However, the initially-large but then-attenuated interaction shows the contribution of feature overlap to the sequentially modulated effect in the all-trial analysis, before feature overlaps were removed. No language-group difference was observed on the congruency effect involving all trials or complete-change trials.

We then examined the effect of English proficiency on sequential modulation task performance among bilinguals using a median-split procedure. The bilingual participants were classified into high-proficient and low-proficient groups first based on their self-reported English proficiency (SEP), and then based on their objectively measured English proficiency (OEP). Individuals with a proficiency score at the median were excluded from the analysis, resulting in 21 participants in the low-proficient group and 20 participants in the high-proficient group based on SEP and 21 participants in each of the groups based on OEP.

SEP-based Results

The same mixed-factor ANOVAs were conducted on accuracy and on RT first on all trials, then on complete-change trials only, with SEP as the between-subjects factor. In the accuracy analysis, the main effect of SEP and interactions involving SEP were not significant in either the all-trial analysis or the complete-change trial analysis.

The RT analysis revealed a three-way interaction of previous congruency, current congruency, and proficiency on complete-change trials, F (1, 39) = 6.44, p = .02; see *Table 3-3* for a summary of means shown by low-proficiency and high-proficiency bilinguals on complete-change trials. The conflict effect exhibited by low-proficient bilinguals after an incongruent trial was numerically larger than the conflict effect after a congruent trial, thus eliminating the sequencing effect altogether. In comparison, the high-proficient bilinguals demonstrated a larger conflict effect after a congruent trial as well as a smaller conflict effect after an incongruent trial, thus fully exhibiting the sequencing effect. The low-proficient bilinguals showed faster RTs, as well as lower accuracies, than high-proficient bilinguals in all conditions (except for the RT in Incongruent-Incongruent), but these group differences did not reach statistical significance.

Table 3-3. *Mean Response Time (ms) and Accuracy with Respect to Previous Congruency and Current Congruency on Complete-Change Trials Shown by Self-Reported English Proficiency Level for Bilinguals (nBL = 41) in the Sequential Modulation Task.*

Previous	**Current**	**Low SEP RT**	**High SEP RT**	**Low SEP ACC**	**High SEP ACC**
Congruent	Congruent	595.49 (20.99)	602.90 (21.51)	.938 (.008)	.949 (.008)
	Incongruent	673.38 (21.33)	689.96 (21.85)	.822 (.017)	.841 (.018)
	Conf. Effect	77.89	87.06		
Incongruent	Congruent	604.60 (19.92)	616.72 (20.41)	.930 (.010)	.942 (.010)
	Incongruent	690.28 (23.06)	670.69 (23.63)	.859 (.017)	.876 (.018)
	Conf. Effect	85.68	53.97		

OEP-based Results

The same mixed-factor ANOVAs were conducted on accuracy and on RT first on all trials, then on complete-change trials only, with objectively assessed English proficiency (OEP) now as the between-subjects factor. See *Table 3-4* for a summary of accuracy means for all and complete-change trials. The proficiency groups differed in accuracy for all trials, F (1, 40) = 5.09, p = .03, as well as for complete-change trials, F (1, 40) = 5.42, p = .03; the high-proficient group (M = .91, SE = .01) performed more accurately than the low-proficient group ($M \approx .88$, SE = .01) in both analyses.

The RT analysis showed neither the main effect of OEP nor interactions involving OEP were significant in either the all-trial analysis or complete-change-trial analysis.

Table 3-4. *Mean Accuracy with Respect to Previous Congruency and Current Congruency on All Trials and Complete-Change Trials Shown by Objectively Assessed Proficiency Level for Bilinguals (nBL = 41) in the Sequential Modulation Task.*

		All Trials		*Complete-Change Trials*	
Previous	**Current**	**Low OEP ACC**	**High OEP ACC**	**Low OEP ACC**	**High OEP ACC**
Congruent	Congruent	.937 (.007)	.962 (.007)	.927 (.008)	.956 (.008)
	Incongruent	.828 (.014)	.861 (.014)	.804 (.016)	.847 (.016)
Incongruent	Congruent	.920 (.008)	.944 (.008)	.923 (.010)	.945 (.010)
	Incongruent	.841 (.014)	.881 (.014)	.839 (.016)	.888 (.016)

While SEP correlated with OEP somewhat strongly for bilinguals, the outcomes for using one or the other measure to classify high-proficient vs. low-proficient bilinguals yielded markedly different results. Most notably, SEP-based analyses showed low-proficient bilinguals failing to exhibit a pattern of modulated control from trial-to-trial in terms of RT, and OEP-based analyses revealed the same individuals to perform less accurately than the high-proficient individuals. These results converged in demonstrating that high-proficiency bilinguals performed at a superior level compared with their low-proficiency counterparts on conflict monitoring as assessed by the sequential modulation task.

Feature Overlap Effect

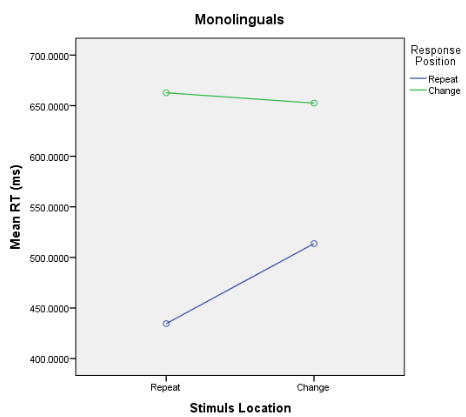
We next focused on the contributions of feature overlap, comparing trials on which there was a repetition versus alternation of the location of the stimulus, and a repetition or alternation of the required response. Mean RTs and accuracies with respect to repetition versus change of the stimulus location and response are shown separately by language group in *Table 3-5*.

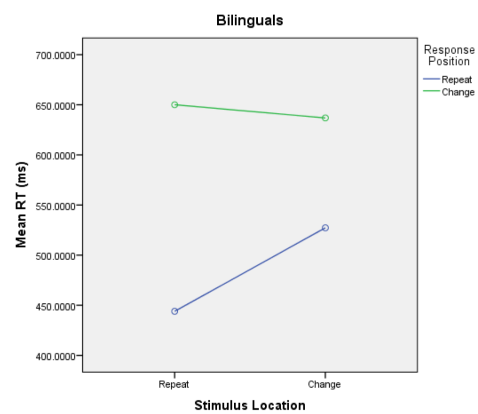
A 2 (location repetition/location alternation) x 2 (response repetition/response alternation) x 2 (monolingual/bilingual) mixed-factor ANOVA was conducted on all trials. There was a main effect of stimulus location, F (1, 73) = 205.93, $p < .001$, a main effect of response position, F (1, 73) = 571.95, $p < .001$, and a highly significant interaction, F (1, 73) = 478.16, $p < .001$. There was a non-significant response by language group interaction, F (1, 73) = 2.77, $p = .10$, with monolinguals performing numerically faster on response repetitions (M_{ML} = 474.10, SE = 9.44; M_{BL} = 486.62, SE = 7.92) and bilinguals performing numerically faster on response alternations (M_{ML} = 657.64, SE = 16.56; M_{BL} = 646.30, SE = 13.97). See *Figure 3.3* for a visual illustration of the RT data.

Table 3-5. *Mean Response Time (ms) and Accuracy Separated by whether the Location and Response were a Repetition or Change from the Previous Trial (nML = 31, nBL = 44) in the Sequential Modulation Task.*

Location	**Response**	**ML RT**	**BL RT**	**ML ACC**	**BL ACC**
Repetition	Repetition	434.49 (10.05)	444.24 (8.44)	.99 (.004)	.98 (.003)
	Change	662.79 (16.20)	652.60 (13.60)	.88 (.010)	.86 (.009)
Change	Repetition	513.71 (9.80)	529.01 (8.23)	.95 (.006)	.94 (.005)
	Change	652.50 (17.25)	640.00 (14.48)	.91 (.008)	.89 (.007)

Figure 3.3. Mean Response Time (ms) with Respect to whether the Location and Response were a Repetition or Change from the Previous Trial.





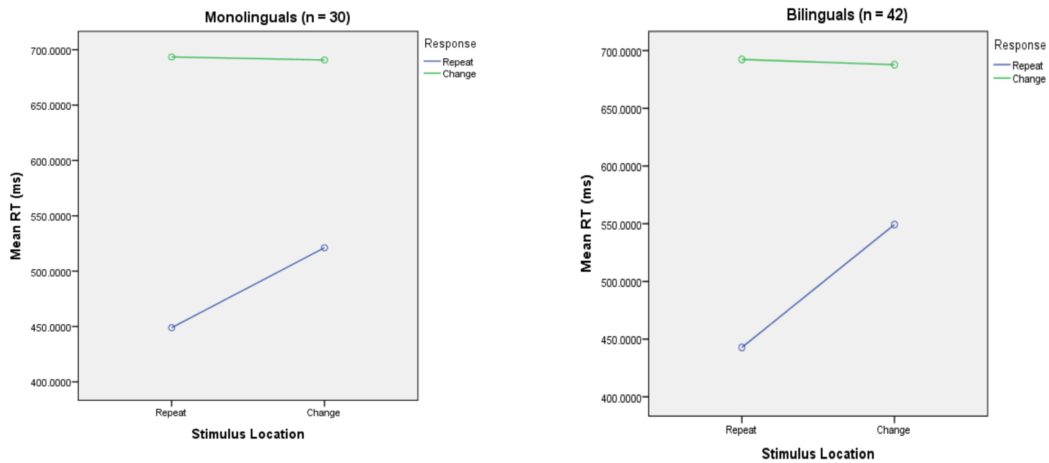
The main effects of stimulus location and response, as well as their interaction, once again replicated Akçay and Hazeltine (2007). Then we assessed the effects of feature overlap apart from the influence of congruency by including only Incongruent-Incongruent (II) trial sequences in the subsequent analysis, for a total of up to 224 trials if no trials were eliminated. Negative priming trials were likewise removed, in step with the data processing procedure adopted by Akçay and Hazeltine (2007). Due to one monolingual and two bilinguals missing valid trials in the location-repetition and response-repetition condition, the sample sizes of the language groups were 30 and 42, respectively, in the current analysis. See *Table 3-6* for mean RTs and accuracies with respect to the repetition and change of stimulus location and response for Incongruent-Incongruent trial sequences displayed by language group.

A 2 (location repetition/ location alternation) x 2 (response repetition/response alternation) x 2 (monolingual/bilingual) mixed-factor ANOVA revealed a main effect of stimulus location, F (1, 70) = 45.38, $p < .001$, a main effect of response position, F (1, 70) = 544.71, $p < .001$, and a significant interaction, F (1, 70) = 53.08, $p < .001$. The three-way interaction involving language group was not significant, F (1, 70) = 2.00, $p = .16$. See *Figure 3.4* for a visual illustration of the RT data.

Table 3-6. *Mean Response Time (ms) and Accuracy Separated by whether the Location and Response were a Repetition or Change from the Previous Trial for Incongruent-Incongruent Trial Sequences (nML = 30, nBL = 42) in the Sequential Modulation Task.*

Location	**Response**	**ML RT**	**BL RT**	**ML ACC**	**BL ACC**
Repetition	Repetition	448.93 (21.40)	442.82 (18.08)	.97 (.020)	.97 (.017)
	Change	693.51 (17.06)	692.33 (14.42)	.85 (.015)	.84 (.013)
Change	Repetition	521.16 (11.70)	549.29 (9.89)	.94 (.017)	.92 (.009)
	Change	690.77 (18.43)	687.74 (15.57)	.87 (.013)	.86 (.011)

Figure 3.4. Mean Response Time (ms) with Respect to whether the Location and Response were a Repetition or Change from the Previous Trial for Incongruent-Incongruent Trial Sequences.



The interaction between stimulus location and response showed the dependence of stimulus features upon response and vice versa. This effect was obtained in Incongruent-Incongruent (II) trial sequences, or trials completely devoid of congruency transitions, demonstrating the contribution of feature overlap to the sequencing effect, although the pattern of the interaction was markedly different from that previously observed for two-choice tasks (Akçay & Hazeltine, 2007; Hommel, 2004). Because the language groups were shown not to differ, they were collapsed for the present discussion. Complete-repetition trials were the fastest (445.88 ms), followed by location-change response-repetition trials (535.23 ms). Complete-change (689.86 ms) and location-repetition response-change trials (692.92 ms) were slower and did not differ from each other. In our results the two slower conditions from Akçay & Hazeltine (2007) are reversed, but the general pattern remained; complete-repetition and complete-change trials were not performed faster than partial-repetition trials, contrary to the prediction of the feature integration account (Hommel et al., 2004).

SEP-based Results

The same mixed-factor ANOVAs were conducted on accuracy and on RT first on all trials, then on II trial sequences only, with SEP as the between-subjects factor. For the accuracy analysis, the main effect of SEP and interactions involving SEP were not significant in either the all-trial analysis or the II sequenced trial analysis.

The RT analysis revealed a three-way interaction of previous congruency, current congruency, and proficiency on II-sequenced trials, F (1, 37) = 4.33, p = .05; the high-proficiency group performed complete-repetition trials and complete-change trials more rapidly than the low-proficiency group. See *Table 3-7* for a summary of the mean RTs and accuracies with respect to the repetition and change of stimulus location and response shown separately by proficiency group.

Table 3-7. *Mean Response Time (ms) Separated by whether the Location and Response were a Repetition or Change from the Previous Trial for Incongruent-Incongruent Trial Sequences in the Sequential Modulation Task for Low vs High SEP Bilinguals.*

Location	**Response**	**Low SEP-BL**	**High SEP-BL**	**Low ACC**	**High ACC**
Repetition	Repetition	440.38 (14.18)	416.90 (14.55)	.976 (.029)	.950 (.030)
	Change	690.36 (22.38)	693.13 (22.96)	.824 (.019)	.844 (.020)
Change	Repetition	546.83 (15.90)	546.43 (16.31)	.910 (.015)	.925 (.015)
	Change	697.90 (23.50)	673.16 (24.12)	.853 (.016)	.863 (.016)

OEP-based Results

The same mixed-factor ANOVAs were conducted on accuracy and on RT first on all trials, then on II-sequenced trials only, with OEP now as the between-subjects factor.

The proficiency groups differed in accuracy for all trials, $F(1, 40) = 4.81$, $p = .03$, with the high-proficient group ($M = .93$, $SE = .01$) outperforming the low-proficient group ($M = .90$, $SE = .01$). The proficiency groups did not significantly differ in accuracy for II-sequenced trials, $F(1, 40) = 2.91$, $p = .10$.

The RT analysis showed that neither the main effect of OEP nor interactions involving OEP were significant in either the all-trial analysis or the II-sequenced trial analysis.

Overall, we replicated nearly all of the findings by Akçay & Hazeltine (2007) with some modifications in our task parameters. We found support for both conflict monitoring and feature integration playing a role in sequencing effects, which is consistent with prior research (Notebaert, Gevers, Verbruggen, & Liefooghe, 2006). Aside from monolinguals performing marginally faster on response repetitions and bilinguals performing marginally faster on response alternations, no statistically meaningful differences were observed between language groups. Within bilinguals, those of higher English proficiency exhibited speed and accuracy advantages over those lower in English proficiency, and high-proficient bilinguals appeared to recruit the conflict monitoring system more effectively than low-proficient bilinguals when analyses were based on self-reported proficiency scores. A speed-accuracy tradeoff seemed to be present in both language groups; the overall task RT and accuracy correlated at $r(30) = .41$, $p = .03$ for monolinguals and $r(41) = .29$, $p = .07$ for bilinguals, after the removal of one bilingual outlier by visual inspection of the scatterplot.

3.7.3 Standard Simon and Arrow Simon

Data from individuals with an overall task accuracy that was three *SDs* from their respective language group means or more extreme were excluded. This led to the exclusion of 1 individual from each language group in standard Simon and the exclusion of 2 bilinguals in arrow Simon, or three individuals (1 monolingual and 2 bilinguals) across both tasks. Error trials, correct trials with RTs of less than 250 ms or three *SD*s above a given individual's RT mean were excluded from the RT analysis. This resulted in the elimination of 484 correct trials in standard Simon and 663 correct trials in arrow Simon, or 1.84% and 1.97% of all trials, respectively. Single-trial-type mini-blocks served as a buffer between the mixed blocks and were excluded from the analysis.

The language groups were compared on the RT and accuracy of congruent trials, incongruent trials, and the Simon effect. See *Table 3-8* for a summary of the pattern of means by language group.

Table 3-8. *Mean Response Time (ms) and Accuracy of Congruent Trials, Incongruent Trials, and the Simon Effect by Language Group (nML = 36, nBL = 45).*

	ML RT	**BL RT**	**ML ACC**	**BL ACC**
Standard Simon				
Congruent	419.37 (9.50)	423.51 (8.50)	.953 (.009)	.934 (.008)
Incongruent	442.36 (9.87)	449.02 (8.82)	.929 (.010)	.914 (.009)
Simon Effect	22.99	25.51		
Arrow Simon				
Congruent	463.91 (8.70)	455.88 (7.78)	.973 (.006)	.957 (.006)
Incongruent	501.08 (9.25)	486.76 (8.28)	.913 (.011)	.895 (.010)
Simon Effect	37.17	30.88		

Accuracy Data

A 2 (standard Simon/arrow Simon) x 4 (more-congruent/more-incongruent/low-switch/high-switch) x 2 (congruent/incongruent) x 2 (monolingual/bilingual) mixed-factor ANOVA was conducted on accuracy. There was no main effect of task; the two Simon tasks were performed with comparable accuracy (.932 and .934). There were main effects of block type and trial type. The four block types arranged in an ascending order of accuracy were more-congruent (*M* = .905, *SE* = .006), more-incongruent (*M* = .933, *SE* = .006), high-switch (*M* = .944, *SE* = .004), and low-switch (*M* = .951, *SE* = .004), with all pairwise comparisons being significantly different from one another. Consistent with existing research, congruent trials (*M* = .954, *SE* = .004) were performed more accurately than incongruent trials (*M* = .913, *SE* = .006). The overall language-group difference approached significance, F (1, 79) = 3.63, p = .06; monolinguals (.942) tended to perform the tasks more accurately than bilinguals (.925).

In addition, block type and trial type each interacted with task and the two interactions will be described in turn. For the block type by task interaction, more-congruent blocks (*M* = .945, *SE* = .007) and low-switch blocks (*M* = .954, *SE* = .005) were performed more accurately than more-incongruent blocks (*M* = .938, *SE* = .006) and high-switch blocks (*M* = .932, *SE* = .006) in the standard Simon task, with the latter two conditions not differing from each other. In arrow Simon, high-switch blocks (*M* = .956, *SE* = .004) were performed more accurately than low-switch (*M* = .949, *SE* = .004) and more-congruent block (*M* = .950, *SE* = .004) statistically, with high-switch and more incongruent blocks (*M* = .950, *SE* = .005) not differing from each other. In other words, the block types performed comparatively more accurately in one task became the block types performed less accurately in the other task. Inspection of the trial type by task interaction showed that congruent trials of arrow Simon (*M* = .965, *SE* = .004) were performed more accurately than the congruent trials of standard Simon (*M* = .943, *SE* = .006), whereas accuracy on the incongruent trials did not differ between tasks.

The three-way interaction of task, block type, and trial type was significant, F (2.03, 160.43) = 11.01, $p < .001$. Follow-up analyses conducted by task further revealed that, in standard Simon, congruent trials were performed more accurately than incongruent trials only in the more-congruent (.964 vs .870) and low-switch blocks (.968 vs .941). Incongruent trials were performed more accurately than congruent trials in the more-incongruent blocks (.945 vs .909), and performance between trial types did not differ in the high-switch blocks (.933 vs .931). In arrow Simon, on the other hand, congruent trials were consistently performed more accurately except for the more-incongruent blocks, in which incongruent trials (.958) were performed more accurately than congruent trials (.920).

RT Data

Next, the same 2 (standard Simon/arrow Simon) x 4 (more-congruent/more-incongruent/low-switch/high-switch) x 2 (congruent/incongruent) x 2 (monolingual/bilingual) mixed-factor ANOVA was conducted on RT. All the main effects other than language group were significant; standard Simon (*M* = 433.57, *SE* = 6.40) was performed faster than arrow Simon (*M* = 476.91, *SE* = 5.90); the four block types ranging from the fastest to the slowest were low-switch (*M* = 430.38, *SE* = 5.04), more-congruent (*M* = 456.76, *SE* = 5.53), more-incongruent (*M* = 461.93, *SE* = 5.78), and high-switch (*M* = 471.88, *SE* = 6.15), with all pairwise comparisons being significantly different from one another; and congruent trials (*M* = 440.67, *SE* = 5.46) were performed faster than incongruent trials (*M* = 469.81, *SE* = 5.63).

Similar to the accuracy data, block type and trial type each interacted with task in different ways. For the block type by task interaction, more-congruent blocks (*M* = 411.26, *SE* = 6.74) and low-switch blocks (*M* = 405.13, *SE* = 5.61) were performed faster than more-incongruent blocks (*M* = 433.06, *SE* = 6.95), which were in turn faster than the high-switch blocks (*M* = 460.80, *SE* = 7.51) in standard Simon. In arrow Simon, the trend largely remained the same except the more-congruent blocks (*M* = 447.45, *SE* = 6.12) were performed more rapidly than low-switch blocks (*M* = 454.84, *SE* = 5.57). The high-switch blocks produced the largest RT of the four block types across the Simon tasks. The data supported our hypothesis of the high-switch condition being the most cognitively demanding in terms of the overall RT in standard Simon as well as arrow Simon. With regard to the trial type by task interaction, the magnitude of the Simon effect in the standard Simon task (24.25 ms) was smaller than the Simon effect in the arrow Simon task (34.02 ms). Our RT data showed that the main effects of task and trial type, as well as the faster performance of low-switch blocks versus high-switch blocks all replicated Bialystok (2006). In addition, our data support the view that the arrow Simon is a task of greater perceptual conflict than standard Simon, both in terms of the overall RT and the Simon effect.

The four-way interaction of task, trial type, block type, and language group was significant, F (3, 237) = 3.55, p = .02. The three-way interaction involving trial type, block type, and language group was then examined in the Simon tasks separately. The three factors only interacted in arrow Simon, F (2.29, 180.62) = 4.29, p = .01, but not in standard

Simon. The Simon effect exhibited by monolinguals in arrow Simon ranged from 116.19 ms in more-congruent blocks to -30.55 ms in more-incongruent blocks, with 30.53 ms and 31.80 ms for low-switch and high-switch blocks in between. Bilinguals showed a similar pattern of Simon effects in the four block types, except at a markedly smaller magnitude in the more-congruent blocks, where the Simon effect was 93.43 ms, or nearly 23 ms smaller than that observed for monolinguals. The present pattern of Simon effects elicited by different block types is consistent with the literature on the proportion of congruent trials positively corresponding to the magnitude of the Simon effect (Stürmer et al., 2002; Costa et al., 2009); the more-congruent blocks elicited the largest effect, the more-incongruent blocks the smallest effect, and the 50% congruent blocks (low-switch and high-switch) in between.

The reduced Simon effect by bilinguals in the more-congruent blocks was further examined with simple effects analyses conducted on each block type; the trial type by language group interaction was significant only in the more-congruent blocks (and not in the other three block types), F (1, 79) = 5.80, p = .02. Bilinguals exhibited a smaller Simon effect (93.43) than monolinguals (116.19) in this particular block type, as a result of bilinguals outperforming monolinguals on incongruent trials (M_{ML} = 547.27, SE = 11.33; M_{BL} = 523.59, SE = 10.14), but not on congruent trials (M_{ML} = 431.08, SE = 8.86; BL M_{BL} = 430.16, SE = 7.93).

To further assess bilinguals' apparent RT advantage on the incongruent trials of the more-congruent blocks, we examined speed-accuracy tradeoffs by language group. Monolinguals showed a marginally significant tradeoff on the arrow Simon task overall, r (36) = .31, p = .06, whereas bilinguals showed a highly significant tradeoff effect much stronger in magnitude, r (45) = .57, p < .001. In more-congruent blocks where a bilingual speed advantage was observed on the incongruent trials, no speed-accuracy tradeoff existed for monolinguals, r (36) = .10, p = .57, whereas it did exist for bilinguals, r (45) = .33, p = .03. In fact, the difference in accuracy on the incongruent trials of the more-congruent blocks trended towards significance between language groups, t (79) = 1.69, p = .09; *ML* = .828 and *BL* = .774. Thus bilinguals' RT advantage on the incongruent trials was qualified by a slight accuracy disadvantage on the same trials.

Bialystok (2006) reported a three-way interaction involving task, trial type, and switch type (low switch/high switch), where a reverse Simon effect (i.e., incongruent trials being faster than congruent trials) was observed in the high-switch condition of the standard Simon but disappeared altogether in the arrow Simon. We were unable to replicate either of these effects. We were also unable to replicate the finding of the language groups performing similarly in the low-switch condition but bilinguals demonstrating a speed advantage in the high-switch condition. While a reduced Simon effect was observed for bilinguals in the condition with the largest conflict effect as a result of bilinguals' faster performance on the incongruent trials, it may have largely stemmed from their attenuated accuracy relative to monolinguals.

We then examined the effect of English proficiency among bilinguals using a median-split procedure. The bilingual participants were first classified into high-proficient and low-proficient groups based on their self-reported English proficiency (SEP), and then on their objectively measured English proficiency (OEP). Individuals with a proficiency score at the median were excluded from the analysis. This resulted in 22 and 20 participants in the low- and high-proficient groups, respectively, based on self-reported English proficiency, and 22 and 21 participants, respectively, based on objectively measured English proficiency.

SEP-based Results

The same mixed-factor ANOVAs were conducted on accuracy and on RT, with SEP as the between-subjects factor. The main effect of SEP and interactions involving SEP were not significant in either the accuracy analysis or RT analysis.

OEP-based Results

The same mixed-factor ANOVAs were conducted on accuracy and on RT, with OEP as the between-subjects factor. The three-way interaction of task, trial type, and proficiency approached significance in the accuracy analysis, F (1, 41) = 3.40, p = .07; low-proficiency bilinguals tended to exhibit a smaller difference in accuracy between trial types on standard Simon ($M_{Congruent}$ = .918, SE = .01; $M_{Incongruent}$ = .909, SE = .01) and a larger difference in accuracy between trial types on arrow Simon ($M_{Congruent}$ = .955, SE = .01; $M_{Incongruent}$ = .888, SE = .02) relative to their high-proficiency counterparts (standard Simon: $M_{Congruent}$ = .955, SE = .01, $M_{Incongruent}$ = .888, SE = .02; arrow Simon: $M_{Congruent}$ = .957, SE = .01, $M_{Incongruent}$ = .899, SE = .02).

The RT analysis revealed a significant three-way interaction involving task, block, and proficiency, F (3, 123) = 3.56, p = .02. Follow-up analyses of task and proficiency were then conducted for each of the four block types. The low-proficient and high-proficient bilinguals did not differ in their performance on the low-switch, high-switch, and more-congruent blocks, but the task by proficiency interaction was significant for the more-incongruent blocks, F (1, 41) = 6.36, p = .02. The RT difference between the two Simon tasks was smaller in low-proficient bilinguals (13.74 ms) than in high-proficient bilinguals (55.79 ms), as a result of the high-proficient bilinguals performing faster on standard Simon and slower on arrow Simon.

3.8. Discussion

In the present study, we explored the question of whether bilinguals possess an advantage over monolinguals from the perspective of conflict monitoring theory, both in the classical sense of intra-trial conflict modulated on a trial-to-trial basis, as well as in an extended form of the theory in which the anterior cingulate cortex is viewed as encoding information about mental effort or task difficulty. We presented the results of three behavioral tasks: a four-choice sequential modulation task to allow the empirical disambiguation of the contributions of conflict monitoring from the contributions of its theoretical alternative – feature integration, and two variants of the Simon task developed to examine performance under varying levels of difficulty as indexed by the overall RT and Simon effects of different magnitudes. For each task, we conducted between-group comparisons based on a categorical assignment of language groups and within-group comparisons of bilinguals based on their self-rated and objective measures of English proficiency.

Monolingual and Bilingual Comparisons

Across the three experimental tasks, we observed a modest trend for monolinguals to achieve higher accuracy rates than bilinguals. While bilinguals showed reduced RTs numerically in certain conditions, these outcomes did not reach statistical thresholds and were further qualified by the group's slightly higher error rates. In the only instance where a bilingual advantage in the Simon effect was observed in the more-congruent blocks of the arrow Simon task, a highly significant speed-accuracy tradeoff was detected for bilinguals on block type, r (45) = .53, $p < .001$, as well as on the incongruent trials within the blocks, r (45) = .33, $p = .03$, thus accounting for their reduced RT on the incongruent trials. The language groups exhibited differential patterns of speed-accuracy tradeoffs across tasks as well as from each other; monolinguals showed a slight tradeoff effect on the sequential modulation task, bilinguals a significant tradeoff effect on the arrow Simon task, and neither group exhibited a speed-accuracy tradeoff on the standard Simon task.

These results highlight the importance of interpreting RTs in conjunction with accuracy (Salthouse, 2010), which is not always a given in research reports. For instance, the Cantonese-English group outperformed the other two groups in RT while also showing the lowest accuracy of all three groups (Bialystok et al., 2005). These accuracy data were displayed but not discussed. By and large, monolinguals and bilinguals did not differ on conflict monitoring or its theoretical alternative of feature integration, nor did language-group differences emerge on the more challenging task (arrow Simon relative to standard Simon) or task conditions (overall RT of the high-switch blocks and the Simon effect of the more-congruent blocks).

High-proficient and Low-proficient Bilinguals

We compared performance of high-proficient and low-proficient bilinguals based on their self-rated English proficiency (SEP) and objective measure of English proficiency (OEP). The two methods of classifying proficiency converged in revealing low-proficient bilinguals performing poorer than high-proficient bilinguals; low-proficient bilinguals did not exhibit modulated control on a trial-by-trial basis and achieved lower accuracy on the conflict monitoring account as assessed by the sequential modulation task. When examined from the perspective of the feature integration account, high-proficient bilinguals performed complete-repetition and complete-change trials more rapidly than the low-proficient bilinguals and achieved higher accuracy overall.

The performance of the two proficiency levels on the Simon tasks was less straightforward. Compared to their high-proficiency counterparts, low-proficient bilinguals performed congruent and incongruent trials more similarly in accuracy on standard Simon and more dissimilarly on arrow Simon. Compared to low-proficient bilinguals, on the other hand, high-proficient bilinguals were faster on the more-incongruent blocks of the standard Simon task and slower on the same condition of the arrow Simon task.

While a majority of the studies to date have compared monolinguals and bilinguals, Mishra (2015) advocates a research strategy of comparing different levels of bilingualism, be it on proficiency or switching, particularly in linguistic contexts where “pure” monolinguals do not exist. Our current findings support a notable difference between high-proficient and low-proficient bilinguals, although the classification of proficiency was based on English only, which was arguably an inaccurate proxy for degrees of bilingualism.

Notwithstanding the limitation, the findings of performance as a function of proficiency extend research based on proficiency from participants who are homogeneous bilinguals with a shared L2 (Hommel et al., 2011; Khare et al., 2012; Singh & Mishra, 2013) to heterogeneous bilinguals including both immigrants and non-immigrants. Instead of assessing proficiency in L2 with a vocabulary test or reading comprehension exercises administered in each of bilinguals’ languages, we assessed proficiency in the bilinguals’ shared English language with an aggregate measure of four linguistic domains.

SEP (Subjective English Proficiency) and OEP (Objective English Proficiency)

The inclusion of an objective English proficiency measure yielded some notable results. The measure of objective proficiency (OEP) moderately correlated with self-report proficiency (SEP) in the entire sample ($r = .44$, $p < .001$), with the association being further strengthened among the bilinguals ($r = 59$, $p < .001$). These outcomes are consistent with the moderate to high correlations between self-reports and objective measures typically reported in the literature (Marian, Blumenfeld, & Kaushanskaya, 2007). Bilinguals were able to evaluate their English proficiency more accurately than their monolingual peers, which is also in accordance with previous reports pointing to bilinguals' better metalinguistic awareness (Bialystok, Majumder, & Martin, 2003; Campbell & Sais, 1995; Galambos & Hakuta, 1988). While the psychometric properties of the OEP used in our study need to be further established, the objective assessment did positively correlate with letter fluency at r (84) = .33, $p = .002$, and also with category fluency at r (84) = .38, $p < .001$, whereas SEP did not correlate with either verbal fluency outcome. See Appendix C for materials, procedure, and scoring method of the letter and category fluency tasks not reported here. The critical difference between the two methods of assessing English proficiency is that OEP showed monolinguals to be higher on English proficiency, whereas SEP showed the language groups to be identical. Although OEP and SEP showed a moderate correlation, the difference between using one classification method or the other has implications for determining language-group comparability and thus potentially the interpretations of results.

Chapter 4

The Probe on Creative Performance

4.1. Introduction

Creativity is sparked and sustained by the meeting of multiple perspectives and varied experiences. This would suggest that individuals who are deeply versed in two or more languages (and associated cultures) might—all else being equal—demonstrate greater originality, ideational flexibility, and conceptual reach. Yet, despite the recent surge of scientific research on how bilingualism influences broader aspects of cognition, and a long-standing understanding that bilingualism might be positively associated with creativity (e.g., for review, see Kharkhurin, 2012; Ricciardelli, 1992; Simonton, 2008), the question of whether and how extensive experience in two or more languages shapes creativity remains open empirically and theoretically.

Empirically, studies reporting a positive relation between bilingualism and creativity (e.g., Bruck, Lambert, & Tucker, 1976; Kharkhurin, 2007, 2008; Lambert, Tucker, & d'Anglejan., 1973) are interspersed with a few studies that find no or little relation for subcomponents of creativity, such as the ability to activate multiple unrelated concepts or work through their interrelations (e.g., Kharkhurin, 2009), and many further studies that point to potential mediating and moderating influences. Among such factors are the level of proficiency in the two languages (e.g., Carringer, 1974; Kessler & Quinn, 1987; Lee & Kim, 2011) and the extent to which assessments of creativity are themselves dependent on language rather than modalities such as visual-spatial or figural processing (e.g., Kharkhurin, 2010). More generally, relatively few studies have examined the contribution of individual difference and personality characteristics, such as openness to experience, in the context of bilingualism and creativity, though one study reported no relation between extensiveness of multicultural experiences and openness to experience (e.g., Leung & Chiu, 2010; see also Leung, Maddux, Galinksy, & Chiu, 2008).

Theoretically, if bilingualism relative to monolingualism is associated with increases in creativity, what might be the cognitive mechanism(s) undergirding such an increase? From a cognitive perspective, creativity can be seen as a process through which people retrieve and recombine previous knowledge and experiences in new ways. Thus individuals who can access and integrate diverse knowledge systems will experience enhanced levels of creativity (Amabile, 1996; Finke et al., 1992; see also Gocłowska & Crisp, 2014; Tadmor, Galinsky, & Maddux, 2012; Tadmor & Tetlock, 2006). For instance, beyond bilingualism per se, research indicates that bi-culturals and mixed-race individuals with well-integrated cultural and racial identities, respectively, perform better on a variety of creative tasks than do individuals whose identities are less well-integrated (Cheng, Sanchez-Burks, & Lee, 2008; Viki & Williams, 2014).

In the current study, we sought to take a more configural approach, involving an integrated examination of objective and subjective measures of creative performance in relation to not only language group, but also other known contributors to creativity, including personality factors (e.g., openness to experience) and fluid reasoning ability. We hypothesized that, on non-language related visual and figural tasks, bilinguals compared to monolinguals would show evidence of enhanced creativity. In addition, we aimed to assess the relative importance of language group (monolingual vs. bilingual) in combination or interaction with personality traits in performance of creativity. From the perspective of the "Four P" model of creativity (Rhodes, 1961) – according to which creativity involves the conjunctive influences of traits or characteristics of the person, the "product" (what is made), the process (how), and press or environment – we focus especially on characteristics of the person, and the product.

4.1.1. Personal characteristics related to creativity

The term person conveys "information about [the] personality, intellect, temperament, physique, traits, habits, attitudes, self-concept, value systems, defense mechanisms, and behavior" of the creator (Rhodes, 1961, p. 307). In a sample of 269 adolescent and adult students at an English center in Brazil, bilinguals were found to score higher on a non-verbal intelligence test than monolinguals (Mendonça & Fleith, 2005). In an Australian sample, early bilingual adults who had arrived in Australia by age 6 also scored higher on a shortened version of the Raven's Advanced Progressive Matrices than monolingual adults (Tao, Marzecova, Taft, Asanowicz, & Wodniecka, 2011). However, findings that point to a bilingual advantage on measures of fluid reasoning or intelligence are indeed rare, as intellectual ability mostly functions as a control variable in the literature, in order to ensure that any between-group differences are not attributable to differing cognitive ability rather than bilingualism.

The personality correlates of highly creative individuals comprise a constellation of both favorable and unfavorable traits that interact with one another and with situational factors. Some of these traits include: imaginativeness, flexibility, enthusiasm, tolerance of ambiguity, preference for complexity, independence, playfulness, openness to experience, intrinsic motivation, as well as restlessness, nervousness, and low self-control (Barron & Harrington, 1981; MacKinnon, 1962, 1965; Gough & Heilbrun, 1965; Gough, 1979). A now widely accepted construct describing human personality is the five-factor model

(FFM), which characterizes personality variations along the five dimensions of Extraversion, Openness, Conscientiousness, Neuroticism, and Agreeableness. A meta-analysis by Feist (1998) suggested that though scientists and artists varied in the degree to which they were extraverted and conscientious, both groups shared a high level of openness, which is the trait mostly strongly tied to creativity among the Big Five (Dollinger et al., 2004; McCrae, 1987).

Extant work on the personality of bilinguals focuses on whether bilinguals exhibit cross-language differences in personality. For instance, German and Spanish versions of the Neuroticism Extraversion Openness-Five-Factor Inventory were administered to two groups of late bilinguals of these two languages (Veltkamp, Recio, Jacobs, & Conrad, 2013). The researchers found that regardless of individuals' first language, both groups scored higher on Extraversion and Neuroticism when tested in Spanish and scored higher on Agreeableness when tested in German. The result illustrated the presence of cultural frame shifts that were consistent with the cultural norms of the presently used language. However, effects of language on personality shift could be weak and influenced by other factors. Chen and Bond (2010) recruited fluent Chinese-English bilinguals in Hong Kong and used self-reports and behavioral observation to assess self-perceived personality and personality as perceived by others in Chinese and in English. The bilinguals were perceived to be more extraverted, open, and assertive in English than in Chinese, matching their self-derived profiles, but the effect interacted with the ethnicity of the interviewer; cross-language differences existed with Chinese interviewers but not Caucasian interviewers. Personality variables are seldom compared *between* monolinguals and bilinguals, perhaps due to a lack of theoretical motivation for such inquiries. An exception to the norm entailed the investigation of a specific, creativity-related trait, Tolerance of Ambiguity, among monolingual, bilingual, and multi-lingual adults (Dewaele & Li, 2013). Multi-linguals scored the highest on this trait, followed by bilinguals who in turn were higher than monolinguals.

To assess aspects of the creative "Person," we included measures of fluid intelligence (Cattell, 1940) and openness to experience (DeYoung, Quilty, & Peterson, 2007), as well as an adjective checklist containing personality attributes that are characteristically endorsed by creative individuals (Gough, 1979). In addition, we included an objective measure of English proficiency not only because it is a relevant attribute in the discussion of monolingual versus bilingual performance, but also varying degrees of language proficiency have been found to contribute differentially to creativity-related thinking processes (Hommel, Colzato, Fischer, & Christoffels, 2011; Kharkhurin, 2011).

4.1.2. Features of the creative work (the "product")

From the "4-P" approach to creativity, product refers to an idea embodied in "words, paint, clay, metal, stone, fabric, or other material" and communicated to others (Rhodes, 1961, p. 309). The standard view of a creative product meets two essential criteria: novel and useful (Barron, 1955; Stein, 1953). Laboratory assessments of creative production often involve domain experts and trained raters evaluating performance outcomes meeting these criteria, operationalized by interrelated components such as the number of ideas generated (fluency), the number of categories to which the ideas belong (flexibility), and the overall novelty of the ideas (originality) (Guilford, 1950), to name a few. In two separate studies by Kharkhurin (2008, 2009), the Abbreviated Torrance Test for Adults (ATTA), comprising one verbal and two figural paper-and-pencil activities, was administered to college-aged participants in the U.S., in the UAE, and in Iran. The earlier study conducted with the U.S. sample revealed that bilinguals outperformed monolinguals on subcomponents of fluency, flexibility, and elaboration of the task, but not on originality. The later study conducted with the UAE and Iranian samples, however, showed the opposite; bilinguals scored higher than monolinguals on originality but not the other subcomponents. The author attributed these conflicting outcomes to environmental differences of economic, political, social, cultural, educational, and intellectual aspects between the samples, which purportedly influenced creative potential. Language group was also found to interact with the modality of assessment on creative production. On the same ATTA measure, monolinguals performed comparatively better on the verbal activity whereas bilinguals did better on the figural activities (Kharkhurin, 2010).

In our assessment of the creative product, we first included an abridged version of Torrance Tests of Creative Thinking (TTCT) with both verbal and figural components (Torrance, 1974). Secondly, in order to reflect indicators of creative output outside a laboratory setting, as might be indicated by records of accomplishments or honors, ratings of existing products by experts and non-experts, and self-report inventories of achievements (Colangelo, Kerr, Hallowell, Huesman, & Gaeth, 1992; Hocevar & Bachelor, 1989; MacKinnon, 1962; Torrance, 1972), we included a questionnaire assessing ten domains of creative endeavors (Carson, Peterson, & Higgins, 2005). Thirdly, because ideation fluency has been suggested as one of the best indicators of creativity (Feldhusen & Goh, 1995; Runco, 1990), we incorporated an additional test of ideation fluency involving figural processing (Ruff, 1988) aside from the fluency measures that can be obtained from TTCT.

In summary, the main aims of the current study were to examine objective and subjective measures of creative performance in relation to not only language group (monolingual vs. bilingual), but also personality factors and fluid reasoning ability. The relative importance of language group in combination or in interaction with personality traits will also be assessed.

4.2. Method

Materials

Openness/Intellect – The combined term “Openness to Experience/Intellect” refers to two aspects of the Openness to Experience domain of the five-factor model of personality. Openness and Intellect represent related but separate trait dimensions relating primarily to aesthetics and ideas respectively (DeYoung et al., 2007). Each aspect is measured by 10 items, with items answered on a 5-point Likert scale ranging from strongly disagree to strongly agree.

Creative Personality Scale (CPS) – The Creative Personality Scale is a list of 30 adjectives developed from the Adjective Check List to identify creative talent within the individual (Gough, 1979). Eighteen of the items that previous research found to be consistently endorsed by the creative personality receive 1 point when checked; the remaining 12 items receive -1 point when checked. The range of scores on the CPS is therefore from -12 to +18.

Torrance Tests of Creative Thinking (TTCT) – The original TTCT consists of 6 verbal activities and 3 figural activities, requiring a total of 90 minutes for test administration (Torrance, 1974). We selected 3 of the verbal activities (product improvement, improbable situation, and asking questions) based on suggestions from the task manual for an abridged administration, and Activities 2 and 3 of form A from the figural TTCT. The verbal and figural components were each allotted 15 minutes in test administration.

Ruff Figural Fluency Test (RFFT) – This is a neuropsychological test designed to discriminate healthy subjects from those with traumatic head injury, and to explore whether figural fluency would be reduced in proportion to the severity of the injury (Ruff, 1988). The test consists of five different matrices of random dot patterns. For each set of matrices, participants were asked to generate, within 1 minute, as many unique patterns as possible by connecting the dots with straight lines.

The Creative Achievement Questionnaire (CAQ) – The CAQ is a 96-item self-report measure that requires participants to indicate the extent to which their creative achievements have been recognized across 10 domains (visual arts, music, dance, architectural design, creative writing, humor, inventions, scientific discovery, theater and film, and culinary arts) (Carson et al., 2005).

4.3. Results

Different pairs of student raters were trained to assess the creativity of the TTCT-Verbal, TTCT-Figural, RFFT, and the English-speaking and English-writing tasks. The intra-class correlations of rater scores were .990, [.98, .99] for TTCT-Verbal fluency, .968, [.91, .99] for TTCT-Verbal originality, .997, [.996, .998] for TTCT-Figural fluency, .967, [.92, .98] for TTCT-Figural originality, .999, [.998, .999] for RFFT fluency, .866, [.79, .91] for English-speaking, and .832, [.73, .89] for English-writing. The average score of the two raters was used for each of the measures.

The Creative Person

Independent-samples *t*-tests were conducted between language groups on CCF, English proficiency, the Openness/Intellect subscale (of the BFAS), and CPS. See *Table 4-1* for a summary.

Table 4-1. *Language Group Means, SDs, Independent-Samples t-tests, and Confidence Intervals on Variables of the Person.*

	ML	SD	BL	SD	*t*(82)	*p*	95% CI
CCF	30.08	3.77	28.81	4.22	1.44	.154	[-.49, 3.03]
Eng. Prof.	23.27	3.97	21.46	3.68	2.16	.034*	[.14, 3.47]
Open/Intellect	3.57	.52	3.57	.57	.03	.978	[-.24, .24]
Open	3.49	.73	3.63	.70	-.87	.388	[-.45, .18]
Intell	3.65	.50	3.51	.66	1.09	.277	[-.12, .40]
CPS	5.05	3.77	4.04	4.09	1.17	.247	[-.72, 2.74]

Note. ML = monolinguals, BL = bilinguals; CI = confidence interval; CCF = Cattell Culture Fair; CPS = Creative Personality Scale. * $p < .05$.

Monolinguals scored higher than bilinguals on the objective measure of English proficiency by an average of 1.81 points out of a total of 35 points possible, 95% CI [.14, 3.47].

Pearson correlations of the aforementioned variables were then computed separately for the two language groups. See *Table 4-2* and *Table 4-3* for a summary.

Table 4-2. *Pearson Correlations of CCF, English Proficiency, Openness/Intellect, and CPS for Monolinguals (n = 37).*

	1	2	3	4
1. CCF	-			
2. Eng. Prof.	.062	-		
3. Open/Intellect	-.062	.190	-	
4. CPS	-.131	.295[b]	.448**	-

Note. [b] $p = .08$. ** $p < .01$.

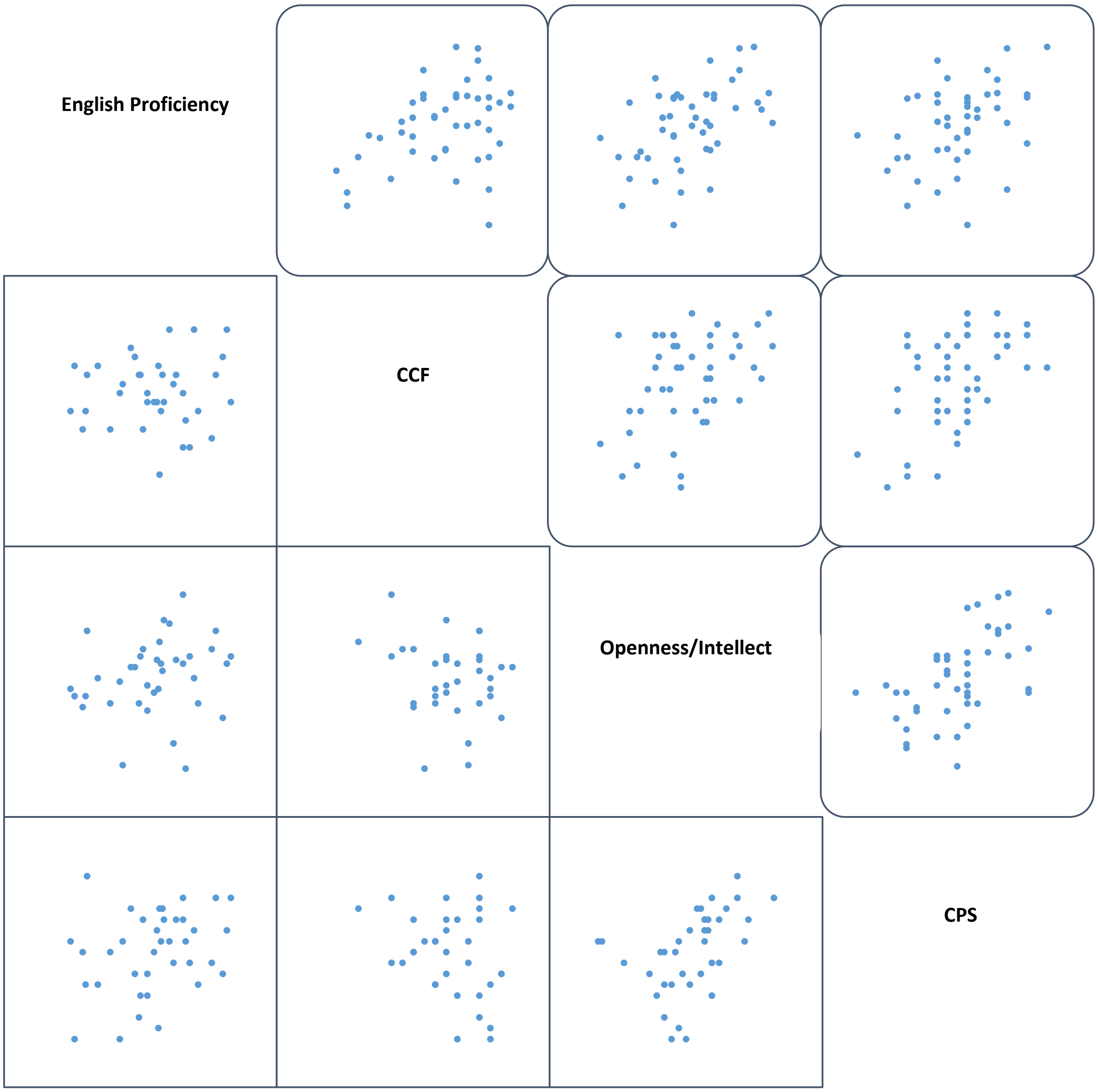
Table 4-3. *Pearson Correlations of CCF, English Proficiency, Openness/Intellect, and CPS for Bilinguals (n = 47).*

	1	2	3	4
1. CCF	-			
2. Eng. Prof.	.355*	-		
3. Open/Intellect	.403**	.476**	-	
4. CPS	.507**	.400**	.532**	-

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

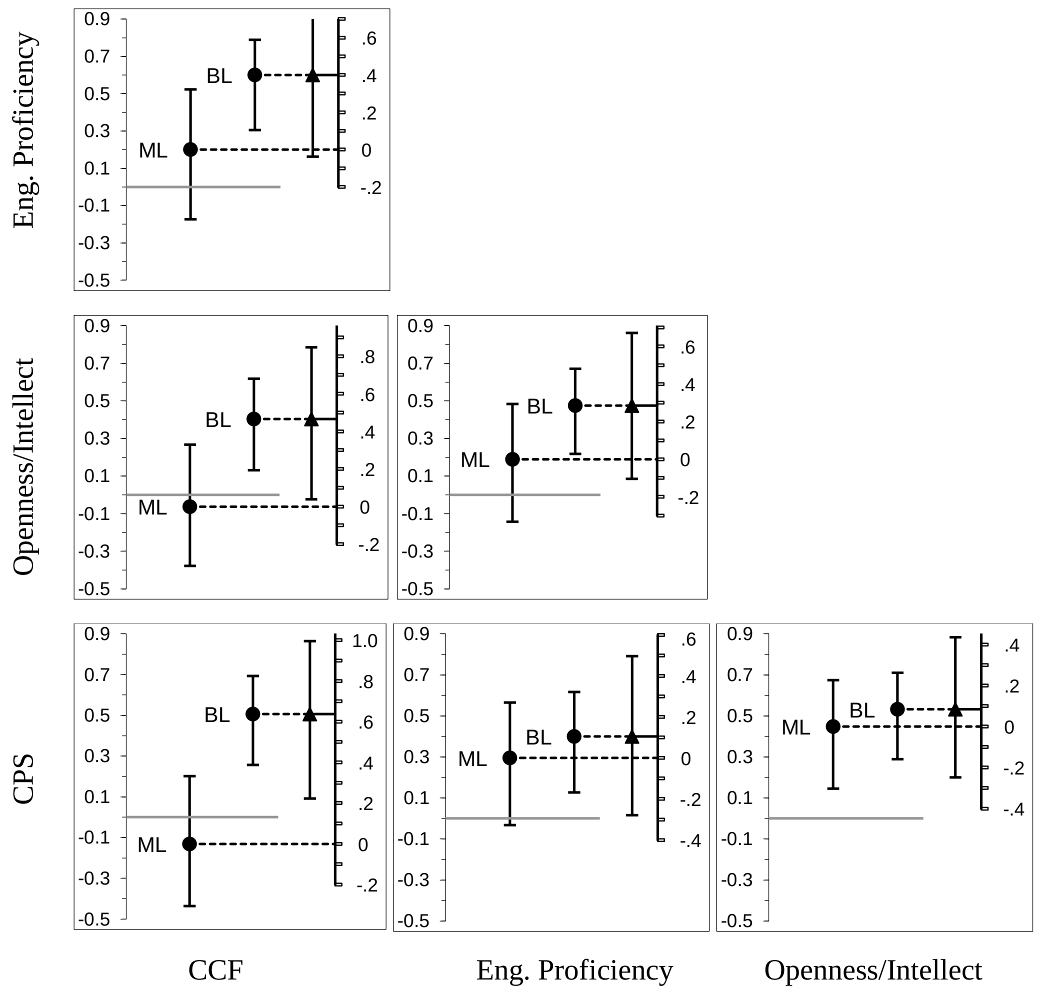
The most striking difference between the two sets of correlation matrices was the magnitude as well as the strength of the correlations. While the variables were generally weakly, if at all, associated for monolinguals with the exception of the correlation between Openness/Intellect and CPS, all of pairwise correlations exhibited greater magnitudes and all reached conventional statistical thresholds in bilinguals. For a visual illustration of the variable relations, see the scatterplot matrix of these variables in *Figure 4.1*.

Figure 4.1. Scatterplot matrix of English Proficiency, CCF, Openness/Intellect, and CPS. Monolinguals are represented by data points in matrices with straight edges in the lower left, and bilinguals are represented by data points in matrices with round corners in the upper right.



These correlations were then compared between language groups with 95% confidence intervals. See *Figure 4.2*.

Figure 4.2. Correlations of monolinguals (ML) and bilinguals (BL) with 95% confidence intervals (CIs). The difference between ML and BL is plotted as the solid triangle on a floating difference axis, with its 95% CI.



As evidenced by *Fig. 4.2,* larger correlations were shown by bilinguals overall. Correlations involving CPS were fairly similar between language groups (except for the CPS-CCF pair), and the largest group difference was observed in correlations involving CCF. Specifically, the correlation of English proficiency and CCF was larger for bilinguals than monolinguals by .29, [-.13, .69]; the correlation of Openness/Intellect and CCF was larger for bilinguals than monolinguals by .47, [.04, .85]; and the correlation of CPS and CCF was larger for bilinguals than monolinguals by .64, [.22, 1.00].

Because of the strong correlations exhibited by bilinguals on the "person" variables, a partial correlation of English proficiency and CCF was conducted controlling Openness/Intellect and CPS in order to identify potential mediators. The result showed a correlation coefficient reduced in strength and magnitude, $r = .14$, $p = .353$, compared with the initial correlation coefficient of $r = .36$, $p = .014$ before the effect of Openness/Intellect and CPS was removed. These data suggest that Openness/Intellect and CPS may mediate the relationship between CCF and English proficiency.

To estimate the effects of CCF on English proficiency directly as well as indirectly through the two aspects of Openness/Intellect and CPS as mediators, a multiple mediation analysis was conducted using ordinary least squares path analysis. See *Figure 4.3* for a conceptual diagram of the model, and *Table 4-4* for a summary of the statistics.

Figure 4.3. A conceptual diagram of a multiple mediation analysis with CCF as the antecedent, English proficiency as the consequent, and Openness, Intellect, and CPS as mediators.

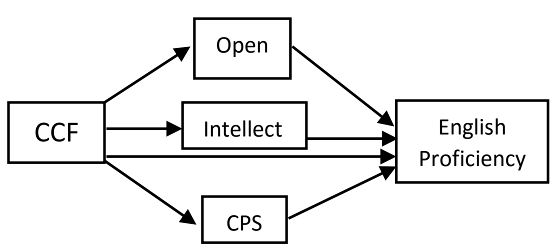


Table 4-4. *Regression Coefficients, Standard Errors, and Model Summary Information for the Personality Influence Parallel Multiple Mediator Model Depicted in Figure 3.*

	Consequent											
	M_1 (Intellect)			M_2 (Open)			M_3 (CPS)			Y (Proficiency)		
Anteced.	Coeff.	SE	*p*	Coeff.	SE	*p*	Coeff.	SE	*p*	Coeff.	SE	*p*
X (CCF)	a_1 .052	.022	.023	a_2 .056	.023	.019	a_3 .491	.125	.0003	*c'* .138	.136	.319
M_1 (Intell.)	--	--	--	--	--	--	--	--	--	b_1 1.600	.943	.097
M_2 (Open)	--	--	--	--	--	--	--	--	--	b_2 .740	.772	.343
M_3 (CPS)	--	--	--	--	--	--	--	--	--	b_3 .097	.160	.549
Constant	i_{m1} 2.011	.640	.003	i_{m2} 2.000	.678	.005	i_{m3} -10.106	3.626	.008	i_y 8.814	4.603	.062
	R^2 = .110			R^2 = .116			R^2 = .257			R^2 = .278		
	F (1, 45) = 5.585			*F* (1, 45) = 5.874			*F* (1, 45) = 15.541			*F* (1, 45) = 4.040		
	p = .023			*p* = .019			*p* = .0003			*p* = .0073		

As can be seen in *Table 4-4*, higher CCF scores led to higher scores on the Intellect aspect (a_1 = .052), the Openness aspect (a_2 = .056), as well as the CPS (a_3 = .491) than lower CCF scores, and higher scores on Intellect, Openness, and CPS in turn may result in higher English proficiency (or b_1 = 1.600, b_2 = .740, and b_3 = .097, respectively), although the path coefficients of b_1, b_2, and b_3 were not significant. Bias-corrected bootstrap confidence intervals were constructed for each of the indirect effects. Based on 5000 bootstrap samples, the interval for the effect mediated by Intellect (a_1b_1 = .083) was above zero (.014 to .243), whereas the intervals for the effects mediated by Openness (a_2b_2 = .042) and CPS (a_3b_3 = .048) straddled zero (-.012 to .149 and -.110 to .210, respectively). There was no evidence that CCF influenced English proficiency independent of all the mediators (c' = 1.009, p = .319).

The mediation analysis suggests that higher CCF led to higher Intellect, which in turn resulted in higher English proficiency. But independent of the aspect of Intellect, there was no evidence of CCF influencing English proficiency through Openness or CPS.

To further assess the Openness/Intellect subscale in relation to language group, a two-way mixed ANOVA was conducted on subscale scores with aspect (Openness/Intellect) as the within-subject factor and language group as the between-subject factor. No main effect of language group emerged, nor a statistically significant aspect by language group interaction, F (1, 82) = 2.99, p = .088. Monolinguals (M = 3.65, SE = .10) scored numerically higher on Intellect than bilinguals (M = 3.51, SE = .09), whereas bilinguals (M = 3.63, SE = .10) scored numerically higher on openness to experience than monolinguals (M = 3.49, SE = .12).

The Creative Product

Independent-samples *t*-tests were conducted on TTCT-V, TTCT-F, CAQ, and RFFT between language groups. See *Table 4-5* for a summary.

Table 4-5. *Language Group Means, SDs, and Independent-Samples t-tests on Creative Products.*

	ML	SD	BL	SD	*t*(82)	*p*	95% CI
TTCT-V-Flu	39.45	11.76	35.16	10.52	1.76	.082	[-.56, 9.13]
TTCT-V-Orig	26.77	10.12	23.82	9.11	1.40	.164	[-1.23, 7.13]
TTCT-F-Flu	21.59	6.99	22.70	6.63	-.74	.460	[-4.08, 1.86]
TTCT-F-Orig	14.08	5.95	15.54	5.10	-1.21	.229	[-3.86, .94]
CAQ	9.27	6.90	14.79	17.01	-2.02[a]	.047*	[-10.97, -.07]
RFFT	16.75	5.23	17.41	5.36	-.57	.573	[-2.98, 1.66]

Note. ML = monolinguals, BL = bilinguals; CI = confidence interval; Flu = fluency; Orig = originality; CAQ = Creative Achievement Questionnaire; RFFT = Ruff Figural Fluency Test.
* $p < .05$. [a] degrees of freedom was 82 except for CAQ, which was 63.70 when equal variances were not assumed.

Monolinguals generated on average 4.29 more acceptable responses than bilinguals on the verbal component of Torrance, [-.56, 9.13], and scored lower than bilinguals on the CAQ by an average of 5.52 points, 95% CI [-10.97, -.07]. Pairwise correlations of the aforementioned variables were then obtained separately for the two language groups. See *Table 4-6* and *Table 4-7* for a summary.

Table 4-6. *Pearson Correlations of TTCT-V, TTCT-F, CAQ, and RFFT for Monolinguals (n = 37).*

	TTCT-V	TTCT-F	CAQ	RFFT
TTCT-V	-			
TTCT-F	.470**	-		
CAQ	.283	-.116	-	
RFFT	.264	.563**	-.221	-

Note. ** *p* < .01. Torrance-V and Torrance-F were correlated on fluency.

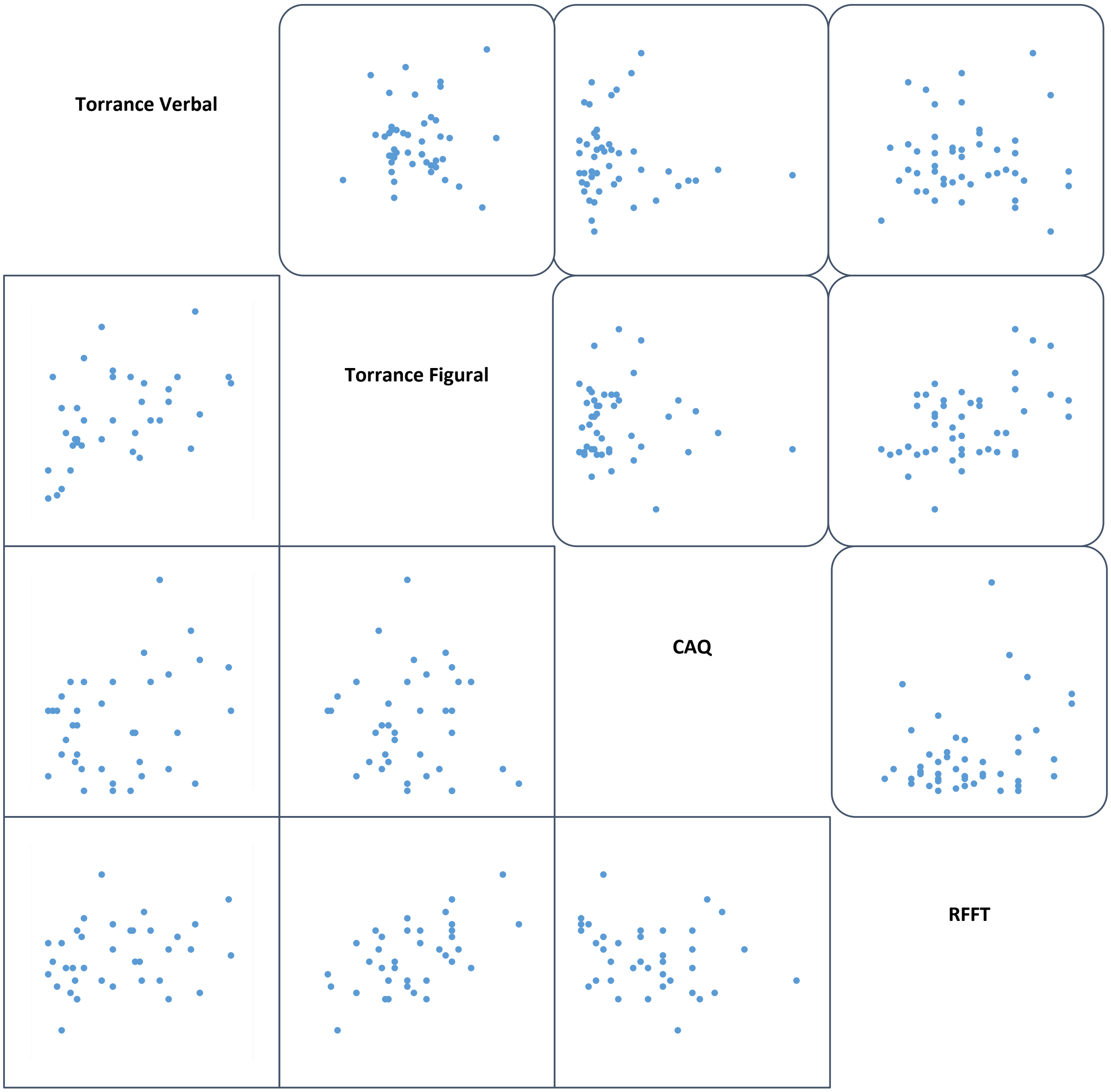
Table 4-7. *Pearson Correlations of TTCT-V, TTCT-F, CAQ, and RFFT for Bilinguals (n = 47).*

	TTCT-V	TTCT-F	CAQ	RFFT
TTCT-V	-			
TTCT-F	.038	-		
CAQ	-.094	-.050	-	
RFFT	-.023	.439**	.256	-

Note. ** *p* < .01. Torrance-V and Torrance-F were correlated on fluency.

The verbal and figural components of Torrance were positively correlated at $r = .47$ for monolinguals and at a much smaller magnitude of $r = .04$ for bilinguals. TTCT-F and RFFT were correlated to a similar degree between the language groups, whereas CAQ and RFFT correlated in opposition directions for monolinguals and bilinguals. To see a visual illustration of the variable relations, refer to the scatterplot matrix of these variables in *Figure 4.4*.

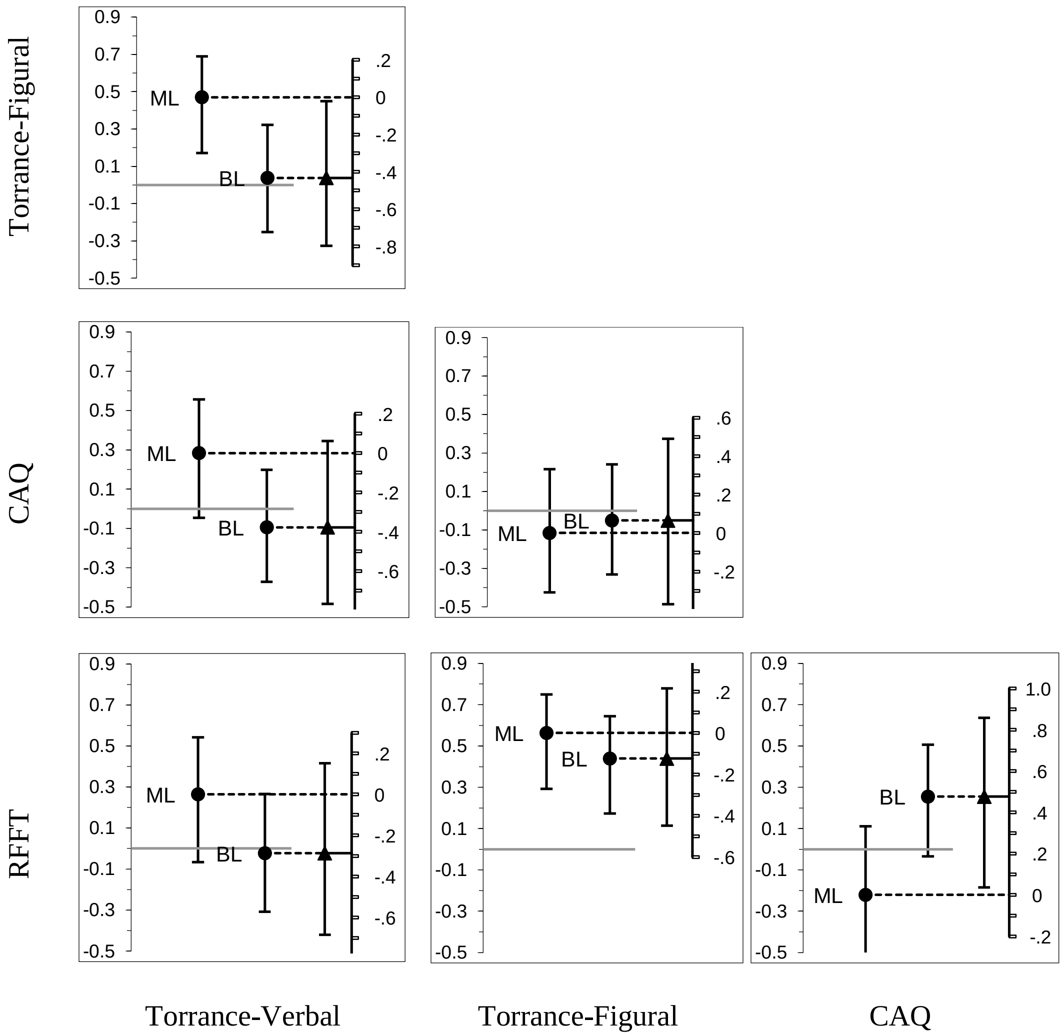
Figure 4.4. Scatterplot matrix of the “product” variables. Monolinguals are represented by data points in matrices with straight edges in the lower left, and bilinguals are represented by data points in matrices with round corners in the upper right.



In the scatterplots involving CAQ, we observed a number of bilinguals exhibiting scores low on CAQ but comparatively higher on the corresponding measures of TTCT-V, TTCT-F, and RFFT, thus forming funnel-shaped distributions in the scatterplots. The primary difference between CAQ and the other product measures lies in the former being an index of the creative products generated by an individual in real life, whereas the other three measures tap into one's creative ability or potential in a laboratory setting. The disjointed connection between these two domains of creative production, particularly among the bilinguals, will be revisited with more detail in the Discussion.

The correlations of the four creative-product measures are compared between language groups with 95% confidence intervals in *Figure 4.5.*

Figure 4.5. Correlations of monolinguals (ML) and bilinguals (BL) with 95% confidence intervals (Cis). The difference between ML and BL is plotted as the solid triangle on a floating difference axis, with its 95% CI.



These graphs reveal that for correlations involving Torrance-Verbal, monolinguals showed numerically larger correlation coefficients than bilinguals. Specifically, the correlation of Torrance-Figural and Torrance-Verbal was larger for monolinguals by .43, [.02, .80]; the correlation of CAQ and Torrance-Verbal was larger for monolinguals by .37, [-.07, .77]; and the correlation of RFFT and Torrance-Verbal was larger for monolinguals by .29, [-.15, .69]. Correlations involving Torrance-Figural (except for Torrance-Verbal) were fairly comparable between language groups.

In contrast to how the variables of the creative person were correlated in *Fig. 2*, we note a trend for monolinguals to show larger correlation coefficients on the creative-product variables, although not as clear cut as the person-variable correlations were for bilinguals. Correlations involving CAQ seem to behave differently from correlations involving other product measures, both evidenced by monolinguals' lower correlation of CAQ and RFFT in light of their generally higher correlations compared to bilinguals', and the funnel-shaped distributions formed by CAQ and other product measures for bilinguals.

Aside from examining the correlation pattern matrix of creative products within and between language groups, a three-way mixed-factor repeated-measures ANOVA was conducted on TTCT scores. Modality (verbal/figural) and creativity dimension (fluency/originality) served as within-subject factors and language group (ML/BL) was the between-subject factor. The main effect of group was not significant, F (1, 82) = .65, p = .42; however, there was a significant modality by group interaction, F (1, 82) = 4.78, p = .032. Whereas the monolinguals (M = 33.11, SE = 1.67) outperformed bilinguals (M = 29.49, SE = 1.48) on the verbal component, bilinguals (M = 19.12, SE = .87) showed superior performance on the figural component (M = 17.84, SE = .98). These results were in accordance with Kharkhurin (2010), which showed a bilingual advantage in non-verbal creativity and a monolingual advantage in verbal creativity.

Person and Product Interactions

To examine how "person" variables bear upon "product" variables, a median split was conducted on each of the "person" variables of English proficiency, CCF, Openness/Intellect, and CPS by language group, forming groups of monolingual low (ML Low), bilingual low (BL Low), monolingual high (ML High), and bilingual high (BL High). Individuals with values at the median on any of the variables were excluded from further analyses, then the effect of each "person" variable on products of TTCT, CAQ, and RFFT was separately analyzed.

Effect of English proficiency on creative products

After excluding 1 monolingual and 1 bilingual at the median English proficiency of their respective language groups, a mixed-factor three-way repeated measures ANOVA was conducted on Torrance scores, with modality (verbal/figural) and creativity dimension (fluency/originality) as within-subject factors and English proficiency by language group (ML Low/BL Low/ ML High/BL High) as the between-subject variable. Results revealed a modality by group interaction, F (3, 78) = 2.75, p = .048. For Torrance-Verbal, the largest difference in performance was observed between BL Low and ML High, with the former producing 5.65 fewer acceptable responses on average than the latter, [-12.13, .82]. BL High and ML Low performed in the middle, producing 4.16 fewer responses, [-10.64, 2.31] and 2.85 fewer responses, [-9.70, 4.01] than ML High, respectively. Results showed that monolinguals high on English proficiency performed the best (M = 34.56, SE = 2.44) and bilinguals low in English proficiency performed the worst (M = 28.90, SE = 2.15), while low monolinguals (M = 31.71, SE = 2.44) and high bilinguals (M = 30.39, SE = 2.15) performed in between. For Torrance-Figural, on the other hand, the groups did not differ statistically, F (3, 78) = 1.28, p = .287, nor was there a dimension by group interaction, F (3, 78) = 1.57, p = .205. ML High produced 1.55 fewer responses on average than BL High,

[-5.28, 2.19], who produced 1.22 fewer responses on average than ML Low, [-4.95, 2.52], who in turn produced .70 fewer responses than BL Low, [-4.43, 2.04].

What can be deduced from these findings is that the combination of English proficiency and language group affected performance on the verbal and figural components of TTCT differentially. Monolinguals high on English proficiency tended to outperform bilinguals low on English proficiency on the verbal component, whereas the four groups performed comparably on the figural component of TTCT.

Next, one-way ANOVAs were conducted on CAQ and RFFT with English proficiency by language group (ML Low/BL Low/ML High/BL High) as the between-subject variable. No statistically significant differences were detected among the groups on CAQ, F (3, 78) = 1.01, p = .395, or on RFFT, F (3, 78) = .40, p = .75.

Effect of CCF on creative products

Three monolinguals and 5 bilinguals with the median CCF scores of their respective language groups were eliminated from the analysis. Then a mixed-factor three-way repeated measures ANOVA was conducted on Torrance scores, with modality (verbal/figural) and creativity dimension (fluency/originality) as within-subject factors and CCF by language group (ML Low/BL Low/ML High/BL High) as the between-subject variable. Results revealed a modality by group interaction, F (3, 72) = 3.07, p = .033. For Torrance-Verbal, the largest difference in performance was observed between BL Low and the other three groups. Specifically, BL Low scored lower than ML Low by an average of 7.15 points, [-13.67, -.63], lower than BL High by an average of 7.40 points, [-13.72, -1.08], and lower than ML High by an average of 7.58 points, [-14.10, -1.05]. Results showed that bilinguals low on CCF (M = 25.84, SE = 2.13) performed the verbal component of TTCT worse than monolinguals low on CCF (*M* = 32.99, *SE* = 2.48), bilinguals high on CCF (*M* = 33.24, *SE* = 2.35), and monolinguals high on CCF (*M* = 33.41, *SE* = 2.48). For Torrance-Figural, on the other hand, the groups did not differ statistically, F (3, 72) = 1.30, p = .280, nor was there a dimension by group interaction, F (3, 72) = .13, p = .941. ML Low scored lower than BL Low by an average of 2.25 points, [-6.15, 1.65], lower than ML High by an average of 2.50 points, [-6.68, 1.68], and lower than BL High by an average of 4.01 points, [-8.08, .06].

What can be deduced from these findings is that the combination of CCF and language group affected performance on the verbal and figural components of TTCT differentially. Bilinguals low on CCF performed most poorly on the verbal component, whereas the four groups performed more comparably on the figural component of TTCT.

Next, one-way ANOVAs were conducted on CAQ and RFFT with CCF by language group (ML Low/ML High/BL Low/BL High) as the between-subject variable. No statistically significant differences were detected among the groups on CAQ, F (3, 72) = 1.95, p = .129, or on RFFT, F (3, 72) = 1.62, p = .191.

Effect of Openness/Intellect (O/I) on creative products

One monolingual and 3 bilinguals with the median O/I scores of their respective language groups were eliminated from the analysis. A mixed-factor three-way repeated measures ANOVA was conducted on Torrance scores, with modality (verbal/figural) and creativity dimension (fluency/originality) as within-subject factors and O/I group (ML Low/BL Low/ML High/BL High) as the between-subject variable. Results showed a significant dimension by group interaction, F (3, 76) = 3.95, p = .011; monolinguals low on O/I exhibited the largest score difference between dimensions of fluency and originality, primarily due to their relatively lower originality score compared to the other three groups.

Then one-way ANOVAs were conducted on CAQ and RFFT with O/I by language group (ML Low/BL Low/ML High/ BL High) as the between-subject variable. The groups statistically differed on CAQ, F (3, 76) = 3.82, p = .013; bilinguals high on O/I scored the highest on CAQ, by an average of 13.34 points, [-24.68, -2.00] than ML Low; by an average of 9.74 points, [-20.61, 1.14] than BL Low; and by an average of 8.39 points, [-19.73, 2.95] than ML High. No statistically significant differences were detected among the groups on RFFT, F (3, 76) = .35, p = .787.

Effect of CPS on creative products

Five monolinguals and 3 bilinguals with the median CPS scores of their respective language groups were eliminated from the analysis. A mixed-factor three-way repeated measures ANOVA was conducted on Torrance scores, with modality (verbal/figural) and creativity dimension (fluency/originality) as within-subject factors and CPS group (ML Low/BL Low/ML High/BL High) as the between-subject variable. Neither the main effect of CPS group, F (3, 72) = .89, p = .451, nor interactions involving CPS group were statistically significant.

Then one-way ANOVAs were conducted on CAQ and RFFT with CPS by language group (ML Low/BL Low/ML High/BL High) as the between-subject variable. The groups statistically differed on CAQ, F (3, 72) = 4.05, p = .010; bilinguals high on CPS scored the highest on CAQ; by an average of 12.53 points, [-23.52, -1.54] than ML Low; by an average of 10.89 points, [-21.42, -.35] than BL Low; and by an average of 5.91 points, [-

17.74, 5.92] than ML High. No statistically significant differences were detected among the groups on RFFT, F (3, 72) = 1.46, p = .233.

In brief summary of the interactions of "person" variables and creative products, bilinguals low on English proficiency or CCF performed the poorest and monolinguals high on English proficiency performed the best when the creative assessment was primarily conducted in the verbal modality. Bilinguals high on personality factors of O/I and CPS and monolinguals low on the same factors scored the highest and lowest on the CAQ, respectively.

The main results from the sections of "person," "product," and person-product interactions were as follows: 1) CCF was significantly correlated with English proficiency as well as personality variables in bilinguals, but not in monolinguals. The significant correlation between CCF and English proficiency was greatly reduced, however, when controlling for personality factors of O/I or CPS; 2) monolinguals outperformed bilinguals on creative tasks dependent on the verbal modality, whereas bilinguals outperformed monolinguals on creative tasks dependent on figural processing; 3) as a group, bilinguals scored higher than monolinguals on CAQ, a scale reflecting overt behaviors of creative performance in real life; 4) bilinguals low on English proficiency or CCF would be at particular risk for performing poorly on verbal creative tasks, and monolinguals low on creativity-related personality traits would be likely to underperform in real-life creative endeavors.

4.4. Discussion

In the present study, we investigated selective characteristics of the "person" and "product" from the perspective of the "Four P" model of creativity both between and within language groups. In addition, we examined "person" and "product" interactions in combination with language group in order to develop a nuanced understanding of the relations between language experiences and creative performance.

Perhaps the most surprising finding pertained to English proficiency being correlated with CCF in bilinguals but not in monolinguals. A monolingual's basic competence in his or her native language is unrelated to fluid reasoning ability, as we would expect. However, for bilinguals who had acquired both (or all) of their languages by the age of 8 and with above-average self-rated abilities in these languages, English proficiency was significantly correlated with CCF at a moderately high magnitude. Furthermore, the Intellect aspect of the O/I subscale (but not the Openness aspect) was found to mediate the relationship between CCF and English proficiency, providing support for Intellect and Openness being distinct aspects (DeYoung, Quilty, Peterson, & Gray, 2010). The mediation outcome suggests that among bilinguals, higher fluid reasoning ability led to more pronounced interests in truth and ideas (cognitive behavioral traits more closely related to Intellect) rather than beauty perceived by the senses (traits more closely related to Openness to Experience), which in turn resulted in higher English proficiency.

In terms of data patterns observed within and between language groups, the "person" variables of English proficiency (self-report or objectively measured), fluid intelligence, and personality factors were significantly correlated with one another in bilinguals, whereas the same correlational pattern was absent in monolinguals. Pertaining to the "product" variables, the fluency measure of TTCT-V correlated with the fluency measure of TTCT-F only in monolinguals at .47, which was consistent with the correlation of .36 obtained between full scales of TTCT-V and TTCT-F in a sample of college students

(Clapham, 2004). Bilinguals showed no correlation between verbal and figural fluency scores on the Torrance, and both language groups exhibited comparable correlations between the figural measures of TTCT-F and RFFT. These results, taken together, suggest that creativity measures dependent on verbal processing may not reflect the same norming standards for bilinguals as they do for monolinguals.

Bilinguals *as a group* scored higher than monolinguals on CAQ, which is a self-report scale measuring creative endeavors and achievements in real life. The higher group mean of bilinguals, however, was largely driven by a few individuals as evidenced by the group's smaller median, i.e., 8 vs. 9, and larger standard deviation, i.e., 17.01 vs. 6.90, than observed for monolinguals. Regarding CAQ's relations with other creative-product measures, it weakly and positively correlated with a verbal creativity measure (TTCT-V) in monolinguals and with a figural fluency measure (RFFT) in bilinguals. For the most part, though, CAQ as a measure of realized creativity correlated little with the creative potential assessed in the laboratory, which is consistent with the finding of Zabelina and Robinson (2010). A selective few of the bilinguals demonstrated an extraordinary level of creative output in real life, a phenomenon not observed in the monolinguals.

In addition, we replicated Kharkhurin (2010), which demonstrated a bilingual advantage in non-verbal creativity and a monolingual advantage in verbal creativity on the Abbreviated Torrance Tests for Adults (ATTA). The monolinguals in our study outperformed the bilinguals on the verbal component of Torrance collapsing fluency and originality subscales, whereas the bilinguals outperformed the monolinguals on the figural component of Torrance collapsing fluency and originality subscales. The replication was obtained even though criterion-referenced indicators were used by Kharkhurin (2010) for creative performance and we adopted norm-referenced indicators.

With regards to whether bilinguals are more creative than monolinguals, the answer is likely “it depends” – on the characteristics of the monolingual or bilingual person, as well as the type of creative assessment. Generally, monolinguals outperform bilinguals on tasks of verbal creativity and bilinguals outperform monolinguals on tasks of non-verbal creativity. Combining language-group status with the “person” variables revealed that bilinguals low on English proficiency or CCF would be at particular risk for performing poorly on creative tasks dependent on verbal processing, whereas monolinguals low on O/I and CPS would tend to underperform in real-life creative endeavors.

Chapter 5

General Discussion and Conclusion

In very broad strokes, monolinguals and bilinguals did not differ on executive functions or conflict monitoring (or its theoretical alternative of feature integration), whereas the language groups demonstrated differential advantages on creativity tasks administered in different modalities. Specifically, monolinguals performed better on a standardized test of verbal creativity and bilinguals on a standardized assessment of figural creativity. More fine-grained analyses revealed either a lack of or inconsistent differences between high-proficient and low-proficient bilinguals across experimental tasks; cognitive and personality characteristics of the monolingual or bilingual individual, in conjunction with the type of creative assessment, influenced the level of creative output; and the Intellect aspect of the Openness/Intellect subscale mediates the relationship between bilinguals' fluid reasoning and English proficiency.

A rather interesting pattern that emerged from the three previous chapters was that the EF variables correlated more strongly among one another in bilinguals than in monolinguals, and the same between-group difference was observed in the pairwise correlations of "person" variables of English proficiency, fluid intelligence, and personality measures. These two sets of variables are not ostensibly similar, as one set pertained to tasks and the other set to participant characteristics. Yet stronger inter-correlations were obtained with regards to both of these sets among bilinguals than monolinguals. At the same time, bilinguals did not exhibit a correlation between scores on the verbal versus figural components of the Torrance Test of Creative Thinking, as one would expect for those measures, whereas monolinguals did. These results seem to indicate that monolinguals and bilinguals indeed exhibit different data patterns but it is not entirely clear what this difference means.

The college-aged monolingual and bilingual participants in the current study were comparable to one another on demographic, socioeconomic, and cognitive variables. Group differences were found, however, with regards to years of education and scores on the anxiety scale of the BSI. Bilinguals as a group had approximately 1 more year of formal education because of the greater number of graduate students in this subgroup. More years of formal education, however, did not appear to aid the bilinguals in performing better than their monolingual peers. In addition, bilinguals reported higher scores on the anxiety scale, which reflects symptoms usually associated with high manifest anxiety and is a good indicator of current stress levels (Derogatis & Melisaratos, 1983). Few studies involving language-group comparisons have actually reported such anxiety data; Murphy (1990) compared state and trait anxiety between monolingual suburban and bilingual urban children and found no group differences on either anxiety type. To further explore the potential effect of anxiety on task performance in the current study, the monolingual and bilingual individuals were matched on anxiety scores and included in the same set of robust analyses used to investigate EF differences between language groups, as well as the same set of analyses used to examine conflict monitoring vs. its theoretical alternative of feature integration between language groups. Notably, the results remained largely unchanged. It is unclear if the language-group difference in anxiety is due to bilingualism, some "third variable," or reflects a Type I error (especially given the number of pairwise comparisons considered across the demographic variables).

We included general assessments of video gaming and music experience because training involving either of these variables has been found to enhance performance on some EF tasks (Bavelier, Green, Pouget, & Schrater, 2012; Bugos, Perlstein, McCrae, Brophy & Bedenbaugh, 2007; Merono et al., 2011). The monolingual and bilingual participants in our study did not differ on these variables, though it should be noted that the measures used did not differentiate between active video game experience and non-active video gaming experience. When incorporating both music training and language experience, researchers in a recent study found reduced local and global switch costs in musicians compared with non-musicians but no advantage for bilinguals relative to monolinguals (Moradzadeh et al., 2015).

The participation rate of bilinguals in the laboratory portion of the study was nearly double that of monolinguals. This differential participation rate likely reflects bilinguals' greater interest in studies that "explore the relations between language use and various kinds of task performance" — as mentioned in the invitation email.

Because a randomized group design cannot be utilized in experiments comparing monolingual and bilingual persons, it is important to consider whether the language groups are comparable in all possible respects except for language experiences per se. In the current study, the inclusion of a wide range of demographic, socioeconomic, and cognitive variables was an effort to identify potentially confounding or nuisance variables that could bias study outcomes. The language groups were found to differ on formal years of education and the anxiety scale of the BSI, but these differences did not appear to have affected experimental outcomes in any observable manner. One could argue that "language experiences per se" is itself an oxymoron as language does not exist in a vacuum but is intimately intertwined with culture, worldview, and values; any study investigating the effect of language cannot but also encompass the effects of those closely associated concepts. But would we expect culture, worldview, and values to exert an influence on the performance of domains of EF, conflict monitoring, and creativity? Our speculation is that creative performance might be more directly affected than the other two domains, which could be a reason for our observing a bilingual advantage in the creative domain but not in the other two domains.

The bilingual advantage found in non-verbal creativity (and the monolingual advantage in verbal creativity) replicated Kharkhurin (2010), who used criterion-referenced scoring with English monolingual and a homogenous group of Russian-English bilingual college students in the U.S., whereas we adopted norm-referenced scoring with English monolingual and a heterogeneous group of bilingual college students. Our general null findings between language groups in domains of executive function and conflict monitoring are consistent with the recent summaries of findings involving a number of non-verbal interference and switching tasks (Paap et al., 2014), as well as the Simon and flanker tasks published within the last five years (Hilchey et al., 2014).

Some researchers caution against a simple binary approach to comparing language groups (Mishra, 2015; Titone et al., 2015). Instead, the research strategy of comparing different levels of bilingualism on the dimensions of proficiency, switching frequency, etc. is advocated particularly in linguistic contexts where “pure” monolinguals do not exist. On the one hand, a more fine-grained approach of this sort has the potential of yielding rich and detailed information about the effects of various types of bilingual experiences, as illustrated by our data on the performance difference between bilinguals of high vs. low

English proficiency on the sequential modulation task for instance. On the other hand, the lack of language-group differences that we observed across a wide range of executive function and cognitive control tasks seems to also speak to limits on the scope and/or magnitude of the advantage that may be conferred by bilingualism-related experiences. After all, if an advantage can only be realized under very specific and limiting conditions, the practical significance of these experiences would undoubtedly be diminished.

Proponents of a bilingual advantage have argued that the dearth of studies supporting a language-group difference in the college-aged population is because young adults, both monolingual and bilingual, are functioning in their cognitive prime, which makes detecting differences difficult (Kroll & Bialystok, 2013). By this line of reasoning, we should expect more children and older-adult studies to show a bilingual advantage than young-adult studies. Indeed, of the studies listed by Paap et al. (2014), 10/45 experiments, or about 22% of the experiments with young adults as participants reported a bilingual advantage, whereas 47% (9/19) and 43% (3/7) with children and older adults as participants did, respectively. The number of significant findings reported for studies of children and older adults indeed is almost twice that for young adults, which likely means one of two things.

One possible interpretation is that given the potential threat of the file-drawer problem and publication bias, the actual percentage of significant findings will be lower than currently reported, especially as the number of publications on the children and older adult populations approaches that involving young adults. A second possible interpretation of the extant data is simply that the bilingual advantage is real and is more evident in children and older adults than in young adults. This claim would be in partial agreement with Valian (2015), who has concluded that the data demonstrating cognitive benefits of bilingualism for children and young adults are currently inconclusive, but a bilingual advantage seems to be present among older adults. Contributing to such a narrative are the findings of Abutalebi et al. (2014), who examined the decrease of cerebral gray-matter volume in healthy aging, and found bilinguals to exhibit significantly increased gray-matter volume in left temporal pole—postulated to act as an amodal semantic hub—whereas the monolingual brain showed more age-related gray-matter decreases. Also, retrospective studies conducted with samples in Canada (Toronto and Montreal), India (Hyderabad), and Belgium (Ghent and Brussels) show a significant effect of speaking two or more languages in delaying the onset of Alzheimer's disease by up to 4-5 years (Alladi et al., 2013; Freedman et al. 2014; Woumans et al., 2014). Valian (2015), however, questions how representative such retrospective studies are of the general monolingual and bilingual populations, since many individuals with or without cognitive impairments would never appear at a memory clinic. She further contends that prospective studies with community samples that follow individuals over time do not tend to show advantages of bilingualism.

The internal consistency of the English proficiency assessment used in the current study was low, or .56 measured with Cronbach's alpha. Replications conducted on other bilingual samples with a more valid language assessment tool would be imperative to shedding light on associations between language proficiency and task performance. A broader and closely related issue is how to best operationally define and assess bilingualism in controlled experiments. Proficiency, dominance, degree of bilingualism, age of acquisition, frequency of language use, and the context of language use are just a few variables used by researchers to characterize their bilingual samples. However, well-validated and standardized instruments for measuring language proficiency, language dominance, and degree of bilingualism are largely lacking. Gollan and colleagues (Gollan et al., 2012; Sheng et al., 2013) have contributed to this body of work by designing a naming test specifically for bilinguals and assessing the convergence and divergence between subjective and objective measures of proficiency and dominance in homogenous bilingual samples of children, young adults, and older adults. But what about bilingual groups characterized by considerable language heterogeneity? Is there a viable way to objectively assess proficiency in a heterogeneous group of bilinguals, especially given the likelihood of one language being more proficient than the other (Gathercole & Thomas, 2009)? Measuring objective English proficiency was a step we took toward addressing this methodologically and empirically complex, yet practically important question. We understand, however, that the method is far from ideal as some of the bilinguals in the sample English would have been, whereas for others (i.e., 7 bilinguals in the sample) it would not have been, their more proficient language.

Our findings, alongside the recent developments in the field, do not unequivocally demonstrate that bilingualism yields a definitive advantage or no advantage for cognitive functioning. Collectively, our findings suggest that there are complex interrelations among language use and higher-order problem-solving abilities, including enduring personality traits revolving around cognitive exploration. We conclude, that, at least for young adults, there is no uniform overall cognitive advantage conferred by bilingualism, but facility in two or more languages can beneficially influence measures of nonverbal creativity.

Bibliography

Abutalebi, J., Canini, M., Della Rosa, P. A., Sheung, L. P., Green, D. W., & Weekes, B. S. (2014). Bilingualism protects anterior temporal lobe integrity in aging. *Neurobiology of Aging, 35*, 2126-2133.

Abutalebi, J., Della Rosa, P. A., Green, D. W., Hernandez, M., Scifo, P., Keim, R., … Costa, A. (2012). Bilingualism tunes the anterior cingulate cortex for conflict monitoring. *Cerebral Cortex*, *22,* 2076-2086.

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*, 207-245.

Akçay, Ç., & Hazeltine, E. (2007). Conflict monitoring and feature overlap: Two sources of sequential modulation. *Psychonomic Bulletin & Review, 14*, 742 – 748.

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A., … Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology, 81*, 1938-1944.

Amabile, T. M. (1996). Creativity in context: Update to the social psychology of creativity. Boulder, CO: Westview Press.

Antón, E., Duñabeitia, J. A., Estévez, A., Hernández, J. A., Castillo, A., Fuentes, L. J., … Carreiras, M. (2014). Is there a bilingual advantage in the ANT task? Evidence from children. *Frontiers in Psychology, 5:* 398.

Banich, M. T. (2009). Executive function: The search for an integrated account. *Current Directions in Psychological Science, 18,* 89-94.

Barron, F. (1955). The disposition towards originality. *Journal of Abnormal and Social Psychology, 51*, 478-485.

Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, *32*, 439-476.

Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience, 35*, 391–416.

Bialystok, E. (1986). Factors in the growth of linguistic awareness. *Child Development, 57*, 498-510.

Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition.* Cambridge, UK: Cambridge University Press.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60,* 68-79.

Bialystok, E., Craik, F. I. M., & Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, *45,* 459-464.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, Y., Gunji, A., & Pantev, C. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage*, *24,* 40-49.

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*, *32*, 1341-1354.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Linguistics, 24*, 27-44.

Biedroń, A. (2010). Personality factors according to the Five Factor model in gifted foreign language learners. In G. Perez-Bustamante, K. Phusavat, & F. Ferreira (Eds.), *Proceedings of the IASK International Conference Teaching and Learning 2010, Seville, Spain, 29 November – 01 December 2010* (pp. 148-154). International Association for the Scientific Knowledge.

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in stimulus-stimulus inhibition. *Bilingualism: Language and Cognition, 17*, 610-629.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624-652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*, 539-546.

Biedroń, A. (2011). Personality factors as predictors of foreign language aptitude. *Studies in Second Language Learning and Teaching, 1*, 467-489.

Bruck, M., Lambert, W. E., & Tucker, G. R. (1977). Cognitive consequences of bilingual schooling: The St. Lambert project through grade six. *Linguistics, 187*, 13-33.

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging and Mental Health, 11*, 464–471.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Ed.), *Methodology of frontal and executive function* (pp. 81-116). Hove, HK: Psychology Press.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*, 215-222.

Byron, K., & Khazanchi, S. (2011). A Meta-analytic investigation of the relationship of state and trait anxiety to performance on figural and verbal creative tasks. *Personality and Social Psychology Bulletin, 37*, 269-283.

Campbell, R., & Sais, E. (1995). Accelerated metalinguistic (phonological) awareness in bilingual children. *British Journal of Developmental Psychology, 13*, 61-68.

Carringer, D. C. (1974). Creative thinking abilities of Mexican youth: The relationship of bilingualism. *Journal of Cross-Cultural Psychology, 5*, 492–504.

Carson, S. H., Peterson, J. B., & Higgins, D. M. (2005). Reliability, validity, and factor structure of the Creative Achievement Questionnaire. *Creativity Research Journal, 17*, 37-50.

Cattell, R. B. (1940). A culture-free intelligence test: Part I. *Journal of Educational Psychology*, *31*, 161-179.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.

Cattell, R. B. (1973). Manual for the Cattell culture fair intelligence test. Champaign, IL:

Institute for Personality and Ability Testing.

Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin, 36*, 1514-1528.

Cheng, C.-Y., Sanchez-Burks, J., & Lee, F. (2008). Connecting the dots within: Creative performance and identity integration. *Psychological Science*, *19*, 1178-1184.

Christoffels, I. K., de Haan, A. M., Steenbergen, L., van den Wildenberg, W. P. M., & Colzato, L. S. (2014). Two is better than one: Bilingual education promotes the flexible mind. *Psychological Research, 79*, 371-379.

Clapham, M. M. (2004). The convergent validity of the Torrance Tests of Creative Thinking and Creativity Interest Inventories. *Educational and Psychological Measurement, 64*, 828 – 841.

Coderre, E. L., & van Heuven, W. J. B. (2014). Electrophysiological explorations of the bilingual advantage: Evidence from a Stroop task. *PLoS ONE, 9*, e10324

Colangelo, N., Kerr, B., Hallowell, K., Huesman, R., & Gaeth, J. (1992). The Iowa inventiveness inventory: Toward a measure of mechanical inventiveness. *Creativity Research Journal, 5*, 157-163.

Collette, F., Van Der, L. M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. *Human Brain Mapping, 25*, 409–423.

Colzato, L. S., Bajo, M. T., Wildenberg, W. v. d., Paolieri, D., Nieuwenhuis, S., Heij, W. L., & Hommel, B. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 302-312.

Colzato, L. S., Zmigrod, S., & Hommel, B. (2013). Working memory updating predicts individual differences in updating stimulus-response episodes. *Visual Cognition, 21*, 13-22.

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.

Cromdal, J. (1999). Childhood bilingualism and metalinguistic skills: Analysis and control in young Swedish-English bilinguals. *Applied Psycholinguistics*, *20*, 1-20.

De Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, *26*, 99-107.

Decker, S. L., Hill, S. K., & Dean, R. S. (2007). Evidence of construct similarity in executive functions and fluid reasoning abilities. *International Journal of Neuroscience 117*, 735-748.

Derogatis, L. R., & Melisaratos, N. (1983). The brief symptom inventory – An introductory report. *Psychological Medicine, 13*, 595-605.

Dewaele, J.-M., & Li, W. (2013). Is multilingualism linked to a higher tolerance of ambiguity? *Bilingualism: Language and Cognition, 16*, 231-240.

DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facts and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology, 93*, 880-896.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.

Dollinger, S. J., Urban, K. K., & James, T. A. (2004). Creativity and openness: Further validation of two creative product measures. *Creativity Research Journal, 16*, 35-47.

Dörnyei, Z. (2005). The Psychology of the Language Learner: Individual differences in second language acquisition. Mahwah NJ: Lawrence Erlbaum Associates.

Duñabeitia, J. A., Hernández, J. A., Antón, E., Macizo, P., Estévez, A., Fuentes, L. J., & Carreiras, M. (2013). The inhibitory advantage in bilingual children revisited. *Experimental Psychology, 61*, 234–251.

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: An update. *Frontiers in Psychology, 5*, [article 1001].

Egner, T. (2007). Congruency sequence effects and cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 380-390.

Emmorey, K., Luk, G., Pyers, J. E., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals. *Psychological Science, 19*, 1201-1206.

Engel de Abreu, P. M. (2011). Working memory in multilingual children: Is there a bilingual effect? *Memory, 19*, 529-537.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor model theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 145-199). New York, NY: Elsevier.

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review, 2*, 290-309.

Feldhusen, J. F., & Goh, B. E. (1995). Assessing and accessing creativity: An integrative review of theory, research, and development. *Creativity Research Journal, 8*, 231-247.

Festman, J. & Münte, T. F. (2012). Cognitive control in Russian-German bilinguals. *Frontiers in Psychology, 3*: 115.

Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions, 6*: 5.

Finke, R. A., Ward, T. B., & Smith, S. M. (1992). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.

Fisk, J., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology, 26*, 874–890.

Freedman, M., Alladi, S., Chertkow, H., Bialystok, E., Craik, F. I. M., Phillips, N. A., … Bak, T. H. (2014). Delaying onset of dementia: Are two languages enough? *Behavioral Neurology, 2014*: 808137.

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*, 172-179.

Galambos, S. J., & Hakuta, K. (1988). Subject-specific and task-specific characteristics of metalinguistic awareness in bilingual children. *Applied Psycholinguistics, 9*, 141-162.

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., … Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage, 53,* 1272-1278.

Garbin, G., Sanjuan, A., Forn, C., Bustamante, J. C., Rodriguez-Pujadas, A., Belloch, V., … Ávila, C. (2010). Bridging language and attention: Brain basis of the impact of bilingualism on cognitive control. *NeuroImage, 53,* 1272-1278.

Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition, 12,* 213-237.

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*: 11.

Geary, D. C., Cormier, P., Goggin, J. P., Estrada, P., & Lunn, M. C. E. (1993). Mental arithmetic: A componential analysis of speed-of-processing across monolingual, weak bilingual, and strong bilingual adults. *International Journal of Psychology, 28*, 185-201.

Goclowska, M. A., & Crisp, R. J. (2014). How dual-identity processes foster creativity. *Review of General Psychology, 18,* 216-236.

Goetz, P. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition, 6,* 1-15.

Goldman, M. C., Negen, J., & Sarnecka, B. W. (2014). Are bilingual children better at ignoring perceptually misleading information? A novel test. *Developmental Science, 17*, 956-964.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish-English bilinguals. *Bilingualism: Language and Cognition*, *15,* 594-615.

Goral, M., Campanelli, L., & Spiro, A. (2015). Language dominance and inhibition abilities in bilingual older adults. *Bilingualism: Language and Cognition, 18*, 79-89.

Gough, H. G. (1979). A Creative Personality Scale for the Adjective Check List. *Journal of Personality and Social Psychology*, *37*, 1398-1405.

Gough, H. G., & Heilbrun, A. B., Jr. (1965). *The Adjective Check List Manual*. Palo Alto, CA: consulting Psychologists Press.

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

GRE: The Official Guide to the Revised General Test, 2nd Edition (2012).

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5,* 444-454.

Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, *13*, 622-635.

Hayes, A. F. (2013). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach methodology in the social sciences. New York: The Guilford Press.

Hedden, T., & Yoon, C. (2006). Individual differences in executive processing predict susceptibility to interference in verbal working memory. *Neuropsychology, 20*, 511–528.

Hernandez, A. E., & Kohnert, K. J. (2015). Investigations into the locus of language-switching costs in older bilinguals. *Bilingualism: Language and Cognition, 18*, 51-64.

Hernández, M., Martin, C. D., Barceló, F., & Costa, A. (2013). Where is the bilingual advantage in task-switching? *Journal of Memory and Language, 69*, 257-276.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625–658.

Hilchey, M. D., Saint-Aubin, J., Klein, R. M. (2014). Does bilingual exercise enhance cognitive fitness in non-linguistic executive processing tasks. In J. W. Schwieter (Ed.), *Cambridge Handbook of Bilingual Processing*. Cambridge University Press.

Hocevar, D., & Bachelor, P. (1989). A taxonomy and critique of measurements used in the study of creativity. In J. A. Glover, R. R. Ronning & C. R. Reynolds (Eds.), *Handbook of Creativity* (pp. 53-75). New York: Plenum Press.

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition, 5*, 183-216.

Hommel, B., Colzato, L. S., Fischer, R., & Christoffels, I. K. (2011). Bilingualism and creativity: Benefits in convergent thinking com with losses in divergent thinking. *Frontiers in Psychology, 2*:273.

Hommel, B., Proctor, R. W., & Vu, K. L. (2004). A feature-integration account of sequential effects in the Simon task. *Psychological Research*, *68*, 1-17.

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia, 44,* 2017–2036.

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

Kessler, C., & Quinn, M. E. (1987). Language minority children's linguistic and cognitive creativity. *Journal of Multilingual and Multicultural Development, 8*, 173-186.

Khare, V., Verma, A., Kar, B., Srinivasan, N., & Brysbaert, M. (2013). Bilingualism and the increased attentional blink effect: Evidence that the difference between bilinguals and monolinguals generalizes to different levels of second language proficiency. *Psychological Research, 77*, 728-737.

Kharkhurin, A. V. (2007). The role of cross-linguistic and cross-cultural experiences in bilinguals' divergent thinking. In I. Kecskes & L. Albertazzi (Eds.), *Cognitive Aspects of Bilingualism* (pp. 175-210).

Kharkhurin, A. V. (2008). The effect of linguistic proficiency, age of second language acquisition, and length of exposure to a new cultural environment on bilinguals' divergent thinking. *Bilingualism: Language and Cognition*, *11*, 225-243.

Kharkhurin, A. V. (2009). The role of bilingualism in creative performance on divergent thinking and invented alien creatures tests. *Journal of Creative Behavior, 43*, 59-71.

Kharkhurin, A. V. (2010). Bilingual verbal and nonverbal creative behavior. *International Journal of Bilingualism*, *14*, 211-226.

Kharkhurin, A. V. (2011). The role of selective attention in bilingual creativity. *Creativity Research Journal, 23,* 239-254.

Kharkhurin, A. V. (2012). *Multilingualism and Creativity*. Bristol: Multilingual Matters.

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*, *26*, 640-648.

Klein, R. (2015). Is there a benefit of bilingualism for executive functioning? *Bilingualism: Language and Cognition, 18,* 29-31.

Kousaie, S., & Phillips, N. A. (2012a). Conflict monitoring and resolution: Are two languages better than one? Evidence from reaction time and event-related brain potentials. *Brain Research, 1446*, 71–90.

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

Kousaie, S., Sheppard, C., Lemieux, M., Monetta, L., & Taler, V. (2014). Executive function and bilingualism in young and older adults. *Frontiers in Behavioral Neuroscience, 8*: 250.

Kroll, J. F., & Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology, 25*, 497–514.

Lambert, W. E., Tucker, G. R., & d'Anglejan, A. (1973). Cognitive and attitudinal consequences of bilingual schooling. *Journal of Educational Psychology, 65*, 141-159.

Lee, H., & Kim, K. H. (2011). Can speaking more languages enhance your creativity? Relationship between bilingualism and creative potential among Korean American students with multicultural link. *Personality and Individual Differences, 50,* 1186-1190.

LeFevre, J., Fast, L., Skwarchuk, S., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*, 1753-1767.

Lemmon, C. R., & Goggin, J. P. (1989). The measurement of bilingualism and its relationship to cognitive ability. *Applied Psycholinguistics, 10*, 133–155.

Leung, A. K., & Chiu, C. (2010). Multicultural experience, idea receptiveness, and creativity. *Journal of Cross-Cultural Psychology*, *41*, 723-741.

Leung, A. K., Maddux, W. W., Galinsky, A. D., & Chiu, C. Y. (2008). Multicultural experience enhances creativity: The when and how. *American Psychologist, 63*, 169-181.

Liu, H., Rossi, S., Zhou, H., & Chen, B. (2014). Electrophysiological evidence for domain-general inhibitory control during bilingual language switching. *PLoS ONE*, *9*: e110887.

Lu, C-H., & Proctor, R. W. (1995). The influence of irrelevant location information on performance: A review of the Simon and the spatial Stroop effects. *Psychonomic Bulletin & Review*, *2*, 174-207.

Luk, G., Anderson, J. A. E., Craik, F. I. M., Grady, C., & Bialystok, E. (2010). Distinct neural correlates for two types of inhibition in bilinguals: Response inhibition versus interference suppression. Brain and *Cognition, 74*, 347-357.

Lyons-Warren, A., Lillie, R., & Hershey, T. (2004). Short- and long-term spatial delayed response performance across the lifespan. *Developmental Neuropsychology, 26*, 661-678.

MacKinnon, D. W. (1962). The nature and nurture of creative talent. *American Psychologist, 17*, 484-495.

MacKinnon, D. W. (1965). Personality and the realization of creative potential. *American Psychologist, 20*, 273-281.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940-967.

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*, 81-93.

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention contruct. Neuropsychology, 24, 222-243.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of Personality and Social Psychology, 52*, 1258-1265.

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Fackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain: Proficiency in a second language and age at acquisition affect grey-matter density. *Nature, 431*, 757.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*, 270–291. doi:10.1037/a0028228

Mishra, R. K. (2015). Let's not forget about language proficiency and cultural variations while linking bilingualism to executive control. *Bilingualism: Language and Cognitive, 18*, 39-40.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49-1000.

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

Mohades, S. G., Struys, E., Van Schuerbeek, P., Baeken, C., Van De Craen, P., & Luypaert, R. (2014). Age of second language acquisition affects nonverbal conflict processing in children: An fMRI study. *Brain and Behavior, 4*, 626-642.

Mor, B., Yitzhaki-Amsalem, S., & Prior, A. (2014). The joint effect of bilingualism and ADHD on executive functions. *Journal of Attention Disorders, 19*, 527-541.

Moradzadeh, L., Blumenthal, G., & Wiseheart, M. (2015). Musical training, bilingualism, and executive function: A closer look at task switching and dual-task performance. *Cognitive Science, 39*, 992-1020.

Morales, J., Calvo, A., & Bialystok, E. (2013). Working memory development in monolingual and bilingual children. *Journal of Experimental Child Psychology, 114*, 187-202.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology, 25*, 531-546.

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E. G., Cepeda, N. J. & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22*, 1425–1433.

Mosty, N., Lefever, S., & Ragnarsdóttir, H. (2013). Parents' perspectives towards home language and bilingual development of preschool children [Special Issue 2013—Research and School Practice. University of Iceland, School of Education]. Netla—*Online Journal on Pedagogy and Education*. Retrieved from http://netla.hi.is/serrit/2013/rannsoknir_og_skolastarf/006.pdf

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences, 15*, 453-459.

Murphy, D. G. (1990). Comparison of scores of bilingual urban and monolingual suburban elementary-school-children for 2 measures of intelligence. *Psychological Reports, 67*, 1375-1378.

National Council of State Supervisors for Languages. (2008).

Obonsawin, M., Crawford, J., Page, J., Chalmers, P., Cochrane, R., & Low, G. (2002). Performance on tests of frontal lobe function reflect general intellectual ability. *Neuropsychologia, 40*, 970-977.

Paap, K. R. (2014). The role of componential analysis, categorical hypothesizing, replicability and confirmation bias in testing for bilingual advantages in executive functioning. *Journal of Cognitive Psychology, 26*, 242-255.

Paap, K. R., & Greenberg, Z. I. (2013). There is not coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232-258.

Paap, K. R., Johnson, H. A., & Sawi, O. (2014). Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology, 26*, 615-639.

Paap, K. R., & Sawi, O. (2014). Bilingual advantages in executive functioning: Problems in convergent validity, discriminant validity, and the identification of the theoretical constructs. *Frontiers in Psychology, 5*: 962.

Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of Clinical and Experimental Neuropsychology, 33*, 456-470.

Phillips, L. H. (1997). Do “frontal tests” measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Ed.), Methodology of frontal and executive function (pp. 191-213). Hove, UK: Psychology Press.

Poarch, G. J., & van Hell, J. G. (2012). Executive functions and inhibitory control in multilingual children: Evidence from second-language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology, 113*, 535-551.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition*, *13*, 253-262.

Ratiu, I., & Azuma, T. (2015). Working memory capacity: Is there a bilingual advantage? *Journal of Cognitive Psychology*, *27*, 1-11.

Rhodes, M. (1961). An analysis of creativity. *The Phi Delta Kappan*, *42*, 305-310.

Ricciardelli, L. A. 1992. "Creativity and Bilingualism." *The Journal of Creative Behavior 26*, 242–254.

Rodríguez-Pujadas, A., Sanjuán, A., Fuentes, P., Ventura-Campos, N., Barrós-Loscertales, A., & Ávila, C. (2014). Differential neural control in early bilinguals and monolinguals during response inhibition. *Brain and Language, 132*, 43-51.

Rodríguez-Pujadas, A., Sanjuán, A., Ventura-Campos, N., Román, P., Martin, C., Barceló, F., et al. (2013). Bilinguals use language-control brain areas more than monolinguals to perform non-linguistic switching tasks. *PLoS ONE*, *8(9)*, e73028.

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*, 225-237.

Ruff, R. M. (1988). Ruff Figural Fluency Test professional manual. Odessa, FL: Psychological Assessment Resources.

Runco, M. A. (1990). Implicit theories of ideational creativity. In M. A. Runco & R. S. Albert (Eds.), Theories of creativity (234-252). Thousand Oaks, CA: Sage Publications, Ind.

Ryskin, R. A., Brown-Schmidt, S., Canseco-Gonzalez, E., Yiu, L. K., & Nguyen, E. T.

(2014). Visuospatial perspective-taking in conversation and the role of bilingual experience. *Journal of Memory and Language, 74*, 46-76.

Salthouse, T. A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology, 19,* 532-545.

Salthouse, T. A. (2010). Is flanker-based inhibition related to age? Identifying specific influences of individual differences on neurocognitive variables. *Brain and Cognition, 73*, 51-61.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology, 132*, 566-594.

Salthouse, T. A., & Davis, H. P. (2006). Organization of cognitive abilities and neuropsychological variables across the lifespan. *Developmental Review, 26*, 31-54.

Salthouse, T. A., & Siedlecki, K. L. (2007). Efficiency of route selection as a function of adult age. *Brain and Cognition, 63*, 279-287.

Salvatierra, J., & Rosselli, M. (2010). The Effect of Bilingualism and Age on Inhibitory Control. *International Journal of Bilingualism 15*, 26-37.

Schecter, S. R., Sharken-Taboada, D., & Bayley, R. (1996). Bilingual by choice: Latino parents' rationales and strategies for raising children with two languages. *The Bilingual Research Journal, 20*, 261-281.

Sheng, L., Lu, Y., & Gollan, T. H. (2013). Assessing language dominance in Mandarin-English bilinguals: Convergence and divergence between subjective and objective measures. *Bilingualism: Language and Cognition, 17*, 364-383.

Shimamura, A. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28*, 207-218.

Simon, J. R., & Small, A. M. (1969). Processing auditory information – Interference from an irrelevant cue. *Journal of Applied Psychology, 53*, 433-435.

Simonton, D. K. (2008). "Bilingualism and Creativity." In An Introduction to Bilingualism: Principles and Processes, edited by J. Altarriba and R. R. Heredia, 147–166. New Jersey: Erlbaum.

Singh, N., & Mishra, R. K. (2013). Second language proficiency modulates conflict-monitoring in an oculomotor Stroop task: Evidence from Hindi-English bilinguals. *Frontiers in Psychology, 4*: 322.

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

Stein, M. I. (1953). Creativity and culture. *Journal of Psychology, 36*, 31-322.

Stürmer, B., Leuthold, H., Soetens, E., Schröter, H., & Sommer, W. (2002). Control over location-based response activation in the Simon task: Behavioral and electrophysiological evidence. *Journal of Experimental Psychology. Human Perception and Performance, 28*, 1345-1363.

Sylvester, C. Y., Wager, T. D., Lacey, S. C., Hernandez, L., Nichols, T. E., Smith, E. E., & Jonides, J. (2003). Switching attention and resolving interference: fMRI measures of executive functions. *Neuropsychologia, 41*, 357–370.

Tadmor, C. T., Galinsky, A. D., & Maddux, W. W. (2012). Getting the most out of living abroad: Biculturalism and integrative complexity as key drivers of creative and professional success. *Journal of Personality and Social Psychology, 103*, 520-542.

Tadmor, C. T., & Tetlock, P. E. (2009). Acculturation strategies and integrative complexity: The cognitive implications of biculturalism. *Journal of Cross-Cultural Psychology, 40*, 105-139.

Titone, D., Pivneva, I., Sheikh, N. A., Webb, N., & Whitford, V. M. (2015). Doubling down on multifactorial approaches to the study of bilingualism & executive control. *Bilingualism: Language and Cognition, 18*, 43-44.

Tao, L., Marzecova, A., Taft, M., Asanowicz, D., & Wodniecka, Z. (2011). The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology, 2*: 123.

Torrance, E. P. (1972). Predictive validity of Torrance Tests of Creative Thinking. *Journal of Creative Behavior, 6*, 236-252.

Torrance, E. P. (1974). Torrance Tests of Creative Thinking: Norms Technical Manual. Lexington, MA: Ginn.

Tse, C. S., & Altarriba, J. (2014). The relationship between language proficiency and attentional control in Cantonese-English bilingual children: Evidence from the Simon, Simon Switching, and working memory tasks. *Frontiers in Psychology, 5*: 954.

Tse, C. S., & Altarriba, J. (2015). Local and global task switching costs in bilinguals who vary in second language proficiency. *American Journal of Psychology, 128,* 89-106.

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition, 18*, 3-24.

van der Sluis, S., de Jong, P. F., & van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence, 35*, 427–449.

Van Hell, J., & Dijkstra, A. (2002). Foreign language knowledge can influence native language performance. *Psychonomic Bulletin & Review, 9*, 780–789.

Veltkamp, G. M., Recio, G., Jacobs, A. M., & Conrad, M. (2013). Is personality modulated by language? *International Journal of Bilingualism, 17*, 496-504.

Verbruggen, F., Logan, G. D., & Stevens, M. A. (2008). STOP-IT: Windows executable software for the stop-signal paradigm. *Behavior Research Methods, 40*, 479-483.

Viki, G. T., & Williams, M. L. J. (2014). The role of identity integration in enhancing creativity among mixed-race individuals. *Journal of Creative Behavior*, *48,* 198-208.

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research, 78*, 803–820.

Weissberger, G. H., Gollan, T. H., Bondi, M. W., Clark, L. R., & Wierenga, C. E. (2015). Language and task switching in the bilingual brain: Bilinguals are staying, not switching, experts. *Neuropsychologia, 66*, 193-203.

Wilcox, R. R. (2012). Introduction to robust estimation and hypothesis testing (3rd ed). Oxford, UK: Elsevier.

Wilcox, R. R., & Schönbrodt, F. D. (2014). The WRS package for robust statistics in R (version 0.24). Retrieved from http://r-forge.r-project.org/projects/wrs/

Willcutt, E. G., Pennington, B. F., Boada, R., Tunick, R. A., Ogline, J., Chhabildas, N. A., & Olson, R. K. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology, 110*, 157–172.

Wimmer, M. C., & Marx, C. (2014). Inhibitory processes in visual perception: A bilingual advantage. *Journal of Experimental Child Psychology, 126*, 412-419.

Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M., & Duyck, W. (2015). Bilingualism delays clinical manifestations of Alzheimer's disease. *Bilingualism: Language and Cognition, 18*, 568-574.

Wuhr, P. & Ansorge, U. (2005). Exploring trial-by-trial modulations of the Simon effect. *The Quarterly Journal of Experimental Psychology*, *58A*, 705-731.

Yow, W. Q., & Li, X. (2015). Balanced bilingualism and early age of second language acquisition as the underlying mechanism of a bilingual executive control advantage: Why variations in bilingual experiences matter. *Frontiers in Psychology*, *6*: 164.

Zabelina, D. L., & Robinson, M. D. (2010). Creativity as flexible cognitive control. *Psychology of Aesthetics, Creativity, and the Arts, 4*, 136-143.

Appendix A

Session 1

Consent form
SAM (Self-assessment mannequin)
Baseline RT
Demographic questionnaire
Antisaccade
Torrance tests of creative thinking - Verbal
Spatially-delayed response task
Language assessment: Speaking
Arrow Simon

BREAK

Number letter
Language assessment: Writing
Personality scale
Flankers
Ruff figural fluency test
Tone monitoring
Language assessment scale

Session 2

SAM (Self-assessment mannequin)
BSI (Brief Symptoms Inventory)
Baseline RT
Language assessment: Reading
Flanking color patches
Adjective checklist
Sequential modulation task
Torrance tests of creative thinking - Figural

BREAK

Letter memory
Language assessment: Listening
CCF (Cattell Culture Fair)
Stop signal
Trails A & B
Letter and category fluency
Standard Simon
Forward Corsi blocks
Creative achievement questionnaire
Debriefing form

Appendix B

Table 2-7. *Task-Specific t-Tests between Language Groups (nML = 37, nBL = 47).*

	Language Group				95% CI for Mean Difference		
	Monolingual		Bilingual				
Task	M	SD	M	SD		*t*	*p*
Anti	.87	.07	.85	.09	-.02, .05	.87	.386
SSRT	-264.48	33.88	-271.79	41.14	-9.36, 23.97	.87	.386
TM	.80	.16	.77	.18	-.04, .11	1.00	.320
LM	.59	.18	.60	.18	-.09, .07	-.28	.778
NL	-382.83	237.12	-403.59	287.58	-.95.81, 137.32	.35	.724
Trails	-2.15	.66	-2.50	.85	.02, .69	2.10	.039*
SDR	-980.56	241.71	-1101.09	228.42	18.07, 222.99	2.34	.022*
Corsi	.70	.11	.67	.13	-.02, .09	1.38	.172
Flanker	-489.75	42.95	-485.73	58.29	-26.81, 18.76	-.36	.717
FColors	-387.31	38.29	-386.40	48.07	-20.17, 18.35	-.09	.925

Appendix C

Letter Fluency

Participants were asked to say as many words as they could think of that begin with a given letter of the alphabet. The words given were to exclude proper nouns, numbers, and the same word with a different suffix. The letter C was presented first, and then the letter L, and one minute was allotted per letter. Participant responses were recorded via online audio-recording software.

One point was given for each legitimate response. The sum of scores for both letters constituted the total score for letter fluency.

Category Fluency

Participants were asked to say as many words as they could think of that belong in a given category. Four different categories were presented one after another with one minute allotted per category: animals, parts of a house, things that make sound, and things people put in their pocket. Participant responses were recorded via online audio-recording software.

One point was given for each legitimate response. The sum of scores for all four categories constituted the total score for category fluency.