OXFORD UNIVERSITY PRESS

Archives of Clinical Neuropsychology 37 (2022) 826–838

Archives of CLINICAL NEUROPSYCHOLOGY

Neuropsychological Test Performance Among Native and Non-Native Swedes: Second Language Effects

Jacob Stålhammar[1,*], Per Hellström[2], Carl Eckerström[1,3], Anders Wallin[1]

[1] *Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*

[2] *Department of Clinical Neuroscience, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden*

[3] *Department of Immunology and Transfusion medicine, Sahlgrenska University Hospital, Region Västra Götaland, Sweden, Gothenburg, Sweden*

*Corresponding author at: Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, PO SU Sahlgrenska, SE-413 45, Gothenburg, Sweden. Tel.: +46(0)707 20 53 67. *E-mail address:* jacob.stalhammar@neuro.gu.se (J. Stålhammar).

Received 18 November 2019; revised 12 March 2020; Accepted 16 June 2020

Abstract

Objective: We aimed to study second language effects on neuropsychological (NP) test performance.
Method: We administered an NP test battery in Swedish to 322 healthy community dwelling participants, recruited through the Gothenburg Pilot phase of the Swedish CArdioPulmonary BioImage Study (SCAPIS Pilot). All participants were conversationally fluent Swedish speakers (237 native, 85 non-native, mean age 61.1 years). We compared the NP scores of native and non-native participants. We also investigated the influence of (a) age of arrival to Sweden, (b) majority language family of the birth country, and (c) proficiency in Swedish as assessed with a 30 item Boston naming test (BNT).
Results: Native speakers obtained better results on all NP tasks with a verbal component, whereas no significant differences were seen on completely nonverbal tasks (Rey complex figure). For non-native speakers, lower age at arrival to Sweden, arrival from a country where Swedish was also spoken, or arrival from a country with a majority language closer to Swedish, were all linked to better NP scores. Dichotomizing by BNT showed that normally-to-highly proficient non-native speakers obtained better scores.
Conclusions: Second language effects may contribute to misclassification of non-native speakers. Assumptions of fluency based on short conversations may be misleading. A proficiency assessment with BNT may improve NP score interpretation among non-native speakers.

Keywords: Non-native speaker; Neuropsychological test; Test administration; Language effects; Second language, bilingualism

Introduction

The number of speakers with any degree of multilingualism now constitutes more than half of the world's population (Bialystok, 2017; Bialystok, Craik, & Luk, 2012). According to Statistics Sweden (the Swedish census bureau) the proportion of non-native Swedes has risen from 0.7% in the 1900's to currently 19.6% (Sweden_Statistics, 2020). From a neuropsychological (NP) test perspective, administration of tests to individuals in their second language (L2) may produce invalid scores (Strauss, Sherman, & Spreen, 2006). The use of an interpreter should be used judiciously because interpreter use can increase total scores, and/or contribute to a larger score variability on verbal tests (Casas, Guzman-Velez, Cardona-Rodriguez, et al., 2012). For dementia assessment, specific cross-cultural tests have shown promise, for example the Rowland Universal Dementia

https://doi.org/10.1093/arclin/acaa043 Advance Access publication 28 July 2020

Assessment Scale (RUDAS, (Rowland, Basic, Storey, & Conforti, 2006)) as well as combinations of tests (Nielsen, Segers, Vanderaspoilden, et al., 2019). Bilingualism has been hypothesized to be cognitively advantageous (Bialystok et al., 2012); however, both advantages and disadvantages to second-language acquisition have been raised (Rivera Mindt, Arentoft, Kubo Germano, et al., 2008). Also, there is no generally accepted definition of bilingualism, even if many criteria suggest comparisons between native and non-native language skills (Adesope, Lavin, Thompson, & Ungerleider, 2010). As most hospitals have limited resources for comparative language assessment, NP assessments of multilingual speakers may have to rely on qualitative judgments of conversational fluency (e.g., "speaks good Swedish"). Thus, the usefulness of information gathered in a naturalistic setting should be evaluated.

Assessing Bilingual Proficiency

An appearance of fluency may conceal a limited vocabulary (e.g., an actor's lines), and, conversely, understanding may exceed defective pronunciation or grammar; vocabulary size differs not only between native and non-native speakers, but also depending on context (compare e.g., a surgeon to an auto mechanic). Depending on the field, linguistic proficiency may be assessed in different ways. For professional/educational use there are guidelines for achievement (e.g., the Common European Framework of Reference for Languages), or tests (e.g., Test of English as a Foreign Language). For clinical investigation into bilingualism, both self-report, for example preference and scales (Gasquoine, Croyle, Cavazos-Gonzalez, & Sandoval, 2007), NP tests—for example shortened versions of the Boston naming test (BNT) (Erdodi, Jongsma, & Issa, 2017)—and combinations of both (Rosselli, Ardila, Santisi, et al., 2002) have been used.

Bilingualism denotes proficiency in two languages, L1 (native) and L2 (non-native). The cognitive effects of bilingualism have been investigated from the 1920s' and onward, with early research often neglecting to account for factors that have later been shown to be relevant, such as rural–urban differences, gender, socioeconomics and proficiency (Adesope et al., 2010; Peal & Lambert, 1962). Several methods to measure the balance between L1 and L2 have been suggested; for example dividing word flow in L1 by word flow in L2 (Erdodi et al., 2017; Peal & Lambert, 1962; Shishkin & Ecke, 2018; Suarez, Gollan, Heaton, Grant, & Cherner, 2014). However, when dividing word flow between L1 and L2, current L1 capacity should be considered, as first language (L1) maintenance has been shown to be a factor (Shishkin & Ecke, 2018). Several bilingual varieties have been suggested (balanced, early, late, etc.), but, as noted by Adesope et al. in their review of 2010, information to describe or code the category "bilingualism" was often lacking (Adesope et al., 2010), an omission that has also been criticized by others (Erdodi et al., 2017).

Although an attempt to define bilingualism would be a worthwhile endeavor, this is outside the scope and purpose of the present study. We will only investigate effects of Swedish language NP-test administration among native and non-native Swedish speakers. To our knowledge, this has not been studied previously on this scale in Sweden, and it is a subject of great importance to clinical neuropsychologists who encounter an increasing number of patients with Swedish as a second language.

Objectives

Our objective was to investigate performance differences in healthy native vs. non-native Swedish speakers in a naturalistic setting (i.e., comparable to clinical procedure) with a Swedish language administrated NP test battery. We anticipated that non-native speakers would present second language effects in lower NP scores in tests with a prominent verbal component (e.g., confrontation naming, word flow), but not in tests with a less salient verbal component (e.g., trail making test, Stroop). We further wanted to evaluate if possible differences could be investigated by commonly available information (e.g., age of arrival in Sweden, or proficiency assessed with the Boston naming test).

Methods/Material

All participants in this work were recruited via the Swedish CArdioPulmonary BioImage Study (SCAPIS). SCAPIS was designed to investigate cardiopulmonary disease, with a main aim to describe a demographically representative cohort of 30,000 men and women, between the ages of 50 and 64 years (Bergstrom, Berglund, Blomberg, et al., 2015). As a feasibility exploration, a smaller pilot study (SCAPIS pilot) was completed in Gothenburg, Sweden in 2012, inviting 2,243 participants, recruiting 1,111, both from areas of high and low socioeconomic status (Bergstrom et al., 2015). Although cognition was not a primary focus of the main SCAPIS study, for the SCAPIS pilot section 1,089 participants were invited for additional cognitive assessment and 322 were recruited. These 322 participants, both native and non-native Swedish speakers, were administered NP tests in Swedish, under naturalistic conditions. Figure 1 describes invitation and selection steps for the cognition examinations added to the SCAPIS pilot study.

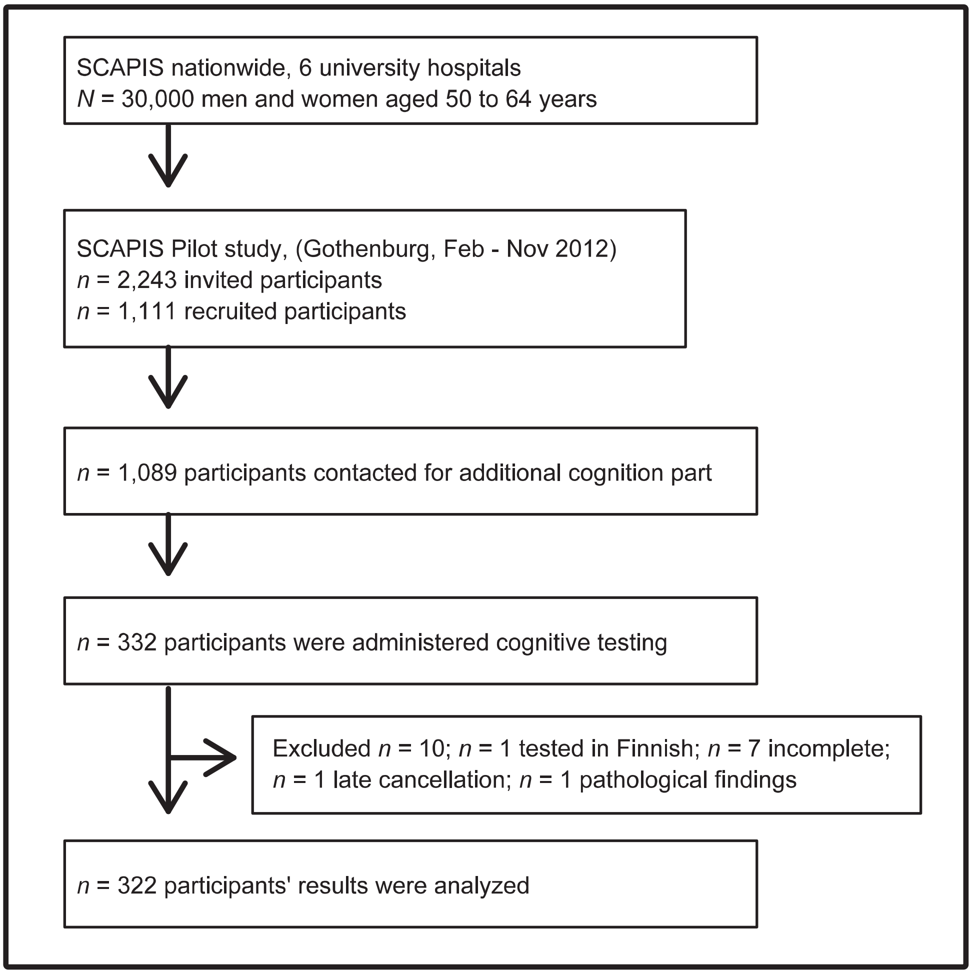

Fig. 1. Selection steps for cognition assessments added to the SCAPIS pilot. SCAPIS, the Swedish CArdioPulmonary BioImage Study.

Table 1. Participant demographics

	Native Swedish speakers	Non-native Swedish speakers	*p*
n	237	85	
% Women	53.2%	51.8%	0.825*
Age *M* (*SD*) range	61.1 (3.3) 54–67	61.1 (3.5) 54–66	0.857**
Education *M* (*SD*) range	13.5 (2.8) 8–24	13.0 (3.5) 5–26	0.255**
Clock drawing *M* (*SD*) range	4.9 (0.4) 3–5	4.8 (0.4) 3–5	0.173**

Note: Comparison of native and non-native Swedish speakers, *Prob > ChiSq and **Student's T-test. SD = standard deviation.

The main SCAPIS study applied no exclusion criteria "except the inability to understand written and spoken Swedish for informed consent" (Bergstrom et al., 2015). To this sub study of SCAPIS pilot participants we added a brief interview to enable pre-test discovery of reasons for exclusion due to lack of conversational fluency. The interview included questions about education, main profession, hearing, eyesight, and cognitive complaints (profession and remarks data not shown). Only minor remarks surfaced (e.g., no examples of major trauma, none of major substance abuse, none of severe psychiatric illness).

Of the 322 recruited participants, 237 were native and 85 were non-native Swedish speakers. As also seen in Table 1, there were no examples of illiteracy; all participants (non-native and native) had attended school, between 5 and 26 years of total education. With only minor remarks noted from interviews, exclusion criteria for further analyses were: test administration in another language than Swedish (Finnish, $n = 1$); incomplete administrations ($n = 7$); late cancelation ($n = 1$); and pathological findings ($n = 1$) determined by the clock-drawing test (CDT) and NP scores.

Table 2. The 85 non-native Swedish speakers' 34 countries of birth

Country of birth	*n*	%	Country of birth (cont.)	*n*	%
Finland	22	25.9	Brazil	1	1.2
Iran	9	10.6	Burma (Englishman)	1	1.2
Bosnia	4	4.7	Cape Verde	1	1.2
Chile	4	4.7	Czech Republic	1	1.2
Hungary	4	4.7	Denmark	1	1.2
Macedonia	3	3.5	England	1	1.2
Poland	3	3.5	Eritrea	1	1.2
Serbia	3	3.5	Gambia	1	1.2
Yugoslavia	3	3.5	Iceland	1	1.2
Austria	2	2.4	India	1	1.2
Iraq	2	2.4	Italy	1	1.2
Lebanon	2	2.4	Kosovo	1	1.2
Norway	2	2.4	Netherlands	1	1.2
Turkey	2	2.4	Pakistan	1	1.2
Asia (not specified)	1	1.2	Portugal	1	1.2
Belarus	1	1.2	Romania	1	1.2
Bolivia	1	1.2	Thailand	1	1.2
			Sum	85	100.0

Note: Table is sorted first by number of participants, second by countries, and ordered alphabetically.

Consistent with a naturalistic setting, no formal assessment of the non-native speakers' native language skills were performed, yet countries of birth and age of arrival to Sweden were noted. Table 2 shows the 34 countries of birth of the non-native Swedish speakers.

Ethics

SCAPIS was approved by the ethics committee at Umeå University (Bergstrom et al., 2015). The addition of the cognitive battery in the SCAPIS pilot sub study was approved by the regional ethics board of the University of Gothenburg, diary number: Dnr 734–13. Written informed consent was obtained from all participants.

Neuropsychological Tests

After the interview, a battery of 11 NP tests was administered. The test battery was composed of commonly used tasks (e.g., trail making tests A and B (Partington & Leiter, 1949) from (Strauss et al., 2006)) and tasks featured in a presentation of a brief cognitive assessment battery (CAB) (Nordlund, Pahlsson, Holmberg, Lind, & Wallin, 2011). However, as the NP testing included administration in a second language, instructions and scoring procedures were slightly revised compared to Nordlund et al. (2011). All tests were administered in their entirety (no initiation/termination rules were used), and both verbatim and synonym answers were recorded for the memory test. The 11 NP tests rendered a total of 22 scores, in the following domains:

Simple Speed/Attention

Trail making test A (TMT A), timed pencil-tracing of a numbered path (Partington & Leiter, 1949) from (Strauss et al., 2006); Stroop test Victoria version, part 1, timed verbal naming of 24 colored dots (Strauss et al., 2006); Rey complex figure (RCF), timed pencil-copy of a printed figure (Strauss et al., 2006).

Divided (Executive) Attention

Stroop test Victoria version part 2, timed naming of words' print color, Part 3 naming of color words' print color (Strauss et al., 2006); TMT B, timed pencil-tracing of numbers and letters, 1-A-2-B, etc. (Strauss et al., 2006); symbol digit modalities Test (SDMT), pencil in numbers in rows of blank squares, guided by nine corresponding symbols (CAB version (Nordlund et al., 2011), same symbols but different numbers compared to the original (Strauss et al., 2006)); parallel serial mental operations (PASMO), following a recital of the Swedish alphabet (28 letters A-Ö), recite letters-numbers, A-1, B-2 etc., throughout the Swedish alphabet (Nordlund et al., 2005), similar to oral TMT B (Strauss et al., 2006) but longer.

Learning/Memory

Short story memory test with repetition (“learning effect”), same text as in CAB (Nordlund et al., 2011) but with revised scoring, recording both verbatim and synonym answers (e.g. both “plane” and “airplane,” where the original version might only have score correct for “plane”) for both immediate and delayed recall. Verbatim and synonym scorings were separately recorded and for brevity labeled “MEMO”; RCF immediate, delayed recall, and recognition (Strauss et al., 2006).

Visuo-Construction

RCF figure copy, pencil copying of a printed figure, scoring the copy strategy and number of correctly copied items (Strauss et al., 2006). The clock-drawing test, five-point scoring (Shulman, 2000), was administered, but used only for exclusion and descriptive purposes.

Verbal

The Token test CAB 6 item version (Nordlund et al., 2011), similar to the Token test (Strauss et al., 2006) but shorter. Subjects are verbally instructed to reposition one of eight plastic “tokens” in relation to the remaining seven. In the controlled oral word association test, subjects are asked to say as many words as possible beginning with a given letter in 1 min (COWAT FAS, (Strauss et al., 2006)). In the category fluency test “animal naming,” subjects are asked to name as many animals as possible in 1 min (Strauss et al., 2006). A 30 item BNT-CAB confrontation naming task (Nordlund et al., 2011), featuring redrawn images from the original BNT (Kaplan, Goodglass, Weintraub, & Segal, 2001) was also administered.

Test Administration

All tests were administered in Swedish by psychologists in training, a trained researcher or a health coach, all under the supervision of licensed psychologists. Scores were rechecked by licensed psychologists. Patients used their normal glasses and/or hearing aids, but no formal assessments of vision or hearing were performed. One participant reported impaired color perception, however, tests depending on color vision (Stroop) indicated normal results for this participant. Interview time was indicated as equal as there were no native/non-native differences by presence/absence of minor remarks (e.g., “I use reading-glasses”, or “my memory is not what it used to be”). Rater analysis indicated no differences with respect to participant proportions of native/non-native, gender, or years of education, per rater.

Statistical Analyses

Results are presented both in raw and normalized form. Analyses were made on raw scores before conversion to T-scores (mean 50, *SD* 10). Two normalized/transformed score comparisons are provided: first a comparison using native Swedish speakers’ means and standard deviations (illustrating native speaker norms on non-native speakers); and second; effect sizes (Hedges’ g: group mean differences divided by a pooled *SD*, incorporating group differences in n and standard deviations). No imputation was used, but cases of administrations to fewer than 90% of a group (per specific task) are indicated, and n or *SD* difference related effects are also reflected in the effect size. Binary categories (e.g., gender) were compared with Chi-square tests at $p < .05$. Continuous data from the groups were compared with the two tailed Student’s T-test. Three decimals were used for p-values, to enable comparison to a Bonferroni correction $p < .05/22 = .002$. Education (total number of years) was reported separately, and not adjusted for, because firstly in three of four analyses there were no group differences in education (this was only seen in the birth country group analysis), and secondly, because the geographical location for primary education varied (Discussion). In the age-of-arrival analysis we compared non-native speakers only, and contrasted groups of early arrivers (<20 years of age) to late arrivers (≥20 years of age).

In the birth country language group analysis we contrasted the entire set of native speakers to non-native speakers grouped into five language groups by the majority language of the birth country; Germanic (Austria, Norway, Netherlands, Iceland, England, Burma (Englishman), Denmark, $n = 9$); Italic-Romance (Chile, Bolivia, Brazil, Cape Verde, Italy, Portugal, Romania, $n = 10$); Finno-Ugric (Finland, Hungary, $n = 26$), Slavic (Bosnia, Serbia, Poland, Yugoslavia, Macedonia, Belarus, Czech Republic, Kosovo, $n = 19$), and Other (Turkey, Thailand, Pakistan, Lebanon, Iran, Iraq, India, Gambia, Eritrea, Asia (not specified) $n = 21$).

In the BNT analysis we dichotomized by the native speaker norms at 1 *SD* below the mean (≤25 or >25 points ≈T41, rounded to the nearest word) to evaluate a normal-to-high vs. below-normal cutoff for naming proficiency.

Table 3. Native and non-native neuropsychological scores (test order: effect size)

	Native (n = 237)		Non-native (n = 85)		Effect size	
Neuropsychological test	Mean (*SD*)	T	Mean (*SD*)	T	*p*	(Hedges' *g*)
BNT (CAB)	27.0 (2.2)	50.0	20.8 (5.5)	21.8	<.001	1.8
Stroop Victoria 1	13.4 (2.6)	50.0	19.7 (5.7)	26.0	<.001	1.7
Stroop Victoria 2	17.4 (3.5)	50.0	25.1 (6.3)	28.1	<.001	1.7
Token Test (CAB)	5.7 (0.6)	50.0	4.1 (1.9)	22.0	<.001	1.5
Stroop Victoria 3	25.8 (6.2)	50.0	35.5 (12.1)	34.4	<.001	1.2
Animal Naming	25.2 (6.0)	50.0	18.0 (5.9)	38.0	<.001	1.2
COWAT FAS	41.7 (12.3)	50.0	28.2 (11.4)	39.0	<.001	1.1
TMT B**	81.9 (29.1)	50.0	123.3 (70.2)	35.8	.001	1.0
PASMO**	72.7 (25.9)	50.0	101.0 (44.4)	39.1	<.001	1.0
MEMO direct synonym	9.3 (3.1)	50.0	6.6 (3.1)	41.2	<.001	0.9
MEMO direct verbatim	6.8 (2.9)	50.0	4.6 (2.8)	42.4	<.001	0.8
TMT A	35.9 (12.4)	50.0	46.1 (17.1)	41.7	<.001	0.7
MEMO delayed synonym	11.6 (3.4)	50.0	9.1 (3.6)	42.6	<.001	0.7
MEMO delayed verbatim	8.5 (3.3)	50.0	6.3 (3.3)	43.0	<.001	0.7
SDMT (CAB) correct	43.4 (8.8)	50.0	37.4 (8.7)	43.1	<.001	0.7
SDMT (CAB) errors	0.5 (1.0)	50.0	1.0 (1.5)	45.2	.006	0.4
RCF copy	33.2 (2.9)	50.0	32.1 (4.2)	46.3	.033	0.3
RCF recognition	20.4 (2.0)	50.0	19.7 (2.5)	46.7	.023	0.3
RCF copy time	185.9 (86.5)	50.0	213.7 (98.4)	46.8	.041	0.3
RCF immediate	19.3 (5.5)	50.0	17.7 (6.4)	47.0	.031	0.3
RCF delayed	18.7 (5.4)	50.0	17.5 (6.6)	47.7	.302	0.2
RCF strategy (A1.B2.C3)	1.4 (0.6)	50.0	1.5 (0.7)	48.6	.121	0.1
Mean T		50.0		39.1		
Memory test sub analyses						
Synonym–verbatim direct	2.5 (1.4)	50.0	2.0 (1.5)	46.3	.007	0.4
Synonym–verbatim delayed	3.1 (1.6)	50.0	2.9 (1.5)	48.5	.236	0.1
Synonym/verbatim direct	145% (36%)	50.0	165% (81%)	55.6	.031	0.4
Synonym/verbatim delayed	147% (45%)	50.0	172% (86%)	55.6	.014	0.4
Verbatim learning effect	1.7 (2.6)	50.0	1.6 (2.1)	49.5	.659	0.1
Synonym learning effect	2.3 (2.6)	50.0	2.4 (2.1)	50.4	.723	0.0

Note: BNT = Boston naming test; CAB = cognitive assessment battery; COWAT, controlled oral word association test, letters F, A, S; TMT = trail making test; PASMO = parallel serial mental operations; **, <90% test administration (Discussion); MEMO = short story memory test; SDMT = symbol digit modalities test; RCF = Rey complex figure test; T-score (mean 50, *SD* 10). All (n = 237) native speakers' scores used for norms.

Results

First, we list a comparison between native and non-native NP scores in order of the most different to the least different, by effect size.

Table 3 shows lower mean NP scores for non-native speakers on all tasks, and Bonferroni corrected significantly lower scores for all except for the Rey complex figure tasks. Further, the results of the "Memory test sub analyses" show that allowing synonym answers led to higher mean story test memory scores for all participants. However, as non-native participants obtained significantly lower scores for their initial verbatim answers, their synonym gains were proportionally larger.

Analysis by Age of Arrival to Sweden

Table 4 shows non-native speakers dichotomized by age of arrival to Sweden. Non-natives with earlier arrival (<20 years of age) had higher scores than those with later.

Analysis by Grouped Countries of Birth

Although no analysis of native languages was possible, analysis via simplified grouping of participants per majority language family in their birth country was performed. Table 5 shows that subjects with native languages belonging to two language families, namely the Germanic and the Finno-Ugric, had NP results closer to the native Swedish speakers' results.

Table 4. Age of arrival in Sweden <20 or ≥20 (order: effect size)

	Arrival < 20 years of age (*n* = 22)		Arrival ≥20 years of age (*n* = 63)		Effect size	
	Mean (*SD*)	T	Mean (*SD*)	T	*p*	(Hedges' *g*)
Stroop Victoria 1	16.2 (4.6)	39.1	21 (5.6)	20.9	<.001	0.9
Animal Naming	21.5 (5.7)	43.8	16.7 (5.4)	35.9	.002	0.9
BNT (CAB)	23.9 (5.6)	36.0	19.7 (5.1)	16.7	.004	0.8
Stroop Victoria 2	21.8 (4.7)	37.4	26.3 (6.4)	24.7	.001	0.7
MEMO direct verbatim	5.9 (3.2)	46.8	4.2 (2.5)	41.1	.035	0.6
MEMO direct synonym	7.7 (3.7)	44.8	6.3 (2.9)	40.2	.127	0.5
MEMO delayed verbatim	7.2 (3.8)	46.1	5.9 (3.1)	42.1	.149	0.4
MEMO delayed synonym	10.1 (4.5)	45.6	8.8 (3.2)	41.8	.236	0.4
RCF strategy (A1, B2, C3)	1.7 (0.8)	55.5	1.4 (0.6)	50.6	.120	0.4
COWAT FAS	31.2 (10.9)	41.4	27.1 (11.5)	38.1	.144	0.4
PASMO**	91.9 (39.8)	42.6	110.2 (48.1)	35.5	.250	0.4
TMT B**	103.2 (42.1)	42.7	129.9 (76.3)	33.5	.060	0.4
SDMT (CAB) errors	0.6 (0.9)	49.1	1.1 (1.6)	43.9	.070	0.3
Token Test (CAB)	4.5 (1.6)	30.8	3.9 (1.9)	20.1	.128	0.3
Stroop Victoria 3	33.1 (10.9)	38.2	36.3 (12.5)	33.0	.254	0.3
RCF recognition	20.2 (2.2)	48.9	19.6 (2.6)	45.9	.304	0.2
RCF copy	31.7 (4.6)	44.9	32.2 (4)	46.6	.657	0.1
RCF copy time	221.1 (110.6)	45.9	211.1 (94.5)	47.1	.707	0.1
RCF immediate	18.2 (7.6)	48.0	17.5 (5.9)	46.7	.703	0.1
TMT A	44.8 (13.7)	42.8	46.6 (18.3)	41.4	.634	0.1
SDMT (CAB) correct	37.2 (10.4)	42.9	37.4 (8.1)	43.2	.915	0.0
RCF delayed	17.6 (7)	47.9	17.4 (6.5)	47.6	.934	0.0
Mean T-score		43.2		38.0		
Demographic analysis						
% Women	68.2%		46.0%		.071^	
Age	61.3 (3.9)		61.0 (3.4)		.758	0.1
Education	12.3 (4.2)		13.3 (3.2)		.293	0.3
Clock drawing	4.8 (0.4)		4.8 (0.4)		.810	0.0
Age of arrival to Sweden	14.7 (5.5)		30.3 (8.1)		<.001	2.1
Years of Swedish residency	46.6 (5.0)		30.6 (8.9)		<.001	1.9

Note: BNT = Boston naming test; CAB = cognitive assessment battery; COWAT = controlled oral word association, letters F, A, S; TMT = trail making test; PASMO = parallel serial mental operations; ** = < 90% test administration (Discussion); MEMO = short story memory test; SDMT = symbol digit modalities test; RCF = Rey complex figure test; ^ = Chi square; T = T-score (mean 50, *SD* 10). All (*n* = 237) native speakers' scores used for norms.

Analysis by BNT Proficiency

Table 6 shows results from dichotomizing by proficiency assessed with BNT (the test that showed the largest differences between native and non-native speakers, Table 3). The grand mean of BNT normal-to-high scoring non-native speakers was T48.6, but with Stroop 1 and 2 almost 1 standard deviation below the mean. However, non-native speakers with BNT scores below T41 presented a grand mean of T36.5, with many NP scores 2–3 *SD* below the mean.

Discussion

The present study showed that non-native but conversationally fluent Swedish speakers from 34 different countries scored significantly lower as a group when tested in Swedish, not only in NP tests with a large verbal component, but in all tests except for purely visuoconstructive tests. Analysis of the influence of non-native speakers' arrival before or after 20 years of age showed higher scores for earlier arrivers. Secondly, participants with native languages more distant to Swedish, in terms of vocabulary, and numerical and alphabetical symbols, presented lower scores than participants with more closely related native languages. Thirdly, using BNT as a proficiency test showed that normal-to-high proficiency non-native speakers performed on par with native speakers, whereas low-proficiency non-native speakers performed worse.

Comparison to Other Studies

Lower mean BNT scores even for proficient non-native speakers was described by Roberts, Garcia, Desrochers, and Hernandez (2002) who also found different item difficulty order between native and both bilingual groups (Roberts et al.,

Table 5. Normalized NP scores by grouped countries of birth (same order as Table 3)

Variable mean T-score (Hedges' *g*)/Language family group	Native Swedes (*n* = 237)	Germanic (*n* = 9)	Italic-Romance (*n* = 10)	Finno-Ugric (*n* = 26)	Slavic (*n* = 19)	Other (*n* = 21)
BNT (CAB)	T50 (0)	T49 (−0.2)	T25 (−2.4)	T36 (−1.3)	T16 (−3.0)	T-4 (−4.7)
Stroop Victoria 1	T50 (0)	T46 (0.4)	T21 (2.7)	T29 (1.9)	T21 (2.5)	T21 (2.6)
Stroop Victoria 2	T50 (0)	T43 (0.7)	T22 (2.7)	T29 (1.9)	T28 (2.1)	T23 (2.6)
Token test (CAB)	T50 (0)	T50 (0.0)	T14 (−2.9)	T33 (−1.3)	T12 (−3.0)	T8 (−3.0)
Stroop Victoria 3	T50 (0)	T35 (1.4)	T30 (1.9)	T39 (1.0)	T31 (1.8)	T33 (1.4)
Animal naming	T50 (0)	T42 (−0.8)	T36 (−1.4)	T41 (−0.9)	T36 (−1.3)	T35 (−1.5)
COWAT FAS	T50 (0)	T47 (−0.3)	T35 (−1.5)	T41 (−0.9)	T39 (−1.1)	T35 (−1.5)
TMT B**	T50 (0)	T49 (0.1)	T20 (2.3)	T44 (0.5)	T32 (1.6)	T30 (1.8)
PASMO**	T50 (0)	T49 (0.1)	T41 (0.9)	T35 (1.4)	T42 (0.8)	T32 (1.8)
MEMO direct synonym	T50 (0)	T49 (−0.1)	T42 (−0.8)	T44 (−0.6)	T38 (−1.3)	T38 (−1.2)
MEMO direct verbatim	T50 (0)	T51 (0.1)	T43 (−0.7)	T45 (−0.5)	T38 (−1.2)	T39 (−1.1)
TMT A	T50 (0)	T47 (0.3)	T39 (1.0)	T46 (0.4)	T40 (1.0)	T37 (1.2)
MEMO delayed synonym	T50 (0)	T45 (−0.5)	T41 (−0.9)	T47 (−0.3)	T40 (−1.0)	T40 (−1.0)
MEMO delayed verbatim	T50 (0)	T48 (−0.2)	T40 (−1.0)	T48 (−0.2)	T39 (−1.1)	T39 (−1.1)
SDMT (CAB) correct	T50 (0)	T46 (−0.4)	T45 (−0.5)	T44 (−0.6)	T41 (−0.9)	T42 (−0.8)
SDMT (CAB) errors	T50 (0)	T52 (−0.1)	T46 (0.4)	T47 (0.3)	T42 (0.7)	T43 (0.7)
RCF copy	T50 (0)	T53 (0.3)	T46 (−0.4)	T49 (−0.1)	T48 (−0.2)	T39 (−1.0)
RCF recognition	T50 (0)	T53 (0.3)	T40 (−1.0)	T47 (−0.2)	T49 (−0.2)	T44 (−0.5)
RCF copy time	T50 (0)	T50 (0.0)	T49 (0.1)	T45 (0.5)	T45 (0.5)	T48 (0.2)
RCF immediate	T50 (0)	T55 (0.4)	T45 (−0.5)	T46 (−0.4)	T48 (−0.2)	T46 (−0.4)
RCF delayed	T50 (0)	T58 (0.8)	T46 (−0.4)	T46 (−0.4)	T49 (−0.1)	T45 (−0.5)
RCF strategy (A1, B2, C3)	T50 (0)	T48 (−0.1)	T54 (0.4)	T53 (0.2)	T51 (0.2)	T50 (0.0)
Mean *T*	50.0	48.6	36.9	42.2	37.4	34.7
Demographic analysis						
Mean *n* (by administered tests)	235.8	8.9	9.6	25.4	18.2	19.1
% Women	53.2	55.6	60.0	65.4	52.6	28.6
Age	61.1	63.7	60.6	62.6	60.5	58.7
Education	13.5	15.1	13.3	12.3	12.7	13.3
Clock drawing	4.9	5.0	4.8	5.0	4.5	4.8
CDT T-score, (Hedges' *g*)	T50 (0.0)	T54 (0.4)	T48 (−0.2)	T54 (0.3)	T40 (−0.9)	T48 (−0.3)
Arrival to Sweden	native	23.7	27.4	22.9	27.7	29.9
Years of Swedish residency	native	40.0	33.2	39.7	32.9	28.9

Note: BNT = Boston naming test; CAB = cognitive assessment battery; CDT = clock-drawing test; COWAT = controlled oral word association, letters F, A, S; TMT = trail making test; PASMO = parallel serial mental operations; ** = <90% test administration (Discussion); MEMO = short story memory test; SDMT = symbol digit modalities test; RCF = Rey complex figure test. All (*n* = 237) native speakers' scores used for norms. T-scores (mean 50, *SD* 10) rounded to even numbers for readability.

2002). Potentially invalid scores (e.g., adversely affected results among participants diverging from normative samples, such as more recent US residents) were described by Strauss et al. (2006). Lower non-native English speaker scores were seen in TMT A and category fluency (Kisser, Wendell, Spencer, & Waldstein, 2012). Previously, language effects in Stroop have been found (Rosselli et al., 2002), however, the complexities of interpreting a bilingual Stroop effect have also been raised (Bialystok, Craik, Green, & Gollan, 2009). Still, our findings of non-native speaker having lower scores also in Stroop 1 (color naming) lead us to suspect that even seemingly minimal verbal components may create NP test artifacts. Also, in light of the hypothesis of higher bilingual executive scores (Bialystok et al., 2012), a find of no superior mean scores is somewhat surprising. However, other studies may be only partly comparable.

First, the previously mentioned common lack of explicit bilingualism definitions makes the validity of comparisons uncertain (Adesope et al., 2010). Second, when several factors contribute, many combinations may affect NP scores, and in the case of bilingualism this list is long: Listening in a noisy environment has been found to be easier in L1 than in L2 (Golestani, Rosen, & Scott, 2009); high L2-proficiency benefits listening comprehension in reverberating environments (Sorqvist, Hurtig, Ljung, & Ronnberg, 2014); foreign languages appear to sound faster (Bosker & Reinisch, 2017); bilinguals make more tongue-twisting errors than monolinguals (Gollan & Goldrick, 2012); L2 slow-down may be located in late stages of speech production (Broos, Duyck, & Hartsuiker, 2019); L2 vocabulary size is often smaller (Bialystok et al., 2012; Luk & Bialystok, 2013); grammar is better in a language learned early (Hartshorne, Tenenbaum, & Pinker, 2018); declarative and procedural memory may be related to different L2 learning stages (Hamrick, Lum, & Ullman, 2018); emotional memory is more strongly tied to L1 than to L2 (Dylman & Bjärtå, 2019). Furthermore, if bilingualism means higher levels of proficiency, selecting bilinguals might

Table 6. Non-native participants by BNT at or above 25 points (order: effect size)

Neuropsychological test	BNT = < 25 mean *n* = 63 (*n* range = 15–66)		BNT > 25 mean *n* = 19 (*n* range = 17–19)		*p*	Effect size (Hedges' *g*)
	Mean (*SD*)	T Cf native (T Cf > 25)	Mean (SD)	T Cf native (T Cf > 25)		
BNT (CAB)	18.9 (4.8)	13.3 (−22.8)	27.3 (1.2)	51.3 (50)	<.001	2.0
Animal naming	16.4 (5.0)	35.4 (37.0)	23.4 (5.4)	46.9 (50)	<.001	1.4
COWAT FAS	25.2 (10.3)	36.5 (35.0)	38.5 (8.9)	47.3 (50)	<.001	1.3
MEMO delayed verbatim	5.4 (3.1)	40.4 (35.4)	9.2 (2.6)	51.9 (50)	<.001	1.3
MEMO delay synonym	8.3 (3.4)	40.3 (38.1)	11.9 (3.0)	50.9 (50)	<.001	1.1
MEMO direct verbatim	4.0 (2.5)	40.3 (40.3)	6.7 (2.8)	49.7 (50)	<.001	1.0
Token test (CAB)	3.7 (1.9)	15.2 (27.4)	5.4 (0.8)	45.7 (50)	<.001	1.0
StroopVictoria2	26.4 (5.8)	24.3 (40.1)	20.6 (5.9)	41.0 (50)	<.001	1.0
RCF delayed	16.1 (6.2)	45.3 (40.5)	21.9 (6.1)	55.8 (50)	.001	0.9
StroopVictoria1	20.8 (5.6)	21.7 (38.4)	15.9 (4.3)	40.6 (50)	<.001	0.9
RCF immediate	16.5 (5.9)	44.8 (41.4)	21.8 (6.2)	54.5 (50)	.003	0.9
TMT B**	135.9 (75.0)	31.5 (74.8)	82.1 (21.7)	49.9 (50)	<.001	0.8
MEMO direct synonym	6.1 (3.1)	39.6 (41.4)	8.5 (2.8)	47.4 (50)	.004	0.8
RCF recognition	19.4 (2.6)	44.8 (39.8)	21.0 (1.6)	53.1 (50)	.002	0.7
StroopVictoria3	37.2 (12.2)	31.7 (42.7)	29.7 (10.2)	43.8 (50)	.011	0.6
SDMT (CAB) correct	36.3 (8.5)	41.9 (44.2)	41.3 (8.6)	47.5 (50)	.033	0.6
RCF copy	31.6 (4.4)	44.5 (39.2)	33.9 (2.2)	52.7 (50)	.002	0.6
PASMO**	113.5 (47.1)	34.2 (44.1)	90.0 (40.1)	43.3 (50)	.142	0.5
TMT A	48.0 (18.2)	40.2 (42.3)	39.7 (10.8)	47.0 (50)	.016	0.5
SDMT (CAB) errors	1.1 (1.6)	44.0 (44.0)	0.5 (1.0)	49.6 (50)	.056	0.4
RCF copy time	222.2 (104.9)	45.8 (55.7)	184.7 (65.7)	50.1 (50)	.066	0.4
RCF strategy (A1, B2, C3)	1.5 (0.7)	48.2 (47.7)	1.4 (0.5)	50.0 (50)	.431	0.2
MEAN T-scores		36.5 (39.4)		48.6 (50)		
Demographic analysis						
% Women	48.5		63.2		.257^	
Age	60.8 (3.4)		61.9 (3.6)		.223	
Education	12.8 (3.2)		13.9 (4.2)		.281	
Clock drawing	4.7 (0.5)	46.5 (n/a)	5.0 (0.0)	53.8 (50)	<.001	0.6
Age of arrival to Sweden	28.5 (9.2)		18.5 (9.6)		<.001	
Years of Swedish residency	32.3 (10.0)		43.4 (8.3)		<.001	

Note: BNT = Boston naming test; CAB = cognitive assessment battery; COWAT = controlled oral word association, letters F, A, S; TMT = trail making test; PASMO = parallel serial mental operations; ** = < 90% test administration (Discussion); MEMO = short story memory test; SDMT = symbol digit modalities Test; RCF = Rey complex figure test; ^ = Chi square; T = T-score (mean 50, *SD* 10). All (*n* = 237) native speakers' scores used for norms. The effect size (Hedges' *g*) was calculated between the two non-native groups, BNT ≤ 25, and BNT > 25.

imply selecting those with superior scores, generally. Better executive scores among participants with superior scores are not uncommon in NP measurements. Although not the focus of this study, specification of bilingualism appears necessary to evaluate the concept.

Comparison to Swedish Studies

Many studies of multilingualism address English and Spanish (e.g., the USA) or English and French (Canada). Not many studies involve Swedish. One recent study in Finland (Finland's two official languages are Finnish and Swedish) assessed one hundred Finnish children and suggested that no specific norms were needed for Finnish-Swedish bilingual children (Karlsson, Soveri, Rasanen, et al., 2015). Yet, perhaps the largest study in Sweden addressing similar questions to date is that of the Betula material (Ljungberg, Hansson, Andres, Josefsson, & Nilsson, 2013) (*n* = 178), and this study found advantages in bilinguals. However, in the Betula material the bilinguals were native Swedes, living in Sweden, who had studied English in school for approximately 7 years (Ljungberg et al., 2013). Our findings contradict the use of identical norms, but we investigated multilingualism in older participants with different countries of birth. We also studied participants with only minor health complaints (if any). More research is needed.

Cutoff at 20 Years of Age Upon Arrival to Sweden

We found different NP scores by dichotomizing by arrival before or after a cutoff age of 20 years. A cutoff age (instead of e.g., a regression) was selected for two reasons: First, Sweden has 9 years of compulsory primary schooling starting at 6–7 years, with 2–4 years of optional upper secondary/high school (gymnasium). A cutoff at 20 years of age would dichotomize between those more or less likely to have attended Swedish schools, and thus learned the Swedish alphabet first. The geographical location of primary schooling was expected to be more relevant than the total length of education in NP tests using the alphabet. Second, the capacity for language learning (measured by an online grammar test on 669,498 native and non-native English speakers) was recently suggested to begin to decline from 17.4 years of age (Hartshorne et al., 2018). In our study the mean arrival age for non-native speakers with high-normal BNT scores was 18.5 years, 10 years earlier than that of non-native speakers with BNT ≤ 25 at 28.5 years. Thus, our results support the importance of the age of acquisition (Bylund, Abrahamsson, Hyltenstam, & Norrman, 2019; Hartshorne et al., 2018). Although a slightly lower cutoff age might have been interesting, our cutoff at 20 years of age was deemed acceptable.

Native Language Family

We found that participants who had emigrated from countries where the majority language is more closely related to Swedish (Germanic), or where Swedish is also spoken (Finland in Finno-Ugric) performed better on NP tests. However, a closer look also revealed that the language families with the highest NP-scores were also those whose members had arrived in Sweden at the youngest ages (Germanic at 23.7 years, Finno-Ugric at 22.9), whereas the group with the lowest NP scores (Other) arrived at the oldest age (29.9 years). However, although gender proportions, age or education were not significantly different in any of the other analyses, sub grouping into native language families produced group differences in gender, age and education. However, here these differences were only listed (i.e., not controlled for), due to minimal information on participants' native language(s), as well as a lack of information on socioeconomic status, factors that have previously been shown to be important (Adesope et al., 2010; Peal & Lambert, 1962). To properly disentangle the effects of factors such as language family distance from those of age of arrival, further studies will be needed with more participants, preferably with matched features of education (e.g., length, grades, geographical location of primary schooling), comparable ages of arrival, more information on native language proficiency, and more information on acculturation and socioeconomic factors.

BNT Proficiency

We found assumptions of language skills from fluency in short conversations to be insufficient and split the group of non-native speakers in BNT normal-to-high (>25 of 30 points) and BNT low (≤ 25 of 30 points) proficiency, a level approximately one standard deviation below the mean of our native participants. Our BNT (CAB, 30 item) native speaker mean score was 27.0, and while our BNT (CAB) was a redrawn version with only 30 pictures, the same motives as in the standard 60 item BNT were used. A 30 item test cannot be directly compared to a 60 item, but using the published list of the percentage of correct answers per item from Swedish BNT norms (Tallberg, 2005) a mean score of 27.2 would be suggested for these 30 items. Thus, as the mean BNT (CAB) score was found to be very similar to the published Swedish BNT norms (Tallberg, 2005), our cutoff was deemed acceptable. Also, BNT may be influenced by level of education, but in this shorter version such effects were expected to be smaller. However, please see Tallberg (2005) for a more in-depth discussion on the many aspects on how education may affect BNT responses. Further research to corroborate our finding of the usefulness of BNT as a language proficiency test in non-native speakers is needed.

Mass Significance and Cases of Below 90% Test Administration

We found large differences between native and non-native NP scores. A Bonferroni correction for 22 tests means a *p*-value of .05/22 = .002, and *p*-values were given with three decimals to enable comparison—in addition to Hedges' *g* effect size calculations (which incorporates *n* and *SD* per respective group and comparison). However, two tasks were administered to fewer than 90% of the non-native speakers; PASMO (Nordlund et al., 2005), a "verbal TMT B" administered to $n = 32$ of $n = 85$, and TMT B (Partington & Leiter, 1949, from (Strauss et al., 2006)), administered to $n = 77$ of $n = 85$. Both these tests incorporate letter/digit shift in numerical/alphabetical order, on condition that the alphabet was correctly recited prior. Although PASMO includes "å, ä, ö" (letters particular to the Swedish alphabet), TMT B stops at "L." However, although a majority of the 34 represented countries use Latin-based alphabets, similar differences between native and non-native speakers were seen not only in PASMO (which requires vocalization) but also in TMT B (which does not require vocalization). From this we hypothesize: First, even if a non-native speaker recites the Swedish alphabet correctly, native recital speed is not guaranteed (similar to the find of slower Stroop naming of colors). Second, slower scores with *no* required vocalization (TMT B) suggest that recital-speed factors are not restricted to vocalized tasks. Third, not only alphabetical order is important, but also choice of alphabetical and/or numerical symbols. Further studies on second language effects in silent tasks, incorporating investigation into multilingualism with respect to alphabet and numbers systems (e.g., Hindu-Arabic vs. Arabic) would be very interesting.

Clinical Strategy

We found that information gathered in a naturalistic clinical setting (i.e., demographic information + confrontation naming) could be used to further NP assessment. However, although age of acquisition was found important, scores among participants arriving before 20 years of age were not found sufficiently similar to native speakers' to automatically use native norms without further investigation. Also, we discourage simple assumptions of bilingualism from conversational fluency, but instead recommend BNT proficiency assessment. Further, based on the TMT B results, we caution against alphabet-based tasks (and neuropsychologists should be careful to take note of a patient's native language *and* writing system). However, perhaps most importantly, our analysis of both original and modified CAB memory test administration indicated that allowing synonym recall (e.g., "plane" not only "airplane") in a memory test likely gives more valid results for all, native and non-native alike.

Limitations

Although our groups were equivalent with regard to gender, education and age, regrettably we lacked data concerning other potentially important variables such as bilingual status, current L1 proficiency (and the age of learning), acculturation, and socioeconomic status. We also had more native speakers than non-native speakers. Item per item analysis of the exact answers in BNT would have been interesting but was not available due to the undetailed registration of the responses (i.e., "correct" or "false"). Still, this may be one of the largest studies of NP administration in Swedish as a second language, and hopefully, our findings will be useful to guide interpretation of common NP tests.

Future Studies

The field of language effects in neuropsychological testing is both important and growing. Clinicians need to learn more, not only about fluency and proficiency, but also of suspected overdiagnosis from common tests of speed and attention.

However, future studies should also address socioeconomic status, attitude toward the second language community, acculturation (e.g., participation in culture of L1 and L2 communities), feelings of conflict between L1 and L2 communities, awareness of ridicule or shame from using one language etc. Studying factors that contribute to integration will be essential. Our participants had lived in Sweden for a mean of 34.8 years, and yet presented large differences in Swedish proficiency.

Conclusion

Second language effects were clear in all neuropsychological tests except for purely visuoconstructive tests. Non-native speakers presented lower scores in all Swedish language administered NP tests with a verbal component. As even our highest performing non-native speakers did not obtain superior mean scores we found no cognitive advantage among middle-aged late multilingual speakers.

However, earlier age of acquisition was clearly associated with better scores, and the Boston naming test may be useful to guide clinical interpretation also of tests perhaps thought of as only minimally "verbal."

Funding

This work was supported by Roos af Hjelmsäter testamentsfond. The main funding body of The Swedish CArdioPulmonary bioImage Study (SCAPIS) is the Swedish Heart and Lung Foundation. The study is also funded by the Knut and Alice Wallenberg Foundation, the Swedish Research Council and VINNOVA (Sweden's Innovation agency). In addition, the SCAPIS Pilot study received support from the Sahlgrenska Academy at University of Gothenburg and Region Västra Götaland. We are very grateful to all the participants in this study and the staff at the SCAPIS test center in Gothenburg.

Author Contributions

Jacob Stålhammar formulated research questions, performed statistical analyses, wrote drafts and revisions of the manuscript. Per Hellström, Carl Eckerström and Anders Wallin participated in the statistical analyses, writing and revising of the manuscript. The authors declare that there is no conflict of interest.

Acknowledgements

Eva Bringman and Marie Johansson oversaw participant logistics. Jacob Stålhammar and Dr Arto Nordlund (deceased) prepared the neuropsychological test battery, the database, and supervised raters. Judith Klecki performed the English language editing.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research*, *80*, 207–245. doi: 10.3102/0034654310368803.

Bergstrom, G., Berglund, G., Blomberg, A. et al. (2015). The Swedish CArdioPulmonary BioImage study: Objectives and design. *Journal of Internal Medicine.*. doi: 10.1111/joim.12384.

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

Bialystok, E., Craik, F. I., Green, D. W., & Gollan, T. H. (2009). Bilingual minds. [review]. *Psychol Sci Public Interest*, *10*(*3*), 89–129. doi: 10.1177/1529100610387084.

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

Bosker, H. R., & Reinisch, E. (2017). Foreign languages sound fast: Evidence from implicit rate normalization. *Frontiers in Psychology*, *8*, 1063. doi: 10.3389/fpsyg.2017.01063.

Broos, W. P., Duyck, W., & Hartsuiker, R. J. (2019). Monitoring speech production and comprehension: Where is the second-language delay? *Quarterly Journal of Experimental Psychology (2006)*, *72*(*7*), 1601–1619. doi: 10.1177/1747021818807447.

Bylund, E., Abrahamsson, N., Hyltenstam, K., & Norrman, G. (2019). Revisiting the bilingual lexical deficit: The impact of age of acquisition. *Cognition*, *182*, 45–49. doi: 10.1016/j.cognition.2018.08.020.

Casas, R., Guzman-Velez, E., Cardona-Rodriguez, J. et al. (2012). Interpreter-mediated neuropsychological testing of monolingual Spanish speakers. *The Clinical Neuropsychologist*, *26*(*1*), 88–101. doi: 10.1080/13854046.2011.640641.

Dylman, A. S., & Bjärtå, A. (2019). When your heart is in your mouth: The effect of second language use on negative emotions. *Cognition and Emotion*, *33*(*6*), 1284–1290. doi: 10.1080/02699931.2018.1540403.

Erdodi, L. A., Jongsma, K. A., & Issa, M. (2017). The 15-item version of the Boston naming test as an index of English proficiency. *The Clinical Neuropsychologist*, *31*(*1*), 168–178. doi: 10.1080/13854046.2016.1224392.

Gasquoine, P. G., Croyle, K. L., Cavazos-Gonzalez, C., & Sandoval, O. (2007). Language of administration and neuropsychological test performance in neurologically intact Hispanic American bilingual adults. [research support, non-U.S. Gov't]. *Archives of Clinical Neuropsychology*, *22*(*8*), 991–1001. doi: 10.1016/j.acn.2007.08.003.

Golestani, N., Rosen, S., & Scott, S. K. (2009). Native-language benefit for understanding speech-in-noise: The contribution of semantics. *Bilingualism (Cambridge, England)*, *12*(*3*), 385–392. doi: 10.1017/S1366728909990150.

Gollan, T. H., & Goldrick, M. (2012). Does bilingualism twist your tongue? *Cognition*, *125*(*3*), 491–497. doi: 10.1016/j.cognition.2012.08.002.

Hamrick, P., Lum, J. A. G., & Ullman, M. T. (2018). Child first language and adult second language are both tied to general-purpose learning systems. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(*7*), 1487–1492. doi: 10.1073/pnas.1713975115.

Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. doi: 10.1016/j.cognition.2018.04.007.

Kaplan, E., Goodglass, H., Weintraub, S., & Segal, O. (2001). *Boston Naming Test* (, pp. 78757–76897). Austin, TX: Pro-Ed, 8700 Shoal Creek Blvd.

Karlsson, L. C., Soveri, A., Rasanen, P. et al. (2015). Bilingualism and performance on two widely used developmental neuropsychological test batteries. *PLoS One*, *10*(*4*), e0125867. doi: 10.1371/journal.pone.0125867.

Kisser, J. E., Wendell, C. R., Spencer, R. J., & Waldstein, S. R. (2012). Neuropsychological performance of native versus non-native English speakers. *Archives of Clinical Neuropsychology*, *27*(*7*), 749–755. doi: 10.1093/arclin/acs082.

Ljungberg, J. K., Hansson, P., Andres, P., Josefsson, M., & Nilsson, L. G. (2013). A longitudinal study of memory advantages in bilinguals. *PLoS One*, *8*(*9*), e73029. doi: 10.1371/journal.pone.0073029.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology (Hove, England)*, *25*(*5*), 605–621. doi: 10.1080/20445911.2013.795574.

Nielsen, T. R., Segers, K., Vanderaspoilden, V. et al. (2019). Validation of a European cross-cultural neuropsychological test battery (CNTB) for evaluation of dementia. [multicenter study validation studies]. *International Journal of Geriatric Psychiatry*, *34*(*1*), 144–152. doi: 10.1002/gps.5002.

Nordlund, A., Pahlsson, L., Holmberg, C., Lind, K., & Wallin, A. (2011). The cognitive assessment battery (CAB): A rapid test of cognitive domains. *International Psychogeriatrics*, *23*(*7*), 1144–1151. doi: 10.1017/S1041610210002334.

Nordlund, A., Rolstad, S., Hellstrom, P., Sjogren, M., Hansen, S., & Wallin, A. (2005). The Goteborg MCI study: Mild cognitive impairment is a heterogeneous condition. *Journal of Neurology, Neurosurgery, and Psychiatry*, *76*(*11*), 1485–1490. doi: 10.1136/jnnp.2004.050385.

Peal, E., & Lambert, W. E. (1962). *The relation of bilingualism to intelligence*. Washington: American Psychological Association.

Rivera Mindt, M., Arentoft, A., Kubo Germano, K. et al. (2008). Neuropsychological, cognitive, and theoretical considerations for evaluation of bilingual individuals. *Neuropsychology Review*, *18*(*3*), 255–268. doi: 10.1007/s11065-008-9069-7.

Roberts, P. M., Garcia, L. J., Desrochers, A., & Hernandez, D. (2002). English performance of proficient bilingual adults on the Boston naming test. *Aphasiology*, *16*(*4–6*), 635–645. doi: 10.1080/02687030244000220.

Rosselli, M., Ardila, A., Santisi, M. N. et al. (2002). Stroop effect in Spanish-English bilinguals. [clinical trial]. *Journal of the International Neuropsychological Society*, *8*(*6*), 819–827. doi: 10.1017/s1355617702860106.

Rowland, J. T., Basic, D., Storey, J. E., & Conforti, D. A. (2006). The Rowland universal dementia assessment scale (RUDAS) and the Folstein MMSE in a multicultural cohort of elderly persons. [comparative study]. *International Psychogeriatrics*, *18*(*1*), 111–120. doi: 10.1017/S1041610205003133.

Shishkin, E., & Ecke, P. (2018). Language dominance, verbal fluency, and language control in two groups of Russian–English bilinguals. *Languages*, *3*(*3*), 27.

Shulman, K. I. (2000). Clock-drawing: Is it the ideal cognitive screening test? [review]. *International Journal of Geriatric Psychiatry*, *15*(*6*), 548–561. doi: 10.1002/1099-1166(200006)15:6<548::aid-gps242>3.0.co;2-u.

Sorqvist, P., Hurtig, A., Ljung, R., & Ronnberg, J. (2014). High second-language proficiency protects against the effects of reverberation on listening comprehension. *Scandinavian Journal of Psychology*, *55*(*2*), 91–96. doi: 10.1111/sjop.12115.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary* (3rd ed.). Oxford; New York: Oxford University Press.

Suarez, P. A., Gollan, T. H., Heaton, R., Grant, I., & Cherner, M. (2014). Second-language fluency predicts native language stroop effects: Evidence from Spanish-English bilinguals. *Journal of the International Neuropsychological Society*, *20*(*3*), 342–348. doi: 10.1017/S1355617714000058.

Sweden_Statistics (2020). Folkmängd efter födelseland 1900-2019, be0101_folkmangd-fodelselander-1900-2019. *Population statistics.* Retrieved June 11, 2020, from. http://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningens-sammansattning/befolkningsstatistik/pong/tabell-och-diagram/helarsstatistik--riket/folkmangd-efter-fodelseland/.

Tallberg, I. M. (2005). The Boston naming test in Swedish: Normative data. *Brain and Language*, *94*(*1*), 19–31. doi: 10.1016/j.bandl.2004.11.004.