Ultimate Attainment in Late Second Language Acquisition:

Determined by Age of Arrival or Individual Socio-Psychological Differences?

Janna Brummel

3217833

Master's thesis

Linguistics: The Study of the Language Faculty

Utrecht University

June 2012

Supervisor: Dr. Merel Keijzer

Second reader: Dr. Sharon Unsworth

Index

1. Introduction

Anecdotally, people seem to experience more difficulties in acquiring a new, second, language as they become older; children seem to be more likely to achieve a level of ultimate attainment in their second language that is comparable to that of native speakers of that particular language, whereas adults might not be as successful. These assumptions are represented in the form of the critical period hypothesis (Lenneberg 1967), which proposes that second language learners will experience increasing difficulty if they start learning a second language after puberty. While this hypothesis has gained immense popularity within the field of second language acquisition, it has not always been validated in research due to the identification of successful late second language learners. Furthermore, the prominent focus on the relationship between age of acquisition or arrival and level of attainment in the second language has been questioned, as learners differ on much more dimensions than merely the age they started learning a new language (Muñoz & Singleton 2011). These differences among individuals, such as educational level or length, working memory capacity, language aptitude or intelligence, may also account for variation among late second language learners (Dörnyei 2006). However, the exact influence of these predictors has not yet been widely studied.

This thesis will explore the validity of the critical period hypothesis by comparing linguistic performances of late Dutch-English bilingual migrants and English monolinguals. Most importantly, this study investigates the additional impact of individual socio-psychological factors on second language performances of Dutch migrants in Melbourne, Australia. In doing so, this study aims to determine which individual differences, in addition to age of arrival, are predictive of second language success of late learners of English in a migrant context. As a consequence, this study may help to uncover a blueprint of the successful late second language learner.

First, in section 2, a theoretical framework for this study will be provided. In this framework, the critical period hypothesis and its consequences for late second language acquisition will be discussed, followed by an overview of studies which argue that individual differences might play a more dominant role in late second language acquisition than is typically assumed by proponents of the critical period hypothesis. Based on this theoretical framework, research questions and hypotheses for the current study will be formulated in section 3. The methodology employed to answer the research questions is provided in section 4. Results are presented in section 5, followed by a discussion and conclusions in section 6.

2. Theoretical framework

This section aims to discuss two prominent lines of research in the field of late second language acquisition. In section 2.1, the critical period hypothesis (Lenneberg 1967) will be introduced and discussed. Roughly, this hypothesis argues that late second language learners should experience significantly more difficulty in the acquisition of their second language than early second language learners. The critical period hypothesis and age effects have maintained the interest of many researchers in the field of second language research, even predominantly so (Muñoz & Singleton 2011). Due to this prevalent focus on the effect of age on second language acquisition, the assumption seems to be maintained that learners only differ on the basis of age of acquisition while individual learner characteristics that could similarly influence the degree of obtained second language success are often overlooked. In section 2.2, these individual differences among learners, including working memory, level and length of education, fluid intelligence, language aptitude and first language proficiency, that may be predictive of successful late learners' second language performances will be identified and discussed.

2.1. Critical period hypothesis

2.1.1. Critical periods

Generally, most behavioral domains show an increasing developmental competence since the majority of behavioral skills, e.g. social, reading or presentation skills, do not decrease with age but increase with practice (Johnson & Newport 1989, Newport 1991). However, the opposite pattern can be observed in domains in which critical periods are found. A critical period for learning can be characterized as a limited time span, typically occurring early in life, during which the learning potential is at its peak. In subsequent development, this ability to learn decreases and eventually is bounded by a predictable terminus (Birdsong 2005).

The existence of critical periods has often been observed in developmental biology. Cited examples include a number of bird species who can only successfully acquire birdsong during a limited period in early life (Marler 1991). Furthermore, the development of binocularity in cats, monkeys and man can only occur during a limited time span (Almli & Finger 1987) and a critical period can also be stipulated for imprinting in ducklings (De Villiers & De Villiers 1978). However, critical periods are not unique to biology and similar effects are often said to be observable in both first and second language acquisition.

2.1.2. A critical period in language acquisition

Penfield & Roberts (1959) and Lenneberg (1967) were the first to extend the notion of a critical period to the field of language acquisition by proposing that "a child's brain has a specialized capacity for learning language – a capacity that decreases with the passage of years" (Penfield & Roberts 1959, p. 240). Since hardly any cases have been reported in which children did not receive any language input until after the critical period, the evidence in favor of age effects in first language acquisition is mostly derived from clinical investigations.

Such investigations have, for example, shown that feral children, who were raised in isolation from human contact, such as Genie and Chelsea (Curtiss 1988), often largely fail to acquire the grammar of their first language at an older age, i.e. 13 years and early thirties respectively, due to the absence of normal input during the critical period. However, it has been rightfully claimed that these results could have been influenced by the child's social isolation and experiential deprivation (Mayberry 1993). To avoid these confounding factors, studies were conducted on the level of attainment reached in the first language acquisition of American Sign Language (ASL) by adults who were first exposed to ASL during late childhood (Mayberry 1993). When deaf children are born to hearing parents without any knowledge of ASL or become deaf postlingually, their age of first exposure to ASL often occurs towards the end of or after the critical period due to normally hearing parents' lack of signing knowledge, educational policies in the 1960s (i.e. oralism, a movement prohibiting gesturing and signing in early schooling) and late enrolment at schools for deaf children. The late exposure has been found to heavily impede first language acquisition of ASL on all linguistic levels and late learners do not seem to catch on when compared to adults who acquired ASL at an earlier age despite 28+ years of exposure to the language (Mayberry & Eichen 1991, Mayberry 1993). Finally, the observation of age effects in aphasia recovery patterns has also been claimed to support the existence of a critical period in language acquisition since adults are not guaranteed to relearn language after the onset of aphasia while children generally are (Lenneberg 1967).

Whereas this evidence only seems to argue in favor of a critical period for first language acquisition, Lenneberg (1967) also introduced a critical period hypothesis for second language acquisition, suggesting that second language learning is often increasingly more difficult after puberty. Furthermore, it is proposed that "automatic acquisition from mere exposure to a given language seems to disappear after this age, and foreign languages have to be taught and learned through a conscious and labored effort" (Lenneberg 1967, p. 176). This statement does not only match the frequently observed negative correlations between age and level of attainment in a second language (Birdsong 2005), but also the popular belief that younger is better when it comes to learning languages (DeKeyser 2000). Since Lenneberg's (1967) proposal, many studies on second

language acquisition have taken the critical period hypothesis as their vantage point whenever non-native outcomes were achieved (Moyer 2004), not taking into account individual differences, as will be addressed in section 2.2.

2.1.3. Geometric and temporal features of a critical period

Critical periods in biology and language acquisition share geometric features (Van Boxtel 2005). Typically, critical periods have an onset, a peak or plateau, a gradual offset and a terminus after which the function flattens, as illustrated by Fig. 1. During the onset, the sensitivity to environmental stimuli essential for the development of a specific behavioral ability increases until it reaches a peak or plateau. For the duration of this peak, it is possible for the learner to reach the highest level of attainment on the basis of the environmental stimuli. Subsequently, this ability to learn decreases in the offset until it reaches a terminus marking the end of the critical period.

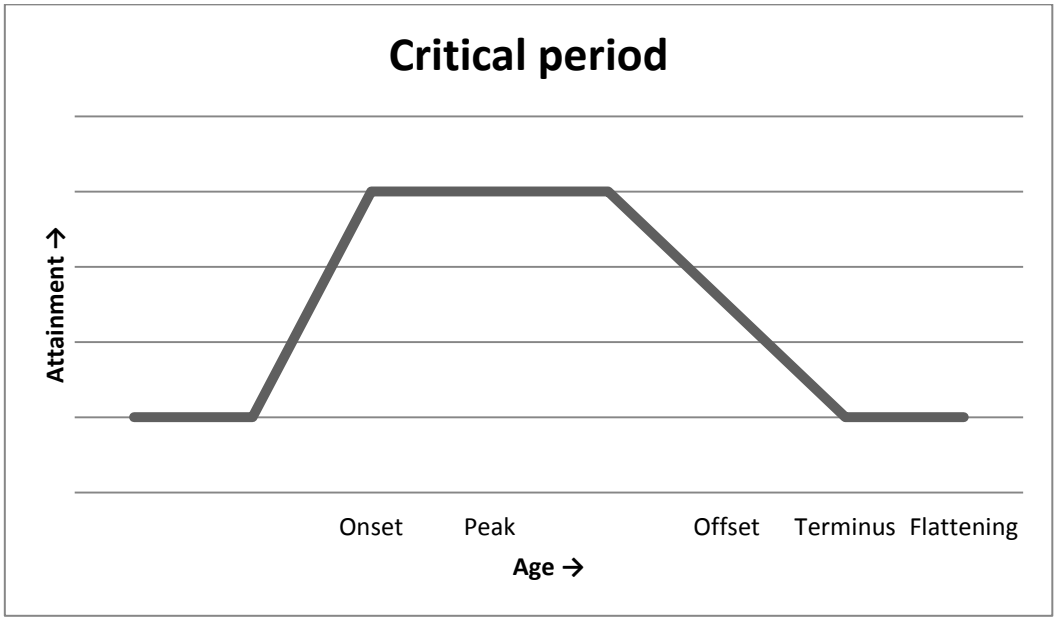

Figure 1: Geometric features of a critical period, adapted from Birdsong (2005).

It should be noted that some studies prefer the term sensitive period, which differs from a critical period with regards to abruptness of offset and the degree of sensitivity afterwards (Newport et al. 2001). This thesis will focus on the critical period, rather than a sensitive period, as defined in this section and in Fig. 1.

A critical period hypothesis for first and second language acquisition can only be established if the critical period geometry would follow the function as shown in Fig. 1, including an identifiable discontinuity or cut-off point in the declining slope at the terminus (Bialystok & Miller 1999). Furthermore, once the process of language acquisition commences after the terminus (generally

considered to be at least before or around puberty (Bialystok & Miller 1999), as will be further discussed below), the learner would not be expected to reach a level of native-like ultimate attainment. However, both these essential patterns are often not found in studies on second language acquisition of phonology and morphosyntax (Bongaerts 2005). A number of alternative age functions for attainment in second language acquisition have been proposed, as illustrated by Fig. 2. As can be seen, Fig. 2.A, Fig. 2.B and Fig. 2.C all differ from Fig. 2.D since they all lack the required discontinuity characteristic of a critical period. Fig. 2.A represents a function in which attainment declines with age after maturation has ended. Fig. 2.B is similar to Fig. 2.A since, in this function, attainment declines with age before the maturational process has been completed. Both Fig. 2.A and 2.B are unbounded age of acquisition – attainment functions, so no flattening can be observed. Fig. 2.C represents a continuous linear function in which level of attainment declines with age of acquisition. Even though Fig. 2.C seems to suggest a negative correlation between proficiency levels and age of acquisition, a similar effect could also be caused by a number of alternative factors such as educational background, cross-linguistic influence, motivation, attitude and length of exposure (Bialystok & Miller 1999), which will be further discussed in section 2.2. Lastly, Fig. 2.D represents a traditional critical period as presented in Fig. 1 since it features all the required characteristics of an offset and flattening.

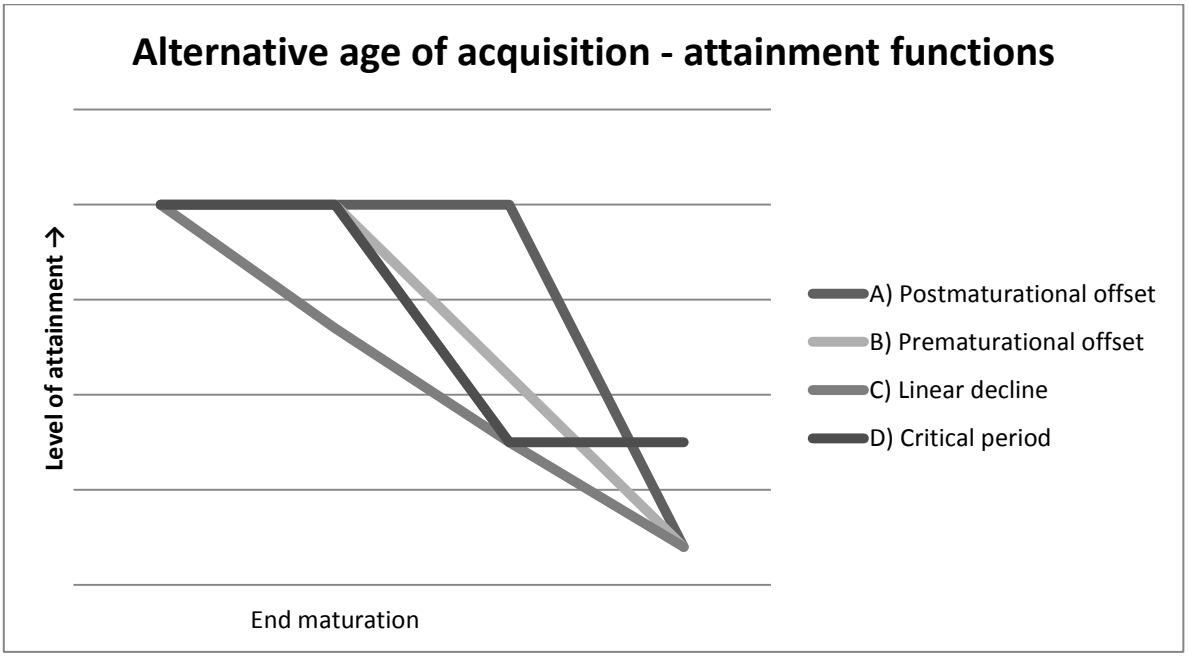

Figure 2: Unbounded alternative age functions for the critical period hypothesis as adapted from Birdsong (2005). Note that no onsets are represented in this figure.

Previous second language research has attempted to provide insights into the temporal features of the critical period by proposing a range of ages for the offset. In doing so, the focus has

often been on phonology and morphosyntax since not all linguistic components seem to be equally susceptible to a critical period. For example, the acquisition of vocabulary and semantics does not seem to be determined by a critical period (Singleton 2005, Slabakova 2006), presumably since both elements can be learned through explicit learning mechanisms. Furthermore, a distinction can also be made between phonology and morphosyntax on the basis of differences in frequently observed age effects (Seliger 1978), in the sense that the offset of second language phonological acquisition seems to commence earlier than the offset of the second language acquisition of morphosyntax (Singleton 2005). Consequently, "many critical periods, successive and perhaps overlapping, lasting probably throughout one's lifetime, each closing off different acquisition abilities" (Seliger 1978, p. 16) can be observed.

However, it is still extremely difficult to determine the end point of the critical period for language acquisition since a wide range of ages has been proposed (Bates et al. 1997). For phonology, offsets have been proposed between age 1 and puberty, and for morphosyntax, offsets have been proposed between age 9 and puberty (Singleton 2005). Hence, the closing point of the critical period is typically considered to be at least before or around puberty (Bialystok & Miller 1999).

2.1.4. Explanations for the critical period

Generally, maturational constraints are considered to be the main underlying cause for the critical period in language acquisition (Hyltenstam & Abrahamsson 2003). Various proposals have been put forward regarding the nature of these maturational constraints involved in the critical period.

Early accounts by Penfield & Roberts (1959) and Lennenberg (1967) both relate the critical period to neurobiology. Whereas Penfield & Roberts (1959) argue that age effects can be attributed to a decreasing plasticity of the maturing brain, Lennenberg (1967) proposes the lateralization of language functions to be the underlying cause of the critical period. Alternative explanations, such as the maturation of brain cells, myelination[1], the localization of particular language functions and differences in spatial representations in the brain, have also been proposed (Seliger 1978, and for a review, see Singleton 2005).

A different approach, focusing on cognitive-developmental factors, is taken by DeKeyser (2000), who claims that the critical period hypothesis is constrained by implicit learning mechanisms. This claim is supported by the observations that children rely heavily on implicit learning mechanisms while learning languages whereas adults cannot and rely solely on explicit learning mechanisms.

[1] During myelination, axons of neurons become progressively wrapped by glial cells and this reduces plasticity (Singleton 2005).

Furthermore, Hakuta et al. (2003) suggest that age effects can be viewed as a consequence of non-language specific cognitive maturation and decline.

Another explanation suggests that, after puberty, the language learner becomes aware of differences in ease of learning when compared to the situation in childhood and this influences the learning process (Rosansky 1975). Other studies relate the critical period to affective-motivational factors such as the strengthening of the affective filter, inhibition caused by the Freudian super-ego, the hardening of ego-boundaries, narcissism that causes the adult language learner to feel ashamed when learning a new language and social and psychological changes after puberty (for a review, see Singleton 2005).

In brief, whereas most studies agree on the involvement of maturational constraints in the critical period hypothesis in second language acquisition, no consensus has been reached with regards to the nature of these constraints.

2.1.5. Support for a critical period in second language acquisition

In one of the most influential studies investigating the relationship between age effects and ultimate attainment in second language morphosyntax, Johnson & Newport (1989) aimed to provide a better understanding of the nature of the critical period hypothesis by investigating 46 second language learners of English whose first language was either Chinese or Korean. Among these participants, 23 had arrived in the United States before age 15 while the other half of the learners had arrived in the United States after age 17. Length of exposure for the two groups of participants was fairly even at 9.8 and 9.9 years on average. In addition to the second language learners, 23 monolingual speakers of American English were included in the study as control subjects. Morphosyntactic knowledge was assessed using a grammaticality judgment task that consisted of 12 English syntactic structures (including rule types such as subcategorization, wh-questions, pronominalization and past tense), which were tested in 276 sentence items. The results suggested a strong relationship between age of arrival and ultimate attainment ($r = -.77$, $p < .01$). Subsequent analyses were based on smaller age groups (3-7, 8-10, 11-15, 17-39) to look at the relationship between age of arrival and ultimate attainment in more detail. Only a subset of the early arrivals, i.e. participants in the youngest age group, performed similarly to the native speaker controls while participants from all other age groups performed significantly below the monolingual controls. Consequently, this analysis suggests that the second language morphosyntactic proficiency of participants who arrived in the United States after age 7 decreases as age of arrival increases and a level of attainment comparable to that of native

speakers cannot be achieved[2]. Subsequently, participants were redivided into an early and late arrival group to study the effects of age of arrival before and after puberty. For the early arrivals (with an age of arrival between 3 and 15 years), age of arrival correlated significantly ($r = -.87$, $p < .01$) with ultimate attainment while no significant correlation was observed for the late arrivals (with an age of arrival between 17 and 39 years), as shown in the scatterplots in Fig. 3. These results are suggestive of a flattening out of the age of acquisition – ultimate attainment function since age effects do not play a prominent role among the late second language learners, a finding consistent with the critical period hypothesis.

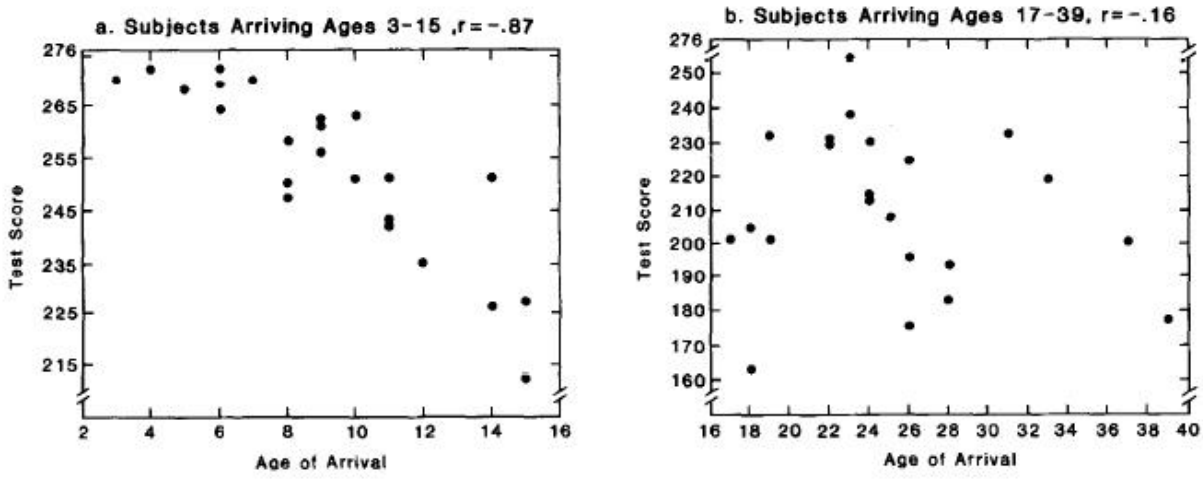

Figure 3: The age of arrival-ultimate attainment functions for early and late arrivals from Johnson & Newport (1989).

As a result, the performance of the late learners seems to be influenced by individual differences due to the great degree of variability and lack of linearity in the relevant scatterplot, as shown in Fig. 3.b. Hence, Johnson & Newport (1989) argue that their critical period hypothesis function demonstrates an early linear decline from age 7 onwards with a greater decline around puberty.

DeKeyser et al. (2010) aimed to contribute to the discussion on the nature and shape of the age of acquisition – ultimate attainment function by presenting two parallel large-scale studies of adult Russian learners of either English or Hebrew as a second language. By studying learners with the same first language background but different second languages, the role of cross-linguistic influence and the generalizability of results can be investigated. The participants of this large-scale study, 138 Russian immigrants who had all lived in North America or Israel for a minimum of 8 years, were presented with an English grammaticality judgment task adapted from Johnson & Newport (1989) or a Hebrew grammaticality judgment task while a Russian language aptitude test was administered for all subjects. Results for both groups of second language learners revealed similar

[2] It should be noted that this conclusion was based on results from a grammaticality judgment task, a task of which the reliability and validity have been questioned in recent discussions (Altenberg & Vago, 2004, Mandell 1999).

patterns since the learners' age of acquisition - ultimate attainment functions revealed patterns consistent with the traditional notion of a critical period (including a discontinuity and flattening out of the function). The evidence in favor of a critical period was provided by significant correlations between age of acquisition and ultimate attainment for participants with an age of acquisition below 18 years old ($r = -.71$, $p = < .01$ for migrants in North America, $r = -.51$, $p = < .05$ for migrants in Israel) while no significance was reached in the correlations for the groups of subjects who were 18 to 40 years old or above 40 years old at age of acquisition when age at testing was partialed out. When a closer look was taken at the youngest age group in order to determine the position of the steepest decline of the function, age 12 was found to be that location, consistent with the proposed cut-off point of puberty by Johnson & Newport (1989). Additionally, in the adult second language learners, aged between 18 and 40 at the onset of learning, a significant effect of language aptitude was observed, a result matching the conclusions drawn by DeKeyser (2000), as will be discussed below. Consequently, DeKeyser et al. (2010) support the critical period hypothesis for the second language acquisition in morphosyntax as proposed by Johnson & Newport (1989).

While Johnson & Newport (1989) and DeKeyser et al. (2010) both support the critical period hypothesis in their studies on morphosyntax relying on highly similar materials to study the phenomenon, Birdsong & Molis (2001) did not replicate these results. In their study, Birdsong & Molis (2001) replicated Johnson & Newport's (1989) study by applying the exact same methods and administering the same materials used in Johnson & Newport (1989) while focusing on a new population of 61 native speakers of Spanish who became second language learners of English after migrating to the United States at either an early age (below 16 years) or at a later age (17 years and above). Contrary to the findings of Johnson & Newport (1989), the results of Birdsong & Molis (2001) suggested that for the early arrivals, no significant correlation could be observed between age of arrival and ultimate attainment ($r = -.24$, $p = .22$) while strong age effects were found in the grammaticality judgment performance of the late arrivals ($r = -.69$, $p < .0001$). The findings of Birdsong & Molis (2001) are not consistent with the critical period hypothesis, but, instead, seem to reflect an age of acquisition – ultimate attainment function with a postmaturational offset, as shown in Fig. 2.A. Birdsong & Molis (2001) is not the only study to have concluded that its results do not match the critical period hypothesis (see Bongaerts 2005 and Muñoz & Singleton 2011 for a review) and at the moment no consensus has been reached with regards to this issue. However, various researchers have slightly shifted away from questioning the existence of the critical period hypothesis by focusing on differences among second language learners (e.g. on the basis of language aptitude, working memory, level of education) that could influence second language performance, rather than solely studying the effect of age of acquisition or arrival (Muñoz & Singleton 2011).

2.1.6. Practical implications

Since the 1950s and 1960s, research on the critical period hypothesis for second language acquisition has heavily influenced language educational policies (Stern 1983, Singleton 2007). Language policy makers have often relied on literature suggesting that children have more success in learning a second language than adults do to advocate and justify early second language instruction programs (DeKeyser & Larson-Hall 2005, Singleton 2005). On the other hand, adult second language learners might become discouraged and give up on learning a foreign language for the same reasons (DeKeyser & Larson-Hall 2005).

However, DeKeyser (2000) warns for overstating age effects in second language acquisition. Children and adults rely on different learning mechanisms so the critical period hypothesis does not necessarily suggest that children will achieve second language success in classroom settings since their strengths include learning languages implicitly, not explicitly. Conversely, late second language learners seem to rely on explicit learning mechanisms and, hence, their strengths lie in the acquisition of second language vocabulary and rule-based languages. Johnson & Newport (1989) also stress the differences between classroom learning and immersion, proposing the reliance on explicit learning mechanisms might be the cause for unsuccessful early second language instruction. Consequently, the critical period hypothesis for second language acquisition should be used with caution when it comes to language educational policies.

2.2. Individual socio-psychological factors

Recently, the validity of the critical period hypothesis has been questioned by various researchers who have (partly) argued against age effects in second language acquisition, and, instead, attribute variation among second language learners to additional factors. These factors, including foreign language aptitude (DeKeyser 2000, Abrahammson & Hyltenstam 2008), working memory (McDonald 2006, DeKeyser & Koeth 2010, Juffs & Harrington 2011) and level and/or length of education (Hakuta et al. 2003), are, despite their predictive abilities of second language success (Dörnyei 2006), often undermined in second language research due to a prevalent focus on the critical period hypothesis and group effects (Muñoz & Singleton 2011). However, studying these socio-psychological factors in addition to age of acquisition or arrival might provide insights into second language teaching and the psychology of the second language learner.

2.2.1. Length and level of education

Intuitively, people with a higher degree of education tend to use more complex and diverse syntax and vocabulary in language production. These intuitions seem to be verified by studies focusing on the effect of the language input provided by parents with a high socioeconomic status (a measure comprising education, occupation and income) on child language development (Vasilyeva & Waterfall 2011). Some researchers, including Flege et al. (1999) and Hakuta et al. (2003), argue that the effect of socioeconomic status, and level and length of education in particular, is not restricted to first language acquisition but extends also to the domain of second language acquisition.

In their study, Flege et al. (1999) aimed to examine the critical period hypothesis by investigating the influence of age of arrival and additional factors on the second language morphosyntactic performance of a migrant population. The morphosyntactic knowledge of 240 Korean second language learners of English was assessed using the grammaticality judgment task designed by Johnson & Newport (1989), including both rule-based (i.e. relatively simple rules with widespread application, such as adding *–ed* to create a past tense in **The man paints his house yesterday*) and lexically-based sentence structures (i.e. irregular and ungeneralizable aspects of English morphosyntax dealing with particles, prepositions with verbs or idiosyncratic features of verbs, such as adding *for* in **The farmers were hoping rain*). The results of the grammaticality judgment task revealed that scores decreased as the participants' age of arrival increased, according to a pattern consistent with a critical period ending around age 12. However, as Korean immigrants with an earlier age of arrival had generally received more years of education in the United States, a significant and high correlation between the factor age of arrival and the learners' received amount of education was observed ($r = .92$, $P = < .01$), and this multicollinearity caused the two factors to be confounded. When variation in age of arrival was controlled for, statistical analyses revealed that length of education had a significant and independent influence on the learners' performance on rule-based morphosyntactic structures, while language use patterns influenced the immigrants' performance on lexically-based morphosyntactic structures. Consequently, the apparent age effects on morphosyntax seem to be related to variations in length of education and language use, rather than variations in age of arrival, and thus do not wholly match the critical period hypothesis.

Similar results were obtained in Hakuta et al.'s (2003) study on the existence of the critical period hypothesis in self-assessed English language proficiency measures provided by 2,016,317 Spanish immigrants and 324,444 Chinese immigrants in the 1990 U.S. Census (U.S. Department of Commerce 1995). When these data were modeled on the factor of level of education, this factor explained 42.21% of the variance observed in the second language proficiency of Chinese-speaking immigrants and 26.22% of the variance in the second language proficiency of Spanish-speaking

immigrants. Adding the factor age of arrival accounted for another 9.32% of explained variance for Chinese-speaking immigrants and 6.32% of explained variance for Spanish-speaking immigrants. Consequently, both level of education and age of arrival contributed significantly to explaining the variance in the self-reported second language measures. Thus, these data reveal that besides age of arrival, level of education is the most significant predictor in the self-observed second language success of Spanish-speaking and Chinese-speaking migrants. When mapping the regression of English proficiency on age of arrival, the regression line declined in a linear fashion rather than matching the pattern of a critical period with a close around puberty, a finding consistent with the function in Fig. 2.C.

2.2.2. Working memory

Working memory can be defined as a limited capacity system which temporarily stores and manipulates information required for performing a range of complex cognitive tasks including comprehension, learning and reasoning (Baddeley 1992, 2000). According to one of the most commonly used models, proposed by Baddeley (2000), working memory consists of a central executive component which interacts with a phonological loop, episodic buffer and visuo-spatial sketchpad, all of which subsequently transmit information to be stored in the long term memory, as shown in Fig. 4.

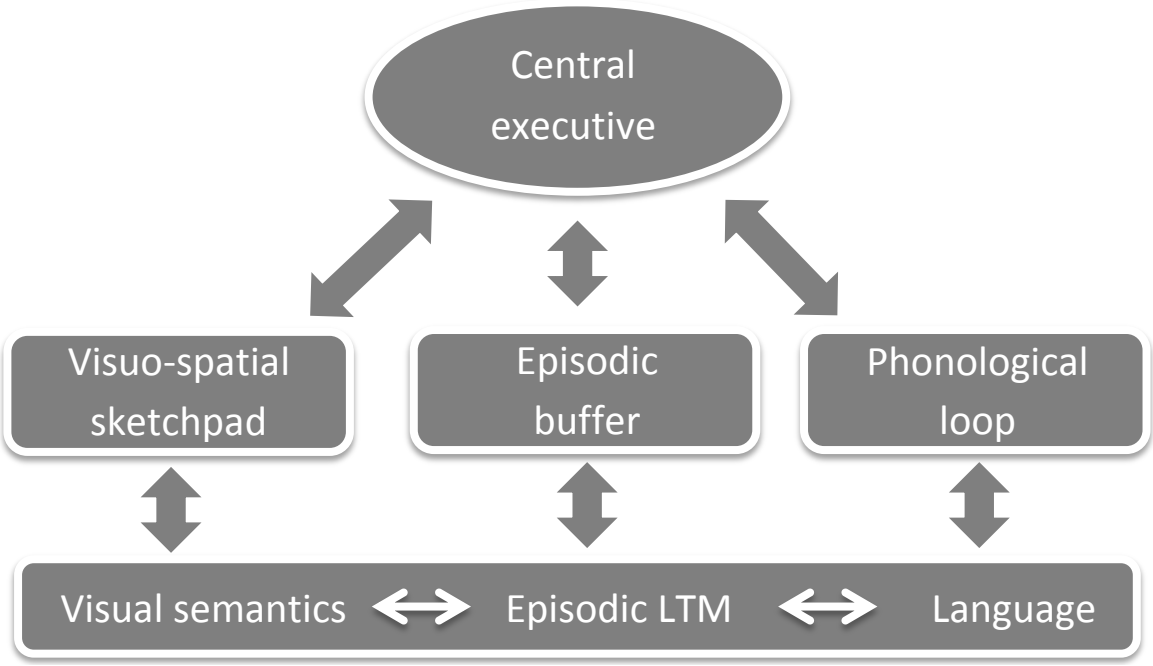

Figure 4: Baddeley's (2000) multi-component working memory model.

While the phonological loop is responsible for storing and rehearsing phonological information from input and transforming verbal cues into phonological information, the visuo-spatial sketchpad is responsible for processing and storing visual and spatial information. The episodic buffer integrates information from various sources in order to create a temporary representation. These three 'slave systems' are all controlled by the central executive component which coordinates the flow of information to the other components. The functions of the central executive and the 'slave systems' reflect the direct relation between working memory and language. The phonological loop in particular has been claimed to have evolved to support language learning since this component of working memory is thought to be essential in vocabulary acquisition (Baddeley et al. 1998). As such, differences in working memory capacity among learners might have a direct impact on second language success, as reflected in research by McDonald (2006), Van den Noort et al. (2006) and Service et al. (2002).

One of the research questions posed in McDonald (2006) was whether the frequently observed late second language learners' poor grammaticality judgment performances could be attributed to a low second language working memory capacity rather than age of arrival. In order to test this, 50 native speakers of English and 50 late second language learners were administered an oral grammaticality judgment task (consisting of a subset of sentence types used in Johnson & Newport (1989)) to assess linguistic performance and an English size judgment task[3] to examine working memory capacity. When the two groups of participants were compared, results revealed that the late second language learners of English had a significantly lower working memory capacity than the native speakers of English. Similarly, the late second language learners also performed poorer on the grammaticality judgment task than the native speakers did. Furthermore, grammaticality judgment performance and reaction times both correlated significantly with working memory capacity scores for the late second language learners, even when the factor age of arrival was controlled for, suggesting that working memory is related to second language proficiency.

Van den Noort et al. (2006) aimed to determine whether an interaction between working memory and language proficiency could be observed in multilinguals' performance on both language proficiency and working memory measures. More specifically, native speakers of Dutch who were late second and third language learners of German and Norwegian respectively were asked to complete a self-assessment and translation test to provide a measure of language proficiency.

[3] In a size judgment task In this task, participants are given a list of words to remember (i.e. store) and reorder on the basis of size (i.e. process) and subsequently report this order. For example, participants would hear a set of nouns (such as 'goldfish', 'pig', 'needle') which they were asked to order on the basis of size from smallest to largest (i.e. 'needle', 'goldfish', 'pig'). This task was preferred over a reading span task as size judgment tasks do not rely on syntax.

Subsequently, they were administered a reading span task, forward and backward digit span tasks and a letter-number ordering test in all languages to provide measures of the working memory capacity per language (since the reading span and backward digit span tasks were included in this study as well, these tasks will be explained in more detail in section 4.2.2. and 4.2.3.). The language proficiency measures revealed that the participants were good speakers of their second language German and fair speakers of their third language Norwegian in both objective and subjective measures. These differences in language proficiency amongst the languages spoken by participants were reflected in performance on the working memory measures. Results of the reading span tasks and backward digit span tasks both revealed that significant differences were found between the performances in the first, second and third languages of the participants. The forward digit span task only resulted in a significant difference between the performance in the first and third language while the letter-number ordering task revealed significant differences between the first language on the one hand, and the second and third languages on the other. Altogether, these results suggest that the working memory capacity of a, more mastered, native language is larger than that of later learned, less mastered, foreign languages. Furthermore, results of the reading span task and backward digit span task suggest that as language proficiency decreases, working memory capacity similarly decreases. Unfortunately, no correlations between language proficiency and working memory measures or statistical analyses including both the factors working memory and age of acquisition were calculated. So while the authors argue that working memory performance differences are due to language proficiency, an alternative explanation focusing on age of acquisition as the determining factor cannot be rejected.

In their comparison of two groups of late second language learners of English, Service et al. (2002) also found that working memory capacity seems to increase as language proficiency increases, as university students majoring in English outperformed students majoring in psychology in both language proficiency tasks and a reading span task. When the working memory capacities of the learners' first and second languages were compared, significant differences could be found for the psychology majors but not for students of English. In their comparison of studies, Van den Noort et al. (2006) seem to suggest that their findings contradict those by Service et al. (2002) since their advanced second language learners' working memory capacity did differ significantly from the working memory capacity in their native language. Alternatively, it might also be the case that the advanced learners in Service et al. (2002) were in fact more native-like than the learners in Van den Noort et al. (2006), which would support the assumption of language proficiency increasing as working memory increases (or vice versa). However, as Service et al. (2002) also excluded correlations between language proficiency and working memory measures, and did not control for

possible age effects (nor did they provide detailed information of the participants), more research in this field is needed.

2.2.3. Language aptitude

The concept of language aptitude refers to a largely innate and relatively stable talent for learning languages. Language aptitude is considered to be a composite measure consisting of various cognitive abilities. In his influential line of work, Carroll (1981) proposed that these abilities include a phonetic coding ability, grammatical sensitivity, rote-learning ability for second language materials and an inductive language learning ability. These cognitive abilities can be assessed with the widely used Modern Language Aptitude Test (MLAT; Carroll & Sapon 1959). Subsequently developed language aptitude tests, such as the Pimsleur Language Aptitude Battery (PLAB; Pimsleur 1966), the Defense Language Aptitude Battery (DLAB; Peterson & Al-Haik 1976) and, more recently, the CANAL-F (Grigorenko et al. 2000) and Llama Language Aptitude Tests (LLAMA; Meara 2005) all expanded on the MLAT (Caroll & Sapon 1959). However, these successors do not seem to be convincing alternatives to the MLAT (Carroll & Sapon 1959) (DeKeyser & Koeth 2010) and, thus, language aptitude is frequently studied within a relatively dated theoretical framework. Furthermore, ever since the development of the MLAT (Carroll & Sapon 1959), language aptitude levels have been automatically equated with scores on language aptitude tests (Dörnyei 2006) despite uncertainties of what the construct of language aptitude actually entails.

Recently, the concept of language aptitude has been reconsidered in light of new developments in the field of cognitive psychology and a renewed interest in the relationship between aptitude and second language acquisition (Dörnyei 2006). One of the directions taken in these new lines of research was initiated by Miyake & Friedman (1998) who claim that language aptitude should be equated with working memory since both concepts encompass a language analytic ability, memory ability and phonetic coding ability. This suggestion has been supported by experimental data from Robinson (2002), who found a moderately high correlation between working memory capacity, as measured via a reading span task, and language aptitude ($r = .35$, $p. < .05$). Despite this reconsideration of the relation between language aptitude and working memory, language aptitude is still widely used in research as it is considered to be a predictor of second language success (Dörnyei 2006) in, for example, the works of DeKeyser (2000), Abrahamsson & Hyltenstam (2008), DeKeyser et al. (2010) and Hopp & Schmid (2011).

DeKeyser (2000) assessed the occurrence of age effects and the influence of language aptitude in the grammaticality judgment performances of 57 Hungarian immigrants, of which 42 subjects were older than 16 at age of arrival and the remaining 16 subjects were younger than 16

when they migrated. The administered grammaticality judgment test was a shortened version (including 200 instead of 276 items) of the test used by Johnson & Newport (1989). To measure language aptitude, participants were administered a Hungarian adaptation (Ottó 1996) of the Words in Sentences subtest of Carroll & Sapon's (1959) MLAT. When these two measures were correlated, a significant effect of language aptitude ($r = .33$, $p < 0.5$) was observed for the late second language learners while this was not the case for migrants who were younger than 16 years old at age of arrival. DeKeyser (2000) interprets this finding as evidence that late second language learners will only achieve a level comparable to that of native speakers and can circumvent age effects if they have a high level of language aptitude.

Abrahamsson & Hyltenstam (2008) primarily aimed to replicate the results by DeKeyser (2000) in order to confirm whether a high degree of language aptitude is a crucial requirement in reaching a near-native level of second language proficiency in late learning. In doing so, Abrahamsson & Hyltenstam (2008) investigated 42 advanced, near-native, second language learners of Swedish of different first language backgrounds by administering an adapted version of the Swansea Language Aptitude Test (Meara et al. 2003) and a grammaticality judgment test. All subjects were pre-selected on the basis of perceived nativeness in a foreign accent rating procedure but when analyzing grammaticality judgment test performances, the majority of advanced late second language learners performed below the native-speaker range. These results suggest that language aptitude is not a sufficient learner characteristic by itself in the acquisition of a native-like level and intuition of second language grammar. Despite these results, Abrahamsson & Hyltenstam (2008) do argue that having a high level of language aptitude is advantageous in late second language learning.

As discussed in section 2.1.6, DeKeyser et al. (2010) studied critical period effects as well as language aptitude effects cross-linguistically. As in DeKeyser (2000), an adapted and shortened version of the grammaticality judgment task from Johnson & Newport (1989) was administered in order to assess the morphosyntactic knowledge of Russian migrants in the United States while Russian migrants in Israel were administered a Hebrew grammaticality judgment task created specifically for the study. To measure language aptitude levels, all participants were administered verbal sections of the Russian version of the Inter-University Psychometric Entrance Test (National Institute for Testing and Evaluation 2001). For subjects with an age of acquisition between 18 and 40 years old, significant correlations (Russian migrants in the United States: $r = .44$, $p < .05$; Russian migrants in Israel: $r = .45$, $p < .01$) between scores on the language aptitude test and grammaticality judgment task were found, replicating results of DeKeyser (2000) in demonstrating that language aptitude is essential for second language success in late learners.

In their study, Hopp & Schmid (2011) investigated age effects in perceived foreign accents of 40 late second language learners and 40 first language attriters, as rated by 130 listeners.

Additionally, Hopp & Schmid (2011) analyzed their data on the basis of predictors such as language use and language aptitude. As their test battery did not contain a measure of language aptitude, Hopp & Schmid (2011) estimated language aptitude levels in their participants through C-test scores on the basis of frequently observed strong correlations between such language proficiency measures and language aptitude scores (Dörnyei & Skehan 2003). For the late second language learners, results from the foreign accent rating procedure correlated significantly ($r = .49$, $p < .001$) with scores from the C-test, a finding consistent with DeKeyser (2000) and DeKeyser et al. (2010).

2.2.4. First language proficiency

Poor performances in second language tasks are frequently related to age of arrival and acquisition or to the individual differences discussed in section 2.2. However, linguistic performance in the second language can also be influenced by general language processing skills (Hulstijn & Bossers 1992). More specifically, if an individual finds word recognition or phonological processing rather difficult in the first language, similar difficulties are likely to hold for the second language. In a most extreme case, these assumptions are reflected in the Linguistic Coding Deficit Hypothesis (Sparks & Ganschow 1991), which assumes that learners with deficits in coding phonological, syntactic and semantic aspects of language will experience difficulties in second language learning. Consequently, the concept of first language proficiency should not be overlooked in second language research, as argued for by Hulstijn & Bossers (1992).

Hulstijn & Bossers (1992) aimed to examine the relationship and interaction between first and second language proficiency in the domains of speaking and reading. To assess second language speaking, 65 Dutch secondary school students were given first and second language cloze tests to read out loud while filling in the blanks simultaneously. Measures extracted from this procedure included reading speed, disfluencies and words supplied. Analyses revealed that participants' speaking skills were less proficient in their second language than in their first, although it should be noted that in this study, speaking skill were equated to reading from script. These two measures of proficiency correlated significantly for reading speed, disfluencies and words supplied ($r = .81$, $p < .001$, $r = .71$, $p < .001$ and $r = .29$, $p < .01$ respectively), suggesting that performance in the first language was predictive of reading in the second language. In a second experiment, second language reading comprehension was assessed in 50 late second language learners. Subjects were provided with two texts in their first and two texts in their second language on the basis of which they were asked to answer 64 multiple-choice questions. Additionally, subjects were administered a test of second language knowledge examining both grammar and vocabulary skills. Participants performed significantly better on the first language reading comprehension texts than on texts in the second

language. In an analysis of predictor variables in second language reading, a moderate correlation (r = .59, p-value not given) was observed between reading comprehension scores in the two languages. However, a stepwise regression analysis revealed that both first language reading skills and second language knowledge contributed significantly to second language performance. More specifically, second language knowledge contributed nearly four times the predictive value of first language reading since these predictors had β-values of 0.19 and 0.73 respectively. Consequently, both experiments in Hulstijn & Bossers (1992) revealed that individual differences in first language proficiency can partly account for differences in second language proficiency levels.

2.2.5. Fluid intelligence

According to Cattell (1963), general intelligence consists of two cooperating factors: crystallized and fluid intelligence. Crystallized intelligence encompasses previously acquired knowledge and skills that have become crystallized with experience (e.g. vocabulary and general knowledge). Conversely, fluid intelligence encompasses analytical reasoning abilities required to solve problems dealing with new information (e.g. pattern recognition, puzzle solving and abstract reasoning). As such, both types of intelligence seem to underlie various aspects of language acquisition. While crystallized intelligence is thought to remain stable throughout life, fluid intelligence is said to decline from the late teens onwards (Cattell 1963, Wang 1998). Consequently, early learners of a second language might be in a more advantageous position than late second language learners due to differences in fluid intelligence levels. Unfortunately, the relationship between fluid intelligence and age effects in second language acquisition has hardly been studied, with the exception of Wang (1998) who focused solely on late second language learners.

To study the relationship between intelligence and second language acquisition, Wang (1998) investigated 30 Mandarin-speaking migrants in Canada, among which 15 were, rather unconventionally, classified as early second language learners of English, with ages of arrival between 25 and 35 years, and another 15 were considered late second language learners of English, with ages of arrival between 40 and 55 years. To assess fluid intelligence, Wang (1998) administered the Quantitative Reasoning and Similarities subtests of Wechsler's (1941) Measure of Adult Intelligence. When the two learner groups were compared on the basis of scores on these tests, the earlier arrivals significantly outperformed the later arrivals. The participants' levels of fluid intelligence also correlated significantly with their second language performances on an elicited imitation test (.53, $p < .001$), oral interview (.66, $p < .001$) and oral fluency measure (.49, $p < .005$). Consequently, fluid intelligence, amongst additional factors, is suggested to significantly predict second language outcomes in two groups of late second language learners.

3. Research questions

According to the critical period hypothesis, language learning abilities decrease with age until puberty, after which a flattening out of the period can be observed. As such, learners with an age of acquisition after the terminus of the critical period are predicted to never reach a level of ultimate attainment in their second language that is comparable to the level reached by native speakers of that particular language. The critical period hypothesis is supported by the frequently observed negative correlations between the factors age of acquisition or arrival and level of ultimate attainment in the second language in previous research (Birdsong 2005). However, more recently, some studies seem to refrain from solely focusing on the impact of age of acquisition or arrival on the achieved levels of ultimate attainment in late second language acquisition. Instead, the influence and predictive abilities of the subjects' individual differences, such as language aptitude, working memory, fluid intelligence, level and length of education and first language proficiency, on late second language acquisition are taken into consideration. Studying these socio-psychological factors in addition to age of acquisition or arrival might provide insights into the psychology of the second language learner and might lead to more effective second language teaching methods.

One of the aims of this study is to assess the impact of such additional factors on second language learning. While this study does also examine the tenability of the critical period hypothesis by investigating whether late learners' second language proficiency actually reached a similar level as native speakers of that second language, the focus will primarily be on which characteristics can be identified to be predictive of second language performance in order to provide a blueprint of the successful late second language learner. In looking at 62 late Dutch-English bilingual migrants and 61 English monolinguals in Australia, this study attempts to answer the following research questions:

1) Can late second language learners in a migrant context reach a level of second language proficiency comparable to native speakers?
2) Which individual socio-psychological differences, in addition to age of arrival, are predictive of second language success in late second language learners of English in a migrant context?

The supporting evidence for the critical period hypothesis discussed in section 2.1.5. is mostly based on late second language learners' performances on a single task, i.e. a grammaticality judgment task. Consequently, the scope of these studies is rather limited as all evidence is derived from performances in one domain of linguistics only (i.e. morphosyntax). To provide more compelling evidence for the verification of the critical period hypothesis, studies aiming to verify or reject the critical period hypothesis should base their assumptions on performances on multiple linguistic tasks

in various domains (Abrahamsson & Hyltenstam 2009). Only if a single late second language learner can be found who performs within a native speaker range on linguistic assessments in multiple domains, the critical period hypothesis can be rejected (Long 1990). The present study will therefore similarly reject the critical period hypothesis if an individual late second language learner can be found who performed within native speaker range on a C-test, grammaticality judgment task, Fluency in Controlled Association task and Peabody Picture Vocabulary Test-4 (Dunn & Dunn 2007). By adopting this methodology, which will be further discussed in the next section, late learners' performances in the linguistic domains of second language morphosyntactic knowledge, second language receptive vocabulary knowledge, general second language proficiency and second language verbal fluency were assessed. Due to the length of residence reported among the participants and the typological proximity of their first and second languages, it was hypothesized that at least one late second language learner would have reached a level of ultimate attainment comparable to that of native speakers on all linguistics tasks in the test battery.

Furthermore, it was expected that individual socio-psychological factors would contribute significantly to second language performance. Since researchers have not often included multiple factors in their studies, as was discussed in section 2.2, it is rather difficult to estimate which factor was expected to be the predominant contributor. However, on the basis of the widespread availability of evidence, it could be hypothesized that working memory and language aptitude would probably play a significant role in late second language acquisition.

4. Methodology

In section 4.1., the subjects of this study, Dutch late second language learners of English, i.e. bilinguals, and native speakers of English, i.e. monolinguals, will be introduced. Subsequently,section 4.2. provides an overview of the materials used in this study, followed by section 4.3 describing the general procedure. Lastly, section 4.4 will introduce the statistical methods on the basis of which results were obtained that are reported in section 5.

The materials and data analyzed and discussed in this thesis are part of a much more elaborate test battery. More specifically, the materials in this thesis are a subset of the NWO VENI scheme research project Crossing Communication Borders: First Language Reversion in Healthy, Aging Dutch Migrants in Australia led by dr. M.C.J. Keijzer (project number 016.104.017). This NWO VENI project's research aims and questions were not to explore the critical period hypothesis in second language acquisition, but rather to investigate first language reversion in elderly bilingual migrants. However, the data provided by the project was deemed suitable to answer the research questions formulated for this study.

4.1. Subjects

The two groups of subjects in this study consisted of late Dutch-English bilinguals and English monolinguals. The bilingual subjects consisted of Dutch migrants who moved to Australia after age 15. As such, their age of arrival is considered equal to their age of acquisition. The majority of monolingual subjects grew up in Australia while some of the monolinguals grew up in other English-speaking countries (including the United Kingdom (Scotland, England), Ireland and South Africa). These two groups of subjects can be further divided into three subgroups based on age, resulting in the groups 40-45, 60-65, 75+, in order to study language development across the life span. More specifically, the group including the youngest subjects (40-45) served as a control group to be compared with the two groups of elderly subjects (60-65 and 75+) on the basis of cognitive measures. In total, 123 participants were investigated, as demonstrated in Table 1.

Table 1: Overview of the subject groups and their age categories.

Group	40-45	60-65	75+	Total
Dutch migrants	17	21	24	62
English controls	20	23	18	61
Total	37	44	42	123

The bilingual subjects were recruited via advertisements in Dutch magazines and newspapers that circulated in Melbourne, Australia, or via word of mouth and social media. Recruitment for the majority of monolingual subjects took place in Melbourne, Australia, and mainly occurred via word of mouth of the already recruited bilingual subjects. As such, bilingual subjects often recruited monolingual family members or friends to participate in the study. When subjects applied to join the research project, they were pre-sent a screening questionnaire (see section 4.2.1.) which served as the basis to determine inclusion in one of the subject groups. The screening questionnaire featured simple questions on educational background, age, gender, region of birth, self-reported frequency of use of the first and second language, vitality, color blindness, use of medication and drugs, history of seizures and handedness. Potential subjects who indicated that they were cognitively impaired, (had) used medicine or drugs extensively, (had) suffered from degenerate diseases, strokes or color blindness were excluded from the research project. Similarly, participants who failed to submit a complete data set or were incapable of traveling to the campus of Monash University in Melbourne were excluded as well.

Those subjects who matched all inclusion criteria were orally administered a second, longer, questionnaire at the beginning of each testing session (see section 4.2.1). This questionnaire touched upon issues such as personal background, changes in first and second language proficiency, social networks, language and self-rated language proficiency. The demographic characteristics relevant for this study, including age at testing, length of education, length of residence and age of arrival, gathered via the two questionnaires are provided in Tables 2 and 3.

Table 2: The monolingual and bilingual subjects' demographic characteristics.

Bilinguals (n=62)			
Factor (in years)	**Mean**	**Standard deviation**	**Range**
Age at testing	64.59	14.96	38 – 86
Length of education	15.97	4.51	7 – 33
Age of arrival	26.80	9.64	13 – 61
Length of residence	37.67	20.51	1 – 61
Monolinguals (=61)			
Factor (in years)	**Mean**	**Standard deviation**	**Range**
Age at testing	59.75	13.88	32 – 87
Length of education	15.98	4.01	8 – 30

The monolingual and bilingual subjects did not differ significantly on the basis of age at testing or length of education. As for gender, of the bilingual subjects, 42% were male and 58% were female. A similar gender distribution was observed in the monolingual subjects of which 39% were male and 61% were female.

4.2. Materials

The tests employed to answer the research questions formulated in section 3 can be divided into two categories; language proficiency measures and individual socio-psychological factor tests, as shown in Table 3. A number of tasks were administered with the use of a 15 inch HP EliteBook 8530p laptop and a Sony plugin power microphone. Each task will be discussed in more detail below.

Table 3: Overview of the test battery.

Language proficiency tests		Socio-psychological tests	
Construct	**Test**	**Construct**	**Test**
General language proficiency	Dutch and English C-tests	**Working memory capacity**	Reading span task (Van den Noort et al. 2008), backward digit span task
Grammatical knowledge	Grammaticality judgment task	**Language aptitude**	C-test, as previously employed by Hopp & Schmid (2011)
Verbal fluency	Fluency in Controlled Association (rapid naming)	**Fluid IQ**	Short form of Advanced Progressive Matrices (Bors & Strokes 1998)
Receptive vocabulary knowledge	Peabody Picture Vocabulary Test-4 (Dunn & Dunn 2007)	**Level and length of education**	Sociolinguistic questionnaire
		Age of arrival and length of residence	Sociolinguistic questionnaire

4.2.1. Sociolinguistic questionnaires

Sociolinguistic questionnaires are commonly employed in (second) language acquisition research to collect background information for each subject. These data are often included in statistical analyses to account for variation in the data or used as the basis for factor analyses (e.g. Schmid & Dusseldorp 2010). In this study, two sociolinguistic questionnaires were administered. The first, shorter, questionnaire served as a screening device before testing and therefore contained simple questions on educational background, age, gender, region of birth, self-reported frequency of use of the first and second language, vitality, use of medication, history of seizures and handedness. The complete screening questionnaire can be found in Appendix A. Bilingual subjects and a few monolingual controls were pre-sent this questionnaire via regular mail or e-mail, but as many monolingual controls were recruited in Australia, this questionnaire was administered at the beginning of the testing session for those who had not yet completed it. In total, 91 bilinguals and 7 monolinguals submitted a screening questionnaire before leaving for Melbourne, Australia. On the basis of these screening questionnaires, 62 bilinguals and all of the monolinguals were selected.

The second, longer, questionnaire was administered orally at the beginning of each testing session. By starting the testing session with a conversation between researcher and subject, subjects

were expected to feel more at ease during the remainder of the session. This questionnaire touched upon issues such as personal background, changes in first and second language proficiency, social networks, language and self-rated language proficiency. The questionnaire for the bilingual population contained 27 questions, of which the majority were discussed in Dutch and part were answered in English. This switch was incorporated to study language switching and inhibition patterns among the bilingual subjects. As questions on language change and differences between first and second language proficiency did not pertain to the monolingual subjects, their questionnaire was shorter by 18 questions. Both the monolingual and bilingual versions of the sociolinguistic questionnaire can be found in Appendix B. All interviews were recorded digitally with a Sony plugin power microphone attached to a laptop computer and the computer program Audacity run under Linux Ubuntu.

4.2.2. Reading span task

The subjects' working memory capacity was assessed through the administration of a reading span task; a relatively complex task of verbal working memory (Juffs & Harrington 2011) in which the subject is asked to read a series of 2 to 6 sentences out loud while simultaneously comprehending the content and remembering the final word of each of those sentences in free recall order. Due to this design, the reading span task is argued to tap into both the processing and storage components of working memory. The version included in this study was developed by Van Den Noort et al. (2008), who introduced various improvements when compared to the original reading span task (Daneman & Carpenter 1980) and its successors. Besides creating a standardized computerized version of the reading span task in four languages (including Dutch and English), Van Den Noort et al. (2008) carefully controlled for sentence length, lexical frequency, semantic relations, concreteness of sentence-final words, the number of syllables and letters per sentence and the plausibility of the sentences. The reading span tasks created by Van Den Noort et al. (2008) consist of 100 sentences.

Due to time limitations, this test battery included a subset of those 100 sentences in a standardized short form consisting of the first three sets (i.e. 60 sentences) as programmed in Zep experiment software run under Linux Ubuntu. These 60 sentences were distributed over 3 series consisting of 20 sentences each, presented in sets of 2, 3, 4, 5 or 6 sentences, in random order. Sitting at a distance of approximately 60 cm. from a laptop screen, subjects were instructed orally: they were asked to read the sentences out loud as they were presented on a laptop screen one-by-one while simultaneously focusing on the content of the sentences as they would be asked multiple text comprehension questions after each series of sentences. Subjects were informed that they would be given a maximum of 6.0 seconds to read each sentence and they were advised to start

reading as soon as possible. This instruction was followed by 2 trial sentences. After each set of sentences, the word 'recall' was printed on the screen as a cue for the subjects to start freely recalling as many sentence-final words as possible. The recalled words were simultaneously recorded by the researcher on a scoring form in the order produced by the subject. Each correctly recalled word would be given a score of 1 and the reading span score constituted all correct items added together, in line with Van den Noort et al. (2008). Consequently, the maximum score to be obtained was 60, matching the number of sentences in the test. During the mini breaks, which were slotted in between each set of 20 sentences, two comprehension questions were asked, but these items were not included in the scoring procedure. They merely served as distractors to add to processing load. In this study, bilingual subjects were asked to complete this test in both Dutch and English while monolingual subjects only completed the test in English.

4.2.3. Backward digit span task

Apart from the reading span task, working memory capacity was also measured by means of a backward digit span task. Both tasks tap into the storage and processing components of working memory, but when compared to the reading span task, the backward digit span task can be considered to be less complex while it also minimizes the use of linguistic knowledge due to its reliance on digits (Juffs & Harrington 2011). Hence, studying the correlations between these two working memory measures, as well as the correlations with dependent variables, might provide insights into similarities (as found in Van den Noort et al. 2006, discussed in the section 2.2.2) or differences between the measurements made by these tasks.

In this study's visual backward digit span task, subjects were presented with a series of random digits (chosen from digits 1 to 9, with no more than 2 double digits appearing in succession) which they were asked to recall in backward order after they disappeared from the laptop screen. Sitting at a distance of approximately 60 cm. from a laptop screen, subjects were instructed orally and presented with a practice trial consisting of 3 digits which disappeared from the screen after two seconds. Subjects were instructed to either type in the remember digits or to orally recall the digits in backward order while the researcher was required to simultaneously type these in (this decision was based on preference of the subject). In the test, a new digit was added to the initial series of 2 digits after each correct answer (for example, 43 would become 435), and new digits were added as long as no recall errors were made. The backward digit span task either terminated when a recall error was made or when the subject had reached the maximum span of 10. At the end of the task, the total score was shown on the screen. Like the Reading span task, this test was also programmed in Zep experiment software run under Linux Ubuntu.

4.2.4. Short form of Raven's Advanced Progressive Matrices

In this study, the subjects' intelligence was assessed by examining length and level of education as well as by providing an objective measure of fluid intelligence. While level and length of education can be adequately assessed by the inclusion of questions on the sociolinguistic questionnaire, fluid intelligence, or analytic reasoning, was examined through the administration of the short form of Raven's Advanced Progressive Matrices (Bors & Strokes 1998). Since the development of Raven's Standard Progressive Matrices in 1938 (Raven 1938), different versions of this test of fluid intelligence have been created, including Raven's Advanced Progressive Matrices (APM; Raven 1962), which was specifically designed to reliably distinguish between adolescents and adults with above-average intelligence. During this test, subjects are presented with progressively more difficult matrices consisting of visual analogy problems. In the bottom right corner of each matrix, one of nine geometric features is missing and the subject is asked to complete the pattern observable in the matrix by choosing the missing picture from eight possible alternatives. Since the APM has a relatively long administration time (45 minutes), this study included the short form of the APM as developed by Bors & Strokes (1998). The short form contains a subset of the items on the full-length APM and was normed on the basis of the performance of 506 first-year university students. This short form has been shown to correlate well with the full-length version of the APM ($r = .92$, $p < .001$) and is therefore generally considered to be a valid replacement. As the subjects in this study do not necessarily all share above average intelligence levels, the short form of the APM could be considered inappropriate for use in this test battery. However, due to time limitations, the short form of the APM was the only version of a test of fluid intelligence that could be included. Furthermore, preliminary analyses showed that the short form discriminated between the subjects' intelligence levels appropriately. Thus, the procedure introduced by Bors & Strokes (1998) was followed.

Also because of these time limitations, the short form of the APM was one of the tests present to the subjects to complete at home in their own time (see section 4.3.). In this paper-and-pencil version, subjects were instructed to solve 2 practice trials and 12 puzzles by looking across rows and columns to formulate rules which served as the basis to choose the missing eighth entry in the 3 x 3 matrix from a set of alternative answers. The score extracted from the subjects' answers was the total number of correct responses, with a maximum score of 12.

4.2.5. C-test

A C-test is a general language proficiency test consisting of continuous, authentic texts in which parts of words have been deleted. When the test is administered, subjects are asked to complete the gapped words again. In this test battery, bilingual subjects were asked to complete the C-test in both Dutch and English while monolingual subjects only completed the task in English. This way, results from the C-test can be used to assess first language proficiency for both the monolinguals and bilinguals and to examine second language proficiency in the bilingual subjects. Furthermore, results from the C-test were used to estimate language aptitude levels, a method previously applied in Hopp & Schmid (2011) and discussed in section 2.2.3.

The Dutch C-test was derived from Keijzer (2007). The format used in the present study consisted of five texts on a variety of topics, each containing 20 gapped items. The mean number of words in these texts was 74 and more content words (14.4 on average) than function words (5.6 on average) were gapped to increase the task's difficulty. Keijzer's (2007) texts in the C-test were pre-tested on Dutch university students at the Free University of Amsterdam and only those texts on which native speakers scored between 87% and 90% were included in the C-test. The English C-test consisted of five texts on a variety of topics, each containing 20 gapped items. Of the five texts, two were taken from Keijzer (2007) and three were newly developed. All texts were pretested on native speakers of English who were university students in either London, Canada, or Melbourne, Australia, and only those texts on which the native speakers scored between 87% and 90% were included in the C-test. After pretesting, the English C-test featured texts that contained 81.2 words on average in which more content words (14.6 on average) than function words (5.4 on average) were gapped. Both the Dutch and English C-test can be found in Appendix C. For both tests, the maximum score of 100 was based on the number of correctly entered items (in this case, correctly entered items are correct words, correct words with spelling errors or correct alternatives that fit the context). Incorrectly entered items consisted of either an incorrect answer or no answer at all.

Because of time limitations, the C-test was one of the tests pre-sent to the subjects to complete at home in their own time (see section 4.3.). In this paper-and-pencil version, subjects were instructed that they would be presented with five short, but incomplete, texts which they were asked to complete by filling in answers below each text. Furthermore, subjects were told that the texts would increase in difficulty level and that they were not to worry if they could not complete all items. When an item was too difficult, subjects were advised to proceed to the next item. However, after an item was skipped, subjects were asked not to return to this item in the end. Lastly, subjects were explicitly requested to not take more than 5 minutes to complete each text.

4.2.6. Grammaticality judgment task

In a grammatically judgment task, subjects are asked to judge the acceptability of sentences in a given language. This way, information is obtained on language user intuitions on the well-formedness of sentences while simultaneously controlling for the avoidance of certain syntactic structures (a strategy applicable in spontaneous speech). The grammaticality judgment task included in this study mostly focused on cross-linguistic differences between Dutch and English. Constructs under investigation in the grammaticality judgment task were clausal embeddings, verb second, passives, negation and reflexive constructions. In this grammaticality judgment task, subjects were asked to express their linguistic intuitions on a 5-point Likert scale, ranging from "entirely incorrect" to "entirely correct". The complete grammaticality judgment task can be found in Appendix D. The inclusion of 5-point Likert scales was preferred over the use of binary rating scales or magnitude estimation (Sorace 2010) for various reasons. Binary rating scales are quite simple and, therefore, not many statistical analyses can be applied to data collected with the use of these scales while a 5-point scale creates a statistically stronger interval scale (Stevens 1946). Also, when compared to binary scales, 5-point Likert scales allow for more insights into the subjects' judgments of sentences as they can indicate varying degrees of acceptability. However, one cannot possibly know whether the grammatical sensitivities used in this study are limited to such a 5-point Likert scale, a problem that could be solved by magnitude estimation (Bard et al. 1996). In grammaticality judgment tasks that use magnitude estimation, subjects are first asked to assign a number to a modulus sentence. All subsequent test items are then compared to the modulus sentence and the subject can indicate whether a stimulus is *n* times as bad or good as the modulus sentence by multiplying the number assigned to the modulus. There are no limits to the numbers subjects can assign to the sentences and thus magnitude estimation captures all linguistic intuitions of the subject (Sorace 2010). Despite valid arguments in favor of magnitude estimation, a 5-point Likert scale was preferred in this study as magnitude estimation was deemed too complicated for the population of interest in this study (including subjects of 75 years and above). Furthermore, results provided by both magnitude estimation and a 5-point Likert scale correlate well and, thus, are considered to be comparable (Sorace 2010).

Because of time limitations, the grammaticality judgment task was one of the tests pre-sent to the subjects to complete at home in their own time (see section 4.3.). In this paper-and-pencil version, subjects were instructed to judge 30 sentences on their grammaticality. These instructions stressed that subjects were asked to judge form rather than meaning, as well as advising subjects to rely on their first intuitions. Subjects were given one example stimulus and accompanying judgment before the actual task started. The ratings provided by the subjects were converted into scores by

judging whether they provided a correct or incorrect answer. This resulted in a score of a maximum of 30 (matching the number of sentences) and scores for each feature separately.

4.2.7. Peabody Picture Vocabulary Test-4

The Peabody Picture Vocabulary Test (Dunn & Dunn 1959) is a standardized picture naming task designed to evaluate a subject's receptive vocabulary. In this task, subjects are provided with an auditory prompt by the experimenter after which they are asked to match the prompt to one of four displayed pictures. The most recent version of the test, Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn 2007), contains a modernized core vocabulary of the English language and has been standardized for subjects between the ages of 2;3 to 90 years old. Due to this standardization, results are comparable across subjects of various ages. The PPVT-4 can be quickly administered (11 to 12 minutes) and the test's reliability and validity coefficients all have values higher than .9 (PPVT product website). In its original format the PPVT is conducted using an easel containing all the pictures, but in this study a digital version of the PPVT-4 in the form of a presentation in Microsoft Powerpoint was used. In this Powerpoint version, auditory prompts recorded in a British English female voice were included, controlling for consistency of prompts across subjects.

Before administering the PPVT-4, subjects were presented with practice items intended to familiarize the subject with the procedure of the task. They were told that they would hear an auditory prompt while looking at the screen containing four pictures, and that it was the subject's task to match the prompt to the correct picture by either pointing at or saying the number of their answer. After the subject correctly answered two training items, a starting set was determined on the basis of the subject's age. In this study, all subjects were required to start with set 14 (intended for people aged 19 years or above). If subjects made one or no errors, set 14 was determined as both the starting and basal set and subjects would proceed to set 15 and upwards until a ceiling set was reached in which 8 or more errors were made. If more than one error was made in set 14, subjects would go down to the previous set, 13, until one or no errors were made in order to establish the basal set. Once this basal set was determined, sets already administered were skipped, and testing continued until the ceiling set was reached. If fewer than 8 errors were made in the final set of the test, this automatically became the ceiling set and the final item of the set became the ceiling item. All subjects' responses were recorded on PPVT-4 scoring forms. The raw scores of the test were calculated by subtracting the number of errors made from the number of the final item in the ceiling set. These calculated raw scores were subsequently converted to standardized scores with the use of standard conversion tables.

4.2.8. Fluency in Controlled Association task

Rapid naming tasks, such as the Fluency in Controlled Association Task (FICA), are often employed to reliably test verbal fluency and vocabulary size. In the FICA, two measures of verbal fluency are generally included: letter fluency and category fluency. During a letter fluency task, subjects are instructed to name as many words as they can that begin with a particular letter (F, A and S are most commonly employed (Bialystok et al. 2008)). Similarly, during a semantic or category fluency task, subjects are asked to generate as many words possible that belong to a particular semantic category (such as animals or fruits and vegetables). Both tasks are usually conducted within a fixed time span, commonly set at 60 seconds per category or letter. During these tasks, subjects are allowed to name any word that comes to mind with the exception of proper names and place names, repetitions or morphological derivations of a previously named word (e.g. apple, apples), as long as it fits the category or letter.

In this study, both letter and category fluency tasks were included as measures of vocabulary size, indications of language proficiency levels and lexical fluency and to investigate the subjects' lexical networks. After practising with naming three items starting with the letter 't', subjects were asked to firstly name as many animals as possible (category fluency), after which they were asked to name items starting with an 'f' (letter fluency), followed by fruits and vegetable items (category fluency), and, finally, items starting with an 'a' (letter fluency), all within the fixed time limit of one minute. Subjects were instructed not to name proper names, place names, repetitions or morphological derivations and subsequently testing began. During the test, subjects were recorded digitally with a Sony plugin power microphone attached to a laptop computer and the computer program Audacity run under Linux Ubuntu. After calculating the absolute number of correct answers for the verbal fluency tasks, these numbers were converted to eight scores per subjects: the total number of items named for each of the four categories, the total number of items for all categories taken together, the mean number of items averaged for the category fluency task, the mean number of items averaged for the letter fluency task, and the mean number of items for all four categories taken together.

4.3. General procedure

All data were collected in the greater region of Melbourne, Australia, over a period of two months from December 2011 until January 2012. During this period, over 120 subjects were tested, of which half were Dutch-English bilingual migrants and half were English-speaking monolingual controls. The majority of these subjects traveled to Monash University for testing but for some subjects house

visits in various parts of Melbourne were made. All tests were administered by a principal investigator and two research assistants. In general, testing was done one-on-one and could be done in parallel with two researchers testing two people at the same time (this limit of 2 was imposed due to availability of equipment).

The materials and data analyzed and discussed in this thesis are part of a much more elaborate test battery. As such, for bilingual subjects, testing lasted approximately 2.5 hours and for monolingual subjects, testing lasted approximately 1.5 hours. Generally, the length of testing sessions tended to increase as subjects became older due to their speed and need for more instructions and explanations before and during testing. Besides differences in length of testing, the order of tests would also differ for monolingual and bilingual subjects as the latter group completed many tasks in both languages. This difference was the cause of variation in length of testing and also created differences in testing orders to avoid priming effects. Generally, testing orders were kept more or less constant to control for priming and to vary between effortful and less demanding tasks. Sometimes, deviations from the set order of testing would occur based on division of testing equipment or what was thought best for the subjects given particular situations (for example, if a subject struggled with a particular test, another difficult test might have been postponed to make the subject feel more comfortable).

At the end of each session, subjects received a small sum of money for their time and efforts (20 Australian dollars for the monolinguals and 30 Australian dollars for the bilinguals, as their testing session took longer). Because the complete test battery would have been overly long if all tests were to be administered in one session, some tests (i.e. the short form of Raven's Advanced Progressive Matrices, C-test and grammaticality judgment task) were already pre-sent in paper-and-pencil form to subjects who were contacted before testing took place in Melbourne, Australia. As some of the monolingual controls were recruited in Australia, it proved impossible to pre-send the materials to them and they were either sent the materials by regular mail or via e-mail.

4.4 Data analysis

In section 5, an overview will be provided of the results obtained with the materials discussed in section 4.2. The collected data were statistically analyzed in IBM SPSS Statistics 20. For the analyses of results on the grammaticality judgment task, PPVT-4, C-test and FICA, the same procedures were followed. Similarly, results on the socio-psychological tests were analyzed in the same manner. The various stages in the statistical analyses for each of these test categories are described below.

Firstly, for the language proficiency tests (reported in section 5.2) and socio-psychological tests (reported in section 5.1), results were analyzed with descriptive statistical methods. For each of

the tests, the means, standard deviations and ranges were calculated and reported. Subsequently, in analyses of the socio-psychological tests, performances of the monolingual and bilingual subjects were compared with independent samples t-tests or Mann-Whitney tests, depending on the distribution of the data. Whenever multiple scores for a single predictor variable were available (as was the case for working memory capacity and educational levels and length), Pearson and Spearman correlation analyses were run to determine which variables could be reduced. No further statistical analyses were conducted for the individual socio-psychological factors. The second stage in the analyses of the language proficiency measures was to explore differences in the subjects' performances on components of each of the tests, whenever multiple subtests were available (e.g. in the case of the C-test, grammaticality judgment task and FICA), by conducting paired samples t-tests or Wilcoxon tests, depending on the distribution of the data. In these instances, variables were reduced by assessing results of Pearson and Spearman correlation analyses. Thirdly, following standard procedures, the effects of gender and age at testing were examined by conducting independent samples t-tests or Mann-Whitney tests (in the case of gender) and by running one-way ANOVAs and Kruskal-Waillis tests (in the case of the assigned age categories). These four statistical tests were all aimed at identifying which groups' performances differed significantly. Furthermore, Pearson and Spearman correlation analyses were run between age at testing and the language proficiency measures, depending on the distribution of the data.

In the fourth step, the language proficiency performances of the monolingual and bilingual subjects were compared by conducting independent samples t-tests or Mann-Whitney tests. Also, correlation analyses were run between language proficiency measures and the variables unique to the bilingual subject group, i.e. age of arrival and length of residence. Additionally, box plots were provided to compare the distributions of results obtained by both subject groups.

Lastly, and perhaps most importantly, the effects of the individual socio-psychological differences were assessed by creating multiple regression models. These multiple regression analyses were conducted as they can predict outcome variable X (i.e. scores on one of the language proficiency tests) on the basis of independent predictor variables Y_1, Y_2, Y_3, Y_n, etc. (i.e. scores on the socio-psychological tests). In doing so, the created models can indicate how much of the variance in the data is explained by the socio-psychological factors. Before any of the variables were entered in the model, correlations were calculated as only those predictors that correlated significantly with the language proficiency measures, and therefore met the assumption of linearity, were included in the regression models. Subsequently, for each of the language proficiency measures, two regression models were created. The first of which was created with the enter method, i.e. by entering all significantly correlating predictors at the same time. The second, and better fitting, regression model

was created with the backward model, and thus, eliminated all non-significant predictors until a model based solely on significant predictor variables remained.

Before carrying out parametric statistical analyses, such as t-tests, ANOVA and Pearson's correlation coefficient, a number of assumptions need to be met. Most notably, parametric statistical tests require the entered data to be normally distributed (Field 2005, p. 64). However, as will be discussed in the subsequent sections, this assumption was not always met in this study, as was indicated by a significant result on the Shapiro-Wilk test. The non-parametric test equivalents, i.e. statistical methods not relying on the data to be normally distributed, featured in this study included the Mann-Whitney test, Wilcoxon test, the Kruskal-Wallis test and the Spearman's correlation coefficient. Whenever one or multiple assumptions underlying the particular statistical analyses were not satisfied, this was reported and a non-parametric test was adopted. Like parametric statistical tests, multiple regression analyses rely on a number of assumptions that need to be satisfied before actually entering data in the analyses (Field 2005, p. 169). These assumptions were all tested and reported if they were not met.

As mentioned in section 4.3., the majority of monolingual and some of the bilingual subjects were sent the take-home materials (i.e. grammaticality judgment task, C-test, the short form of Raven's Advanced Progressive Matrices) after the testing session took place. As such, these subjects were requested to return these via mail. Unfortunately, at time of writing (31th of May 2012), not all materials have yet been returned, which results in missing data. Combining this reason with others to account for missing data (e.g. missing testing materials, time constraints) can also account for the differing and unequal sample sizes throughout the section on results. As such, for each score, the sample size will be reported. Practically, all missing values were specified in SPSS to ensure that these were ignored by the computer (Field 2005, p. 53). Lastly, not all statistic tests actually require equal sample sizes so the challenges posed by missing data and the unequal sample sizes should not be problematic in this study.

5. Results

In section 5.1., the subjects' performances on the socio-psychological tests will be discussed. For each of the tasks assessing the subjects' individual differences, descriptive statistics will be provided as the data is explored before being entered in subsequent regression models. Also, the monolingual and bilingual subjects' scores will be compared. In section 5.2., the monolinguals' and bilinguals' performances on the language proficiency tasks are discussed. Following standard procedures, descriptive statistics and analyses assessing the effect of age at testing and gender are provided. Subsequently, the monolinguals' and bilinguals' scores are compared and the effects of age of arrival and length of residence are investigated to explore the validity of the critical period hypothesis. Lastly, correlation and multiple regression analyses are conducted to examine the predictive values of the socio-psychological factors.

5.1. Socio-psychological predictor variables

5.1.1. Level and length of education

In this study, measures for length and level of education were both collected via sociolinguistic questionnaires. Table 4 presents the subjects' length of education in years while Fig. 5 provides an overview of the highest completed levels of education for monolingual and bilingual subject groups.

Table 4: Overview of the subjects' length of education.

Group	Mean	Standard deviation	Range
Bilinguals (n=64)	15.97	4.51	7 – 33
Monolinguals (n=60)	15.98	4.01	8 – 30

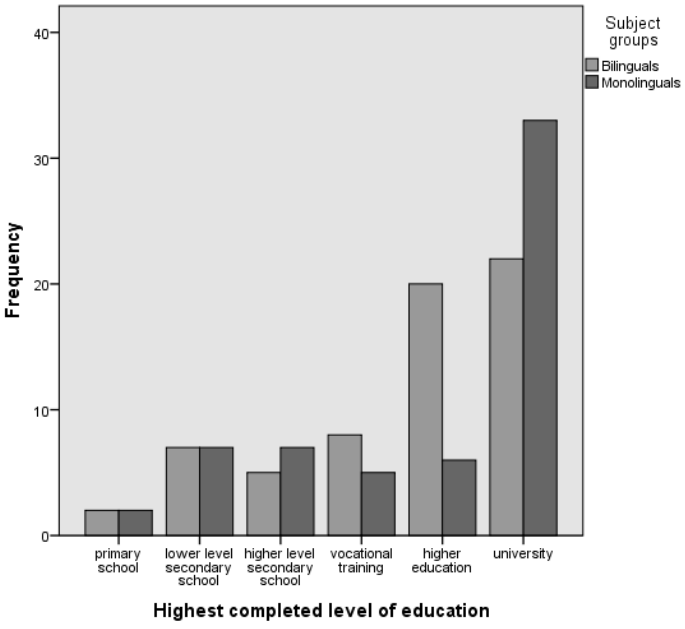

Figure 5: Bar chart of the subjects' level of education.

Since subjects were recruited via advertisements in magazines and newspapers, or via word of mouth and social media, this type of sampling can be considered to be voluntary response sampling. As such, a self-selection bias might arise since subjects have essentially selected themselves. Adopting this type of sampling might, thus, lead to a biased sample that is not representative of the entire target population. However, this sampling method was essential in guaranteeing enough subjects would be tested during the stay in Melbourne,

Australia. Keeping these considerations in mind might explain why the distribution of subject across levels of education is rather skewed. As can be seen in Fig. 5, 57% of the monolingual subjects completed a university degree while 64% of the bilingual subjects completed either higher education or a university degree, and this asymmetry in the data can possibly be attributed to the sampling method of this study. Furthermore, as length and level of education are related (i.e. the highest level of education will generally take longest to complete), normal distributions of the length of education data were not observed for monolingual subjects (w = .953, p = .023) or bilingual subjects (w = .945, p = .007). To further analyze the distributions of the length of education, histograms including a displayed normal curve are presented in Fig. 6 and 7. A Mann-Whitney test determined that the monolinguals' and bilinguals' length of education did not differ significantly (p = .781).

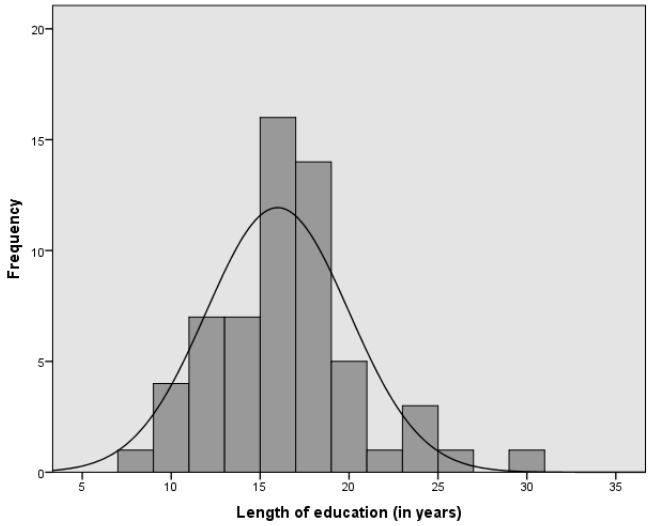

Figure 6: Histogram of the monolingual subjects' length of education.

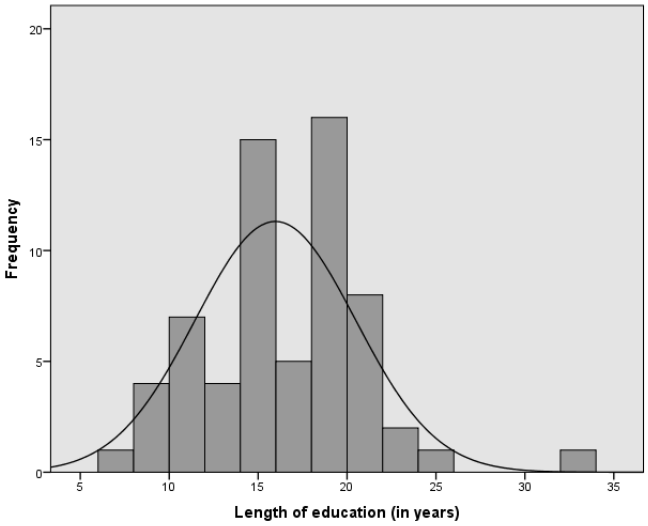

Figure 7: Histogram of the bilingual subjects' length of education.

5.1.2. Working memory capacity

In this study, the subjects' working memory capacity was assessed by administering relatively complex reading span tasks and a slightly simpler backward digit span task. Table 5 presents the results on both types of tasks for monolingual and bilingual subjects.

Table 5: Subjects' scores on the reading span tasks and backward digit span task.

Bilinguals (n = 64)			
	Mean	**Standard deviation**	**Range**
Reading span task L1	35.58	7.43	14 – 51
Reading span task L2	34.56	7.24	18 – 50
Backward digit span task	6.94	1.99	3 – 10
Monolinguals (n = 61)			
	Mean	**Standard deviation**	**Range**
Reading span task L1	36.69	8.06	10 – 54
Backward digit span task	7.00	2.05	3 – 10

The presumed differences in complexity between the two types of working memory capacity tasks were reflected in the finding that 18% of the monolingual and 15.9% of the bilingual subjects scored at ceiling level in the backward digit span task while no subject reached a maximum score in either reading span task. Despite differences in complexity between the two types of tasks, the scores on the backward digit span task correlated significantly with those of the reading span task ($\rho = .282$, $p = .028$) for the monolingual subjects. For the bilingual subjects, scores on the backward digit span task correlated significantly with those of the Dutch ($\rho = .327$, $p = .009$) and English ($\rho = .411$, $p = .001$) reading span tasks. These correlations suggest that the both types of tasks measure a similar construct.

The occurrence of ceiling effects in backward digit span data resulted in a non-normal distribution of the data ($w = .933$, $p = .002$ for bilinguals and $w = .930$, $p = .002$ for monolinguals). Additionally, reading span tasks scores for the monolingual subjects were not normally distributed either ($w = .949$, $p = .013$) while the bilinguals' scores on the Dutch ($w = .980$, $p = .396$) and English ($w = .984$, $p = .582$) tasks were. Due to the observed non-normality of part of the working memory capacity data, the monolinguals' scores on the backward digit span task and the reading span task and the bilinguals' scores on the backward digit span task could not be entered in any parametric statistical analyses. Since the backward digit span task scores and reading span task scores correlated significantly and scores of the former task were not normally distributed for both subject groups, only reading span task scores were included in subsequent statistical analyses. Another argument in favor of the inclusion of reading span task scores rather than backward digit span scores is related to the complexity of measurement scales. As the measurement scale is less complex and detailed for the backward digit span task (varying between 0 and 10) than for the reading span task (varying between 0 and 60), the latter might be able to provide more detailed insights into the subjects' working memory capacity. The non-normality of the reading span task scores for the monolingual subjects did not pose any problems as multiple regression analyses do not require the entered data to be normally distributed.

Furthermore, since the bilingual subjects' scores on the reading span tasks in both languages correlated highly ($r = .802$, $p = .000$) and did not differ significantly ($t = 1.761$, $p = .083$), only reading span task scores in the subjects' first languages (i.e. English for the monolingual subjects and Dutch for the bilingual subjects) were included in regression analyses. A Mann-Whitney test indicated that monolinguals' reading span did not differ significantly from the bilinguals' Dutch or English reading span ($p = .339$ and $p = .062$ respectively).

5.1.3. Language aptitude

Like in Hopp & Schmid (2011), language aptitude levels were estimated by scores on the English C-test. For monolinguals, this predictor variable was not included in the regression analyses as they were not administered a second language C-test. The results of this test are provided in Table 6 and are discussed in more detail in section 5.2.1. Since scores for the English C-test and language aptitude were the same, language aptitude was not included as a predictor variable in the regression analysis including the English C-test as the dependent variable.

Table 6: Bilingual subjects' language aptitude scores, as measured via the English C-test.

Bilinguals (n = 49)	Mean	Standard deviation	Range
Total English C-test score	80.31	15.34	18 – 98

5.1.4. First language proficiency

In this study, the bilinguals' first language proficiency was estimated by the total score on the Dutch C-test. For monolinguals, this predictor variable was not included in the regression analyses. Table 7 reports the mean score, standard deviation and range for the bilinguals' first language proficiency.

Table 7: Bilingual subjects' scores on the Dutch C-test.

Bilinguals (n = 49)	Mean	Standard deviation	Range
Total Dutch C-test score	81.10	18.94	22 – 99

A Shapiro-Wilk test determined that these data were not normally distributed ($w = .801$, $p = .000$) so non-parametrical tests were adopted in statistical analyses.

5.1.5. Fluid intelligence

To assess the subjects' fluid intelligence levels, the short form of Raven's Advanced Progressive Matrices (APM; Bors & Strokes 1998) was administered. In Table 8, the subjects' scores on this task are provided.

Table 8: Subjects' scores on the short form of Raven's Advanced Progressive Matrices (Bors & Strokes 1998).

Group	Mean	Standard deviation	Range
Bilinguals (n = 58)	6.71	3.41	0 – 12
Monolinguals (n = 30)	6.93	3.16	0 – 12

Table 8 shows that the monolingual and bilingual subjects performed within the same range, with some subjects achieving the lowest score possible and others reaching the maximum score of the

task. The bilinguals' scores on the short form of the APM were not normally distributed (w = .954, p = .027) while the monolinguals' scores were (w = .957, p = 263). Results of the Mann-Whitney test did not find significant differences between the monolinguals' and bilinguals' fluid intelligence scores (p = .828).

5.1.6. Age of arrival and length of residence

Table 9 reports on the bilingual subjects' age of arrival and their length of residence in Australia, as measured in years. The subjects' length of residence was calculated by subtracting age of arrival from age at testing. In this study, age of arrival is assumed to be equal to age of acquisition, as most subjects did not receive continuous second language input until they emigrated from the Netherlands to Australia.

Table 9: Overview of the bilingual subjects' age of arrival and length of residence.

Bilinguals (n = 64)	Mean	Standard deviation	Range
Age of arrival	26.80	9.64	13 – 61
Length of residence	37.67	20.51	1 – 61

As can be seen in Table 9, age and arrival and length of residence both seem to have relatively large standard deviations and wide ranges. When these data were analyzed with Shapiro-Wilk tests, no normal distributions could be observed for age of arrival (w = .929, p = .001) or length of residence (w = .874, p = .000), possibly partly due to the implementation of three age groups per subject group. To further analyze the distributions of the age of arrival and length of residence data, histograms including a displayed normal curve are presented in Fig. 8 and 9.

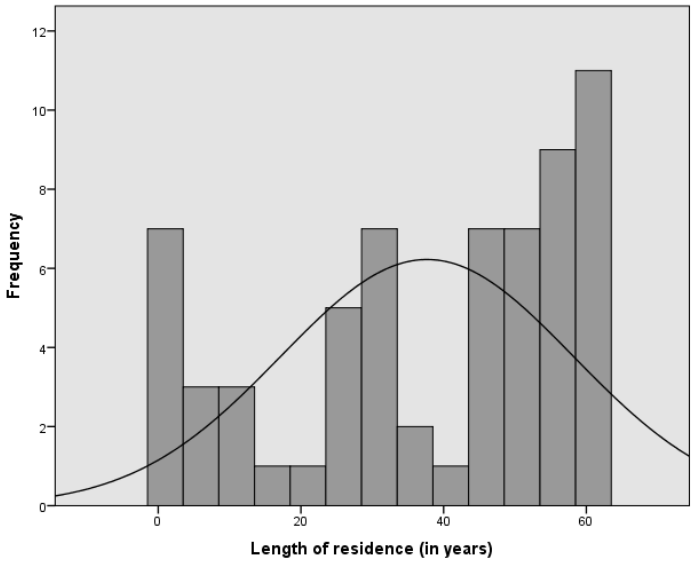

Figure 8: Histogram and normality plot for the subjects' length of residence.

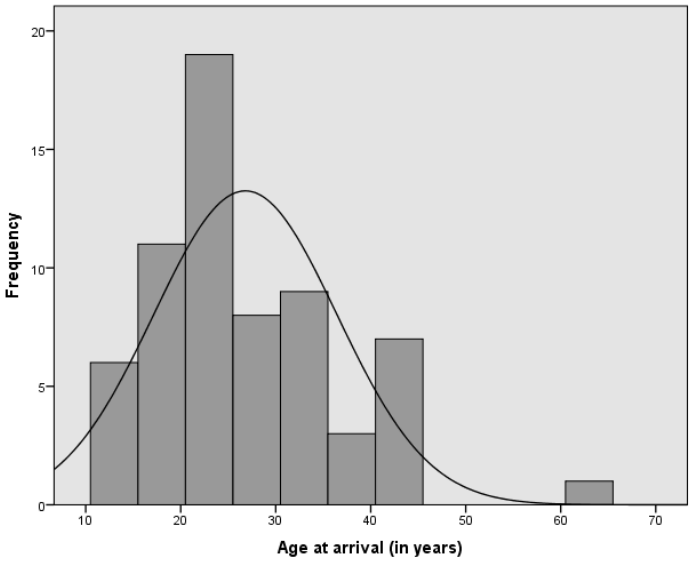

Figure 9: Histogram and normality plot for the subjects' age of arrival.

Other explanations to account for the non-normality in, for example, age of arrival can be related to the age at which decisions of permanent settlement are generally made. In this sense, the outlier subject who migrated at age 61 is an exception to the patterns observed in Fig. 9. The variation in length of residence among subjects seems to be related to the imposed age groups as the histogram displays roughly three peaks. Spearman's correlation coefficient demonstrated that length of residence correlated significantly and positively with age at testing ($\rho = .856$, $p = .000$). This comes as no surprise, as length of residence will generally increase as subjects become older. Due to this high correlation, including both these predictor variables in multiple regression analyses would violate the assumption of no multicollinearity, which is generally assumed with correlations above .800 (Field 2005, p. 175), so it was decided to exclude age at testing as a predictor variable since length of residence can provide more insights with regards to the research aims of this thesis. Length of residence, in a sense, provides an estimation of the amount of second language input the migrants have received. Generally, one would assume that a migrant who has lived in Australia for 10 years would have received less input than a migrant who has resided in the country for 30 years. The insights provided by length of residence can be contrasted and compared with those obtained by the predictor age of arrival. Length of residence can determine whether subjects' second language proficiency kept increasing over the years while age of arrival can provide insights into the existence of a critical period hypothesis. Since length of residence and age of arrival are related, these two predictors correlate significantly as well ($\rho = -.720$, $p = .000$). However, these predictors were not considered to violate the assumption of no multicollinearity, so both age of arrival and length of residence were included in multiple regression analyses.

5.2. Language proficiency tests

5.2.1. C-test

During the C-tests, subjects were asked to fill in gaps of words in 5 continuous texts. These texts increased in difficulty so, generally, text 1 would be expected to be easier than text 5. Per text, subjects could achieve a maximum score of 20. If subjects achieved a maximum score on all texts, they could achieve a maximum score of 100 in total. Only completed C-tests were included in the analyses, and thus, two C-tests were discarded for the bilingual subject group as these two were returned without any answers. Three subjects did, however, decide to leave one of the texts completely unanswered but as they completed the other texts in the C-tests, their scores were included in the analyses while the subjects received a score of '0' on those particular texts. Scores for each individual text, as well as the total scores, are provided in Table 10.

Table 10: Subjects' scores on the C-test.

Bilinguals (n=49)			
Measure	**Mean**	**Standard deviation**	**Range**
Text 1	17.82	2.89	2 – 20
Text 2	14.43	4.61	1 – 20
Text 3	16.61	3.09	4 – 20
Text 4	15.88	4.68	0 – 20
Text 5	15.57	3.69	0 – 20
Total English C-test score	80.31	15.34	18 – 98
Monolinguals (n=25)			
Measure	**Mean**	**Standard deviation**	**Range**
Text 1	19.08	1.91	11 – 20
Text 2	15.52	3.36	7 – 20
Text 3	17.08	2.12	13 – 20
Text 4	17.00	3.43	4 – 20
Text 5	16.44	3.78	0 – 19
Total English C-test score	85.12	9.87	52 – 96

Hardly any scores in Table 10 are normally distributed (with all Shapiro-Wilk test statistics obtaining a $p = .000$ for both monolinguals and bilinguals) for either of the subject groups with the exception of the monolinguals' scores on text 2 ($w = .931$, $p = .093$) and text 3 ($w = .924$, $p = .063$). As such, non-parametric statistical tests were employed to analyze the C-test data. To investigate the differences and similarities between the subjects' performances on the five linguistic constructs under investigation, Wilcoxon tests were administered. These tests revealed that bilinguals performed significantly better on text 1 than on text 2 ($p = .000$), 3 ($p = .002$), 4 ($p = .000$) and 5 ($p = .000$). Additionally, bilingual subjects scores' reflected that they found text 2 significantly more difficult than text 3 ($p = .000$) and 4 ($p = .019$). No significant differences were revealed when comparing performances on text 3 and text 4 ($p = .287$), text 3 and text 5 ($p = .084$) and text 4 and text 5 ($p = .125$). Consequently, bilingual subjects' scores on the C-test texts can be schematically displayed as follows: text 1 >> text 3, 4, 5 >> text 2. The monolingual subjects also performed significantly better on text 1 than on text 2 ($p = .000$), 3 ($p = .002$), 4 ($p = .000$) and 5 ($p = .000$). Furthermore, they experienced significantly more difficulties with text 2 than text 3 ($p = .011$), 4 ($p = .030$) and 5 ($p = .045$). No differences were observed between monolinguals' performances on text 3 and text 4 ($p = .707$), text 3 and text 5 ($p = .985$) and text 4 and text 5 ($p = .610$). As such, the monolinguals' scores can be schematically displayed in the same manner as the bilinguals' scores: text 1 >> text 3, 4, 5 >> text 2.

For reasons of simplicity and redundancy, a number of C-test variables were reduced. Scores on all texts correlated highly and significantly with the total score of the C-test for bilinguals (with correlations ranging between $\rho = .568$, $p = .000$ and $\rho = .896$, $p = .000$) and monolinguals (with correlations ranging between $\rho = .497$, $p = .012$ and $\rho = .782$, $p = .000$). These correlations suggest that scores on the texts separately are predictive of the total scores to be achieved. Hence, only the total scores of the English C-test were included in statistical analyses.

5.2.1.1. Effect of gender

To assess whether gender differences influenced scores on the English C-test, Mann-Whitney tests were administered for each of the subject groups. For both the monolingual and bilingual subjects, gender did not significantly impact on the obtained scores ($p = .147$ and $p = .934$ respectively).

5.2.1.2. Effect of age

For both monolingual and bilingual subjects, Kruskal-Wallis tests demonstrated that their assigned age categories (40-45, 60-65, 75+) did not significantly impact the total scores on the English C-test ($p = .622$ and $p = .218$ respectively). Similarly, Spearman's correlation coefficient analyses did not find significant correlations between age at testing and the C-test scores for the monolinguals ($\rho = .058$, $p = .782$) and the bilinguals ($\rho = -.179$, $p = .217$). These findings suggest that chronological age does not influence C-test scores.

5.2.1.3. Effect of bilingualism in relation to the critical period hypothesis

Since no normal distributions were observed for hardly any of the scores, the Mann-Whitney test was conducted to test for differences between monolingual and bilingual subjects. This test found no significant differences between the subject groups' total scores on the English C-test ($p = .262$). To further explore the differences among the bilingual subjects, correlations between the total score on the C-test and the variables length of residence and age of arrival on the data were investigated. According to the Spearman's correlation coefficient, both age of arrival ($\rho = .008$, $p = .959$) and length of residence ($\rho = -.123$, $p = .400$) did not correlate significantly with total C-test score.

As discussed in section 3, the critical period hypothesis can be rejected if at least one individual second language learner can be observed who performs within the range of native speakers. To test this assumption, Fig. 10 provides box plots for the monolinguals' and bilinguals' total scores obtained on the English C-test. Despite the wider range of variance among the C-test scores of the bilinguals, a substantial part of the bilingual subject group performed within the range of the monolingual subjects and scores obtained by the subject groups did not differ significantly. Surprisingly, the highest C-test score was actually obtained by one of the bilingual subjects. Thereby, the critical period hypothesis can be rejected for the second language performance on the C-test.

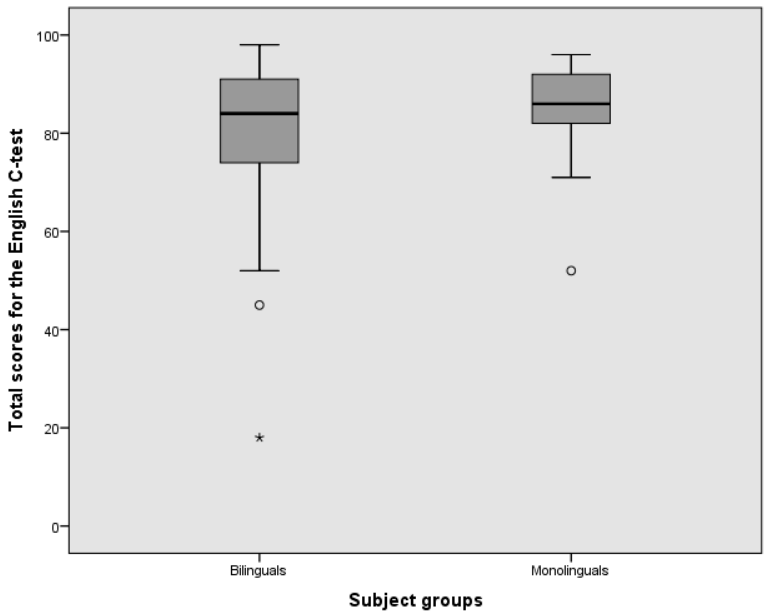

Figure 10: Box plots comparing the monolinguals' and bilinguals' performances on the C-test.

5.2.1.4. Effect of socio-psychological predictor variables

To check the assumption of linear relationships between the variables entered in the regression model, correlations were calculated between the C-test scores on the one hand, and the socio-psychological factors on the other, as shown in Table 11. However, while other analyses included the predictor variable of language aptitude for the bilingual subjects, this factor could not be entered in the current model as this measure was estimated through scores on the English C-test. Only the dependent variables that correlated significantly with the independent variable were entered in the multiple linear regression model.

Table 11: Overview of the correlations between the subjects' performances on the C-test and the socio-psychological factors.

Bilinguals	Working memory	Fluid intelligence	Length of education	First language proficiency	Age of arrival	Length of residence
English C-test	.273	.407*	.330*	.641**	.008	-.123
Monolinguals		**Working memory**		**Fluid intelligence**		**Length of education**
English C-test		.472*		.287		.548**

** p = < .01, * p = < .05.

For the first multiple regression analysis of the bilinguals' performance on the C-test, fluid intelligence, length of education and first language proficiency were entered as predictors in the regression model. The model was significant (F = 21.836 p = .000) and yielded an R^2 of .598, accounting for 60% of the variance in the bilinguals' total scores on the C-test. In this regression model, the predictor variable first language proficiency (β = .789, p = .000) was the only factor with a significant coefficients. To determine the best fitting model, a backward multiple regression analysis was conducted. Results of this analysis are provided in Table 12. This model was significant as well (F

= 62.405 p = .000) and the R^2 of .576 explained 58% of the variance in C-test scores. The only predictor in this analysis was first language proficiency, with a β of .759 (p = .000).

Table 12: Regression model for the bilinguals' performance on the C-test.

	B	Std. error	β	Sig.
(Constant)	32.205	6.325		.000
First language proficiency	.598	.076	.759	.000

R = .759, R^2 = .576, F = 62.405, p = .000.

For the monolingual subjects, working memory capacity and length of education were entered in the regression model. This resulted in a significant model (F = 6.748, p = .005) which predicted 38% of the variance. The only socio-psychological factor with a significant beta coefficient was length of education (β = .421, p = .036). When the second, backward, regression analysis was run to find a better fitting model, significance was reached again (F = 10.339, p = .004). This model accounted for 31% of the variance in C-test scores with an R^2 of .310, as shown in Table 13. With a β of .429, length of education was the only and strongest predictor variable.

Table 13: Regression model for the monolinguals' performance on the C-test.

	B	Std. error	β	Sig.
(Constant)	61.801	7.443		.000
Length of education	1.378	.429	.557	.004

R = .557, R^2 = .310, F = 10.339, p = .004.

5.2.2. Grammaticality judgment task

In Table 14, the subjects' scores on the grammaticality judgment task are presented. For each linguistic construct under investigation (i.e. verb second, passives, negation, reflexives and clausal embeddings), subjects could achieve a maximum score of 6. If subjects achieved a maximum score on all linguistic constructs, they could achieve a maximum score of 30 in total. Only completed grammaticality judgment tasks were included in the analyses, and, thus, one task was discarded for each subject group.

Table 14: Subjects' scores on the grammaticality judgment task.

Bilinguals (n=56)			
Measure	**Mean**	**Standard deviation**	**Range**
Verb second	4.55	1.17	1 – 6
Passives	4.61	1.30	0 – 6
Negation	4.88	1.16	2 – 6
Reflexives	4.05	0.96	2 – 6
Clausal embeddings	3.45	1.22	1 – 6
Total GJT score	21.54	4.14	8 – 27
Monolinguals (n=29)			
Measure	**Mean**	**Standard deviation**	**Range**
Verb second	4.79	1.18	2 – 6
Passives	4.34	1.20	2 – 6
Negation	4.76	1.27	2 – 6
Reflexives	4.24	0.99	2 – 6
Clausal embeddings	3.34	1.01	1 – 5
Total GJT score	21.48	3.79	12 – 27

None of the scores in Table 14 were normally distributed (with Shapiro-Wilk test statistics ranging between w = .837, p = .000 and w = .912, p = .019 for monolinguals' scores and between w = .844, p = .000 and w = .928, p = .002 for bilinguals' scores) for either of the subject groups. As such, non-parametric statistical tests were employed to analyze the grammaticality judgment task data. To investigate the differences and similarities between the subjects' performances on the five linguistic constructs under investigation, Wilcoxon tests were administered. These tests revealed that bilinguals performed significantly better on items containing negation that on items containing verb second (p = .043), reflexives (p = .000) and clausal embeddings (p = .000) while they did not when compared to items containing negation (p = .238). Items containing clausal embeddings were considered significantly more difficult than items containing passives (p = .000), reflexives (p = .002), verb second (p = .000) and negation (p = .000). Lastly, reflexive constructions were also considered more difficult than verb second (p = .005) and passive (p = .002) constructions. Consequently, bilingual subjects' scores for each of the linguistic constructs can be schematically displayed as follows: passives, negation >> verb second >> reflexives >> clausal embeddings. The monolingual subjects shared similar patterns in their linguistic construct scores on grammaticality judgment task. Like the bilinguals, clausal embedding were experienced to be significantly more difficult than reflexives (p = .002), passives (p = .002), verb second (p = .000) and negation (p = .000). Similarly, reflexive constructions received significantly lower scores than items containing negation (p = .036) and verb second (p = .049). Monolinguals did not score significantly differently on items containing passives on the one hand and verb second (p = .118) and negation (p = .091) on the other. Similarly, their performances on items containing verb second and negation did not differ significantly either (p = .917). Consequently, monolingual subjects' scores for each of the linguistic constructs can be schematically displayed as follows: passives, negation, verb second >> reflexives >> clausal

embeddings. As such, both monolinguals and bilinguals considered items containing reflexives and clausal embeddings more difficult than items containing passives, negation or verb second.

For reasons of simplicity and redundancy, a number of grammaticality judgment task variables were reduced. All linguistic construct scores correlated highly and significantly with the total score of the grammaticality judgment task for bilinguals (with correlations ranging between ρ = .606, p = .000 and ρ = .741, p = .000) and monolinguals (with correlations ranging between ρ = .536, p = .003 and ρ = .661, p = .000). These correlations suggest that scores on each linguistic construct are predictive of the total scores to be achieved. Hence, only the total scores of the grammaticality judgment task were included in statistical analyses.

5.2.2.1. Effect of gender

To assess whether gender differences influenced scores on the grammaticality judgment task, Mann Whitney tests were administered for each of the subject groups. For both the monolingual and bilingual subjects, gender did not significantly impact on the obtained scores (p = .256 and p = .577 respectively).

5.2.2.2. Effect of age

For both monolingual and bilingual subjects, Kruskal-Wallis tests demonstrated that their assigned age categories (40-45, 60-65, 75+) did not significantly impact on the total scores on the grammaticality judgment task (p = .172 and p = .105 respectively). Spearman's correlation coefficient analyses did reveal significant correlations between age at testing and the grammaticality judgment task score for both the monolingual subjects (ρ = -.427, p = .021) and the bilingual subjects (ρ = -.354, p = .007). These correlations suggest that as subjects age, their grammaticality judgment task scores decrease.

5.2.2.3. Effect of bilingualism in relation to the critical period hypothesis

Since no normal distributions were observed for any of the scores, the Mann-Whitney test was conducted to test for differences between monolingual and bilingual subjects. This test revealed no significant differences between the subject groups' total scores (p = .748). To further explore the differences between monolingual and bilingual subjects, correlations between the total grammaticality judgment task score and the variables length of residence and age of arrival on the data were investigated. According to the Spearman's correlation coefficient, both age of arrival (ρ =

.057, p = .678) and length of residence (ρ = -.215, p = .111) did not correlate significantly with the grammaticality judgment task scores.

As discussed in section 3, the critical period hypothesis can be rejected if at least one individual second language learner can be observed who performs within the range of native speakers. To test this assumption, Fig. 11 provides box plots for the monolinguals' and bilinguals' total scores obtained on the grammaticality judgment task. As can be seen, the bilinguals and monolinguals share the same median (i.e. 23) and maximum scores (i.e. 27) even though more variation can be observed among the bilinguals' scores. Despite the wider range of grammaticality judgment task scores of the bilinguals, a large majority of this subject group performed within the range of the monolingual subjects, and thereby the critical period hypothesis can be rejected for the second language performance on the grammaticality judgment task, especially since no significant differences were found between the monolinguals' and bilinguals' scores. These findings contrast with those of Johnson & Newport (1989) and Birdsong & Molis (2001) who observed that their subject groups with an age of arrival of 17 years or older differed significantly from monolingual controls with regards to their grammaticality judgment task performances. Possible explanations for these findings will be discussed in section 6.

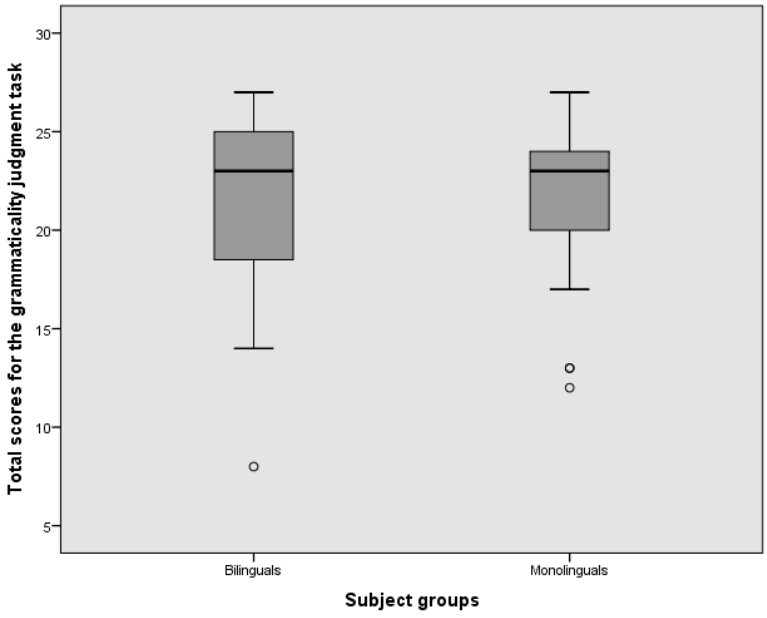

Figure 11: Box plots comparing the monolinguals' and bilinguals' performances on the grammaticality judgment task

5.2.2.4. Effect of socio-psychological predictor variables

To control for the assumption of linearity associated with multiple regression modeling, relationships between the grammaticality judgment task scores and predictor variables were explored and correlations were calculated, as shown in Table 15. Language aptitude, fluid intelligence and first language proficiency correlated significantly with the bilinguals' total scores on the grammaticality judgment task and were therefore entered in the bilinguals' regression models. For monolinguals, fluid intelligence and length of education were entered in their multiple regression models as significant correlations were observed for these variables.

Table 15: Overview of the correlations between the subjects' performances on the grammaticality judgment task and the socio-psychological factors.

Bilinguals	Working memory	Language aptitude	Fluid intelligence	Length of education	First language proficiency	Age of arrival	Length of residence
Total GJT score	.158	.307*	.294*	.206	.390**	.057	-.215
Monolinguals		**Working memory**		**Fluid intelligence**		**Length of education**	
Total GJT score		.238		.440*		.502**	

** p = < .01, * p = < .05.

For the first multiple regression analysis of the bilinguals' performance on the grammaticality judgment task, the predictor variables language aptitude, fluid intelligence and first language proficiency were simultaneously entered in the model. The model was significant (F = 3.063, p = .038) and yielded an R^2 of .173. None of the entered socio-psychological factors contributed significantly to the model. In the backward multiple regression model, however, a significant predictor variable was found in the form of language aptitude (β = .387, p = .007). This significant model (F = 8.085, p = .007), predicting 15% of the variation, is presented in Table 16.

Table 16: Regression model for the bilinguals' performance on the grammaticality judgment task.

	B	Std. error	β	Sig.
(Constant)	12.317	3.216		.000
Language aptitude	.111	.039	.387	.007

R = .387, R^2 = .149, F = 8.085, p = .007.

Again, the same procedures were followed for the monolingual subjects as were applied for the bilingual subjects. Firstly, the significantly correlated predictor variables were entered in the regression model simultaneously. This significant model (F = 6.352, p = .006) accounted for 34% of the variance in grammaticality judgment task scores. The only significant predictor variable in this model was fluid intelligence (β = .392, p = .036). Length of education did not contribute significantly to the model. To create the best fitting regression model, a backward multiple regression analysis was run. Results of this analysis are provided in Table 17. This significant model (F = 9.173, p = .005)

explained 26% of the variance in the monolinguals' grammaticality judgment task scores with an R^2 of .261. The strongest and only predictor in this analysis was fluid intelligence (β = .511, p = .005).

Table 17: Regression model for the monolinguals' performance on the grammaticality judgment task.

	B	Std. error	β	Sig.
(Constant)	16.725	1.701		.000
Fluid intelligence	.662	.219	.511	.005

R = .511, R^2 = .261, F = 9.173, p = .005.

5.2.3. Peabody Picture Vocabulary Test-4

To assess the subjects' receptive vocabulary knowledge, the Peabody Picture Vocabulary Test-4 (PPVT-4; Dunn & Dunn 2007) was administered. Table 18 presents the mean scores, standard deviations and ranges for the subjects' standardized scores on the PPVT-4 .

Table 18: Subjects' scores on the PPVT-4.

Group	Mean	Standard deviation	Range
Bilinguals (n = 61)	103.62	11.63	73 – 129
Monolinguals (n = 61)	111.90	11.92	86 – 142

The standardized scores on the PPVT-4 were normally distributed for both the monolinguals (w = .984, p = .591) and bilinguals (w = .988, p = .835).

5.2.3.1. Effect of gender

To assess whether gender differences influenced scores on the PPVT-4, independent t-tests were administered for each of the subject groups. For the monolingual subjects, gender did not significantly impact on the obtained scores (t = -.877, p = .384). For the bilingual subjects, however, gender did significantly influence scores (t = 2.126, p = .038), as men outperformed women.

5.2.3.2. Effect of age

For monolingual subjects, ANOVAs demonstrated that their assigned age categories (40-45, 60-65, 75+) significantly impacted on their scores on the PPVT-4 (F = 7.136, p = .002). More specifically, post-hoc Gabriel analyses showed that subjects aged between 60 and 65 (p = .001) and those aged 75+ (p = .042) outperformed subjects aged between 40 and 45. For bilingual subjects, no such effect was observed (F = 2.262, p = .113). However, age at testing correlated significantly with test scores for both bilinguals (ρ = .275, p = .032) and monolinguals (ρ = .444, p = .002).

5.2.3.3. Effect of bilingualism in relation to the critical period hypothesis

To assess whether bilingual and monolingual subjects' performances on the PPVT-4 differed significantly, their data were analyzed with independent samples t-tests. Results showed that monolingual subjects outperformed bilingual subjects ($t = -3.817$, $p = .000$). To further explore the differences between monolingual and bilingual subjects, the influences of the variables length of residence and age of arrival on the data were investigated as these were unique to this subject group. For bilingual subjects, significant correlations were observed between length of residence ($\rho = .321$, $p = .012$) and age of arrival ($\rho = -.351$, $p = .006$) on the one hand, and standardized scores on the PPVT-4 on the other. These results suggest that as age of arrival decreases or length of residence increases, scores on the PPVT-4 increase.

As discussed in section 3, the critical period hypothesis can be rejected if at least one individual second language learner can be observed who performs within the range of native speakers. To test this assumption, Fig. 12 provides a box plot for the standardized scores obtained on the Peabody Picture Vocabulary Test. As can be seen, the majority of bilingual subjects do in fact perform within the range of the monolinguals since there is a large overlap between the two box plots for each of the scores, even though the monolinguals outperformed the bilinguals. These findings were expected since no critical period has been previously established for the second language acquisition of vocabulary (Singleton 2005, Slabakova 2006).

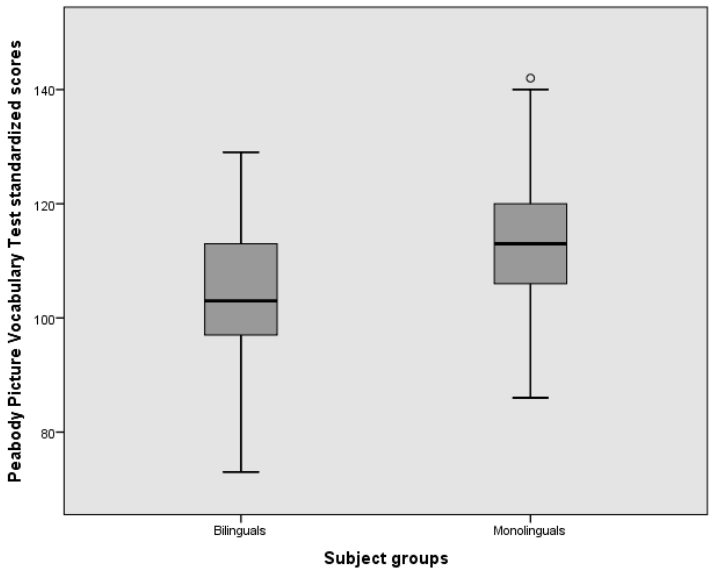

Figure 12: Box plots comparing the monolinguals' and bilinguals' performances on the PPVT-4.

5.2.3.4. Effect of socio-psychological predictor variables

Since linearity between the standardized scores on the PPVT-4 and the socio-psychological predictor variables is one of the assumptions of the multiple linear regression analysis, correlations between these types of variables were calculated for both subject groups and are shown in Table 19. Since only the predictors that correlated significantly with the language proficiency scores were included in the models, the variables first language proficiency, length of education, fluid intelligence and working memory were excluded from the regression models featuring the bilinguals' scores.

Table 19: Overview of the correlations between the subjects' performances on the PPVT-4 and the socio-psychological factors.

Bilinguals	Working memory	Language aptitude	Fluid intelligence	Length of education	First language proficiency	Age of arrival	Length of residence
PPVT-4	.090	.602**	.136	-.151	.112	-.351**	.321**
Monolinguals		**Working memory**		**Fluid intelligence**		**Length of education**	
PPVT-4		.402**		.491**		.632**	

** p = < .01, * p = < .05.

For the first multiple regression analysis of the bilinguals' performance on the PPVT-4, the predictor variables language aptitude, age of arrival and length of residence were entered in the regression model. The created model was significant (F = 7.953, p = .000) and accounted for 36% of the variance in the bilinguals' standardized scores on the PPVT-4. The only socio-psychological factor that contributed significantly to the model was language aptitude (β = .560, p = .000) while length of residence almost approached significance (β = .324, p = .073). Subsequently, a backward multiple regression analysis was conducted to find the best fitting model accounting for the variation in the standardized PPVT-4 scores. Results of this analysis are provided in Table 20. This model was significant (F = 12.190, p = .000) and explained 36% of the variance in the data with an R^2 of .362. The strongest predictor in this analysis was language aptitude, with a β of .555 (p = .000), followed length of residence (β = .303, p = .018).

Table 20: Regression model for the bilinguals' performance on the PPVT-4.

	B	Std. error	β	Sig.
(Constant)	65.533	7.988		.000
Length of residence	.107	.069	.303	.018
Language aptitude	.401	.089	.555	.000

R = .602, R^2 = .362, F = 12.190, p = .000.

For the monolinguals, the same procedure was followed as for the bilingual subjects. Firstly, all predictor variables were entered in the regression model. This resulted in a significant model (F = 4.748, p = .009) which yielded an R^2 of .354, accounting for 35% of the variance. The only predictor variable to contribute significantly to this model was length of education (β = .432, p = .026) while no significant effect of working memory capacity and fluid intelligence was observed. Unsurprisingly, the second, backward, regression analysis aiming to find the best fitting model only included length of

education as a predictor. Results for this analysis are shown in Table 21. This model was significant (F = 11.356, p = .002) and accounted for 29% of the variance in standardized score on the PPVT-4 with an R^2 of .289. With a β of .537, length of education was the only and strongest predictor variable.

Table 21: Regression model for the monolinguals' performance on the PPVT-4.

	B	Std. error	β	Sig.
(Constant)	87.913	8.416		.000
Length of education	1.585	.470	.537	.002

R = .537, R^2 = .289, F = 11.356, p = .002.

5.2.4. Fluency in Controlled Association task

During the Fluency in Controlled Association task (FICA), subjects were asked to name as many animals, fruits and vegetables, lexical items starting with an 'f' and lexical items starting with an 'a' within the fixed time limit of one minute per subtest. Results for these subtests, as well as mean scores per type of subtest and mean scores overall are provided in Table 22.

Table 22: Subjects' scores on the Fluency in Controlled Association Task.

Bilinguals (n=62)			
Measure	**Mean**	**Range**	**Standard deviation**
Animals	21.47	11 – 38	6.12
Fruit and vegetables	18.50	4 – 30	5.73
F	13.23	3 – 24	4.90
A	10.84	3 – 18	3.86
Category fluency	19.98	9.50 – 30	5.22
Letter fluency	11.93	3 – 20.50	3.95
Mean of scores	15.96	8.75 – 24	3.86
Monolinguals (=61)			
Measure	**Mean**	**Range**	**Standard deviation**
Animals	23.15	7 – 40	7.00
Fruit and vegetables	21.97	4 – 32	5.71
F	14.74	3 – 33	6.07
A	12.98	2 – 26	4.71
Category fluency	22.56	5.50 – 34.50	5.65
Letter fluency	13.86	3.50 – 29.50	4.90
Mean of scores	18.21	4.50 – 28.50	4.53

Since all verbal fluency scores presented in Table 22 were normally distributed (with Shapiro-Wilk tests statistics ranging between w = .968, p = .111 and w = .987, p = .782 for the bilinguals' scores and between w = .970, p = .133 and w = .987, p = .787 for the monolinguals' scores), paired samples t-tests could be run to investigate differences and similarities between the subjects' performances on the four subtests. These paired t-tests revealed that bilinguals named significantly more animals than fruits and vegetables (t = 4.150, p = .000) or words starting with an 'f' (t = 10.142, p = .000) or 'a' (t = 14.324, p = .000). Similarly, bilinguals named significantly more fruits and vegetables than words starting with an 'f' (t = 6.120, p = .000) or 'a' (t = 10.361, p = .000) while also naming more words starting with an 'f' than with an 'a' (t = 4.347, p = .000). Consequently, the bilingual subjects' scores on the FICA subtests can be schematically displayed as follows: animals >>

fruits and vegetables >> 'f' >> 'a'. Additionally, the bilinguals' scores on the category fluency tasks are significantly higher than those on the letter fluency tasks (t = 12.429, p = .000). Paired samples t-tests revealed similar results for monolinguals as they named significantly more animals than words starting with an 'f' (t = 9.322, p = .000), or 'a' (t = 12.710, p = .000) while no significant result was revealed when contrasting the number of named animals and fruits and vegetables (t = 1.550, p = .126). In addition, they found naming fruits and vegetables significantly easier than naming words starting with an 'f' (t = 8.152, p = .000) or 'a' (t = 11.185, p = .000), and scores were also significantly higher for naming words starting with an 'f' over those with an 'a' (t = 2.928, p = .005). The monolinguals' scores on the FICA subtests can, just as for the bilinguals, be schematically displayed as follows: animals, fruits and vegetables >> 'f' >> 'a'. Lastly, scores were significantly higher on category fluency tasks than on letter fluency tasks (t = 12.467, p = .000). Consequently, the differences in scores between various verbal fluency measures are almost identical for the monolingual and bilingual subjects.

For reasons of simplicity and redundancy, a number of verbal fluency variables were reduced. As the verbal fluency results for the two category fluency tasks correlated significantly (r = .550, p = .000 for bilinguals and r = .578, p = .000 for monolinguals) and the two letter fluency task did too (r = .548, p = .000 for bilinguals and r = .649, p = .000 for monolinguals), subsequent statistical analyses were conducted on the basis of the means of the category fluency and letter fluency tasks. Additionally, the mean score of all tasks taken together was also included in analyses.

5.2.4.1. Effect of gender

In both monolingual (t = -2.042, p = .046) and bilingual (t = -3.023, p = .004) subject groups, female subjects outperformed male subjects on the category fluency tasks, presumably because of large differences in scores in the category of fruits and vegetables. When the female and male subjects were compared on the basis of their letter fluency and mean scores, no significant results were found (bilinguals: t = -.040, p = .968 and t = -1.989, p = .051 respectively, monolinguals: t = .460, p = .647 and t = 1.500, p = .137 respectively).

5.2.4.2. Effect of age

For the monolingual subjects, ANOVAs demonstrated that their assigned age categories (40-45, 60-65, 75+) significantly impacted on their scores on the letter fluency tasks (F = 6.384, p = .003) and the mean score of all tasks (F = 4.018, p = .023), but not on their scores on the category fluency tasks (F = 2.047, p = .138). More specifically, post-hoc Gabriel analyses demonstrated that during the letter

fluency task, subjects aged between 60 and 65 outperformed the subjects aged between 40 and 45 (p = .003). Similarly, subjects aged between 60 and 65 outperformed those aged 75+ on the mean score of all tasks (p = .054). For bilingual subjects, ANOVAs showed that their assigned age categories (40-45, 60-65, 75+) significantly influenced their scores on the category fluency tasks (F = 11.596, p = .000) and the mean score of all tasks (F = 7.896, p = .001), but not their scores on the letter fluency tasks (F = 1.578, p = .215). Post-hoc Gabriel analyses showed that participants aged between 40 and 45, and 60 and 65 outperformed those aged 75+ during the category fluency tasks (40 – 45: p = .000, 60 – 65: p = .002) and on the mean score of all tasks (40 – 45: p = .003, 60 – 65: p = .005). When investigating the correlations between chronological age and the verbal fluency scores, no significance was observed for the monolinguals' performances (with correlations ranging between ρ = -.187, p = .149 and ρ = -.038, p = .774) while age at testing did correlate significantly with the bilinguals' performances on the category fluency tasks (ρ = -.491, p = .000) and mean score of all tasks (ρ = -.386, p = .002), but not on the letter fluency tasks (ρ = -.106, p = .413). These results suggest that for bilinguals, as subjects age, they seem to be less likely to perform as well as younger subjects. In the monolingual subject group, on the other hand, monolinguals in the age category of 60 to 65 years old performed better than subjects who were younger or older in two of the subtests.

5.2.4.3. Effect of bilingualism in relation to the critical period hypothesis

To assess whether bilingual and monolingual subjects' performances on the FICA differed significantly, their data were analyzed with independent samples t-tests. Results showed that monolingual subjects outperformed bilingual subject on the category fluency tasks (t = -2.625, p = .010), the letter fluency tasks (t = -2.412, p = .017) and mean score of all tasks (t = -2.972, p = .004), suggesting that monolingual subjects were quicker in naming lexical items within the set time limits and had a more expanded vocabulary. To further explore the differences between monolingual and bilingual subjects, the influences of the variables length of residence and age of arrival on the data were investigated. For bilingual subjects, significant, negative, correlations were observed between length of residence and the category fluency scores (ρ = -.406, p = .001) and the mean score of all tasks (ρ = -.303, p = .017), but not between length of residence and the letter fluency scores (ρ = -.054, p = .675). Significance was not reached in correlations between the verbal fluency measures and age of arrival (with correlations ranging between ρ = .127, p = .326 and ρ = -.025, p = .849)

As discussed in section 3, the critical period hypothesis can be rejected if at least one individual second language learner can be observed who performs within the range of native speakers. To test this assumption, Fig. 13 provides three box plots for each of the FICA scores analyzed in this section. As can be seen, the majority of bilingual subjects do in fact perform within the range of the monolinguals since there is a large overlap between the two box plots for each of the scores, even though the two groups did differ significantly. These findings should, however, not come as a surprise. Based on the assumption that the critical period hypothesis does not control the acquisition of vocabulary in a second language (Singleton 2005, Slabakova 2006), bilingual subjects were expected to perform within the range of monolingual subjects.

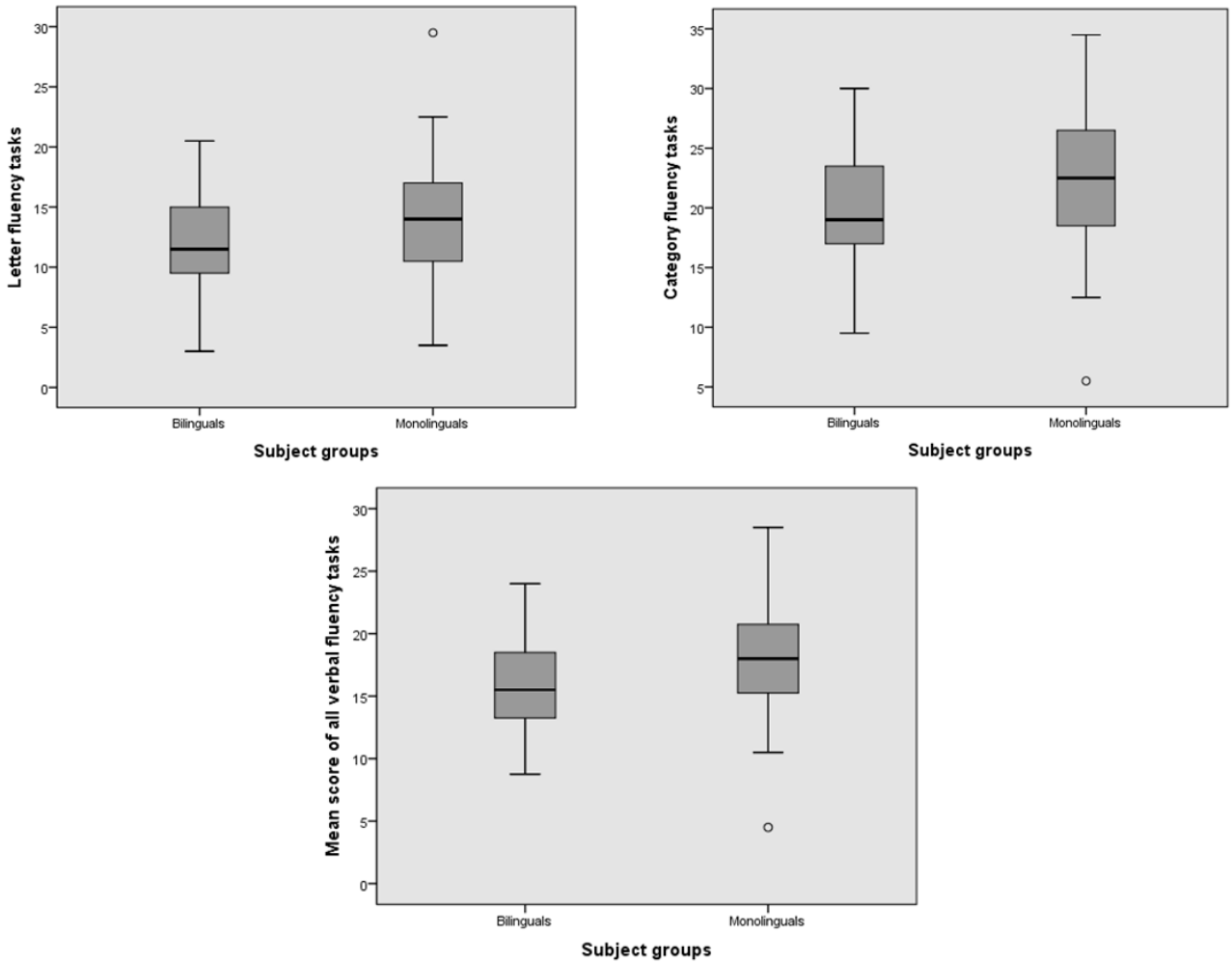

Figure 13: Box plots comparing the monolinguals' and bilinguals' performances on the letter fluency tasks (left), category fluency tasks (right) and mean score of all tasks (bottom).

5.2.4.4. Effect of socio-psychological predictor variables

Since the observation of a linear relationship between the dependent and independent variables to be entered in the multiple regression model is one of the assumptions that need to be checked before conducting the analyses, correlations were calculated between the FICA scores on the one hand, and the socio-psychological factors on the other. As such, only predictor variables that correlated significantly with the FICA scores were included in the model. Results of these correlation analyses for both the monolingual and bilingual subjects are shown in Table 23.

Table 23: Overview of the correlations between the subjects' performances on the FICA and the socio-psychological factors.

Bilinguals	Working memory	Language aptitude	Fluid intelligence	Length of education	First language proficiency	Age of arrival	Length of residence
FICA category	.590**	.259	.312*	.255*	.345*	.140	-.432**
FICA letter	.207	.476**	.224	.194	.339*	.026	-.089
FICA mean	.497**	.369*	.308*	.246	.371*	.081	-.353**
Monolinguals		**Working memory**		**Fluid intelligence**		**Length of education**	
FICA category		.597**		.432*		.580**	
FICA letter		.493**		.325		.522**	
FICA mean		.636**		.484**		.650**	

** p = < .01, * p = < .05.

For the first multiple regression analysis of the bilinguals' performance on the letter fluency tasks of the FICA, the predictor variables first language proficiency and language aptitude were entered in the regression model. The model was significant (F = 8.503, p = .001) and yielded an R^2 of .283, accounting for 28% of the variance in the bilinguals' performances on the letter fluency tasks. Of the two socio-psychological factors entered into the regression, only language aptitude contributed significantly to the model (β = .523, p = .011). Subsequently, a backward multiple regression analysis was conducted to find the best fitting regression model. Results of this analysis are provided in Table 24. This significant model (F = 17.395, p = .000) explained 28% of the variance in the data with an R^2 of .283. Like in the regression analysis created with the enter method, language aptitude was the only predictor in this analysis with a β of .532 (p = .011).

Table 24: Regression model for the bilinguals' letter fluency scores.

	B	Std. error	β	Sig.
(Constant)	.580	2.817		.838
Language aptitude	.143	.034	.532	.000

R = .532, R^2 = .283, F = 17.395, p = .000

` For the first regression model of the monolinguals' letter fluency scores, working memory capacity and length of education were entered as predictor variables simultaneously. This resulted in a significant model (F = 13.220, p = .000) which, with an R^2 of .317, accounted for 32% of the variance in monolinguals' performances on the letter fluency tasks. In this model, both length of education (β = .315, p = .018) and working memory capacity (β = .328, p = .014) were significant predictors. To find

the best fitting model accounting for the variation in the monolinguals' letter fluency scores, a backward multiple regression analysis was conducted. Results of this analysis are provided in Table 25. This model yielded exactly the same results as provided by the enter method.

Table 25: Regression model for the monolinguals' letter fluency scores.

	B	Std. error	β	Sig.
(Constant)	.650	2.665		.808
Length of education	.380	.156	.315	.018
Working memory	.197	.078	.328	.014

$R = .563$, $R^2 = .317$, $F = 13.220$, $p = .000$

Working memory capacity, fluid intelligence, length of education, first language proficiency and length of residence were entered in the multiple linear regression model of the bilinguals' category fluency scores. The model created with the enter method was significant ($F = 6.352$, $p = .000$) and accounted for 44% of the variance within the data with an R^2 of .436. In this model, working memory capacity was the strongest and only significant predictor variable ($\beta = .612$, $p = .001$) while length of residence, first language proficiency, length of education and fluid intelligence did not reach significance. Unsurprisingly, working memory capacity was the only predictor variable in a regression model which was created with the backward method, as displayed in Table 26. Again, the model was significant ($F = 31.096$, $p = .000$) and accounted for 41% of the variance in the data with an R^2 of .409.

Table 26: Regression model for the bilinguals' category fluency scores.

	B	Std. error	β	Sig.
(Constant)	3.768	2.997		.215
Working memory	.470	.084	.639	.000

$R = .639$, $R^2 = .409$, $F = 31.096$, $p = .000$.

For the monolinguals, all the predictor variables were entered in a regression model, together with their scores on the category fluency tasks, resulting in the creation of a significant model ($F = 5.455$, $p = .005$). This regression model explained 39% of the variance in the data with an R^2 of .386. Working memory capacity was the only significant predictor in this model ($\beta = .435$, $p = .030$). Subsequently, a backward regression analysis was conducted to find the best fitting model accounting for the variation in the category fluency task data. The thus created, significant ($F = 24.636$, $p = .000$), model is displayed in Table 27. This model accounted for 36% of the variance in the monolinguals' category fluency task scores with an R^2 of .359. Like in the regression model created with the enter method, working memory capacity was the only significant predictor in this model as well with a β of .580 ($p = .001$).

Table 27: Regression model for the monolinguals' letter fluency scores.

	B	Std. error	β	Sig.
(Constant)	3.762	5.431		.487
Working memory	.520	.138	.580	.001

R = .599, R^2 = .359, F = 24.636, p = .000.

For the first multiple regression analysis of the bilinguals' mean scores on the FICA, the predictor variables working memory capacity, language aptitude, fluid intelligence, first language proficiency and length of residence were entered in the regression model. The model was significant (F = 6.030, p = .000) and yielded an R^2 of .430. Consequently, this model accounted for 43% of the variance in the bilinguals' mean scores on the FICA. The socio-psychological factors that contributed significantly to the model included language aptitude (β = .593, p = .007), working memory capacity (β = .353, p = .034) and length of residence (β = -.446, p = .019). Subsequently, a backward multiple regression analysis was conducted to find the best fitting model accounting for the variation in bilinguals' mean scores on the FICA. Results of this analysis are provided in Table 28. This model was significant (F = 11.028, p = .000) and explained 34% of the variance in the data with an R^2 of .339. The strongest predictor in this analysis was working memory capacity with a β of .427 (p = .002), followed by language aptitude (β = .293, p = .028).

Table 28: Regression model for the bilinguals' mean scores.

	B	Std. error	β	Sig.
(Constant)	1.231	3.298		.711
Working memory	.243	.073	.427	.002
Language aptitude	.080	.035	.293	.028

R = .582, R^2 = .339, F = 11.028, p = .000

Once more, the same regression procedures were followed for the monolingual subjects as had been applied for the bilingual subjects. Firstly, all of the monolinguals' predictor variables were entered in the regression model. This significant model (F = 9.541, p = .000) accounted for 52% of the variance in grammaticality judgment task scores as it yielded an R^2 of .524. The significant predictor variables in this model were length of education (β = .367, p = .028) and working memory capacity (β = .449, p = .012). To create the best fitting regression model, a backward multiple regression analysis was run. Results of this analysis are provided in Table 29. This significant model (F = 14.797, p = .000) explained 52% of the variance in the monolinguals' grammaticality judgment task scores with an R^2 of .523. The strongest predictor in this analysis was working memory capacity with a β of .465 (p = .005), followed by length of education with a β of .375 (p = .020).

Table 29: Regression model for the monolinguals' mean scores.

	B	Std. error	β	Sig.
(Constant)	1.035	3.527		.771
Length of education	.374	.151	.375	.020
Working memory	.311	.101	.465	.005

$R = .723$, $R^2 = .523$, $F = 14.797$, $p = .000$.

5.3. Summary

In sum, the findings presented in section 5 seem to suggest that, for the subjects in this study, the assumptions made by the critical period hypothesis cannot be maintained. On all language proficiency measures, the majority of bilingual subjects performed within the range of the monolingual subjects, and the two groups' performances on the C-test and grammaticality judgment task did not differ significantly. Most importantly, the multiple regression models featured in this section, of which the findings are summarized in Table 30, suggest that the bilingual subjects' language aptitude levels predict outcomes on the majority of language proficiency measures. Additionally, length of residence, first language proficiency and working memory capacity were also found to significantly predict scores on one or more language proficiency tests, while length of residence, fluid intelligence and length of education did not. For the monolingual subjects, length of education and working memory capacity significantly predicted outcomes on language proficiency measures while fluid intelligence did not.

Table 30: Overview of significant predictor variables in the multiple regression models created with the backward method.

Bilinguals						
Socio-psychological predictors	**GJT**	**C-test**	**PPVT-4**	**FICA letter**	**FICA category**	**FICA mean**
Length of residence	No	No	Yes	No	No	No
Age of arrival	No	No	No	No	No	No
Language aptitude	Yes		Yes	Yes	No	Yes
First language proficiency	No	Yes	No	No	No	No
Working memory capacity	No	No	No	No	Yes	Yes
Fluid intelligence	No	No	No	No	No	No
Length of education	No	No	No	No	No	No
Monolinguals						
Socio-psychological predictors	**GJT**	**C-test**	**PPVT-4**	**FICA letter**	**FICA category**	**FICA mean**
Working memory capacity	No	No	No	Yes	Yes	Yes
Fluid intelligence	Yes	No	No	No	No	No
Length of education	No	Yes	Yes	Yes	No	Yes

6. Discussion and conclusions

This section aims to answer the research questions of this thesis, as repeated below, by discussing the results reported in the previous section.

1) Can late second language learners in a migrant context reach a level of second language proficiency comparable to native speakers?
2) Which individual socio-psychological differences, in addition to age of arrival, are predictive of second language success in late second language learners of English in a migrant context?

After these two research questions have been answered, limitations of this present study will be discussed. Subsequently, implications and suggestions for future research will be addressed.

6.1. Critical period hypothesis

One of the aims of this study was to examine the tenability of the critical period hypothesis by investigating whether late learners' second language proficiency actually reached a similar level as native speakers of that second language. As discussion in section 3., the critical period hypothesis can be rejected and the first research question can be positively answered if, minimally, a single late second language learner can be identified who performed within native speaker range on the C-test, grammaticality judgment task, Fluency in Controlled Association task and Peabody Picture Vocabulary Test-4 (Dunn & Dunn 2007). By basing the answer to the first research question on bilinguals' performances in multiple domains of linguistics (i.e. morphosyntax, vocabulary and general language proficiency), more compelling evidence for the verification of the critical period hypothesis can be provided (Abrahamsson & Hyltenstam 2009).

In section 5, the validity of the critical period hypothesis was discussed for each language proficiency measure by analyzing box plots of the monolinguals' and bilinguals' data distributions. Since no critical period has been assumed for the second language acquisition of vocabulary (Singleton 2005, Slabakova 2006), it was expected that almost all bilinguals would perform within native speaker range on the PPVT-4 and FICA despite significant differences between the monolingual and bilingual subjects' scores on these tasks. Similar results were reported for the bilinguals' performances on the C-test and grammaticality judgment task since a large degree of overlap between the monolinguals' and bilinguals' scores was observed on both tasks. However, monolinguals and bilinguals, in fact, performed similarly on these tasks as no significant differences

were observed. Of the 46 bilingual subjects who submitted and completed all second language proficiency tasks, 42 performed within the range of monolinguals on all of these tasks. This can be converted to 91% of successful late second language acquirers, thus refuting the critical period hypothesis. As such, these results seems to justify labeling the subjects in this study as 'bilinguals', as opposed to 'late second language learners'.

Since this study failed to verify the critical period hypothesis, results of the influential Johnson & Newport (1989) study and of its successors were not replicated. In their study, Johnson & Newport (1989) failed to find a single late second language learner who performed within the native speaker range on a grammaticality judgment task. More specifically, none of the subjects who emigrated after age 7 performed within the range of the monolingual controls, and thereby the critical period hypothesis was verified. As such, Johnson & Newport's (1989) results stand in stark contrast with those obtained in the present study. However, this study deviated from Johnson & Newport (1989) in a few crucial manners. Firstly, the migrants in Johnson & Newport (1989) had a length of residence that was much shorter than that of the migrants in this study. More specifically, the average length of residence in Johnson & Newport (1989) was roughly 10 years while the bilinguals in this study had been living in Australia for roughly 38 years on average. Consequently, the migrants in this study were immersed in the country of migration for a longer period of time and were able to receive more foreign language input and practice. Additionally, the second language learners of English in Johnson & Newport (1989) all spoke Chinese or Korean as their native language. Typologically, English and Korean or Chinese are relatively distant, which might make them more difficult to acquire since barely anything is similar in the first and second language. Dutch and English, on the other hand, are typologically relatively close. When compared to Korean and Chinese, the Dutch language is more similar to English with regards to phonology, vocabulary and morphosyntax, which probably benefits second language acquisition. This beneficiary effect of learning a second language that is typologically close is confirmed by Van Wuijtswinkel (1994) who, like this study, found 15 out of 34 native Dutch speaking late second learners of English who performed within the range of native speakers on a subset of the Johnson & Newport (1989) grammaticality judgment task. Similarly, Birdsong & Molis (2001) were able to identify late second language learners of English, with Spanish as their first language, who performed within the range of the native subjects from Johnson & Newport (1989) on a grammaticality judgment task. Consequently, the increased length of residence and the typological proximity of the first and second languages of the bilingual subjects in this study have probably partly contributed to their native like performances despite their late ages of arrival. Additionally, a large number of subjects in the two oldest age groups of this study reported that government migration schemes had encouraged them to culturally assimilate as well as they could. Since many subjects had migrated to Australia on their own or as a couple, and Dutch migrants

were scattered all over Australia rather than living in an isolated community, they were very motivated to learn their second language as quickly as possible so they could find jobs, make friends and blend in in Australian society. As such, the language learners' motivation might also have contributed to their degrees of second language success.

Of the 46 bilinguals who completed all English language proficiency tasks, four did not perform within the range of native speakers of English on one or more of the language measures. Of these four subjects, two performed below the range of monolinguals on the PPVT-4 while another one experienced difficulties with the English C-test. Lastly, one subject performed below the native speaker range on both the C-test and grammaticality judgment task. To explore the reasons for these bilinguals not performing within the monolingual range, their personal characteristics and scores on other tasks were further investigated. The subjects who obtained low scores on the PPVT-4 had both migrated to Australia at a relatively late age, i.e. 34 and 35, and had resided in the country for 7 and 28 years respectively. Since length of residence correlated positively and age of arrival correlated negatively with outcomes of the PPVT, their relatively poor performances on the test can potentially be attributed to these factors. The subject who scored below monolingual range on the English C-test also obtained an outlying score on the Dutch C-test. Since first language proficiency (as estimated by scores on the Dutch C-test) was the only significant predictor in the multiple regression model for the bilinguals' C-test scores, these two performances are likely to be related. Lastly, one of the four subjects experienced difficulties on both the C-test and grammaticality judgment task. As bilinguals' performances on the grammaticality judgment task were predicted by language aptitude (as estimated by scores on the English C-test), this subjects' low score on the grammaticality judgment task seems to be influenced by the obtained C-test score. Furthermore, both C-test and grammaticality judgment task scores correlated significantly with fluid intelligence, on which this subject achieved a relatively low score, performing within the bottom 25% of the bilingual range on the Raven's APM. Consequently, the deviating scores obtained by these four subjects seem to be accounted for by relating them to the analyses focusing on the influences of individual socio-psychological differences on language proficiency measures.

6.2. Individual socio-psychological differences

In sections 5.2.1 through 5.2.4, results were presented with regards to the second language performances of the bilingual subjects and first language performances of the monolingual subjects on the grammaticality judgment task, C-test, PPVT-4 and various components of the FICA. Since this thesis aims at identifying characteristics that are predictive of successful late second language learning, backward multiple regression analyses were conducted, where the second language

performance scores were entered as the dependent variable and where the individual socio-psychological differences were entered as the independent variables. By adopting this statistical method, the relative strength of the predictive values of working memory capacity, length of education, fluid intelligence, language aptitude, first language proficiency, length of residence and age of arrival on late learners' second language performance were assessed. For the monolingual subjects, the predictive values of working memory capacity, fluid intelligence and length of education on their first language performance were assessed. Since factors such as length of residence, age of arrival, language aptitude and first language proficiency were unique to the bilingual subjects group, these differences in socio-psychological predictor variables arose. To contrast and compare the impact of the subjects' individual differences on their language performances, Table 30, providing an overview of significant predictors in the multiple regression models for both subject groups, is repeated below. Whenever a socio-psychological factor was a significant predictor in the best fitting regression model, this is indicated by a 'yes' in Table 30. Whenever a socio-psychological factor was not included in the best fitting model, this is indicated by a 'no' in Table 30. The sections following this table try to account for these results in more detail.

Table 30: Overview of significant predictor variables in the multiple regression models created with the backward method.

Bilinguals						
Socio-psychological predictors	**GJT**	**C-test**	**PPVT-4**	**FICA letter**	**FICA category**	**FICA mean**
Length of residence	No	No	Yes	No	No	No
Age of arrival	No	No	No	No	No	No
Language aptitude	Yes		Yes	Yes	No	Yes
First language proficiency	No	Yes	No	No	No	No
Working memory capacity	No	No	No	No	Yes	Yes
Fluid intelligence	No	No	No	No	No	No
Length of education	No	No	No	No	No	No
Monolinguals						
Socio-psychological predictors	**GJT**	**C-test**	**PPVT-4**	**FICA letter**	**FICA category**	**FICA mean**
Working memory capacity	No	No	No	Yes	Yes	Yes
Fluid intelligence	Yes	No	No	No	No	No
Length of education	No	Yes	Yes	Yes	No	Yes

6.2.1. Effect of individual differences for bilinguals

As can be seen in Table 30, no single socio-psychological factor was identified to be predictive of performances on all second language proficiency tasks. As such, different language proficiency tasks seem to be influenced by different learner characteristics. In this section, the impact of each socio-psychological predictor variable on the language proficiency tasks will be discussed and related to previous research.

The predictor variable to have had the largest impact on the second language proficiency measures was language aptitude. This socio-psychological factor contributed significantly to outcomes on the grammaticality judgment task, PPVT-4, the mean score and the letter fluency tasks of the FICA. Language aptitude was not only a significant predictor in the regression models of these language tasks, but also correlated significantly and positively with scores on these second language proficiency measures (with correlations ranging between $\rho = .307$, $p = .032$ for the grammaticality judgment task and $\rho = .602$, $p = .000$ for the PPVT-4). These correlations all suggest that as language aptitude levels increase, scores on the second language proficiency measures will do so too. As such, language aptitude seems to have a positive effect on the successful late second language acquisition of grammar (as measured via the grammaticality judgment task) and vocabulary size (as measured via both the PPVT-4 and FICA). Previous research has already confirmed a positive effect of language aptitude on the late second language acquisition of grammar (DeKeyser 2000, Abrahmsson & Hyltenstam 2008, DeKeyser 2010) and foreign accent (Hopp & Schmid 2011), but the effect of language aptitude on late second language vocabulary acquisition has, as far as aware, not been studied. Instinctively, language aptitude is likely to positively influence the acquisition of vocabulary size in addition to grammar and accent. If a learner has a particular talent for learning languages, this talent will logically be reflected in the learner picking up new words more easily and quickly than a learner with a lower level of language aptitude. However, it should be noted that, in this study, language aptitude was assessed as an indirect construct with a second language C-test (like in Hopp & Schmid 2011) and not with a conventional language aptitude measure such as the Modern Language Aptitude Test (Carroll & Sapon 1959).

The bilinguals' working memory capacity also had a significant influence on second language proficiency as it was included in the multiple regression models for the mean and category fluency scores of the FICA. Not only was working memory capacity included as a significant predictor in the regression model for both these scores, but it also correlated significantly with the mean scores ($r = .497$, $p = .000$) and category fluency scores ($r = .590$, $p = .000$) of the FICA. Consequently, a late second language learner with a relatively high working memory capacity is likely to perform well on these tasks related to vocabulary size and verbal fluency. This positive influence of working memory capacity is probably related to the design of the FICA. The task requires a subject to name as many words as possible within a certain category or starting with a certain letter. The time limit for each letter or category is one minute. This design requires the subject to quickly process his or her vocabulary to find lexical items that match the given description, and, thus, appeals to the subjects' processing and storage components of working memory. The effect of working memory capacity did not extend itself to other language proficiency measures. Consequently, no significant effect of

working memory capacity on the grammaticality judgment task scores was observed and results of McDonald (2006) were not replicated.

In addition to significantly and positively predicting C-test scores in the regression model, first language proficiency also correlated highly and significantly with the total score on the English C-test (ρ = .641, p = .000), which suggests that an increase in first language proficiency will similarly increase C-test scores. Thus, the assumptions made in Hulstijn & Bossers (1992) were confirmed in the present study for this aspect. It should be noted, though, that the bilinguals' first language proficiency was estimated through their scores on the Dutch version of the C-test. While the C-tests in Dutch and English are not directly comparable because of differences in languages and texts, the significant effect of first language proficiency could potentially be due to the design of the task. More specifically, if a subject finds it very difficult to complete parts of words while simultaneously focusing on coherence and comprehension of the text, these problems might be visible in both the Dutch and English C-test.

Length of residence was included as a significant predictor in the regression model for the PPVT-4. Combining this result with the significant correlation between the two variables suggests (ρ = .321, p = .012) that the longer a learner resides in the country of migration, the more expanded the learner's vocabulary becomes. As no critical period has been previously observed for the late second language acquisition of vocabulary (Singleton 2005, Slabakova 2006), these results support the claims that practice, rather than an early age of arrival, makes perfect in second language vocabulary acquisition.

Lastly, the socio-psychological factors of fluid intelligence, length of education and age of arrival were, despite their significant correlations with most of the language proficiency measures, not significantly included in any of the created multiple linear regression models. Consequently, results by Hakuta et al. (2003) and Wang (1998) were not replicated in this study.

6.2.2. Effect of individual differences for monolinguals

Even though the issues addressed in this section are not required to answer the research questions, the impact of the socio-psychological predictor variable on the language proficiency tasks of the monolingual subjects will be discussed and compared to the bilinguals' results.

The socio-psychological predictor that influenced the results of the most language proficiency tasks was length of education with significant contributions to the regression models for the C-test, PPVT-4, mean scores and letter fluency scores of the FICA. These language proficiency measures are all partly centered on vocabulary acquisition. As such, a longer enrollment in education, correlating with a higher level of education, is expected to lead to an expanded

vocabulary. Significant correlations between length of education and all language proficiency measures (with correlations ranging between ρ = .502, p = .006 for the grammaticality judgment task and ρ = .650, p = .000 for the letter fluency scores of the FICA) support this assumption. Since the monolinguals' length of education correlates significantly with fluid intelligence (ρ = .485, p = .007), the significant contribution of fluid intelligence to the regression model of the grammaticality judgment task scores also supports that a higher level of education or intelligence will increase scores on language proficiency tasks. When compared to the bilinguals' regression models, length of education and fluid intelligence have a much greater impact for the monolinguals than bilinguals, suggesting that the beneficiary effect of intelligence and education is severely limited in second language acquisition. This difference could also potentially be due to the competition of a greater amount of socio-psychological factors relevant for bilinguals. Furthermore, as the significant predictor variables for the bilingual subjects in this study mainly consisted of individual differences unique to this population (i.e. age of arrival, length of residence, language aptitude and first language proficiency), these findings suggest that first and second language acquisition rely on different learner characteristics. As such, it seems to be confirmed that a bilingual is more than simply the sum of two monolinguals (Grosjean 1989).

The monolinguals' working memory capacity was a significant predictor variable in the regression models for all scores on the FICA. Additionally, working memory capacity correlated significantly and rather highly with the FICA scores (with scores ranging between ρ = .493, p = .000 for the letter fluency scores and ρ = .636, p = .000 for the mean scores) in a replication of the results for the bilinguals. As for the bilingual subjects, the impact of working memory capacity is related to the design of the FICA that relies heavily on the processing and storage components of working memory.

6.3. Limitations of the present study

Since the complete test battery would have been overly long if all tests were to be administered in one session, some tests (i.e. the short form of Raven's Advanced Progressive Matrices (Bors & Strokes 1998), C-test, grammaticality judgment task) were, depending on the time of subjects' recruitment, sent to the subjects before or after the testing session, as discussed in section 4.3. and 4.4. As the large majority of bilingual subjects were pre-sent these materials, most of them returned these tests during their testing appointments. However, as the majority of monolingual subjects were recruited while bilingual subjects were tested, the testing materials were given to these subjects afterwards and they were requested to return these materials by mail. Unfortunately, until this point in time (31th of May 2012), not all subjects have returned their materials. To meet this

thesis' deadlines, it was decided to work with the data available at this time while accepting the fact that data was missing and sample sizes would be unequal. Thus, for all results in section 5, sample sizes for each subject group were reported and differed throughout various measures. Not all statistic tests actually required equal sample sizes so the challenges posed by missing data and the unequal sample sizes did not hinder the data analysis procedures itself. However, the reliability of the results on the take-home tasks should not be taken for granted. Since roughly half of the monolingual subjects had returned their take-home tasks at time of writing, their sample size decreased for these tasks and, thus, statistical power decreased as well. Due to this reliability on the subjects themselves, the data was limited. This had direct consequences for the multiple regression analyses of the monolingual subjects since complete data packages were only available for 30 out of 60 subjects, leading to possibly biased data.

Besides the missing data and unequal sample sizes of the short form of APM, C-test and grammaticality judgment task, another disadvantage of these take-home tasks was that subjects could not be controlled while completing these tasks. While the instructions clearly requested the subjects to keep to a certain time limit, it cannot be guaranteed that they actually did or that they did not ask anyone's help, either personally or digitally. However, all these disadvantages did not weigh out against the advantage of having subjects complete these tasks outside testing sessions.

As mentioned both in this section and in section 4.1., the recruitment of monolingual subjects took quite some time. Rather unexpectedly, not many people were willing to participate and to fill the subject groups, virtually all willing participants were included in the study. As such, subjects did not perfectly conform to the set age groups or no equal distribution amongst levels of education could be observed. Since there was no larger pool available to choose from for monolingual subjects, the number of subjects actually matched those who had signed up. For the bilingual subjects, this was not the case, as about 30 willing participants had been rejected on the basis of their screening questionnaires. However, for the bilinguals more diversity in length of residences would have been desirable. As plenty of participants applied for inclusion from the older age groups, subjects aged between 40 and 45 were much more difficult to recruit and as such, some of the subjects had only been in Australia for a year.

This present study is based on the data collected in the NWO VENI research project Crossing Communication Borders: First Language Reversion in Healthy, Aging Dutch Migrants in Australia led by dr. M.C.J. Keijzer (project number 016.104.017). Since this NWO VENI research project's main aims and research questions were not to explore the critical period hypothesis in second language acquisition, but rather to investigate first language reversion, this present study was not designed to include bilingual subjects who migrated to Australia before age 15. Ideally, however, this study would have investigated the second language proficiency performances of this group as well so results

could be fully compared and contrasted with previous research in the field (e.g. Johnson & Newport 1989, DeKeyser 2000, Birdsong & Molis 2001, DeKeyser et al. 2010). Also, by including a subject group of early arrivals, changes in the likelihood of reaching a level of ultimate attainment in a second language as a function of age of arrival could have been studied. In its current design, this study was not able determine which of the age of arrival – ultimate attainment functions in Fig. 2 could be fitted with the data.

6.4. Implications and suggestions for future research

The results of this study have implications for late second language learners aiming to improve their second language proficiency. Firstly, they should not be demotivated by the assumptions that a late age of acquisition will automatically impede the level of ultimate attainment they will achieve. Since this study has rejected the critical period hypothesis for a large proportion of the bilingual subjects, Dutch migrants are generally expected to achieve a level of second language attainment that is comparable to that of native speakers of English, regardless of age of arrival. Furthermore, the majority of subjects in this study seem to have acquired their second language learning through immersion as almost no subjects reported having followed any English language courses after migration. Consequently, late second language learners seem capable of acquiring a language through implicit learning mechanisms, a finding which seems to contradict claims made by DeKeyser (2000).

Results obtained from the multiple regression analyses focusing on the impact of individual socio-psychological differences on second language proficiency measures can be considered less motivating for late second language learners. The individual difference proven to impact most language proficiency scores was language aptitude, a relatively stable and largely innate talent. As such, learners do not seem to be able to significantly improve their language aptitude levels with practice and this can significantly influence the level of ultimate attainment they will reach in their second language. Consequently, second language teaching methods might benefit from offering their late learners with a low level of language aptitude second language aids rather than teaching them second language skills. More specifically, second language learners with a low level of aptitude might appreciate resources such as lists of second language vocabulary and commonly used phrases in particular contexts.

The most important recommendation to be drawn from this study is for future research on late second language acquisition to shift away from the predominant focus on age of arrival as the prime cause for non-native like performances. Instead, studies are advised to gather more information on their subjects' characteristics, as second language learners do not merely differ on

the basis of age of acquisition or arrival and length of residence. Examples of such characteristics were included in this study as individual socio-psychological differences.

Generally, studies aiming to explore the effect of such individual differences on second language acquisition tend to only focus on one potential socio-psychological factor. For example, DeKeyser et al. (2010) only focused on language aptitude and McDonald (2006) only focused on working memory capacity in relation to second language acquisition. These limitations in scope prevent such studies from exploring the competition between and the relative strengths of multiple socio-psychological predictor variables. Eventually, analyzing the impact of multiple factors can potentially create more compelling evidence. More specifically, if a study such as DeKeyser et al. (2010) would have included more socio-psychological predictor variables and language aptitude would have still been the most significant predictor, the evidence in favor of the influence of aptitude would be stronger. As such, future research aiming to explore second language learners' individual differences, which could account for their varying degrees of second language success, should include a multitude of variables, as was done in this study.

Another methodological consideration for future studies to take into account is to include multiple language proficiency measures to determine whether late second language learners achieved native like competence on a range of tasks. Previous research has mainly focused on second language learners' performances on grammaticality judgment task and are, therefore, severely limited in scope and strength. Both the present study and Abrahamsson & Hyltenstam (2009) included linguistic tasks in various domains and required the bilinguals to perform within native speaker range of each of these tasks in order to reject the critical period hypothesis. However, as Abrahamsson & Hyltenstam's (2009) results conflict with those obtained in this study, future research is needed to provide more insights into the verification or rejection of the critical period hypothesis based on these stricter criteria.

References

Abrahamsson, N. & Hyltenstam, K. (2008), The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition* 30(4), 481-509.

Abrahamsson, N. & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Language Learning* 59, 249-306.

Almli, C.R. & Finger, S. (1987). Neural insult and critical period concepts. In M.H. Bornstein (Ed.), *Sensitive Periods in Development: Interdisciplinary Perspectives* (pp. 123-143). Hillsdale, NJ: Lawrence Erlbaum.

Altenberg, E. P. & Vago, R. M. (2004). The role of grammaticality judgments in investigating first language attrition: a cross-disciplinary perspective. In M. S. Schmid, B. Köpke, M. Keijzer, L. Weilemar (Eds.), *First language attrition: Interdisciplinary perspectives on methodological issues* (pp. 105-129). Amsterdam: John Benjamins.

Baddeley, A. (1992). Working memory. *Science* 255(5044), 556-559.

Baddeley, A.D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4, 417–423.

Baddeley, A. D., Gathercole, S.E., Papagno, C. (1998). The phonological loop as a language-learning device. *The Psychological Review* 105, 158–173.

Bates, E., Thal, D., Trauner, D., Fenson, J., Aram, D., Eisele, J., Nass, R. (1997). From first words to grammar in children with focal brain injury. In D. Thal & J. Reilly (Eds.), *Special issue on Origins of Communication Disorders, Developmental Neuropsychology* 13(3), 275-343.

Bard, E.G., Robertson, D., Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language* 72, 32-68.

Bialystok, E. & Miller, B. (1999). The problem of age in second-language acquisition: influences from language, structure, and task. *Bilingualism: Language and Cognition* 2, 127-145.

Bialystok, E., Craik, F.I.M., Luk, G. (2008). Lexical access in bilinguals: effects of vocabulary size and executive control. *Journal of Neurolinguistics* 21(6), 522-538.

Birdsong, D. (2005). Interpreting age effects in second language acquisition. In J. F. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 109-127). Oxford: Oxford University Press.

Birdsong, D. & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44, 1-15.

Bongaerts, T. (2005). Introduction: ultimate attainment and the critical period hypothesis for second language acquisition. *International Review of Applied Linguistics* 43, 259-267.

Bors, D.A. & Stokes, T.L. (1998). Raven's advanced progressive matrices: norms for first-year university students and the development of a short form. *Educational and Psychological Measurement* 58, 382-398.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual Differences and Universals in Language Learning Aptitude* (pp. 83–118). Rowley, MA: Newbury House.

Carroll, J. & Sapon, S. (1959). *The Modern Languages Aptitude Test*. San Antonio, Tx.: The Psychological Corporation.

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

Curtiss, S.R. (1988). Abnormal language acquisition and the modularity of language. In F.J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (pp. 96-116). Cambridge, UK: Cambridge University Press.

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450-466.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22, 499-533.

DeKeyser, R. & Larson-Hall, J. (2005). What does the critical period really mean? In J. F. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 88-108). Oxford: Oxford University Press.

DeKeyser, R., Alfi-Shabtay, I., David, R. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics* 31, 413-438.

DeKeyser, R.M. & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning, Volume II* (pp. 407-425). New York, NY: Routledge.

De Villiers, J. & De Villiers, P. (1978). *Language Acquisition.* Cambridge, MA: Harvard University Press.

Dörnyei, Z. (2006). Individual differences in second language acquisition. *AILA Review* 19, 42-68.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589-630). Malden, MA and Oxford, UK: Blackwell.

Dunn, L. M., & Dunn, D. M. (1959). *Peabody Picture Vocabulary Test.* Circle Pines, MN: AGS.

Dunn, L. M. & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4).* Minneapolis, MN: Pearson.

Field, A. (2005). *Discovering statistics using SPSS (2nd edition).* London: Sage Publications.

Flege, J.E., Yeni-Komshian, G.H., Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language* 41, 78–104.

Grigorenko, E., Sternberg, R., Ehrman, M. E. (2000). A theory based approach to the measurement of foreign language learning ability: the canal-f theory and rest. *Modern Language Journal* 84(3), 390–405.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36, 3–15.

Hakuta, K., Bialystok, E., Wiley, E. (2003). Critical evidence: A test of the critical-period hypothesis for second-language acquisition. *Psychological Science* 14, 31–38.

Hopp, H. & Schmid, M.S. (2011). Perceived foreign accent in L1 attrition and L2 acquisition: the impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, FirstView article.

Hyltenstam, K. & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty & M.H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 539-588). Malden, MA: Blackwell.

Hulstijn, J.H. & Bossers, B. (1992). Individual differences in L2 proficiency as a function of L1 proficiency. *The European Journal of Cognitive Psychology* 4, 341-353.

Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21, 60-99.

Juffs, A. & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching* 44(2), 137–166.

Keijzer, M.C.J. (2007). *Last in first out?: An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada.* Utrecht: Landelijke Onderzoekschool Taalwetenschap (LOT).

Lenneberg, E.H. (1967). *Biological Foundations of Language*. New York: Wiley.

Long, M. H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition* 12, 251–285.

Mandell, P.B. (1999). On the reliability of grammaticality judgment tests in second language acquisition research. *Second Language Research* 15(1) 73-99.

Marler, P. (1991). The instinct to learn. In S. Carey & R. Gelman (Eds.), *The epignesesis of mind: essays on biology and cognition* (pp. 37-66). Hillsdale, NJ: Erlbaum.

Mayberry, R.I. (1993). First-language acquisition after childhood differs from second language acquisition: the case of American sign language. *Journal of Speech and Hearing Research* 36, 1258-1270.

Mayberry, R.I. & Eichen, E.B. (1991). The long-lasting advantage of learning sign language in childhood: another look at the critical period for language acquisition. *Journal of Memory and Language* 30, 486-512.

McDonald, J. L. (2006). Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language* 55(3), 381–401.

Meara, P.M. (2005). *Llama Language Aptitude Tests*. Swansea: Lognostics.

Meara, P., Milton, J., Lorenzo-Dus, N. (2003). *Swansea Language Aptitude Tests (LAT) v2.0.* Swansea, UK:

Lognostics.

Miyake, A. & Friedman, N.P. (1998). Individual differences in second language proficiency: working memory as language aptitude. In A. F. Healy & L. E. Bourne (Eds.), *Foreign Language Learning: Psycholinguistic Studies on Retention and Training* (pp. 339-364). Mahwah, NJ: Lawrence Erlbaum.

Moyer, A. (2004). *Age, accent and experience in second language acquisition: an integrated approach to critical period inquiry*. Clevedon: Multilingual Matters.

Muñoz, C. & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language Teaching*, 44.1, 1-35.

National Institute for Testing and Evaluation. (2001). *Inter-University Psychometric Entrance Test* [in Russian]. Jerusalem: Author.

Newport, E. L. (1991). Contrasting conceptions of the critical period for language. In S. Carey & R. Gelman (Eds.), *The epigenesis of mind: essays on biology and cognition* (pp. 111-130). Hillsdale, NJ: Erlbaum.

Newport, E.L., Bavelier, D., Neville, H.J. (2001). Critical thinking about critical periods: perspectives on a critical period for language Acquisition. In Dupoux, E. (Ed.), *Language, Brain, and Cognitive Development* (pp. 481-502). Cambridge MA: MIT Press.

Ottó, I. (1996). *Hungarian Language Aptitude Test: Words in Sentences.* Unpublished manuscript. Department of English Applied Linguistics, Eötvös Loránd University, Budapest.

Penfield, W. & Roberts, L. (1959). *Speech and Brain Mechanisms*. Princeton, NJ: Princeton University Press.

Peterson, C. & Al-Haik, A. (1976). The development of the Defence Language Aptitude Battery. *Educational and Psychological Measurement* 36, 369–380.

Pimsleur, P. (1966). *Pimsleur Language Aptitude Battery (PLAB)*. New York: Psychological Corporation.

Product - Peabody Picture Vocabulary Test, Fourth Edition (PPVT™-4) (2012). Retrieved from http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAa30700.

Raven, J. C. (1938). *Progressive matrices: a perceptual test of intelligence, 1938, sets A, B, C, D, and E*. London: H. K. Lewis.

Raven, J. C. (1962). *Advanced Progressive Matrices, Set II*. London: H. K. Lewis.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on adult incidental SLA: a replication and extension of Reber, Walkenfeld & Hernstadt (1991). In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (pp. 211-266). Amsterdam: John Benjamins.

Rosansky, E. (1975). The critical period for the acquisition of language: some cognitive developmental considerations. *Working Papers on Bilingualism* 6, 93-100.

Schmid, M.S. & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: the impact of extralinguistic factors. *Second Language Research* 26(1), 125-160

Seliger, H.W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W.C. Ritchie (Ed.), *Second Language Acquisition Research: Issues and Implications* (pp. 11-19). New York, NY: Academic Press.

Service, E., Simola, M., Metsaenheimo, O., Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology* 14, 383–407.

Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics* 43, 269-285.

Singleton, D. (2007). The critical period hypothesis: some problems. *Interlingüística* 17, 48-56.

Slabakova, R. (2006). Is there a critical period for the acquisition of semantics? *Second Language Research* 22, 302-338.

Sorace, A. (2010). Using magnitude estimation in developmental linguistic research. In E. Blom, & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 57-72). Amsterdam: John Benjamins.

Sparks, R.L. & Ganschow, L. (1991). Foreign language learning differences: affective or native language aptitude differences? *Modern Language Journal* 75, 3–16.

Stern, H. H. (1983). *Fundamental concepts of language teaching.* Oxford: Oxford University Press.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science* 103(2684), 677-680.

U.S. Department of Commerce. (1995). *Detailed cross-tabulations of selected language groups for states: 1990* [CD-ROM]. 1990 census of population and housing. Washington, DC: Bureau of the Census, Population Division, Decennial Programs Coordination Branch.

Van Boxtel, S. (2005). *Can the late bird catch the worm? Ultimate attainment in L2 syntax*. Utrecht: Landelijke Onderzoekschool Taalwetenschap (LOT).

Van den Noort, M.W.M. L., Bosch, P., Hughdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist* 11, 289–296.

Van den Noort, M.W.M.L., Bosch, P., Haverkort, M., Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment* 24(1), 35-42.

Van Wuijtswinkel, K. (1994). *Critical period effects on the acquisition of grammatical competence in a second language*. Unpublished B.A. thesis, Department of Applied Linguistics, University of Nijmegen.

Vasilyeva, M. & Waterfall, H. (2011). Variability in language development: relation to SES and environmental input. In S. Neuman and D. Dickinson (Eds.), *Handbook of Early Literacy Research, Volume 3*. (pp 36-48). New York, NY: Guilford Press.

Wang, W. (1998). *Age-related effects in adult second language acquisition: a study of Mandarin speaking learners of English*. Unpublished doctoral dissertation. OISE University of Toronto, Canada .

Wechsler, D. (1941). *The Measure of Adult Intelligence (Section edition)*. Baltimore, MD: Williams & Wilkins.

Appendices

Appendix A: Screening questionnaires

I: Dutch version for the bilingual subjects

The bilingualism and aging project
Proefpersonenformulier (Nederlandse versie) Nederlandse Australiërs

Wat is uw naam?

Bent u een man of een vrouw?
_Man
_Vrouw

Wat is uw geboortedatum?

Hoe oud was u toen u naar Australië verhuisde?

Waar in Nederland bent u geboren en waar bent u opgegroeid?

Zou u zeggen dat u standaard Nederlands spreekt of een dialect?
- standaard Nederlands
- een dialect, namelijk:

Wat is uw huidige beroep of – indien u gepensioneerd bent – wat was uw laatste beroep voor u met pensioen ging?

Wat is uw hoogst afgeronde opleiding?
_ basisschool
_ middelbare school – lager niveau
_middelbare school – hoger niveau
_beroepsopleiding
_ hoger onderwijs
_universiteit

Hoeveel jaar heeft u onderwijs gevolgd, alles bij elkaar genomen?

Indien u getrouwd bent of samenwoont, heeft uw partner een Nederlandse achtergrond?
_ja
_nee anders, namelijk:

Hoe vaak spreekt u nog Nederlands?
_dagelijks
_ wekelijks
_ een paar keer per jaar
_ bijna nooit of nooit

Hoe zou u op dit moment uw Nederlandse taalvaardigheid in schatten?

_ zeer slecht
_ slecht
_ ok
_ goed
_ zeer goed

Hoe zou u op dit moment uw Engelse taalvaardigheid in schatten?
_ zeer slecht
_ slecht
_ ok
_ goed
_ zeer goed

Welke talen spreekt u nog meer, behalve Nederlands en Engels? Kunt u voor elke taal aangeven hoe u uw taalvaardigheid in die taal op dit moment in schat?

Talen	Zeer goede beheersing	Goede beheersing	Ok beheersing	Slechte beheersing	Zeer slecht beheersing

Heeft u ooit een hersenbloeding gehad of een hoofdletsel of lijdt u op dit moment aan een cognitieve ziekte zoals Alzheimer?
_nee
_ja, namelijk:

Zijn uw gezichtsvermogen en gehoor intact?
_ja
_nee (kunt u aangeven waarom u nee heeft geantwoord op deze vraag?):

Bent u kleurenblind?
_nee
_ja

Heeft u ooit (nu of in het verleden) regelmatig drugs of medicijnen gebruikt?
_Nee
_ja, namelijk:

Hoe zou u op dit moment uw fysieke gezondheid in schatten?

- zeer slecht
- slecht
- acceptabel
- goed
- zeer goed

Wat uw is huidige woonsituatie?

- ik woon alleen en zelfstandig
- ik woon alleen in een ouderencomplex
- ik woon met mijn partner zelfstandig
- ik woon met mijn partner in een ouderencomplex

Is uw verblijf in Australië permanent of bent u van plan om weer terug te gaan naar Nederland?

- permanent:
- mijn verblijf is tijdelijk, namelijk: (tot hoe lang denkt u in Australië te blijven?)

Maakt u veel gebruik van services die aangeboden worden voor ouderen binnen uw woonplaats? Dat kan variëren van maaltijdservices tot handwerkclubs en dergelijken.

- nee, helemaal niet
- ja, namelijk:

Bent u rechts of links-handig?
_rechts-handig
_ links-handig

Bent u beschikbaar om getest te worden gedurende de maanden december 2011 en januari 2012?
_ja
_Nee, maar ik ben wel beschikbaar (geef tijdsraam aan):

Kunt u naar Monash University in Melbourne komen om getest te worden?
_ja
_Nee, ik word liever getest in mijn eigen huis vanwege lichamelijke beperkingen

Vermeldt u hieronder alstublieft uw contactgegevens
Adres:
Telefoonnummer:
E-mail:

Kent u nog meer mensen die geïnteresseerd zouden zijn om mee te doen aan dit onderzoek? Zo ja, zou u de namen en contactgegevens van deze mensen hieronder willen vermelden? Deze informatie zal alleen bekend worden gemaakt aan de hoofdonderzoeker van dit project en de privacy van de personen zal altijd gewaarborgd zijn.

Appendix A: Screening questionnaires

II: English version for the bilingual subjects

The bilingualism and aging project
Subject form (English version) Dutch-Australian group

What is your name?

Are you male or female?
_Male
_Female

What is your date of birth?

How old were you when you moved to Australia?

Where in the Netherlands were you born and where did you grow up?

Do you consider yourself to be a speaker of standard Dutch or a dialect speaker?
- standard Dutch
- dialect, namely:

What is your current profession or – if you are retired – what was your last profession before retirement?

What is the highest form of education you completed?
_primary school
_secondary school lower level
_secondary school higher level
_vocational training
_ higher education
_ university

Everything taken together, how many years of formal education have you had?

If you are/were married or live together, does/did your spouse have a Dutch background?
_yes
_no, namely:

How often do you still speak Dutch?
_daily
_ weekly
_ a few times a year
_ hardly ever/never

How would you currently rate your own Dutch language proficiency?

_ very poor
_ poor
_ ok
_ good
_ very good

How would you currently rate your own English language proficiency?
_ very poor
_ poor
_ ok
_ good
_ very good

Apart from Dutch or English, which other languages do you speak? Please indicate for each language how you rate your proficiency in that language:

Languages	Very good proficiency	Good proficiency	Ok proficiency	Poor proficiency	Very poor proficiency

Have you ever suffered from a stroke, head trauma or are you currently suffering from a degenerate disease such as Alzheimer's?
_no
_yes, namely:

Do you have intact hearing and vision?
_yes
_no (could you indicate why you answered no to this question?):

Are you colour-blind?
- no
- yes

Have you ever (now or in the past) used drugs or medication on a regular basis?
_No
_yes, namely:

How would you rate your current physical health?
-Very poor
-poor
-reasonable
-good

-very good

What is your current living situation?

- I live by myself independently
- I live by myself in an assisted living situation (care for the elderly situation)
- I live with my partner independently
- I live with my partner in an assisted living situation (care for the elderly situation)
- Other, namely:
-

Are you making use of community-based services for the elderly (this could vary from meal services to clubs specifically set up for elderly citizens)?

- no
- yes, namely:

Is your stay in Australia temporary or permanent?

- temporary, I intend to stay for…..years
- permanent

Are you right-handed or left-handed?
_right-handed
_ left-handed

Are you available for testing during the months of December 2011 and January 2012?
_Yes
_No, but I am available (indicate time frame):

Are you in a position to come to Monash University in Melbourne for testing?
_Yes
_No, I would like to be tested in my own home because of physical disability

Please provide your contact details below:
Address:
Telephone number:
E-mail address:

Do you know more people who may be interested in participating in this study? If so, could you include their names and contact details below? Please note that these contact details will only be made available to the principle researcher on this project and will not be passed on to others.

Appendix A: Screening questionnaires

III: English version for the monolingual subjects

The bilingualism and aging project
Subject form (English version) Australian group

What is your name?

Are you male or female?
_Male
_Female

What is your date of birth?

Where in Australia were you born and were did you grow up?

What is your current profession or – if you are retired – what was your last profession before retirement?

What is the highest form of education you completed?
_ primary school
_ secondary school lower level
_secondary school higher level
_vocational training
_ higher education
_ university

Everything taken together, how many years of formal education have you had?

Were you raised bilingually or do you speak another language than English on a native-speaker level?
_no
_yes, namely (please indicate the language):

Have you ever lived abroad for a longer period of time (longer than 3 months)?
_no
_yes, namely:

How would you currently rate your own English language proficiency?
_ very poor
_ poor
_ ok
_ good
_ very good

Apart from English, which other languages do you speak? Please indicate for each language how you rate your proficiency in that language:

Languages	Very good proficiency	Good proficiency	Adequate proficiency	Poor proficiency	Very poor proficiency

Have you ever suffered a stroke, head trauma or are you currently suffering from a degenerate disease such as Alzheimer's?
_no
_yes, namely:

Do you have intact hearing and vision?
_yes
_no (could you indicate why you answered no to this question?):

Are you colour blind?
_no
_yes

Have you ever (now or in the past) used drugs or medication on a regular basis?
_No
_yes, namely:

How would you rate your current physical health?
-Very poor
-poor
-reasonable
-good
-very good

What is your current living situation?

- I live by myself independently
- I live by myself in an assisted living situation (care for the elderly situation)
- I live with my partner independently
- I live with my partner in an assisted living situation (care for the elderly situation)
- Other, namely:

Are you making use of community-based services for the elderly (this could vary from meal services to clubs specifically set up for elderly citizens)?

- no
- yes, namely:

Are you right-handed or left-handed?
_right-handed
_ left-handed

Are you available for testing during the months of December 2011 and January 2012?
_Yes
_No, but I am available (indicate time frame):

Are you in a position to come to Monash University in Melbourne for testing?
_Yes
_No, I would like to be tested in my own home due to physical disability

Please provide your contact details below:
Address:

Telephone number:

E-mail address (if you have it):

Do you know more people who may be interested in participating in this study? If so, could you include their names and contact details below? Please note that these contact details will only be made available to the principal researcher on this project and will not be passed on to others.

Appendix B: Sociolinguistic questionnaires

I: Dutch version for the bilingual subjects

Mondeling instructie vooraf:
Ik zou graag willen beginnen met u een paar vragen te stellen over uw situatie en over uw vroegere leven in Nederland en uw huidige leven in Australië. Op die manier hoop ik een goed algemeen beeld van u te kunnen vormen. Veel van de basisinformatie heb ik al, omdat u de vragenlijst heeft ingevuld toen u aangaf interesse te hebben om mee te doen met dit onderzoek. Ik zal soms naar uw antwoorden daarop refereren. Alle vragen zal ik mondeling aan u stellen en ons hele gesprek zal ik opnemen. Ik maak af en toe ook wat aantekeningen tussendoor, maar daar hoeft u zich niets van aan te trekken. Uiteraard zijn er geen goede of foute antwoorden, want het gaat gewoon om uw situatie. Nog iets anders wat belangrijk is: ik wil graag de vragen gerelateerd aan Nederland in het Nederlands stellen en dan overgaan naar het Engels voor de vragen die gaan over uw leven hier in Australie. Dat merkt u vanzelf doordat ik in een andere taal verder ga, maar ik geef u ook wel een seintje. Denkt u dat het allemaal een beetje duidelijk is of heeft u vooraf nog vragen? Zo niet, dan gaan we maar gewoon beginnen.

Sectie 1: Algemene ijsbrekers/uitbreidingen op intake vragenlijst (Nederlands)

1. Wat kunt u zich herinneren van de voorbereidingstijd voor uw vertrek naar Australie? [open vraag]

2. Waarom bent u naar Australië verhuisd of waarom zijn uw ouders verhuisd? [open vraag]

3. Ter opheldering: het is voor ons erg belangrijk om te weten hoeveel jaren opleiding/educatie u heeft gevolgd. Kunt u daarom nog eens vertellen hoe lang u formeel onderwijs heeft gehad in jaren?

Sectie 2: Nederlands (te transcriberen):

4. Vond u het moeilijk om te wennen in het begin? [open vraag]

5. Vond u het moeilijk om om te gaan met de Engelse taal in het begin?

6. Had u voordat u naar Australië kwam al Engelse les gehad, bijvoorbeeld op school of wellicht een avondcursus?
- nee, nooit
- ja, namelijk: (incl. aantal jaren)

7. Hoe lang woont u nu ook alweer in Australie?. Hoe goed schat u uw Engelse taalvaardigheid in toen u aankwam? [laat kiezen uit 5-punts schaal]
- zeer slecht
- slecht
- redelijk

- goed
- zeer goed

8. En hoe was dat 10 jaar later? [laat kiezen uit 5-punts schaal]
- zeer slecht
- slecht
- redelijk
- goed
- zeer goed

9. Hoe schat u dat nu in? [laat kiezen uit 5-punts schaal]
- zeer slecht
- slecht
- redelijk
- goed
- zeer goed

10. En hoe zit dat eigenlijk met het Nederlands? Hoe schat u dat in toen u aankwam in Australie? [laat kiezen uit 5-punts schaal]
- zeer slecht
- slecht
- redelijk
- goed
- zeer goed

11. En hoe was dat 10 jaar later? [laat kiezen uit 5-punts schaal]
- zeer slecht
- slecht
- redelijk
- goed
- zeer goed

12. En hoe zou u uw Nederlandse taalvaardigheid op dit moment inschatten? [laat kiezen uit een 5-punts schaal]
- zeer slecht
- slecht
- redelijk
- goed
- zeer goed

13. Zou u een verschil maken tussen schrijven, lezen, spreken en luisteren als het gaat om uw taalvaardigheid in het Engels en Nederlands? Kunt u aangeven hoe goed u in al deze vaardigheden denkt te zijn in het Nederlands en Engels [laat schema zien indien nodig en zet zelf kruisjes]:

	Zeer goed	Goed	Adequaat	Slecht	Zeer slecht
Nederlands:					

Leesvaardigheid					
Engels: Leesvaardigheid					
Nederlands: Schrijfvaardigheid					
Engels: Schrijfvaardigheid					
Nederlands: Luistervaardigheid					
Engels: Luistervaardigheid					
Nederlands: Spreekvaardigheid					
Engels: Spreekvaardigheid					

14. U heeft het indirect al aangegeven, maar zou u zeggen dat uw Engelse en Nederlandse taalvaardigheid veranderd zijn door de jaren heen? Op wat voor een manier dan precies? [open vraag]

15. Zou u uzelf tweetalig (Nederlands-Engels) willen noemen? Zo ja (of nee) welke van uw twee talen is dan uw beste taal (dominante taal)? [binaire keuze plus open vraag]

16. Ik zou graag meer willen weten over wanneer u Nederlands gebruikt en wanneer Engels en met wie. Ik noem zo een paar contexten op en zou u dan kunnen aangeven of u in die contexten bijna altijd, veel, soms, vrijwel nooit of nooit Engels of Nederlands gebruikt? [laat schema zien indien nodig en zet zelf kruisjes]

Ik spreek Nederlands					
	Bijna altijd	Veel	Soms	Vrijwel nooit	nooit
Met familie					
Met vrienden					
Tegen huisdieren					
Op het werk[4]					
In de kerk					
In winkels					
Op clubs of verenigingen					

[4] Indien proefpersonen gepensioneerd zijn kun je hier vragen hoe dat was voordat zij met pensioen gingen.

Anders, namelijk:					

Ik spreek Engels					
	Bijna altijd	Veel	Soms	Vrijwel nooit	nooit
Met familie					
Met vrienden					
Tegen huisdieren					
Op het werk[5]					
In de kerk					
In winkels					
Op clubs of verenigingen					
Anders, namelijk:					

17. Grofweg genomen, hoeveel procent van uw dagelijks leven gebruikt u Engels en hoeveel Nederlands, zou u zeggen? [binaire vraag]
    a. Engels:
    b. Nederlands:

18. Is dat percentage veranderd door de jaren heen? Hoe en waarom denkt u?

19. Gaat u regelmatig terug naar Nederland voor bezoekjes? Zo ja, wat zijn dan vooral de redenen van die bezoekjes? [open vraag]

20. Behalve de bezoekjes, onderhoudt u regelmatig contact met Nederland? [open vraag]
    a. Hoe vaak precies?

    b. Op welke manier doet u dat: telefoon, brieven, email?

Sectie 3: Engels (te transcriberen):

Intro: because I would now like to talk about your life in Australia I would like to switch to English, as I said I would in the beginning. Could you also answer in English, please?

[5] Indien proefpersonen gepensioneerd zijn kun je hier vragen hoe dat was voordat zij met pensioen gingen op het werk.

21. Have you made many new friends in Australia [open question]

22. Which origin do most of these new friends have? [multiple choice question]
- Dutch
- English (British)
- Other, namely:

23. Do you feel more at ease with the Dutch or the Australian culture? And why? [binary choice plus open question]
- Dutch
- Australian

24. Do you feel more at ease when speaking Dutch or English? [binary choice plus open questions]
- Dutch
- English
    a. Why do you think that is?
    b.
    c. Has this changed over the years?

25. Could you remind me of the background of your (former) partner. [open subquestions]
    a. Did you meet here in Australia?
    b. How did you meet?
    c. What language(s) do you speak with him/her?

26. Do you have children? If so, which language(s) do you speak with them? [open question]

23. I would like to know more about your social network. First of all, do you go outside a lot and meet up with other people or do you mainly stay close to your own home? [binary choice plus open question]
a. I have many contacts outside my home. Things I do include:

b. I mainly stay in on a daily basis and do not have many contacts outside of my home.

24. Do you have any hobbies? What do you like to do in your spare time? [open question, record the number of hobbies and kind of activities]

Sectie 4: afrondende vragen en praktische zaken

27. We have reached the end of the interview. Is there anything you would like to add or ask? [open vraag]

Appendix B: Sociolinguistic questionnaires

II: English version for the monolingual subjects

Section 1: ice breakers/ elaborations on the intake questionnaire

1. Could you tell me/remind me how you would assess your own English language profiency? Why would you say that? [open question]

2. Do you also speak other languages than English? Do you think you can ever be as proficient in a non-native language as in your native language? [in case they do not speak any other language, you can ask whether they would like to learn to speak a second language and then leave the rest of the question as it is]

Section 2: to transcribe

3. Have you ever had any formal foreign language instruction, either at school or perhaps as part of a special course?
- No, never
- Yes, namely (incl. number of years)

4. If we were to return to your English language proficiency, how would you have rated this 10 to 20 years ago? The same? [5 point scale and prompt to expand]
- Very poor
- poor
- ok
- good
- very good

5. Would you say that your English language proficiency has changed over the years? [binary choice plus open question]
- Nee, it hasn't changed
- Yes, it has changed. How would you say it has changed and what could have caused that?

6. Of course, in a way it is strange to ask you about your overall language proficiency, because writing and speaking are not the same as listening or reading. I would therefore like to ask if you could assess your current English language proficiency in each of these areas:

	Very good	Good	Acceptable	Poor	Very poor
ENG: listening					
ENG: speaking					

ENG: reading					
ENG: writing					

7. In general, would you say that English in Australia is changing? Does the younger generation use a different kind of English? What could cause such a change? [open question]

8. I would like to continue with your social life. Do you have many friends and/or acquaintances?

9. Do your normally stay close to home or do you have many contacts outside your home? [binary choice plus open question]

a. I have many contacts outside my home. Things I do include:

b. I mostly stay close to home and I don't have many contacts outside my home

10 Do you have many hobbies? What do you do in your spare time?

Sectie 4: afrondende vragen en praktische zaken

11. We have come to the end of this questionnaire. Is there anything we haven't talked about that you would like to mention or would you like to add anything else?

Appendix C: C-tests
I: Dutch C-test for bilingual subjects

Tekst 1:

Ik houd van Nederland en niet zo'n beetje ook. Waarom ik van het land houd is niet alleen omdat velen van wie ik houd hier leven. Nee, het is me (1) ……dan d (2) ……. De groo (3) …… reden v (4) …… mijn lie (5) …… voor het land ko (6) …… voort u (7) ……het feit dat al (8) …… zo geor (9) ……en syste (10) …… is. Er i (11) ……een systeem e (12) …… het wer (13) ……. Je kan, ni (14) …… zonder twi (15) …… ,maar to (16) …… met dic (17) ……ogen er (18) …… uitgaan d (19) ……het recht zege (20) …….

Antwoorden:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Tekst 2:

Als je reist heb je de kans om te zijn wie je wilt zijn of degene die je echt bent. Dat komt om (1) …… niemand een ste (2) …… op je dr (3) …… . Toen ik n (4) …… het rei (5) …… in Nederland teru (6) ……, werd ik hele (7) …… gek. A (8) …… na vier dagen. A (9) …… ik z (10) …… dat men (11) …… zich opwo (12) …… over een honde (13) …… op de st (14) ……, werd ik pan (15) ……. Dan da (16) …… ik, mens, waar ma (17) …… je je dr (18) ……over? Ik ben ge (19) …… naar de psycholoog ges (20) ……, want ik trok dat echt niet.

Antwoorden:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Tekst 3:

Openlijke narcisten zijn mensen met een opgeblazen gevoel over zichzelf. Ze ei (1) ……. vaak ande (2) ……. aandacht o (3) ……. en ko (4) ……. charmant ov (5) ……., ond (6) ……. het feit d (7) ……. ze wei (8) ……. besef he (9) ……. van de beho (10) ……. van anderen. Verb (11) ……. narcisten zijn weli (12) ……. net z (13) ……. hevig met zichzelf be (14) ……. en ev (15) ……. arrogant a (16) ……. openlijke narcisten, ma (17) ……. ze do (18) ……. dit o (19) ……. een subti (20) ……. manier.

Antwoorden:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Tekst 4:

Het internationaal perscentrum Nieuwspoort discussieert weer eens over de code. De Haagse soci (1) …… waar h (2) …… journaille en de poli (3) …… in een onged (4) ……samenzijn verk (5) ……, hanteert si (6) …… jaar en d (7) …… de ongesc (8) …… regel d (9) …… wat er t (10) …… plekke gez (11) …… wordt ni (12) …… naar bui (13) …… mag ko (14) …… Alt (15) ……, niet her (16) …… mag worden tot de betre (17) …… persoon en pla (18) …… Voorzitter van het bestuur Max de Bok maa (19) …… onlangs pla (20) …… voor Casper Becx, maar het beleid bleef ongewijzigd.

Antwoorden:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Tekst 5:

Prins Claus was een intellectuele gentleman. Voor zijn echtgenote koningin werd had hij een serieuze baan in de ontwikkelingshulp. Na 1980 voe (1) …… hij zich ste (2) …… meer een orna (3) …… van de tr (4) …… Hij raa (5) …… depressief, ge (6) ……, maar we (7) …… nooit meer de ou (8) …… De la (9) …… van een onve (10) …… bestaan a (11) …… prins-gemaal le (12) …… op Claus een gr (13) …… druk, g (14) …… hem het gev (15) …… een ha (16) …… marionet te zijn, een man die acht (17) …… veel lie (18) ……een zelfst (19) …… positie had gehad dan een afge (20) ……

Antwoorden:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Appendix C: C-tests
II: English C-test for monolingual and bilingual subjects

Text 1:

We all live with other people's expectations of us. These are a refl (1)…… of th (2) …… trying to under (3) …… us; th (4) …… are predict (5) …… of wh (6) …… they th (7) …… we will think; d (8) …… and feel. Gene (9) ……, we acc (10) …… the sta (11) …… quo, but these expec (12) …… can be ha (13) …… when they co (14) …… from our fami (15) …… and can be diff (17) …… to ign (18) …… , especially wh (19) …… they come from our par (20) ……

Answers:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Text 2:

Maybe words ***can*** hurt you as much as sticks and stones: Romantic rejection, at least, causes physical pain, according to a new study of brain activity. Past stu (1) .....have sh (2) .....that simu (3) .....social rejection m (4) .....be conn (5) .....to a net (6) .....of brain reg (7) .....that proc (8) .....the mea (9) .....of pain b (10) .....not t (11) .....sensory exper (12) .....itself. Now MRI brain sc (13) .....of peo (14) .....jilted i (15) .....real li (16) .....show activ (17) .....in brain ar (18) .....that a (19) .....actually ti (20) .....to the feeling of pain.

Answers:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Text 3:

[from a US newspaper] The decision to ban soft drinks from elementary and junior high school vending machines is a step in the right direction to help children make better choices when it comes to what they eat and drink. Childhood obe (1)……has bec (2) …… a ser (3) …… problem in th (4) …… country a (5) …… children cons (6) …… more sugar-based fo (7) …… and sp (8) …… less ti (9) …… getting the nece (10) …… exercise. Many par (11) …… have quest (12) …… schools' deci (13) …… to al (14) …… vending machines which disp (15) …… candy and so (16) …… drinks. Many schools, th (17) ……, have co (18) …… to re (19) …… on the mo (20) …… these machines generate through agreements with the companies which make soft drinks and junk food.

Answers:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Text 4:

The incas never used the wheel in any practical manner. Its u (1) .....in to (2) .....demonstrates th (3) .....the prin (4) .....was well-known t (5) .....them, alth (6) .....it w (7) .....not app (8) .....in th (9) .....engineering. The la (10) .....of str (11) .....draft ani (12) ....., as well as st (13) .....terrain a (14) .....dense veget (15) .....issues, m (16) .....have rend (17) .....the wheel impra (18) ..... How th (19) .....placed the enor (20) .....blocks of stones remains a mystery, although the general belief is that they used hundreds of men to push the stones up inclined planes.

Answers:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Text 5:

Global extinction rates may have been overestimated by as much as 160 percent, according to a new analysis. In rec (1) .....decades nume (2) .....studies ha (3) .....predicted th (4) .....habitat destr (5) .....will do (6) .....some 20 to 50 percent o (7) .....Earth's spe (8) .....within 500 ye (9) .....

It's tr (10) .....that ma (11) .....species are st (12) .....dying o (13)....., but the dec (14) .....is happ (15) .....at a slo (16) .....pace than gene (17) .....feared. The b (18) .....news is that sur (19) .....extinctions dri (20) .....by habitat loss remain the critical conversation problem of the 21st century.

Answers:

1:	**11:**
2:	**12:**
3:	**13:**
4:	**14:**
5:	**15:**
6:	**16:**
7:	**17:**
8:	**18:**
9:	**19:**
10:	**20:**

Instruction for subjects

U krijgt in dit taakje vijf korte teksten voorgelegd over verschillende onderwerpen, maar de teksten zijn niet compleet. Na de eerste volledige zin zijn er gaten in de tekst gevallen waarbij steeds delen van woorden zijn weggelaten. Het is aan u om deze woorden aan te vullen en uw oplossingen in te vullen onder de tekst. Er zijn vaak meerdere oplossingen. De teksten lopen op in moeilijkheidsgraad en het is dus goed mogelijk dat u een of meerdere woorden niet aan kunt vullen. Dat is geen probleem. In dat geval laat u dat vakje open. Het is belangrijk dat u niet meer dan 5 minuten per tekst gebruikt om de woorden aan te vullen. Alles wat u na die tijd niet in heeft kunnen vullen laat u open en daarna gaat u door naar de volgende tekst. Voor het onderzoek is het erg belangrijk dat u zich aan de maximale tijd van 5 minuten per tekst houdt. Ook is het van het grootste belang voor de validiteit van dit onderzoek dat u dit taakje zelf uitvoert, zonder hulp van anderen (zelfs niet van uw partner als deze ook mee doet aan het onderzoek). Hartelijk dank voor uw medewerking.

In this task you will be presented with five short texts, about different topics. However, the texts are incomplete: after the first intact sentence, gaps have been created in the texts whereby parts of words have been deleted. Your task is to complete the words and to fill in your answers below each text. It is often the case that several options are possible. The texts increase in difficulty level and so it may well be that you cannot complete one or several items. That is fine. You may then simply leave that item blank and proceed with the next one. It is important that you do not take more than 5 minutes to complete each text. Any items you are unable to fill in after that time will have to be left blank and you are asked to proceed to the next text. It is crucial for this study that you indeed stick rigidly to this time limit of 5 minutes per text. In addition, it is essential for the validity of this study that you do this task on your own, without the help of others (not even from your partner in cases where your partner participates in this study as well). I appreciate your help in this.

Appendix D: English Grammaticality Judgment Task

Is this sentence correct or incorrect?

In this task you will be given sentences that you are asked to judge on their grammaticality. You do not need to pay attention to spelling or punctuation. It is also important that you are not distracted by the content of the sentences: you do not have to evaluate what is said, but how this is formulated. You have 5 options to select in your evaluation of each sentence: entirely correct, largely correct, neutral, largely incorrect, entirely incorrect. In those cases where you think a sentence is either largely or entirely incorrect, please underline those parts of the sentence you think are wrong. You do not have to correct anything. This is an untimed task, but because your initial hunch is usually the best one, we would like to ask you to give your judgment immediately after reading each sentence. It is also of the utmost importance that you complete this task on your own, without the help of others, especially because it is your language intuitions in which we are interested.

Here's an example to start with:

My sister <u>weren't</u> born in 1980

entirely incorrect	largely incorrect	neutral	largely correct	entirely correct
X	○	○	○	○

Now please judge the following sentences:

1. Only yesterday did Dan book his flight.

entirely incorrect	largely incorrect	neutral	largely correct	entirely correct
○	○	○	○	○

2. I already know now that half of my students will have not read the book.

entirely incorrect	largely incorrect	neutra	largely correct	entirely correct
○	○	○	○	○

3. I make it a habit to wash myself every day.

entirely incorrect	largely incorrect	neutral	largely correct	entirely correct
○	○	○	○	○

4. The baby was very much wanted by his parents.

entirely incorrect largely incorrect neutra largely correct entirely correct

5. A number of children, who all came from privileged backgrounds, which is very common in this part of town, because so many diplomats live here, had started a fundraising event.

entirely incorrect largely incorrect neutral largely correct entirely correct

6. Before the hackers can do anything to the company's computer files, the code be cracked.

entirely incorrect largely incorrect neutral largely correct entirely correct

7. You should not have told her your secret, because she can never keep anything to herself.

entirely incorrect largely incorrect neutral largely correct entirely correct

8. Eats your child better now than in the past?

entirely incorrect largely incorrect neutral largely correct entirely correct

9. I don't think it's necessary to pack all those suitcases: you're only going away for a weekend.

entirely incorrect largely incorrect neutral largely correct entirely correct

10. It was not until he was an old man when one day his son, who had moved abroad, told him that he loved him.

entirely incorrect largely incorrect neutral largely correct entirely correct

11. You should prepare yourself for what's coming.

entirely incorrect largely incorrect neutral largely correct entirely correct

12. I got offered clams during the fancy reception.

entirely incorrect largely incorrect neutral largely correct entirely correct

13. No matter how many times I've told her, my mother still doesn't know that I not like strawberries.

entirely incorrect largely incorrect neutral largely correct entirely correct

14. Does Sophie sometimes spend hours in the bathroom?

entirely incorrect largely incorrect neutra largely correct entirely correct

15. Many of the letters that she wrote to her husband that had survived all those years in boxes never reached him.

entirely incorrect largely incorrect neutral largely correct entirely correct

16. The students were given one answer sheet each at the start of the exam.

entirely incorrect largely incorrect neutral largely correct entirely correct

17. I didn’t see no one I knew on my way home.

entirely incorrect largely incorrect neutral largely correct entirely correct

○ ○ ○ ○ ○

18. Before going to bed, Tom sometimes likes to read a book. [WORD ORDER]

entirely incorrect largely incorrect neutral largely correct entirely correct

○ ○ ○ ○ ○

19. Until a few years ago, many of the student dorms that students that attended Monash University lived in were on the other side of Melbourne.

entirely incorrect largely incorrect neutral largely correct entirely correct

○ ○ ○ ○ ○

20. I can amuse me very well on my own.

entirely incorrect largely incorrect neutral largely correct entirely correct

○ ○ ○ ○ ○

21. The children are washed by their mother.

entirely incorrect largely incorrect neutral largely correct entirely correct

○ ○ ○ ○ ○

22. When you compare this author’s earlier and later work in which very different themes are addressed, you’re beginning to wonder whether the novels are actually written by one and the same person.

entirely incorrect largely incorrect neutral largely correct entirely correct

○ ○ ○ ○ ○

23. As usual came James too late.

entirely incorrect largely incorrect neutral largely correct entirely correct

24. When I last visited their place, there was nothing to suggest that they were moving out in less than two weeks.

entirely incorrect largely incorrect neutra largely correct entirely correct

25. For as long as I've come to this restaurant, there's always been a homeless person with a box beside himself.

entirely incorrect largely incorrect neutral largely correct entirely correct

26. He walked into the meeting thinking that if what the board said did not impress him he would quit right there and then.

entirely incorrect largely incorrect neutral largely correct entirely correct

27. He is searched by the police.

entirely incorrect largely incorrect neutral largely correct entirely correct

28. Only by working extremely hard could he satisfy himself.

entirely incorrect largely incorrect neutral largely correct entirely correct

29. A book reads she sometimes, but she more often reads magazines.

entirely incorrect largely incorrect neutral largely correct entirely correct

30. Although I offered to do it, my father insisted on washing himself in spite of his injuries.

entirely incorrect	largely incorrect	neutral	largely correct	entirely correct
○	○	○	○	○

Instruction for subjects

In this task you will be given sentences that you are asked to judge on their grammaticality. You do not need to pay attention to spelling or punctuation. It is also important that you are not distracted by the content of the sentences: you do not have to evaluate what is said, but how this is formulated. You have 5 options to select in your evaluation of each sentence: entirely correct, largely correct, neutral, largely incorrect, entirely incorrect. In those cases where you think a sentence is either largely or entirely incorrect, please underline those parts of the sentence you think are wrong. You do not have to correct anything. This is an untimed task, but because your initial hunch is usually the best one, we would like to ask you to give your judgment immediately after reading each sentence. It is also of the utmost importance that you complete this task on your own, without the help of others, especially because it is your language intuitions in which we are interested.