ABSTRACT

Title of Document:	AGE DIFFERENCES AND COGNITIVE APTITUDES FOR IMPLICIT AND EXPLICIT LEARNING IN ULTIMATE SECOND LANGUAGE ATTAINMENT
	Gisela Granena, Ph.D., 2012
Directed By:	Dr. Michael H. Long, Second Language Acquisition

Very high-level, functional ability in foreign languages is increasingly important in many walks of life. It is also very rare, and likely requires an early start and/or a special aptitude. This study investigated the extent to which aptitude for explicit learning, defined as "analytic ability" and aptitude for implicit learning, defined as "sequence learning ability," are differentially important for long-term L2 achievement in an immersion setting.

A group of 20 native speaker (NS) controls and 100 Chinese-Spanish bilinguals with ages of onset 3-6 ($n = 50$) and > 16 ($n = 50$) participated in the study. Early L2 learners use the same language learning mechanisms as NSs (but still differ in ultimate success), whereas late L2 learners have been claimed to be fundamentally different from NSs in terms of learning mechanisms (and also differ in ultimate success). A set of six L2 attainment measures reflecting a continuum from automatic to controlled use of language knowledge was administered, as well as a battery of six

cognitive tests (four language aptitude subtests, a general intelligence test, and a probabilistic serial reaction time task).

Results confirmed the predicted distribution of cognitive abilities into two main types of aptitudes, interpreted as implicit and explicit. Participants could be high in one, high in both, or low in both. Results further revealed that early and late L2 learners with high aptitude for explicit learning outperformed individuals with low aptitude on tasks that allow controlled use of language knowledge. On these tasks, aptitude for implicit learning also had an effect, but among early L2 learners only. In addition, early and late L2 learners with high aptitude for implicit learning showed greater sensitivity towards agreement violations on the language task at the most implicit end of the continuum. Finally, general intelligence only played a role in late L2 learners' attainment on tasks that allow controlled use of knowledge.

The study concluded that 1) cognitive aptitudes play a role in both early and late L2 learners, 2) different types of cognitive aptitudes have differential effects on L2 outcomes, and 3) individual differences in implicit learning ability are related to L2 attainment in adults.

AGE DIFFERENCES AND COGNITIVE APTITUDES FOR IMPLICIT AND
EXPLICIT LEARNING IN ULTIMATE SECOND LANGUAGE ATTAINMENT

By

Gisela Granena

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Michael H. Long, Chair
Professor Robert M. DeKeyser
Professor Catherine J. Doughty
Professor Steven J. Ross
Professor Jeff MacSwan

Acknowledgements

I would especially like to thank Manne Bylund for all the good times and for his timely support during data collection, Niclas Abrahamsson for supervising my work, Jared Linck for his suggestion of including a probabilistic version of the serial reaction time task, and Luis Jiménez (Universidade de Santiago de Compostela) for his generous help showing me how to create probabilistic sequences.

There are many people who helped me during the intense data-collection period in Madrid. A special thanks to Mª Luisa García Bermejo for making my time in Madrid so enjoyable, to the Education Department at the Universidad Complutense for letting me use their facilities, and to the Chinese community in Madrid for making my participant recruitment and data collection so easy and such fun.

Thank you to all my colleagues and friends at the University of Barcelona, and very especially, to Imma Miralpeix and Natalia Fullana, for all the time we shared as research assistants on the BAF Project, for your friendship, and for always providing me with help when I needed it.

The biggest thanks of all go to my family, to Santi, my aunt Paz, and, very especially, to my dad, for always encouraging me to study abroad, see the world, and think critically.

And finally, to Yucel. Thanks for always being there, despite the distance, and for all the inspiring discussions on SLA that we have had over the years and that started one day in Philadelphia.

This dissertation is yours as much as mine. Thank you all.

Table of Contents

List of Tables

List of Figures

Chapter 1: Review of the Literature

1.1 The Concept of Language Aptitude

Language aptitude is conceptualized as a combination of cognitive and perceptual abilities that are advantageous in second language acquisition (SLA) (Carroll, 1981; Doughty et al., 2007). Carroll (1993) referred to this combination of abilities as "aptitudes" (p. 675) and claimed that they were partly innate, fairly stable and relatively enduring traits. Although experts and laypeople alike would agree on the generic notion of aptitude as a special talent for language, the theoretical construct behind this popular notion has remained somewhat elusive in the SLA field. While there is agreement that language aptitude involves different cognitive abilities, it has been conceptualized in a variety of ways in SLA, each of them with different implications at the measurement level.

One line of research has linked the ability to learn a second language (L2) to first-language (L1) learning skills (Sparks, 1995; Sparks & Ganschow, 1991; Sparks et al., 1995). According to this theory, successful L2 learners have significantly stronger L1 literacy skills (e.g., phonological/orthographic processing, word recognition/decoding). Carroll (1973) also speculated that aptitude could be a residue of L1 learning ability and that rate of L1 acquisition was related to aptitude for L2 learning. Skehan (1990) provided some evidence in this respect in a follow-up study to the Bristol Language Project (Wells, 1985), where he found significant correlations between aptitude measures and L1 indices based on spontaneous speech samples (syntax and complexity of language use).

Other conceptualizations of aptitude consider working memory capacity, responsible for the simultaneous processing and storage of information, as a central component of the construct (Miyake & Friedman, 1998; Sawyer & Ranta, 2001). Although one of the earliest aptitude test batteries, the Modern Language Aptitude Test (MLAT) (Carroll & Sapon, 1959), included memory measures, these were based on theories of memory that preceded these later developments in cognitive psychology, which operationalized working memory as attentional capacity and control. Evidence for this position comes from studies such as Harrington and Sawyer (1992) and Robinson (2002), which have reported correlations between working memory and L2 performance.

Finally, more recent conceptions view language aptitude in a situated manner and distinguish different clusters of aptitude subcomponents according to relevant factors, such as type of learning task and acquisition stage (Robinson, 2001, 2002; Skehan, 1998, 2002). This conceptualization of aptitude results in individually unique language aptitude profiles. L2 learners may have high ability in one aptitude component or complex, but low ability in others. This characterization of aptitude is in line with recent research on aptitude-treatment interaction (ATI), since different ability profiles (e.g., strong memory but weak analytic skills) can be investigated in relation to a given type of instructional treatment or level of L2 attainment.

1.2 Language Aptitude and Instructed SLA

In instructed SLA contexts, aptitude is considered a good predictor of rate, or speed, of L2 learning under intensive conditions (Carroll, 1973). All other things being equal, a high-aptitude L2 learner will be faster to learn and enjoy higher overall

2

foreign language achievement under a variety of instructional approaches. Evidence for this claim can be found in both non-experimental and experimental research. In a survey study, Ehrman and Oxford (1995) reported a strong correlation between aptitude measures and overall learning success, despite the communicative changes in teaching methodology taking place at the time. Harley and Hart (1997) also found that aptitude was related to performance on a variety of L2 measures in an immersion learning program. Research in the laboratory has yielded similar findings, suggesting that aptitude positively affects language learning under a variety of conditions of exposure. De Graaff (1997), Robinson (1997), and Williams (1999) all showed that differences in aptitude, as measured by subtests of the MLAT, resulted in learning differences in implicit and explicit learning conditions. Only learning under the incidental, meaning-focused exposure condition in Robinson (1997) was found to be independent of language aptitude, a result that was replicated by Robinson (2002).

In addition to research showing that aptitude predicts L2 learning in general, and in line with Skehan's (1998) and Robinson's (2002) notion of aptitude profiles, there is some evidence of ATI revealing that different aptitude components may play different roles, depending on instructional treatment (e.g., Erlam, 2005; Sheen, 2007; Wesche, 1981). Sheen (2007), for example, showed that aptitude, operationalized as analytic ability, was more strongly related to achievement with metalinguistic than direct written feedback. An important implication of ATI findings such as Sheen's (2007) is that the predictive power of aptitude in SLA may have to be qualified and investigated in relation to a variety of factors (e.g., instructional treatments,

acquisition stages, aspects of language, and L2 learning environments) in order to fully understand under what type of conditions aptitude is a relevant construct.

1.3 Language Aptitude and Naturalistic SLA

While aptitude determines learning rate in instructed settings (as reflected in short-term differences in language use or performance[1]), the general claim in naturalistic SLA has been that aptitude is related to variation in ultimate level of attainment (i.e., long-term differences in acquisition). Skehan (1989), in fact, argued that aptitude could be even more relevant in naturalistic than instructed learning contexts because of the greater amount of input that the learner has to process and the pressure to discover regularities and make generalizations merely from L2 exposure. However, to date, research on language aptitude has focused primarily on instructed SLA and learning rate and very rarely on naturalistic SLA and ultimate attainment. Given that there is no empirical evidence concerning the extent to which we can generalize findings from one context into the other, both types of research are needed in order to assess fully the predictive power of aptitude in SLA. To date, the few studies that have investigated aptitude in a naturalistic environment have been conducted in the context of tests of the existence of a critical period for language acquisition (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser et al., 2010; Granena & Long, 2010; Harley & Hart, 2002[2]). According to the Critical

---

[1] Long (2005:291) makes a distinction between short-term differences in performance and long-term differences in capacity for acquisition. He argues that these concepts are often confused in studies of rate and ultimate attainment, and that long-term differences in capacity for acquisition are more important for a theory of SLA.

[2] A study that is often cited as evidence of the relationship between aptitude and L2 learning in an informal setting is Reves (1983) (see Skehan, 1998; Abrahamsson & Hyltenstam, 2008). This study is an unpublished Ph.D. dissertation from the Hebrew University of Jerusalem in Israel. The study is not available online and has never been published in refereed or non-refereed journals. In addition, Sawyer

Period Hypothesis (CPH) (Lenneberg, 1967), there are biological/maturational constraints on L2 learning such that post-critical-period L2 learners cannot become nativelike L2 speakers.

Although no adult L2 learner has yet been shown to be entirely nativelike across language domains and tasks in a methodologically robust study (see Long, 2005 for a review), there is evidence that some adult learners can become "near-nativelike" (Hyltenstam & Abrahamsson, 2003) on a variety of linguistic phenomena (e.g., Abrahamsson & Hyltenstam, 2009; Ioup et al., 1994; Novoa et al., 1988). One of the factors considered capable of compensating for maturational constraints and explaining variability in ultimate L2 attainment is language aptitude. DeKeyser (2000) hypothesized that a high degree of language aptitude is a necessary condition for adult L2 learners to reach a level of ultimate attainment in morphosyntax comparable to that of child L2 learners, who attain nativelike command regardless of their language aptitude. DeKeyser (2000) operationalized language aptitude as verbal analytic ability, an ability that is gained by linguistic experience in one's native language, in foreign languages, or linguistics.

The rationale behind DeKeyser's (2000) position is Bley-Vroman's (1988, 1990) Fundamental Difference Hypothesis, according to which there is a qualitative difference between the learning mechanisms of child and adult L2 learners. DeKeyser (2000) argued that, while younger learners learn mostly implicitly, using domain-specific mechanisms, older learners learn mostly explicitly, using problem-solving or domain-general mechanisms, and, therefore, have to rely more on language aptitude.

and Ranta (2001) point out a methodological problem with the study, mainly that the learners were not acquiring Hebrew purely through naturalistic exposure, but had been simultaneously receiving instruction for 6-7 years.

Since individual differences in language aptitude account for variation in explicit language learning ability, and explicit language learning accounts for variation in adult learners' ultimate level of attainment, it follows that language aptitude should explain variation in adult learners' ultimate attainment.

While the numerous studies of language aptitude in instructed SLA have, in general, provided converging evidence of the predictive power of aptitude in L2 learning, the few that have investigated the role of aptitude in naturalistic contexts (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser et al., 2010; Granena & Long, 2010; Harley & Hart, 2002) have yielded mixed findings.

Harley and Hart's (2002) study is usually cited as evidence for the predictive validity of aptitude in naturalistic contexts, even though, due to the extremely short period of naturalistic exposure involved (i.e., three months), the results of the study can only speak to the role of aptitude in rate of L2 learning, not eventual success.[3] The study was based on Harley and Hart's (1997) work on aptitude as a predictor of L2 achievement in French immersion classrooms, which revealed significant positive correlations between memory (defined by the authors as memory for text) and L2 outcomes in early immersion learners and significant positive correlations between analytic ability and L2 outcomes in late immersion learners. Given that the different types of instruction early and late immersion learners were exposed to (i.e., holistic memory-based vs. language analysis) could have affected the results, Harley and Hart (2002) set out to look at the relationship between aptitude and L2 outcomes among

---

[3] In order to assess level of ultimate attainment, a minimum length of residence in the L2-speaking country should be established. In some studies (e.g., DeKeyser, 2000; Abrahamson & Hyltenstam, 2009), length of residence was at least 10 years. In others (e.g., Sorace, 1993; DeKeyser et al., 2010), it was 5 and 8 years. Oyama (1978) found no differences between a group of Italian L2 learners who had been in the U.S. for 5-11 years, and a group that had been for 12-18 years.

adolescent learners after a three-month stay abroad. In addition to two cognitive

measures, a measure of aptitude (operationalized as language analytic ability) and a

measure of memory (operationalized as memory for text), they also administered a

battery of language tests, out of which, after removing the effects of an outlier, only a

sentence-repetition task was found to be related to aptitude. Harley and Hart

concluded that aptitude was a factor related to success in a naturalistic context,

although not consistently so. In fact, the only two measures in the battery that were

administered as pretests, and, therefore, the only measures that could have provided

reliable evidence of the benefits of study abroad, turned out to be unrelated to

aptitude as posttests. Given this limitation of the study, and, given that participants

had been learning French in a classroom context for approximately seven years before

their period of study abroad, the authors cannot discard the possibility that the

significant correlations they found for the sentence-repetition task already existed

before the overseas stay.

DeKeyser (2000) administered an auditory grammaticality judgment test (GJT)

incorporating various elements of morphosyntax to 57 Hungarian speakers of L2

English. He found a significant correlation between GJT scores and language

aptitude, operationalized as verbal ability, among late arrivals ($r = .33$, $p < .05$), but a

non-significant correlation among early arrivals ($r = .07$, *ns*). Those participants that

were late arrivals and scored within the range of child arrivals, or came close, were all

high-aptitude participants. High aptitude was operationalized as being half a standard

deviation above the group's average score, which was 4.7 out of 20. The exception

was a participant who did not have high aptitude (i.e., he was below .46 standard

deviations from the average), but who, nevertheless, scored within the range of child arrivals. This participant scored 3 out 20 on the aptitude test and 186 out of 200 on the GJT. According to DeKeyser, the fact that he was a postdoctoral student in the natural sciences would indirectly suggest that he was of above-average analytic ability and that his aptitude score was not indicative of his true skills.

On the basis of these results, DeKeyser concluded that above-average analytic abilities are required to reach near-native levels in the L2. Long (2007) argued that the language test in DeKeyser (2000), a GJT, could have allowed use of metalinguistic abilities and, therefore, might be measuring the same ability as the aptitude test. Long (2007) further interpreted the lack of a correlation between aptitude and GJT scores among early arrivals as the result of the lack of variance in scores within that group. DeKeyser's GJT was administered by having participants listen to each sentence stimulus twice with a three-second interval between the two repetitions. There was also a six-second interval between sentence pairs. As a result, it was a test that did "not require participants to perform under time pressure" (DeKeyser, 2000: 515). A similar administration format, with slightly shorter intervals, was used by Johnson and Newport (1989) and Birdsong and Molis (2001), i.e., a one- to two-second pause between the first and second readings and a similar interval between sentence pairs. This testing format could have maximized reflection, conscious monitoring, and opportunities for the late arrivals to rely on explicit knowledge. Participants with higher aptitude might have been better able to analyze each stimulus by drawing on their explicit L2 knowledge.

The lack of a correlation among early arrivals could be explained by the narrow range of aptitude/GJT scores (possible floor effect in the aptitude test, as a result of the non-language independence of the test, and ceiling effect in the GJT). The aptitude test administered in the study was a Hungarian version of the Words-in-Sentences subtest in the MLAT. The test was, therefore, measuring L1 verbal analytic ability. However, when asked about their proficiency in Hungarian compared to English, only 22 of the 57 participants reported feeling more comfortable in Hungarian. One of these participants was in the younger group, and 21 were in the older group. This means that the younger group was more homogenous in terms of language dominance, while the older group was more heterogeneous (half of the late acquirers felt more comfortable in the L1 and half either in the L2 or equally in the L1 and L2). Such a distribution is not surprising, given that early acquirers' schooling took place in the L2 and that degree of L2 acquisition tends to correlate with degree of L1 attrition (Yukawa, 1997; Montrul, 2004; Hyltenstam et al., 2009). This could have biased the distribution of aptitude scores in the study by restricting the range of scores in the group of early acquirers. The distribution of high- and low-aptitude participants seems to support this interpretation. There were 15 participants who had an aptitude score half a standard deviation above the mean and who were, therefore, identified as high-aptitude individuals. Only two of these participants were early acquirers. Given that DeKeyser's (2000) study did not include any procedures to screen potential participants as near-native L2 speakers (unlike, for example, Abrahamsson & Hyltenstam, 2008), the fact that almost all high-aptitude participants

were in the group of late acquirers seems to be an artifact of the aptitude instrument used.

An investigation by DeKeyser, Alfi-Shabtay, and Ravid (2010) provided cross-linguistic evidence for the nature of age effects in two parallel studies that looked at the acquisition of English in the U.S. and the acquisition of Hebrew in Israel by native speakers (NSs) of Russian ($n = 76$ and $n = 64$, respectively). The findings for aptitude (operationalized as L1 verbal aptitude and measured by a test comparable to the verbal SAT) showed a significant correlation between ultimate attainment and aptitude for the adult learners, but not for the early learners, replicating the findings inDeKeyser (2000). Specifically, the significant correlation in the two parallel studies in DeKeyser et al. (2010) was found for the 18-40 age of acquisition range ($r = .44$, $p < .05$ and $r = .45$, $p < .01$), but not for the age of acquisition < 18 group ($r = .11$, ns, and $r = -.37$, ns), or > 40 group ($r = .33$, *ns*, and $r = .14$, *ns*). The language measure used in the study followed the same format as that in DeKeyser (2000). The test was untimed and sentences were presented auditorily, twice, with a three-second interval between them. Therefore, similarly to DeKeyser (2000), the testing format used could have influenced the results obtained regarding the relationship between aptitude and GJT scores among adult learners.

To address previous methodological gaps in critical-period studies, Abrahamsson and Hyltenstam (2009) used a multiple-task design covering various language subdomains, L2 knowledge and L2 processing, and perception, as well as production. Their study with L2 speakers of Swedish began with detailed linguistic scrutiny of apparent linguistic nativelikeness. The formal procedure to screen participants into

10

the study was as stringent as the instrumentation they used, and half of the study was devoted to selecting participants who identified themselves as nativelike, and who were also perceived to be nativelike by NS judges. Participants in the final sample were 31 childhood learners with ages of onset ≤ 11 and 10 adult learners with ages of onset ≥ 12. All the late learners were able to score within the NS range on some of the tasks, but, unlike some early learners, not across the whole range of tests employed.

When scores on the aptitude test were considered, Abrahamsson and Hyltenstam (2008) found that the four late learners who were able to score within NS range on the GJT were all above average in terms of aptitude, as measured by the Swansea LAT (Meara et al., 2003). The correlation between GJT scores and aptitude among late learners was moderately positive ($r = .53$), but not significant ($p = .094$), probably due to the small size of the group. The authors further observed that 72% of the early learners who performed within the NS range also had high aptitude. In fact, there was a significant positive correlation between GJT scores and aptitude in the early-learner group ($r = .70, p < .001$). On the basis of these results, Abrahamsson and Hyltenstam (2008) concluded that language aptitude can play a role not only in adult near-native SLA, but also in child SLA. This is a finding that runs contrary to DeKeyser's (2000) hypothesis that aptitude will not be a significant predictor among early learners. It also differs from the results of DeKeyser (2000) and DeKeyser et al. (2010), which showed no relation between proficiency and aptitude among early learners.

In a similar study, Granena and Long (2010) investigated the relationship between language aptitude, as measured by the LLAMA (Meara, 2005), a revised version of the LAT aptitude test used by Abrahamsson and Hyltenstam (2008), and ultimate

morphosyntactic attainment, as measured by an auditory GJT. The results showed no relationship between aptitude and GJT performance. Participants in this study were 65 Chinese speakers of Spanish L2 divided into three groups according to age of onset of L2 learning (≤ 6, 7-15, and ≥ 16). A univariate analysis of variance with the three age groups as a fixed factor and language aptitude as a covariate revealed that, while age group, keeping aptitude constant, was significantly related to GJT scores ($F(2,61) = 21.010$, $p < .001$), aptitude, keeping group constant, was not ($F(1,61) = .816$, $p = .370$). The same analysis also showed no group-by-covariate interaction ($F(2,59) = 1.225$, $p = .301$). Therefore, the effect of aptitude was shown to be comparable in the three groups.

To summarize, research to date regarding the relationship between aptitude and eventual L2 success at a group level in naturalistic SLA has produced mixed findings. While aptitude was not related to ultimate morphosyntactic attainment among early acquirers in studies by DeKeyser (2000) and DeKeyser et al. (2010), Abrahamsson and Hyltenstam (2008) found a relationship between aptitude and GJT performance among those participants that were first exposed to the L2 before age 12. Also, while DeKeyser (2000) and DeKeyser et al. (2010) found a relationship between aptitude and GJT performance among late acquirers, Granena and Long (2010) did not. Finally, research findings have been inconsistent when a battery of tests has been employed, as by Harley and Hart (2002). While DeKeyser (2000), DeKeyser et al. (2010), Abrahamsson and Hyltenstam (2008), and Granena and Long (2010) all investigated the same language domain, morphosyntax, and used the same type of L2 measure, a GJT, the conditions of test administration differed across the studies.

Participants in the studies by DeKeyser (2000) and DeKeyser et al. (2010)

listened to each sentence stimulus in an auditory GJT twice. Abrahamsson and

Hyltenstam (2008) combined the scores of two different GJT modalities, an auditory

(online) and a written (offline), with no time pressure. In the only study that did not

find a relationship between aptitude and ultimate attainment, the test was auditory

(online), and participants had to press a key as soon as they detected an error

(Granena & Long, 2010). Sentences were only played once and, when participants

pressed a key, the computer automatically moved on to the next sentence without a

pause. Therefore, the three studies that reported significant correlations between

language aptitude and morphosyntactic L2 attainment had in common the use of

language measures with offline features, whereas the only study that did not find a

relationship relied on a test with online features.

1.4 Language Aptitude and Controlled vs. Automatic Use of L2 Knowledge

The studies that have reported positive correlations between language aptitude

scores and ultimate morphosyntactic attainment (Abrahamsson & Hyltenstam, 2008;

DeKeyser, 2000; DeKeyser et al., 2010) have in common the operationalization of

language aptitude as verbal or language analytic ability, as well as the use of language

tests or conditions of test administration that allow monitoring and controlled use of

L2 knowledge. Therefore, the measures of language aptitude and ultimate attainment

employed in these studies could have been measuring the same underlying abilities.

This is, in fact, how Long (2007) and Paradis (2009) interpreted DeKeyser's (2000)

findings. Both suggested the possible role of participants' metalinguistic abilities in

affecting the results of the study. Long (2007) noted that aptitude tests and GJTs have

13

in common the fact that they allow use of metalinguistic abilities. Since, in part, they measure the same underlying abilities, "some positive association between the two sets of scores is to be expected" (p. 73). Similarly, Paradis (2009) pointed out that:

> "Very few of the 57 adult Hungarian-speaking immigrants in DeKeyser's (2000) study scored within the range of child immigrants on a grammaticality judgment task, and the few who did had high levels of analytic skills (suggesting that they probably used their metalinguistic knowledge)" (p. 124).

In an instructed language context, Roehr and Gánem (2009) showed that use of metalinguistic L2 knowledge was related to performance on the Words-in-Sentences MLAT subtest, a test that measures grammatical sensitivity, one of the aspects of language analytic ability (Skehan, 1989). The metalinguistic test administered by Roehr and Gánem (2009) included a section with ungrammatical sentences that participants had to correct and explain and a section modeled on the MLAT Words-in-Sentences subtest, which required participants to identify the grammatical role of highlighted parts of sentences. While the study found moderate and significant correlations between MLAT4 (Words-in-Sentences) and the two sections of the metalinguistic test ($r = .45$ and $r = .41$, respectively), correlations between L1/L2 working memory (as measured by reading span tests) and the metalinguistic test were weak and non-significant ($r = .19$ and $r = .13$), as were the correlations between metalinguistic test scores and the other MLAT subtests ($r = .15$, MLAT 1, $r = .11$, MLAT 2, $r = .29$, MLAT 3, and $r = .10$, MLAT 5). Roehr and Gánem (2009) concluded that only the more analytic component of aptitude was related to metalinguistic knowledge, not the more memory-based components.

In the study by Roehr and Gánem (2009), metalinguistic L2 knowledge was positively related to aptitude scores on the same aptitude subtest that correlated with GJT scores in DeKeyser (2000). Using a different language aptitude test and a fully crossed, within-subjects design, Granena (2011a, to appear) found that language aptitude moderated scores on an untimed written GJT, but not on an auditory GJT. Specifically, the study examined two features of GJTs that could have interacted with language aptitude test scores in previous naturalistic SLA studies: test modality (auditory vs. written) and item (sentence) complexity (syntactically simple vs. complex). Granena argued that aptitude, understood as analytic ability, could be related to L2 performance to the extent that the L2 measure includes features that allow L2 learners to make controlled use of L2 knowledge and monitor their performance. If high-aptitude learners approach language as a puzzle-solving task (Skehan, 1998), off-line tests with no time constraints, or metalinguistic tasks with an error correction component, could allow them additional opportunities to rely on their problem-solving and analytic abilities. Granena (2011a, to appear) also argued that language aptitude could interact with test performance to the extent that test items make L2 processing more demanding, given the strengths in the cognitive abilities, such as processing speed and working memory, that information processing draws on, and that high-aptitude individuals are claimed to have.

Participants in the study by Granena (2011a, to appear) were 30 L1 English-L2 Spanish bilinguals with an average length of residence in Spain of 22 years and a group of 15 NS controls. Participants completed an auditory GJT, an untimed written GJT with a correction component, and the LLAMA aptitude test (Meara, 2005). The

two speaker groups were comparable in terms of their language aptitude scores
($t(43)= .003$, $p = .998$). The average score was 50.67 (SD = 11.55) in the NS control
group, and 50.65 (SD = 15.14) in the L2-speaker group, out of a maximum possible
score of 100. Regarding sentence complexity, even though the two speaker groups
scored significantly higher on simple test items, aptitude did not moderate scores in
either group. Regarding modality, both groups scored significantly higher on the
written than on the auditory GJT, but aptitude only moderated difference scores in the
L2-speaker group. In other words, there was an interaction between language aptitude
as a covariate and test modality (auditory GJT vs. untimed written GJT) among L2
speakers, keeping all other factors constant (i.e., target structure, number of items,
sentence complexity) ($\lambda(1,28) = .751$, $p = .005$, $\eta_p^2 = .249$), but not among NSs
($\lambda(1,13) = .952$, $p = .433$, $\eta_p^2 = .048$).

Follow-up analyses with language aptitude as a group variable (high-aptitude = z-
scores > .5, mid-aptitude = -.5 < z-scores < .5, and low-aptitude = z-scores < -.5) and
Bonferroni-adjusted multiple comparisons further showed group differences in the
written GJT ($F(2,27) = 5.694$, $p = .009$, $\eta_p^2 = .297$), between high- and mid- aptitude
L2 speakers and low-aptitude L2 speakers ($p = .013$ and $p = .029$, respectively), but
not in the auditory GJT ($F(2,27) = 1.143$, $p = .334$, $\eta_p^2 = .078$). The correlations
between L2 speakers' aptitude scores and written GJT scores were .438 ($p = .016$)
and .447 ($p = .013$) for simple and complex test items, respectively, while the
corresponding correlations between aptitude scores and auditory GJT scores were
.162 ($p = .392$) and .173 ($p = .360$). No correlation in the NS group had a magnitude
greater than .10. When only ungrammatical items were considered, for which the

built-in error may be the most likely reason for rejection, the correlation between

language aptitude and written GJT scores in the L2-speaker group increased in

magnitude, while the correlation between aptitude and auditory GJT scores decreased,

.495 (p = .005) and .002 (p = .990).

These results were interpreted as indicating a positive association between

language aptitude, understood as analytic ability, and language tests that involve

predominantly controlled use of L2 knowledge by allowing participants time to

reflect on language correctness and language structure. The written GJT gave L2

speakers unlimited time to analyze test sentences and monitor their performance. The

need to provide a correct version of the ungrammatical sentences further encouraged

them to reflect consciously on linguistic structure and sentence correctness. Since

high-aptitude L2 speakers' performance improved significantly on the written test

(with respect to the auditory test), a possible explanation is that they were able to use

the same analytic, metalinguistic abilities that the aptitude test measured.

The auditory GJT, on the other hand, required online processing, which may

minimize monitoring. On this test, high-aptitude L2 speakers did not outperform low-

aptitude participants. In fact, the highest L2-speaker scorer on the auditory GJT was a

low-aptitude individual, an adult acquirer with a length of residence of 30 years who

had arrived in Spain when she was 19 and who reported having started learning

Spanish at university at the age of 18. This L2 speaker obtained an auditory score of

110 out of 128, and she was very close to the lowest scorer in the NS group, a

participant with a score of 113. This result could suggest that the type of aptitude that

the LLAMA test is hypothesized to predominantly measure, a type of aptitude that

relies on explicit cognitive processes such as analytic ability, is not necessary in near native-like attainment in the type of naturalistic context investigated. This finding does not negate the possibility that other cognitive aptitudes could be relevant and account for such high levels of attainment.

The results of Granena (2011a, to appear) run contrary to those studies that have reported a relationship between aptitude and adult L2 learners' ultimate level of attainment in a naturalistic setting (Abrahamsson & Hyltenstam, 2008; DeKeyser, 2000; DeKeyser et al., 2010). They also run contrary to DeKeyser's (2000) claim that only late acquirers with a high level of language aptitude will be able to score within the range of early acquirers. Results are, however, consistent with Granena and Long's (2010) findings and, together, provide converging evidence for the lack of a relationship between aptitude and ultimate level of attainment, as measured by an auditory GJT, from two different L1 populations (English and Chinese) learning the same L2 in a naturalistic context.

The fact that the results reported by Granena (2011a, to appear) and Granena and Long (2010) run contrary to those reported by Abrahamsson and Hyltenstam (2008), DeKeyser (2000), and DeKeyser et al. (2010) suggests that the relationship between language aptitude and adult L2 learners' ultimate level of attainment is more complex than claimed so far on the basis of studies where aptitude has been investigated in single-task designs. A possible explanation for the lack of converging findings is that, even though Abrahamsson and Hyltenstam (2008), DeKeyser (2000), and DeKeyser et al. (2010) all measured ultimate attainment with a GJT, they relied on formats or

conditions of test administration that allowed learners time to reflect on language correctness and make use of metalinguistic abilities.

A question that is relevant to this discussion is what type of GJT, if any, provides a more valid measure of ultimate level of attainment. The answer to this question can vary depending on how linguistic competence is defined. If defined as knowledge that is available in spontaneous language use (Ellis, 2005), online test conditions where performance takes place in real time should provide a more valid measure of language acquisition. From this perspective, an auditory GJT, preferably under timed conditions and involving a single presentation of test items, should be a more valid measure of linguistic competence than a written GJT under untimed conditions, since unpressured tasks encourage a high degree of awareness and maximize the opportunities to access metalinguistic knowledge. Still, both types of GJT require a focus on language forms, since judging the correctness of sentences necessarily entails this. Other language measures involving the ability to handle language within real-time constraints, but with a clearer primary focus on meaning than GJTs would probably be better indicators of L2 learners' linguistic competence.

1.5 Cognitive Aptitudes and Language Learning

Language aptitude was originally understood as a unidimensional construct (e.g., Carroll, 1973) and measured accordingly through a composite of different abilities. Several studies have followed this conceptualization of aptitude that involves across-the-board haves and have-nots by relying on single measures of aptitude conceived as a unitary trait (e.g., Abrahamsson & Hyltenstam, 2008; Bylund et al., 2010; Granena & Long, 2010; Granena, 2011a, to appear). More recent theorizing on aptitude has

called for a multifaceted view of the construct that can result in individually unique L2 aptitude profiles (e.g., Skehan, 1998, 2002; 2012) or "aptitude complexes" (Robinson, 2002). L2 learners may have high ability in one aptitude component or complex, but low ability in others. Different aptitudes, in turn, may moderate L2 learning differently depending on factors such as instructional treatment, L2 learning environment, and type of linguistic feature.

With this goal in mind, Granena (2011b, to appear) ran a Principal Components Analysis (PCA) on the scores of 73 participants on the four subtests of the LLAMA aptitude test (Meara, 2005): vocabulary learning, sound recognition, sound-symbol association, and grammar inferencing. The main research question was whether the LLAMA subtests measured a unitary trait, conceived as language aptitude, or multiple aptitude components. A first analysis was conducted with 63 L1 Chinese-L2 Spanish bilinguals and 10 L1 speakers of Spanish. PCA is an exploratory factor analytic technique that summarizes the interrelationships among a set of original variables in terms of a smaller set of orthogonal (i.e., uncorrelated, in an unrotated solution or a solution with Varimax rotation) or non-orthogonal (i.e., correlated, in a solution with Direct Oblimin rotation) principal components. Its purpose is to reduce data by creating natural sets of composite variables.

An unrotated PCA of the LLAMA resulted in a two-factor solution with loadings (eigenvalues) greater than 1.0. They accounted for 68.497% of the total variance. Three of the LLAMA subtests loaded on a first component ($\lambda = 1.711$) and one subtest on a second component ($\lambda = 1.029$). The three subtests that loaded on the first component with values greater than .3 were vocabulary learning ($\lambda = .711$), sound-

symbol association ($\lambda = .770$), and grammatical inferencing ($\lambda = .759$). The correlations between them were .310 ($p = .003$), between vocabulary learning and sound-symbol association, .317 ($p = .003$), between vocabulary learning and grammatical inferencing, and .414 ($p < .001$), between sound-symbol association and grammatical inferencing. The only subtest that loaded on the second component with a value greater than .3 was the sound recognition test ($\lambda = .944$). The correlations between sound recognition and the other three subtests (vocabulary learning, sound-symbol correspondence, and grammatical inferencing) were close to zero, .141 ($p = .113$), .069 ($p = .279$), and -.026 ($p = .412$), respectively.

PCA has several drawbacks. Extracted components tend to overestimate the patterns of relationships among sets of variables because the analysis does not separate out errors of measurement from shared variance. An alternative exploratory technique is Principal Axis Factoring (PAF). PAF extracts factors only from the variance that variables share in common (i.e., common variance) and not from total variance. Like PCA, the default option is to extract orthogonal factors such that the first factor accounts for the maximum amount of common variance in the data, while the second factor accounts for residual variance after having factored out the influence of the first factor, and so on. In PAF, the total amount of variance accounted for and the variance in the variables explained by each of the factors is lower than in PCA.

The PAF analysis run on the LLAMA also converged on a two-factor solution with vocabulary learning, sound-symbol correspondence, and grammar inferencing loading on a first component, and sound recognition loading on a second component.

As expected, the overall variance accounted for decreased to 35.115%. The amount of variance in the variables explained by each of the factors was also lower, but loadings remained greater than .4. Vocabulary learning, sound-symbol correspondence, and grammar inferencing loaded on a first component with eigenvalues of .516, .598, and .697, respectively. Sound recognition loaded on a second component with a value of .443. Removing the 10 L1 speakers of Spanish and running the analysis with orthogonal (Varimax) or non-orthogonal (Direct Oblimin) rotations made no changes to the reported two-factor structure.

To further confirm the results obtained, sample size was increased to 117 participants. The sample combined L1 Chinese-L2 Spanish bilinguals ($n = 63$), English L1-Spanish L2 bilinguals ($n = 29$), and NSs of Spanish ($n = 25$). The two-factor structure was maintained, but with a second factor that did not quite reach an eigenvalue of 1.0 (it ranged between .909 and .995, depending on the analysis). If only one factor was retained, the three subtests already identified (vocabulary learning, sound-symbol correspondence, and grammatical inferencing) contributed to it with loadings greater than .3, while the contribution of the sound recognition test was negligible and did not reach the recommended threshold of .3.

Regarding the interpretation of the two factors underlying the LLAMA aptitude test, they could be labeled as language analytic ability and phonological sequence learning ability. The analyses suggest that the vocabulary learning subtest is related to the sound-symbol correspondence and grammatical inferencing subtests, but not to sound recognition scores. Therefore, individuals with good phonological sequence learning abilities may not necessarily have good analytic skills, and vice-versa. The

22

three subtests loading on the component that was interpreted as analytic ability have in common the fact that they include a study phase in which participants are given time to work out relations in a dataset. Unlike the sound recognition subtest, therefore, the three subtests loading on analytic ability allow for strategy use and problem-solving techniques. Granena (2011b, to appear) concluded that the LLAMA is mostly a test measuring the analytic component of aptitude and, therefore, aptitude for explicit language learning. Further support for this interpretation comes from the largest loading in the component interpreted as "analytic ability" and which corresponded to grammatical inferencing (LLAMA F). In this subtest, test takers are given time to infer or induce the rules governing a set of language materials presented visually. Therefore, it could be argued that the test is primarily measuring (explicit) inductive language learning ability, an ability that, according to Skehan (1989), should not be separated from grammatical sensitivity. In fact, he reconceptualized inductive language learning ability and grammatical sensitivity as language analytic ability (Skehan, 1998).

While LLAMA F requires test-takers to work out the grammar of an unknown language by means of pictures and short written sentences, LLAMA D measures the ability to discriminate short stretches of spoken language by analogy. As pointed out by Meara (2005), LLAMA D "owes something to Speciale (Speciale, N. Ellis, & Bywater, 2004)" who "suggest that a key skill in language ability is your ability to recognize patterns, particularly patterns in spoken language" (p. 8). Speciale et al.'s (2004) work is based on a strand of cognitive psychology that investigates the implicit induction of phonological sequences (Saffran et al., 1996; Saffran, Johnson,

23

Aslin, & Newport, 1999; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997).

LLAMA D can be, thus, seen as an attempt to measure implicit induction learning ability.

Construct validity in psychological tests involves building a network where tests are related to constructs, constructs are related to other constructs, and finally constructs are related to observables (Cronbach & Meehl, 1955). Empirical evidence for the relationship between the three LLAMA subtests loading on the factor that was interpreted as analytic ability and explicit language learning can be found in work by Yilmaz (2010) and in a re-analysis of Granena (to appear). In his study investigating corrective feedback, Yilmaz (2010) reported moderate-to-strong and significant correlations between the LLAMA grammatical inferencing subtest and posttest scores in a group that received explicit correction, whereas the same correlation was weak and non-significant in a group that received recasts. The re-analysis of Granena (to appear) showed a stronger relationship between untimed written GJT scores and aptitude operationalized as analytic ability than between untimed written GJT scores and aptitude operationalized as the average of the four LLAMA subtests ($r = .58$ vs. $r = .46$). Analytic ability moderated the difference score between auditory and written GJT scores ($p = .008$) in the L2-learner group only. In addition, while the correlation between analytic ability and written GJT scores was .58 ($p = .001$), the corresponding correlation with sound recognition approached zero .034 ($p = .857$).

The two adopted cognitive aptitudes that the LLAMA test is hypothesized to measure (analytic learning ability and sequencing learning ability) are very different in nature. While language analytic ability is gained by linguistic experience in one's

native language, in foreign languages, or linguistics, sequence learning ability can be regarded as a core component of human intellectual skill. In fact, according to N. Ellis (1996), much of language acquisition is sequence learning. This type of learning involves the discovery of language structure by way of statistical properties of the input. In the L1, infants have been shown to be sensitive to sequential dependencies in language. Saffran, Newport, and Aslin, (1996) demonstrated that 8-month-old infants exposed to strings of nonsense syllables were able to detect the difference between three-syllable sequences that appeared as a unit in their learning set and sequences that appeared in random order.

Statistical learning involves implicit learning processes (Perruchet & Pacton, 2006). Little is known about individual differences in this type of learning. According to Reber's earliest work (e.g., Reber, 1989), individual differences in implicit learning should be minimal because implicit forms of learning reflect primitive cognitive systems. Reber (1993) predicted that implicit learning is intelligence quotient (IQ)-independent and age-invariant, a claim that was supported by Reber et al.'s (1991) study, in which no association was found between artificial grammar learning and IQ scores. On the other hand, N. Ellis (1996) argued that individuals differ in their sequencing ability. More recently, Woltz (2003) also claimed that individuals can be expected to differ in implicit cognitive processes, just as they differ on most cognitive measures. Evidence from recent studies seems to support this prediction. Misyak and Christiansen (2012) documented variability in statistical learning performance among adults, as well as correlations between individual differences in statistical learning and L1 abilities. Specifically, statistical learning scores predicted comprehension

accuracy in a self-paced reading task. The authors concluded that individual

differences in statistical learning skills, largely overlooked in previous research, may

be able to account for more language variance than other measures typically used in

individual differences research.

Woltz (2003) also provided evidence of individual differences in repetition and

semantic priming. He suggested that the general domain of implicit cognitive

processes (implicit memory, implicit learning, and procedural knowledge) could be a

fruitful area in which to investigate new aptitude constructs. He argued that exploring

individual differences in implicit learning could result in aptitude constructs that have

minimal overlap with existing ones.

Additional evidence suggesting systematic variation in statistical learning comes

from studies that have investigated Reber's predictions regarding the existence of a

disassociation between implicit/explicit learning and IQ. In a replication and

extension of Reber et al. (1991), Robinson (2002) looked at the relationship between

implicit artificial grammar learning and IQ, as measured by a short form of the

Wechsler Adult Intelligence Scale (WAIS-R), which included three subtests: block

design, vocabulary, and arithmetic. Findings showed a significant negative correlation

between implicit learning and IQ ($r = -.34$, $p < .05$), somehow contrary to Reber et

al.'s (1991) results. Participants with lower scores on the WAIS-R outperformed

those with higher IQ scores. However, the incidental learning condition in the same

study, which involved meaning-based processing, showed no relationship to IQ. This

result would not be contrary to Reber's predictions, since incidental learning shares

with implicit learning the fact that is "unintentional and uncontrolled" (Reber &

Allen, 2000, p. 238). Conversely, other studies have shown that Culture-Fair (non-verbal) IQ test scores were positively related to learning on miniature L2 learning tasks (Brooks, Kempe, & Sionov, 2006; Kempe & Brooks, 2008; Kempe, Brooks, & Kharkhurin, 2010).

A possible explanation to account for these mixed findings could be the actual type of learning that took place in the studies. For example, learning might have not been as implicit as expected in those studies that reported a relationship between IQ or working memory and learning outcomes. As pointed out by Misyak and Christiansen (2012), fluid intelligence has been shown to correlate with artificial grammar learning when participants have been instructed intentionally to search for patterns (e.g., Gebauer & Mackintosh, 2007), but not under more incidental learning conditions. Similarly, Robinson (2002) explained the negative correlation between IQ and implicit learning as indicating participants' adoption of an explicit code-breaking set towards the implicit learning task, a strategy that had more negative effects for those participants with higher IQ scores and who are better at explicit learning. Robinson's explanation was further supported by the correlation between IQ and the explicit learning condition in the study, which was positive and significant.

A clear rationale for the prediction of a lack of a relationship between implicit learning and general measures of intelligence or other cognitive measures that tap explicit processes (i.e., Reber's position) is offered by Woltz (2003). According to Woltz, implicit learning should be unaffected by differences in IQ as measured by currently available intelligence tests, since these measures have been biased towards explicit processes that are attention-driven (e.g., working memory). In support of his

27

argument, Woltz cites studies such as those by Engle, Tuholski, Laughlin, & Conway (1999), which found correlations between IQ and working memory measures (operation span with words, reading span, and counting span), as well as his own (Woltz, 1990, 1999), which showed low correlations between IQ and implicit memory measures of priming. Further support for this claim can be found in Tagarelli, Borges-Mota, & Rebuschat (forthcoming) study, in which working memory predicted learning under the rule-search condition, but not under the incidental condition. Woltz (2003) further argues that explicit, attention-driven memory processes are, to a certain extent, functionally and structurally independent from implicit cognitive processes, even if all can operate in concert in complex learning tasks. Explicit processes require effort and some level of awareness, while implicit processes are revealed in performance facilitation, often with lack of awareness of the original learning event.

With respect to the relationships among aptitude components and language acquisition, the following conclusions can be drawn from the studies reviewed above:

- Language aptitude measures, including those in studies of ultimate L2 attainment to date, have been weighted heavily in favor of explicit cognitive processes (i.e., language analysis) and have overlooked implicit cognitive processes

- There is emerging recognition that individual differences in implicit learning do exist (as even conceded in Reber & Allen, 2000)

- Different cognitive aptitudes measure underlying abilities that are not necessarily correlated with one another

- Different aptitudes relate differently to different measures/aspects of ultimate L2 attainment and language learning

- Research has looked at individual differences in implicit learning, but has overlooked any direct links between differences in implicit learning abilities and variation in language (L1/L2) abilities

1.6 Conclusion

Previous research on language aptitude and long-term L2 achievement has been biased towards measures of language aptitude and language attainment that rely on explicit and attention-driven memory processes. Aptitude measures have focused on analytic ability and explicit induction, while language measures of morphosyntactic attainment have engaged participants in explicit processing by focusing their attention on linguistic structure and language correctness and by relying on testing formats that allow time to think. A positive association between the two sets of measures, therefore, could be partly due to the fact that they measure the same abilities.

Implicit learning and processing and their potential relationship with L2 learning outcomes have been overlooked as an aptitude construct, despite increasing agreement that explicit and implicit learning are two relatively independent learning systems, each with their own sources of individual differences (Kaufman, DeYoung, Gray, Jiménez, Brown, & Mackintosh, 2010). Similarly, previous studies have failed to rely on language measures with a primary focus on meaning that avoid raising the participants' awareness towards the phenomenon investigated, despite agreement that language knowledge can be used in more automatic or controlled processing conditions.

An optimal learning context to start investigating potential relationships between aptitude for implicit learning and L2 outcomes is the naturalistic (i.e., immersion) context. Naturalistic L2 learning environments maximize the opportunities for implicit learning by providing massive input exposure in communicative situations. In this context, cognitive aptitudes for implicit learning could help L2 learners detect complex and noisy regularities.

Chapter 2: Purpose of the Study

This dissertation research was motivated by three main research gaps in the body of literature that has investigated the role of language aptitude in ultimate morphosyntactic attainment. First, the few studies that have looked at the relationship between language aptitude and morphosyntactic L2 attainment have relied on single-task designs, which provide no indication of variation between tasks and that, as a result, may have limited generalizability. Second, they have relied on a single type of L2 measure, the GJT, which is a measure that focuses on accuracy of grammaticality judgment and language structure, and have failed to look at more meaning-based online tasks. Finally, they have relied on language aptitude measures heavily biased in favor of explicit cognitive processes, which could have resulted in positive associations between aptitude and the type of L2 measure employed.

In order to address these gaps, the present study set out to investigate the relationship between different cognitive aptitudes for L2 learning, including general intelligence, and ultimate level of morphosyntactic attainment in early and late L2 learners as measured by tests hypothesized to allow controlled use of L2 knowledge and tests hypothesized to require automatic use of L2 knowledge. A major goal of this dissertation work was to investigate new aptitude constructs in the domain of implicit cognitive processes and their potential role in determining L2 attainment in immersion settings, a relationship that, to my knowledge, no study has addressed before.

It has been argued that analytic ability, memory, and phonetic sensitivity are the three most important components of language aptitude (Carroll, 1964; DeKeyser &

Koeth, forthcoming; Skehan, 1989, 1998). However, studies of ultimate

morphosyntactic attainment have mostly focused on language analytic ability, using

either measures of analytic ability, such as the Words in Sentences MLAT subtest

(e.g., DeKeyser, 2000), or a composite of cognitive abilities that have been heavily

weighted in favor of analytic ability (e.g., Abrahamsson & Hyltenstam, 2008; Bylund

et al., 2010; Granena & Long, 2010). Analytic ability, which, according to DeKeyser

and Koeth (2011), is closely related to verbal aptitude and even general intelligence,

was the cognitive aptitude that DeKeyser (2000) anticipated as a necessary condition

for adult learners to reach high levels of ultimate attainment, due to the explicit

learning mechanisms they are hypothesized by some to rely on. No study of ultimate

L2 attainment to date has investigated multiple cognitive aptitudes following

Skehan's (1989, 1998) and Robinson's (2002) calls for aptitude complexes and

Sternberg's (1985, 1990) notion of 'multiple intelligences'.

In addition to explicit language learning, individuals differ in other domains of

cognitive ability. Although little research exists, recent studies have suggested that

some of these domains may be relevant for implicit language learning (e.g., Woltz,

2003). Individual differences in implicit processes may be especially relevant in

accounting for variation in the learning outcomes of child L2 learners who have

started learning early enough to acquire language without awareness, but late enough

to be affected by age of onset. For example, studies such as by Granfeldt et al. (2007)

have shown that Swedish child L2 learners of French with ages of onset between 3;5

and 6;7 resembled adult learners of French in their use of features such as gender

agreement. Individual differences in implicit learning may also be relevant to account

for variation among adult learners, if, contrary to what has been claimed by some, they are still able to learn an L2 incidentally via unconscious associative mechanisms. Studying aptitudes for implicit language learning in immersion contexts where the L2 is the language of the environment should be more revealing than studying them in instructed contexts where input exposure is limited, since language proficiency develops over a long period of constant exposure to the L2, thereby maximizing the potential involvement of implicit learning processes.

Reconceptualizing existing cognitive-ability constructs as explicit- and implicit-language-learning aptitudes is a worthwhile endeavor, in order to investigate qualitative differences in learning processes between child and adult learners, as well as aptitude-treatment interactions (i.e., the relative contribution of different aptitudes in specific contexts). One difference between children and adults is that adults can learn aspects of the L2 through explicit reflection on linguistic structure. It is a matter of debate, however, whether this is how adults predominantly learn L2 grammar or whether they are still able to learn implicitly, even if this capacity is constrained, due to age. Following DeKeyser's (2000) argument, any relationships between cognitive aptitudes and learning outcomes can be potential evidence for learning processes. As DeKeyser (2000) predicted, cognitive aptitudes that are more likely to play a role in explicit language learning should be necessary for L2 learners to reach a high ultimate level of attainment, if they learn through predominantly explicit mechanisms that draw on domain-general problem-solving abilities. Conversely, even if DeKeyser (2000) did not make such a prediction, cognitive aptitudes that are more likely to play a role in implicit language learning should be necessary for L2 learners to reach a

high ultimate level of attainment, if they learn through predominantly implicit mechanisms.

Specifically, this dissertation research examined the extent to which aptitude for explicit language learning (operationalized as language analytic ability) and aptitude for implicit language learning (operationalized as sequence learning ability) moderate ultimate attainment, as measured by language tasks that involve predominantly controlled and predominantly automatic use of language knowledge. Granena (2011b, to appear) found that analytic ability (as measured by a composite of LLAMA B, E, and F) and phonological sequence learning ability (LLAMA D) were two relatively independent abilities. Granena (2011a, to appear) further showed that analytic ability moderated ultimate attainment only when measured with a task where language performance could be monitored. That study, however, had several limitations. The method included a single online measure that involved predominantly automatic use of L2 knowledge and a single offline measure that involved predominantly controlled use of L2 knowledge. In addition, the two measures were administered in two different modalities (aural and visual). Finally, participants in the study were all late L2 learners with ages of arrival ranging between 17 and 43. Therefore, the present study addressed these limitations by including a multiple-task design and child and adult L2 learners, as well as NS controls.

In order to sample early and late L2 learners, an operationalization of each of the groups was necessary. Unfortunately, there is no consensus in the literature regarding a cutoff point between the two. For morphosyntax, the claimed cutoff age ranges from as young as six years of age (e.g., Paradis, 2009) up to the mid teens (DeKeyser,

2000). There is more agreement, however, regarding the expected shape of the age of onset-ultimate attainment function that the CP predicts, which resembles a stretched "Z" (Johnson & Newport, 1989:79): A peak of enhanced sensitivity, followed by a decline in learning ability, and then by a leveling off marking the end of the offset phase of the CP. Age of onset is expected to predict ultimate attainment in the phase of decline. Before and after the decline, individual differences such as cognitive aptitudes become more relevant as potential factors that can account for the spread in proficiency (Dörnyei, 2005; Dörnyei & Skehan, 2003; R. Ellis, 2004).

Given that the main predictor of ultimate attainment in the phase of decline is age of onset (e.g., Johnson & Newport, 1989), and that the decline becomes clearly visible from around age 6 and expands for a period of roughly 10 to 15 years (DeKeyser et al., 2010), the present study focused on two distinct groups of L2 learners: early child learners who started learning the L2 between ages 3 and 6, and adult learners who started learning the L2 after age 16. By looking at the extremes of the age of acquisition range, an arbitrarily established cutoff point was avoided. Both groups of learners are considered sequential bilinguals and both differ from NSs and simultaneous bilinguals regarding the degree of variability in their linguistic attainment. While there is relative uniformity of learning rate and ultimate success in L1 acquisition, L2 learners show a high degree of variability, across individuals and within learners, even when they start learning an L2 as early as age 3. Such inter- and intra-individual variation could be moderated by cognitive aptitudes. However, learners who acquire the L2 between ages 3 and 6 are not fundamentally different from NSs or simultaneous bilinguals in terms of learning mechanisms, since it has

been claimed that, before age 6, SLA relies on implicit learning (and the younger, the better), while "after age 6 or 7, second language appropriation relies more and more on conscious learning, thus involving declarative memory" (Paradis, 2009:110). In addition, children first manifest metalinguistic behavior, evident when the child has conscious awareness of why a sentence is ungrammatical and can demonstrate this understanding, around age 5 or later (Karmiloff-Smith, 1979).

Unlike early childhood learners, teenagers and young adults may additionally rely on explicit, analytic problem-solving capacities to learn the L2. If the brain of a child L2 learner acquires an L2 much like it acquires the L1, but the brain of an adult learner relies on predominantly explicit mechanisms that draw on domain-general problem-solving abilities, there should be qualitative differences between the two populations, such that different cognitive aptitudes moderate the level of proficiency child and adult L2 learners attain. If, alternatively, fundamental differences in learning mechanisms emerge in early childhood and the FDH (Bley-Vroman, 1988, 1990) applies to child as well as to adult SLA, as argued by Meisel (2009, 2011), the same cognitive aptitudes should moderate ultimate attainment in both child and adult L2 learners.

Chapter 3: Research Questions and Hypotheses

Following DeKeyser's (2000) claim that relationships between individual differences in language aptitude and eventual learning outcomes potentially constitute evidence for differences in underlying learning processes, this study investigated the relationship between different cognitive aptitudes for L2 learning, including general intelligence, and long-term L2 achievement in early and late L2 learners. Six research questions, each of them including three measurable hypotheses, were addressed in this dissertation research. The research questions and hypotheses were the following (see Table 1 for a summary of predictions):

- Research Question 1: To what extent will *early* L2 learners' ultimate attainment on tasks that allow *controlled* use of L2 knowledge be moderated by their cognitive aptitudes?

  Hypothesis 1a. Aptitude for explicit language learning will not moderate early L2 learners' attainment on tasks that allow controlled use of L2 knowledge.

  Hypothesis 1b. Aptitude for implicit language learning will moderate early L2 learners' attainment on tasks that allow controlled use of L2 knowledge.

  Hypothesis 1c. General intelligence will not moderate early L2 learners' attainment on tasks that allow controlled use of L2 knowledge.

- Research Question 2: To what extent will *late* L2 learners' ultimate attainment on tasks that allow *controlled* use of L2 knowledge be moderated by their cognitive aptitudes?

  Hypothesis 2a. Aptitude for explicit language learning will moderate late L2

learners' attainment on tasks that allow controlled use of L2 knowledge.

Hypothesis 2b. Aptitude for implicit language learning will not moderate late L2 learners' attainment on tasks that allow controlled use of L2 knowledge.

Hypothesis 2c. General intelligence will moderate late L2 learners' attainment on tasks that allow controlled use of L2 knowledge.

- Research Question 3: To what extent will NSs' ultimate attainment on tasks that allow *controlled* use of L2 knowledge be moderated by their cognitive aptitudes?

  Hypothesis 3a. Aptitude for explicit language learning will not moderate NS controls' attainment on tasks that allow controlled use of L2 knowledge.

  Hypothesis 3b. Aptitude for implicit language learning will not moderate NS controls' attainment on tasks that allow controlled use of L2 knowledge.

  Hypothesis 3c. General intelligence will not moderate NS controls' attainment on tasks that allow controlled use of L2 knowledge.

- Research Question 4: To what extent will *early* L2 learners' ultimate attainment on tasks that require *automatic* use of L2 knowledge be moderated by their cognitive aptitudes?

  Hypothesis 4a. Aptitude for explicit language learning will not moderate early L2 learners' attainment on tasks that require automatic use of L2 knowledge.

  Hypothesis 4b. Aptitude for implicit language learning will moderate early L2 learners' attainment on tasks that require automatic use of L2 knowledge.

Hypothesis 4c. General intelligence will not moderate early learners'
attainment on tasks that require automatic use of L2 knowledge.

- Research Question 5: To what extent will *late* L2 learners' ultimate attainment
  on tasks that require *automatic* use of L2 knowledge be moderated by their
  cognitive aptitudes?

  Hypothesis 5a. Aptitude for explicit language learning will not moderate late
  L2 learners' attainment on tasks that require automatic use of L2 knowledge.

  Hypothesis 5b. Aptitude for implicit language learning will moderate late L2
  learners' attainment on tasks that require automatic use of L2 knowledge.

  Hypothesis 5c. General intelligence will not moderate late L2 learners'
  attainment on tasks that require automatic use of L2 knowledge.

- Research Question 6: To what extent will NSs' ultimate attainment on tasks
  that require *automatic* use of L2 knowledge be moderated by their cognitive
  aptitudes?

  Hypothesis 6a. Aptitude for explicit language learning will not moderate NS
  controls' attainment on tasks that allow automatic use of L2 knowledge.

  Hypothesis 6b. Aptitude for implicit language learning will not moderate NS
  controls' attainment on tasks that allow controlled automatic use of L2
  knowledge.

  Hypothesis 6c. General intelligence will not moderate NS controls' attainment
  on tasks that allow automatic use of L2 knowledge.

It was predicted that aptitudes that are more relevant for implicit language learning and processing would moderate L2 learners' attainment on tasks that require more automatic use of L2 knowledge. This prediction was made both for early L2 learners who are sequential bilinguals and for adult learners, since adults were still expected to be able to learn implicitly, but not for NSs, whose ultimate attainment is characterized by inter-individual homogeneity and, therefore, predicted to be independent of aptitude. Aptitude for implicit language learning should also moderate early L2 learners' attainment on tasks that require controlled use of L2 knowledge, since early L2 learners were expected to use the same type of knowledge regardless of language task. The nature of this knowledge is hypothesized to be implicit, like NSs' knowledge. Early L2 learners' ultimate attainment, however, is characterized by greater inter-individual variability than NSs and, therefore, was expected to be moderated by language aptitude.

On the other hand, aptitudes that are more relevant for explicit language learning were only expected to moderate adult L2 learners' attainment on tasks that allow controlled use of L2 knowledge. These tasks increase available test time and decrease processing demands; therefore, they provide an opportunity to utilize problem-solving and analytic skills. On these tasks, adult learners can rely on explicit L2 knowledge and compensate for their limited implicit competence. Adult learners with a higher aptitude for explicit language learning should do better as a result of their greater analytic, metalinguistic abilities.

Regarding general intelligence, a debated issue is whether it is related to or independent from language aptitude. Carroll (1981, 1993) argued that aptitude was a

specialized ability beyond general intelligence, whereas Pimsleur (1966) and Oller

and Perkins (1978) considered intelligence a central component of aptitude. Evidence

in support of an independent contribution of language aptitude is the fact that aptitude

correlates more strongly with L2 outcomes than intelligence (Skehan, 1998).

However, aptitude and intelligence "indeed have a significant degree of overlap"

(Skehan, 1998: 208). In this dissertation research, which made a distinction between

cognitive aptitudes for implicit and explicit learning, the same pattern of results was

predicted for general (i.e., fluid) intelligence as for aptitude for explicit language

learning. Therefore, intelligence was considered more relevant for explicit learning,

since it is closely related to analytic ability (DeKeyser & Koeth, forthcoming), and

largely unrelated, or at most weakly related, to performance on implicit learning tasks

(e.g., Gebauer & Mackintosh, 2007; Kaufman et al., 2010; Reber et al., 1991).

It has also been argued that, even though general fluid reasoning ability measures

tap both explicit (attention-driven) and implicit (procedural) cognitive processes,

conventional IQ measures are weighted in favor of explicit processes that require

central executive functioning (Woltz, 2003). This would explain that measures of

general intelligence are highly correlated with working memory measures, but have

low correlations with priming measures (Woltz, 1990, 1999) and with procedural skill

performance beyond the initial stages (Ackerman, 1987, 1988). Further, research on

artificial grammar learning has revealed that fluid intelligence correlates with learning

when participants are instructed to intentionally look for patterns in the training

materials (Gebauer & Mackintosh, 2007), but not under more incidental learning

conditions (Misyak & Christiansen, 2012). Finally, Robinson (2002) reported a

significant, but negative, correlation between IQ and implicit learning of an artificial

grammar, and no significant correlation between IQ and incidental learning involving

meaning-based processing.

Table 1. *Predictions Concerning the Relationship between Cognitive Aptitudes,*
General Intelligence, and Ultimate L2 Attainment

	Automatic L2 Use			Controlled L2 Use		
	Early AO	*Late AO*	*Control*	*Early AO*	*Late AO*	*Control*
Intelligence	No	No	No	No	Yes	No
Explicit Aptitude	No	No	No	No	Yes	No
Implicit Aptitude	Yes	Yes	No	Yes	No	No

Chapter 4: Methodology

4.1 Participants

Participants were 100 Chinese-Spanish bilinguals in Madrid (Spain) and 20 NSs of Spanish ($N = 120$), all of whom were at least 18 years of age at time of testing. The Chinese-Spanish bilingual participants had either immigrated to the country or been born in the country to immigrant parents. Half of them ($n = 50$) were early L2 learners (42% males and 58% females) with ages of onset ranging from 3 to 6. The other half ($n = 50$) were late L2 learners (34% males and 66% females) with ages of onset of 16 and older. Age of onset was operationalized as the beginning of a serious and sustained process of language acquisition as the result of migration or the commencement of a formal Spanish language program. Age of onset, therefore, could differ from age of physical arrival in the country. In this study, when age of onset and age of arrival did not overlap, formal instruction took place in adulthood, after age 16. Therefore, age of first exposure as a result of immersion in the L2-speaking country and age of first instruction still overlapped for the purposes of the current study, where adult L2 learners are defined as those with ages of onset of 16 and older.

Age of onset could also differ from age of physical arrival in the country in the case of early L2 learners, albeit for a different reason. The early L2 learners in the present study arrived in the country at an early age (i.e., from ages 3 to 6) or were born in Spain. In either case, these early L2 learners had been born to Chinese-speaking parents who had immigrated to the country as adults. They had not been born to parents who had themselves been born in Spain. As a result, even those early L2 learners who had been born in Spain had not been immersed in the L2 until a later

age, usually at age 3, in pre-school. Until that age, they were primarily exposed to Chinese and, therefore, can be considered sequential, not simultaneous, bilinguals.

Participants were recruited by advertising in Chinese-Spanish newspapers, by distributing fliers in cultural centers, embassies, and language schools, and by word of mouth in the community. To qualify for the study, participants had to: 1) have Chinese as mother tongue, 2) have lived in Spain for at least 5 years[4], and 3) have an educational level of no less than high school. Participants were informally screened into the study via a telephone interview.[5] A group of 20 NSs of Spanish (50% males and 50% females), born in Madrid and with no less than a high school diploma, served as controls.

All participants completed a detailed biographical questionnaire (see Appendix A). Table 2 summarizes the information regarding age at testing, age of onset, and length of residence for the participants.

---

[4] According to DeKeyser et al. (2010), length of residence "turns out to be unrelated to most dependent measures, provided that it is more than 5 years, and that the dependent measures index basic grammatical proficiency (not purisms, collocations, etc.)" (p. 416).

[5] The inclusion criterion was a score of at least four on a five-point scale that rated participants' degree of native-like pronunciation: 5 Native or near-native pronunciation. No foreign accent. 4 Generally good pronunciation but with occasional non-native sounds. Slight foreign accent. Pronunciation does not interfere with comprehensibility. 3 Frequent use of non-native sounds. Noticeable foreign accent. Pronunciation occasionally impedes comprehensibility. 2 Generally poor use of native-like sounds. Strong foreign accent. Pronunciation frequently impedes comprehensibility. 1 Very strong foreign accent. Definitely non-native. Participants rated with a three on pronunciation were also included in the study if their grammar use was native-like.

Table 2. *Participants' Information*

Group	Age at Testing		Age of Onset		Length of Residence	
	M	*Range*	*M*	*Range*	*M*	*Range*
Control	27.35	20-36				
n = 20	(5.18)					
Early AO	22.38	18-33	4.14	3-6	17.88	11-28
n = 50	(4.45)		(1.23)		(4.49)	
Late AO	29.46	21-50	20.84	16-30	8.42	5-20
n = 50	(6.38)		(4.14)		(3.14)	

Note. Standard deviations appear between parentheses.

Early and late L2 learners were significantly different in terms of age of onset ($t(98)$ = -27.331, $p < .001$) and length of residence ($t(98)$ = 12.207, $p < .001$). Early L2 learners' age of onset was 4 years on average, whereas late L2 learners' age of onset was 20 years on average. Regarding length of residence, the average was 17 years in the early AO group and 8 years in the late AO group. In the late L2 learner group, 12 participants had a length of residence lower than 10 years (between 5 and 10), whereas 38 participants had a length of residence higher than 10[6]. Regarding chronological age at time of testing, late L2 learners were in their late 20's (29 years on average), whereas early L2 learners were in their early 20's (22 years on average). The average age at testing in the NS group was 27. According to Scheffé posthoc tests, early L2 learners were significantly younger than both NSs ($p = .003$) and late

---

[6] Having a length of residence between 5 and 10 years or higher than 10 years in the late L2 learner group did not did not make any difference on any of the morphosyntactic measures in the study. All the comparisons yielded non-significant results with *p* values ranging between .147 and .937. This provides some support to DeKeyser et al.'s (2010) claim that length of residence, provided that it is more than 5 years, is unrelated to measures of grammatical proficiency.

L2 learners ($p < .001$), but NSs and late L2 learners were not significantly different (p = .346). Although the range of ages at testing in the late L2 learner group was 21-50, there were only two participants older than 40 (48 and 50 years old, respectively). The rest of the late L2 learners ($n = 48$) were younger than 40.

Early and late L2 learners also differed in terms of degree of identification with Spanish culture, Chinese literacy skills, Chinese proficiency level, and years of instruction of Spanish as a foreign language. Regarding their identification with Spanish culture, early L2 learners had an average of 3.72 ($SD = 0.70$) on a five-point Likert scale ranging from 1 (i.e., no identification – you do not feel Spanish) to 5 (i.e., total identification – you feel Spanish), whereas late L2 learners had an average of 3.14 ($SD = 0.64$) ($t(98) = 4.323$, $p < .001$). Early L2 learners had significantly lower Chinese literacy skills ($M = 1.46$, $SD = 0.50$) than late L2 learners ($M = 1.98$, $SD = 0.14$) on a two-point scale where 1 indicated oral skills and 2 indicated oral and written skills ($t(97) = -7.031$, $p < .001$). Early L2 learners' proficiency in Chinese on a five-point scale ranging between 1 (basic) and 5 (native-like) was also lower ($M = 3.16$, $SD = 1.45$) than late L2 learners' Chinese proficiency ($M = 4.90$, $SD = 0.51$) ($t(97) = -7.998$, $p < .001$). Finally, regarding years of instruction, late L2 learners had studied Spanish formally for an average of two-and-a-half years ($M = 2.45$, $SD = 1.82$), whereas early L2 learners had not taken any Spanish language courses ($M = 0.0$, $SD = 0.0$). The number of years of instruction ranged between zero and seven in the late AO group, and instruction had usually taken place in the learners' country of origin (China) before arrival in Spain. A total of 19 late L2 learners had received instruction for one year or less than a year ($n = 12$), always upon arrival in Spain, or

no instruction at all ($n = 7$), whereas six late L2 learners had taken between five ($n = 3$) and seven years ($n = 1$) of Spanish. The remaining 25 participants had received instruction for either two ($n = 8$), three ($n = 5$), or four ($n = 12$) years.

Early and late L2 learners did not differ regarding percentage of daily Chinese use ($t(96) = -1.500$, $p = .137$) or percentage of daily Spanish use ($t(96) = 1.713$, $p = .090$). Early L2 learners used 28.5% Chinese ($SD = 15.43$) and 69.80% Spanish ($SD = 15.52$) daily on average, and late L2 learners 34.76% Chinese ($SD = 24.96$) and 62.73% Spanish ($SD = 24.52$) daily on average.

4.2 Design of the Study

The study combined an ex-post-facto design with a repeated-measures experimental design. Groups were compared in four experimentally-manipulated test conditions: 1) A time-pressured visual GJT, 2) a time-pressured auditory GJT, 3) an unpressured auditory GJT, and 4) an unpressured visual GJT. The four tests were administered following a 4x4 balanced Latin square to control for order and carry-over effects (see Table 3). In a balanced Latin square, each condition appears only once in a given ordinal position and no two conditions are juxtaposed in the same order more than once.

Table 3. *Balanced Latin Square Design*

Order 1	1	2	4	3
Order 2	2	3	1	4
Order 3	3	4	2	1
Order 4	4	1	3	2

Overall, the same number of participants ($n = 30$) was randomly assigned to each of the four test orders. Within every group, the same number of participants was also assigned to each test order, as long as the group's sample size allowed that. Thus, in the control group, the same number of participants ($n = 5$) could be assigned to each test order. However, in the early and late AO groups, two test orders had 12 participants each, and the other two had 13 participants each.

In order to discount test ordering effects, there should be no interaction between the four order groups (between-subjects factor) and scores on the four test formats (within-subjects factor). A repeated-measures ANOVA with the four GJTs as the repeated factor and Test Order as the group factor indicated a significant multivariate effect for GJT ($F(3, 114) = 14.639$, $p < .001^7$, $\eta_p^2 = .076^8$, $\lambda = .720$), but no significant two-way interaction between the order in which tests were administered and GJT scores for the sample as a whole ($F(9, 360) = 1.271$, $p = .253$, $\eta_p^2 = .033$, $\lambda = .906$). The interaction was non-significant in each of the three participant groups, as well: controls ($F(9, 57) = 1.064$, $p = .413$, $\eta_p^2 = .181$, $\lambda = .548$), early L2 learners ($F(9, 147)$ $= 1.880$, $p = .063$, $\eta_p^2 = .114$, $\lambda = .695$), and late L2 learners ($F(9, 147) = .674$, $p = .732$, $\eta_p^2 = .044$, $\lambda = .875$).

This experimental design allows testing the variables of interest while keeping all other factors constant. However, it suffers from two limitations. First, it involves variants of just one method (GJTs) and, second, it involves a method that requires focusing participants' attention on language correctness. Therefore, two additional tasks at the extremes of the controlled/automatic use of language knowledge

---

[7] Alpha was set at 0.05 for all inferential tests in this study.
[8] For partial eta squared (η_p^2), a small effect size is $.01 \leq \eta_p^2 < .06$, medium is $.06 \leq \eta_p^2 < .14$, and large is $\eta_p^2 \geq .14$.

continuum were included in the design: A metalinguistic knowledge test and a word monitoring task. These two tasks are hypothesized to tap directly into controlled and automatic use of L2 knowledge, respectively. In a metalinguistic test, participants' attention is directly focused on linguistic structure, correctness and grammatical rules (i.e., explicit declarative facts about language). It requires language analysis rather than intuition about correctness. In a word monitoring task, participants' attention is not focused on the linguistic relationship of interest to the researcher. Participants monitor for a target word in a sentence and focus their attention on meaning comprehension, while the researcher measures sensitivity to grammatical violations.

4.3 Instruments

A battery of 12 tests was administered as part of the study. Six of the tests were language measures hypothesized to lie along a continuum of controlled to automatic use of L2 knowledge: four GJTs (timed visual, timed auditory, untimed visual, and untimed auditory), a metalinguistic knowledge test (at the controlled end of the L2 knowledge use continuum), and a word monitoring task (at the automatic end of the L2 knowledge use continuum). There were also six cognitive measures hypothesized to be aptitudes relevant for either implicit or explicit learning: four verbal language-independent aptitude subtests (the LLAMA aptitude test battery), a non-verbal measure of general intelligence (the GAMA general ability measure for adults), and a non-verbal measure of sequence learning (a probabilistic serial reaction time task).

4.3.1 Language Tests that Require Automatic Use of L2 Knowledge

Timed Auditory GJT (*k* = 60). The timed auditory GJT was a computer-delivered test with sentences presented aurally. Participants indicated whether each sentence was grammatical or ungrammatical by pressing a response button within a fixed time-limit. They were asked to press a key as soon as an error was detected in the sentence. Once participants pressed a key, the computer automatically moved on to the next sentence without a pause. Following R. Ellis (2005), the time-limit for each item was established on the basis of NSs' average response time in a pilot study (*n* = 10). Following R. Ellis, as well, an additional 20% of the time taken for each sentence was added to allow for the slower processing speed of L2 learners. The time allowed for judging each sentence in the timed auditory GJT ranged between 3408.72 milliseconds (3.4 seconds) to 10045.92 (10 seconds) (*M* = 5807.98, *SD* = 1000.76). In terms of target structure, NSs' longest response times were on aspectual contrasts (*M* = 5365.09, *SD* = 1156.64), followed by gender agreement (*M* = 5102.60, *SD* = 471.69), the passive (*M* = 4988.20, *SD* = 432.40), person agreement (*M* = 4892.22, *SD* = 608.58), number agreement (*M* = 4691.73, *SD* = 844.26), and the subjunctive (*M* = 4000.05, *SD* = 714.31).

Each item was scored dichotomously as correct/incorrect, and percentage accuracy scores were calculated for grammatical and ungrammatical items overall, as well as for grammatical and ungrammatical items separately. Percentage scores out of total number of attempts were used due to the relatively high proportion of missing data as a result of the speeded nature of the test (10.61% of total items).

The internal consistency of the test, according to Cronbach's alpha, which measures the rank-order stability of individuals' scores on different items of the test, was .92.

Timed Visual GJT ($k = 60$). The timed visual GJT was a computer-delivered test with sentences presented visually. Participants indicated whether each sentence was grammatical or ungrammatical by pressing a response button within a fixed time-limit. Once participants pressed a key, the computer automatically moved on to the next sentence without a pause. The time limit for each item was also established by adding 20% to NSs' average response time. The time allowed for judging each sentence in the timed auditory GJT ranged between 3590.23 milliseconds (3.5 seconds) to 8587.20 (8.5 seconds) ($M = 5804.37$, $SD = 993.40$). In terms of target structure, NSs' longest response times were again on aspectual contrasts ($M = 5289.30$, $SD = 931.76$), followed by gender agreement ($M = 4942.46$, $SD = 844.04$), the passive ($M = 4930.43$, $SD = 720.62$), number agreement ($M = 4742.05$, $SD = 1122.58$), the subjunctive ($M = 4595.84$, $SD = 625.48$), and person agreement ($M = 4521.77$, $SD = 553.95$).

Each item was scored dichotomously as correct/incorrect, and percentage accuracy scores were calculated for grammatical and ungrammatical items overall, as well as for grammatical and ungrammatical items separately. Percentage scores out of total number of attempts were used due to the relatively high proportion of missing data as a result of the speeded nature of the test (15.67% of total items).

The internal consistency of the test, according to Cronbach's alpha, was .89.

Word monitoring Task (*k* = 120). The word monitoring task was a computer-delivered test with sentences presented aurally, and words to monitor presented visually (i.e., cross-modal modality). Word monitoring is considered an implicit task in the sense that participants' attention is not directed towards the linguistic variable of interest. Participants monitor ongoing auditory language input for a prespecified target word that is presented visually on their computer screen, and press a button when they hear the target word. Target words occur immediately after the relevant target structure in each sentence. The onset of each target word triggers a timing device that is stopped when the participant presses one of the response keys. The reaction time is the duration between the onset of the target word and the time when a response is provided. The test also includes comprehension questions, in order to focus participants' attention on meaning. This dual-task paradigm, which involves simultaneously engaging participants in a second unrelated task while performing the experimental task (i.e., word monitoring), minimizes the application of explicit language knowledge and strategy use (Kilborn & Moss, 1996).

Participants' word monitoring latencies of grammatical and ungrammatical sentences are compared, and delays in monitoring target words in ungrammatical sentences are interpreted as suggesting automatic and involuntary activation of integrated L2 knowledge (Marslen-Wilson & Tyler, 1980). Results in this experimental paradigm are typically analyzed within a repeated-measures design at a group level. In this study, however, what was called a *Grammatical Sensitivity Index* (i.e., GSI) was created by subtracting the response latencies of grammatical items from the latencies of ungrammatical items. This index was a measure of degree of

sensitivity for each individual participant. By providing a continuous numerical value, this index permitted investigation of any relationships between degrees of sensitivity to grammatical violations and cognitive abilities in correlational and factorial statistical analyses. It also permitted computing correlations with other language measures and comparisons in between-subjects analyses.

Two presentation lists (A and B), counterbalanced for grammaticality, were used, with half of the participants in each group randomly assigned to each list. No sentence appeared twice in the same list. A grammatical sentence in one list appeared as ungrammatical in the other, and vice-versa. Each presentation list included 60 target items (10 per target structure), half grammatical and half ungrammatical, and 60 grammatical distracters. The word to monitor appeared in target sentences (i.e., critical items), so that latencies could provide a measure of sensitivity, but it did not appear in distracter sentences. This way, the probability of a word appearing in a sentence was .5. The position of the target word in the distracters varied randomly to prevent participants from anticipating when to respond. The position of the target word in critical items was located immediately after the target structure. In order to assess word monitoring latencies as accurately as possible, split recordings of the target sentences were used. The target structure appeared at the very end of the first half of the sentence, and the timer started at the onset of the second half with the target word (i.e., the word to monitor). This way, the onset of the target word and the timer could be synchronized. This allowed use of the same second half for both the grammatical and ungrammatical version of each item, and control for possible

confounding factors, such as the speed with which different versions of an item were read and recorded.

Half of the test items ($k = 60$) were followed by a comprehension question. All comprehension questions were yes/no questions that participants answered by pressing "A" for yes and "L" for no. Half of the questions required a positive response and half a negative response.

Participants were instructed to monitor ongoing auditory language input for a pre-designated target word that would be presented visually on their computer screen. They were asked to maintain their hands on the keyboard with their index fingers on the *yes* key ("A") and the *no* key ("L"). These two keys were the right-most and left-most keys on the keyboard and allowed participants to rest their wrists on the keyboard table. Participants were instructed to press *yes* as soon as they heard the word that was displayed on the screen or to press *no*, if the sentence finished playing, and they had not heard the word displayed on the screen. They were also instructed to pay attention to the meaning of the sentences, since they would be randomly asked comprehension questions. To respond to comprehension questions, the same *yes*/*no* keys were used. Pressing the *yes* key in the word monitoring portion of the task did not stop the sentence from playing, so that participants would have all the necessary information to answer the comprehension questions.

A total comprehension score was computed on the basis of correct responses to comprehension questions. A cutoff was adopted in order to exclude any participants who were not listening for comprehension. Previous studies carried out in the same framework (Jiang, 2004; 2007) included only participants who had an error rate lower

than 37% (i.e., 63% accuracy level) (Jiang, 2004) or lower than 20% (i.e., 80% accuracy level) (Jiang, 2007). In the present study, a minimum of 75% response accuracy was required (i.e., an error rate lower than 25%). This cutoff is similar to Jiang (2007) and well-above chance-level performance (i.e., 50%).

In order to ensure that participants had been focusing their attention on meaning while performing the task, a minimum of 75% response accuracy was required for each participant to be included in the analysis. Before comparing monitoring latencies, data were checked for outliers, defined as +/- 3 SDs from each individual's mean. Only response times for correctly accepted target words (i.e., hits) were included in the analysis, since failure to monitor the word successfully implied that the task had not been performed correctly.

The reliability of the task, using the split-halves method, was .98.

4.3.2 Language Tests that Allow Controlled Use of L2 Knowledge

Untimed Auditory GJT ($k = 60$). The untimed auditory GJT was a computer-delivered test with sentences presented aurally. Participants were required to indicate whether each sentence was grammatical or ungrammatical by pressing a response button. Unlike its time-pressured counterpart, this test presented each sentence twice before participants were allowed to provide a response. Following DeKeyser (2000) and DeKeyser et al. (2010), each sentence was played twice, with a three-second interval between the repetitions and a six-second interval between sentence pairs.

Each item was scored dichotomously as correct/incorrect, and percentage accuracy scores were calculated for grammatical and ungrammatical items overall, as well as for grammatical and ungrammatical items separately.

55

The internal consistency of the test, according to Cronbach's alpha, was .89.

Untimed Visual GJT ($k = 60$). The untimed visual GJT was a computer-delivered self-paced test with sentences presented visually. Participants were required to indicate whether each sentence was grammatical or ungrammatical by pressing a response button.

Each item was scored dichotomously as correct/incorrect and percentage accuracy scores were calculated for grammatical and ungrammatical items overall, as well as for grammatical and ungrammatical items separately.

The internal consistency of the test, according to Cronbach's alpha, was .85.

Metalinguistic Knowledge Test ($k = 60$). The metalinguistic knowledge test was a computer-delivered, self-paced test with sentences presented visually. This test followed the same format as the untimed visual GJT, but it included an error correction component and a metalinguistic knowledge component in order to encourage use of metalinguistic abilities. Participants were required to indicate whether each sentence was grammatical or ungrammatical and, if ungrammatical, to correct the error and state the grammar rule. Unlike the word monitoring task, the metalinguistic knowledge test was a correction task that focused participants' attention on language forms and analyzed representations (Bialystok, 1986).

Grammatical items were dichotomously scored as correct/incorrect, and ungrammatical items were scored following a system of partial credit (0-3), yielding a maximum of 120 points on the test. One point was given for identifying the sentence as ungrammatical, one point for correcting the error, and one point for providing a statement of the grammar rule.

56

The internal consistency of the test, according to Cronbach's alpha, was .89.

4.3.3 Explicit Language Aptitude Tests

In general, the use of L1- or L2-based cognitive tests can result in confounds between participants' proficiency level and their cognitive capacity. In the case of studies that include both child and adult L2 learners, language-based cognitive measures can be particularly problematic, since degree of L2 acquisition tends to correlate with degree of L1 attrition. A commonly used test of analytic ability is the Words-in-Sentences MLAT subtest. However, this is a test that participants need to take either in their L1 or L2. DeKeyser (2000) administered the test in the participants' L1 (Hungarian). According to the descriptive data, the highest score on the test belonged to the latest arrival (age of arrival = 38). The next highest aptitude scorers were also late arrivals. Conversely, early arrivals, probably with poorer L1 literacy skills, were not able to score as high as late arrivals. The use of L1-based cognitive tests, therefore, can artificially reduce the range of aptitude scores and affect the magnitude of the correlation for early arrivals.

The LLAMA aptitude test (Meara, 2005), on the other hand, is to a large extent independent of test takers' L1 and L2, since it relies on picture stimuli and verbal stimuli based on languages that differ from any languages that test takers are likely to know in practice (i.e., a dialect of a language in Northern Canada and a Central American language). The test includes no instructions for test takers, only for test administrators, who provide them orally to test-takers. Granena (2011b, to appear) showed that three of the LLAMA subtests measured the same underlying aptitude and that this aptitude could be interpreted as analytic ability. In the proposed study,

explicit language learning aptitude, operationalized as analytic ability, will be measured as a composite of three LLAMA subtests: Vocabulary Learning (LLAMA B), Sound-symbol Correspondence (LLAMA E), and Grammatical Inferencing (LLAMA F).

The reliability of the LLAMA test ($k = 90$) in terms of internal consistency according to Cronbach's alpha was .77 (an acceptable research standard is considered to be .70, according to Nunnally & Bernstein, 1994). A total of 74 participants aged 19-47 were sampled. Stability over time (i.e., test-retest reliability) according to a Pearson product-moment correlation was .64 ($p = .002$), based on a subsample of 20 participants from the present study that were tested twice with a two-year period between test and retest (years 2009 and 2011) in order to minimize carryover effects. The internal consistency of the composite score of the three LLAMA subtests ($k = 60$) that loaded on the same factor (LLAMA B, E, and F) was .79 and their average test-retest reliability was .63 ($p = .003$).

Vocabulary Learning -LLAMA B- (Meara, 2005). LLAMA B is a test that measures the ability to learn new words. The words to be learned are presented visually and are real words taken from a Central American language. Each of them is assigned to a target image. Participants have to learn as many words as possible by relating each of them to a target image. There is a timed study phase in which participants click on the different images displayed on the screen. The name of each object is shown in the centre of the panel. Then, the program displays the name of an object and participants have to identify the correct image on the screen. The internal

consistency of LLAMA B ($k = 20$) was .76, according to Cronbach's alpha, and its test-retest reliability was .53 ($p = .016$).

Sound-symbol Correspondence -LLAMA E- (Meara, 2005). LLAMA E is a test that measures the ability to form sound-symbol associations. Participants have to work out the relationship between the sounds they hear (i.e., recorded syllables) and a transliteration of these sounds in an unfamiliar alphabet. There is a timed study phase in which participants click on the different transliterations displayed and try to learn the corresponding sound association. Then, they hear a syllable and have to decide its symbol correspondence by clicking on the right transliteration. The internal consistency of LLAMA E ($k = 20$) was .64, according to Cronbach's alpha, and its test-retest reliability was .60 ($p = .005$).

Grammatical Inferencing -LLAMA F- (Meara, 2005). LLAMA F is a test that measures the ability to induce the rules of an unknown language. Participants have to relate a sentence presented visually on the screen with its picture. There is a timed study phase in which participants click on a series of small buttons displayed on the screen. For each button, a picture and a sentence describing the scene are displayed. In the testing phase, the program shows a picture and two sentences, a grammatical and an ungrammatical one. Participants choose the correct sentence. The internal consistency of LLAMA F ($k = 20$) was .60, according to Cronbach's alpha, and its test-retest reliability was .56 ($p = .010$).

4.3.4 Implicit Language Aptitude Tests

It has been argued that much of language acquisition is sequence learning and that individual differences in the ability to remember verbal strings determine the

acquisition of grammar (N. Ellis, 1996). In the proposed study, implicit language learning aptitude, operationalized as phonological and visual sequence learning ability, will be measured by means of two tests: a sound recognition test (LLAMA D) and a probabilistic serial reaction time task.

Sound Recognition -LLAMA D- (Meara, 2005). LLAMA D is a test that measures participants' ability to recognize patterns in spoken language. According to Meara (2005), this ability should help learners recognize the small variations in endings that languages use to signal grammatical features. The test is based on Speciale, N. Ellis, and Bywater (2004) and on research on implicit induction of phonological sequences (e.g., Saffran et al., 1996). Participants listen to a string of words based on the names of objects in a British Columbian Indian language. They then complete a recognition test and indicate whether they have heard each stimulus previously. Participants who rapidly acquire the phonological sequences of the target items are able to discriminate better between old and new items. The internal consistency of this subtest ($k = 30$), according to Cronbach's alpha, was .63. This coefficient is .07 below the acceptable standard of .70, but not substantially. The test can be considered to have marginal reliability and relatively uniform test items. Test-retest reliability was .61 ($p = .004$).

Serial Reaction Time (SRT) Task. The SRT task is a test that measures participants' implicit sequence learning ability. Unlike other paradigms for studying implicit learning (e.g., Artificial Grammar learning tasks), learning in the SRT is measured online (i.e., during the training phase), which, according to Destrebecqz and Cleeremans (2001), makes it a better measure of implicit learning.

Originally developed by Nissen and Bullemer (1987), the SRT task used in the present study was a probabilistic version created using the same stimuli as Kaufman et al. (2010). The probabilistic SRT task measures participants' sensitivity to high- and low-frequency events. Participants see a visual cue (an asterisk) appear at one of four prescribed locations on a computer screen. The four locations are separated by 1.2 inches and indicated by means of a placeholder. Participants are required to press a key corresponding to the location of the asterisk as fast and accurately as possible by placing their middle and index fingers of each hand on the keys marked "z", "x", ".", and "/", respectively (see Figure 1). Keys "z" and "x" and "." and "/" were adjacent and allowed participants to place their wrists comfortably on the laptop table. No instructions to memorize the series or look for underlying rules are provided.
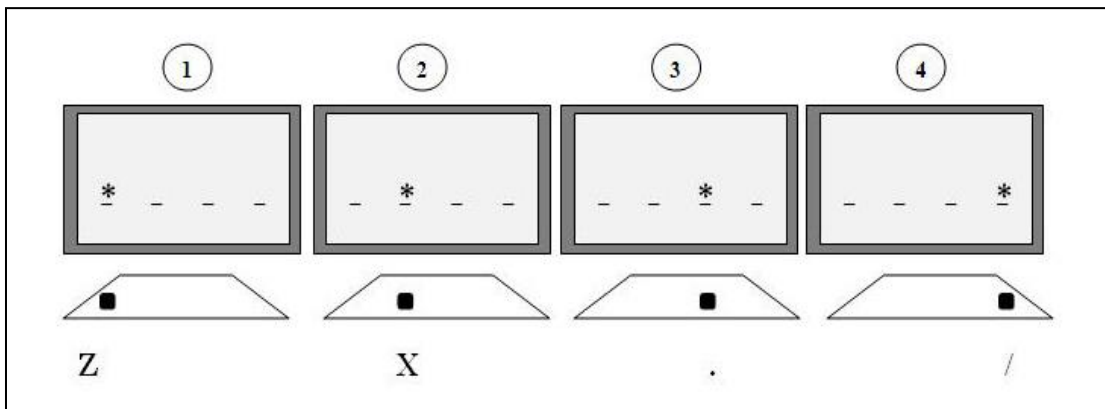
Figure 1. Representation of visual cues and required key-presses in SRT task

The asterisks play out a repeating sequence of positions. This sequence, unlike in the deterministic version of the task, follows a probabilistic order. In every task block, sequence trials are interspersed with control trials. Control trials are incongruent with sequence trials and make it more difficult for participants to explicitly discover the

target sequence. As a result, the task has greater ecological validity, since implicit learning in the real world takes place under conditions of uncertainty (i.e., noise) that make learning probabilistic, rather than deterministic (Jiménez & Vázquez, 2005).

Following Schvaneveldt and Gómez (1998), stimuli were congruent with the target sequence 85% of the time and intermixed with an alternate sequence 15% of the time. The two sequences used to generate either training (A) or control (B) trials had 12 elements each and were balanced for simple location and transition frequency (Reed & Johnson, 1994). They exclusively differed in the second-order conditional information they conveyed. Reed and Johnson gave the sequences of three locations the name of *second order conditionals* (SOCs) (vs. first-order probabilities, where the location of an item is unambiguously predicted by the preceding item with a probability of 1.0). In second-order conditionals, at least two previous locations are needed to predict the next location in the sequence.

The target sequence chosen (Sequence A) was 1-2-1-4-3-2-4-1-3-4-2-3, while the alternate sequence (Sequence B) was 3-2-3-4-1-2-4-3-1-4-2-1. The starting point was randomly chosen for each block. Figure 2 shows how the two sequences are related to each other. Transitions in one sequence respect the second-order conditionals of the other sequence, but lead to different predictions. If a participant is trained in Sequence A, the most likely successor after locations 4-3 would be 2, but on some trials it could be 1, which is the successor of the series 4-3 according to Sequence B. Trials following the alternate control sequence could appear isolated or in small groups (e.g., 3-2-4-1- **2-4-3** -2- **3** -1-2-1- **3-2** -4-1).
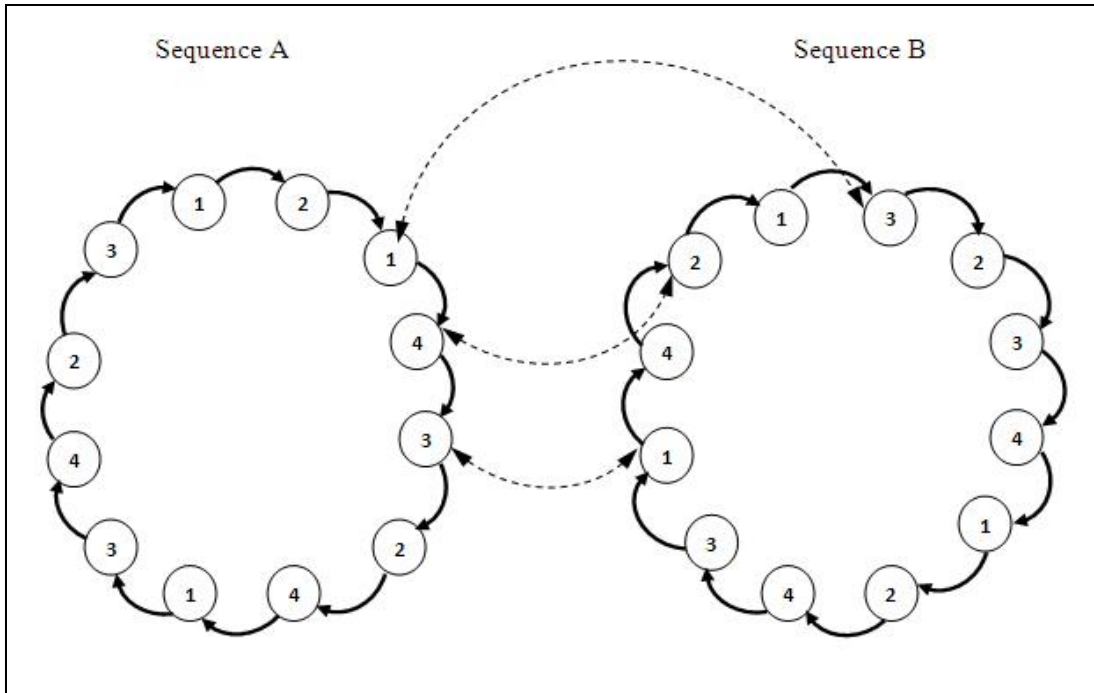
Figure 2. Representation of the two sequences used to generate training trials (A) and control (B) trials

The SRT task started with a practice block that included 14 trials where the likelihood of probable and improbable transitions was the same (.5 probability). After the practice block, participants completed eight training blocks of 120 trials each (960 in total). The task did not include response-stimulus intervals, since there is evidence that explicit learning can take place when people are given 250 or 500-msec to think (Destrebecqz & Cleeremans, 2001). Out of 960 trials, 149 (15.52%) were control trials and 811 were training trials (84.48%). In order to increase the probabilistic nature of the task, the probability of transitions generated from Sequence A and Sequence B also differed from block to block (see Table 4).

Table 4. *Probabilities of Probable and Non-probable Trials in SRT Task*

	Sequence A (Probable Trials)		Sequence B (Improbable Trials)	
	n	%	*n*	%
Block 1	102	85	18	15
Block 2	105	87.5	15	12.5
Block 3	98	81.67	22	18.33
Block 4	108	90	12	10
Block 5	101	84.17	19	15.83
Block 6	94	78.33	26	21.67
Block 7	101	84.17	19	15.83
Block 8	102	85	18	15

All trials were initially randomized within each block and then presented in the same fixed order for each participant. According to Kaufman et al. (2010), this procedure maximizes "the extent to which individual differences reflect trait differences rather than differences in item order" (p. 326). Participants were allowed to take a short rest between blocks.

Accuracy and reaction time in milliseconds were recorded on each trial. Degree of learning was quantified as the average difference in reaction time between correct responses to congruent and incongruent trials (incongruent RT - congruent RT). The larger the difference, the more learning occurred.

At the end of the SRT task, participants were administered a recognition test adapted from Shanks and Johnstone (1999). They were told that they would be presented with short sequences of three elements. They were asked to respond to the asterisks as quickly as possible, and then to provide a rating of how confident they were that the sequence was part of the test they had just taken. The recognition test included 24 three-element sequences (triads) presented in a randomized order for each participant. There were 12 old sequences, constructed following second order conditionals in Sequence A (3-4-$\underline{2}$, 3-1-$\underline{2}$, 1-4-$\underline{3}$, 2-4-$\underline{1}$, 4-2-$\underline{3}$, 1-2-$\underline{1}$, 4-3-$\underline{2}$, 4-1-$\underline{3}$, 2-3-$\underline{1}$, 2-1-$\underline{4}$, 3-2-$\underline{4}$, 1-3-$\underline{4}$), and 12 novel sequences, constructed following second order conditionals in Sequence B (3-4-$\underline{1}$, 3-1-$\underline{4}$, 1-4-$\underline{2}$, 2-4-$\underline{3}$, 4-2-$\underline{1}$, 1-2-$\underline{4}$, 4-3-$\underline{1}$, 4-1-$\underline{2}$, 2-3-$\underline{4}$, 2-1-$\underline{3}$, 3-2-$\underline{3}$, 1-3-$\underline{2}$). It should be noted that, in fact, all sequences had been seen before, but with different probabilities (.85 vs. .15), so the terms "old" and "new" are relative and actually mean "familiar" and "less familiar". Each location and each first-order transition appeared with the same likelihood. The only difference between old and new sequences was second-order conditional information (e.g., transition 3-4 was followed by location 2 in Sequence A and by location 1 in Sequence B). There were also four practice trials containing novel random sequences (1-1-1, 4-4-4, 1-2-3, 3-2-1). Immediately after participants selected the response button for the third element, they were asked to give a confidence rating on a six-point scale, where 1 = I'm sure that this sequence was part of the test; 2 = I'm pretty sure that this sequence was part of the test; 3 = I think that this sequence was part of the test; 4 = I think that this sequence was not part of the test; 5 = I am pretty sure that this sequence was not part of the test, and 6 = I'm sure that this sequence was not part of the test.

The consensus in the sequence learning literature appears to be that if participants are able to discriminate old from new sequences, they have acquired explicit sequence knowledge (Perruchet & Amorim, 1992; Shanks & Perruchet, 2002; Shanks, Wilkinson, & Channon, 2003; Willingham, Salidis, & Gabrieli, 2002). In addition to a measure of (explicit) recognition, the test used in the present study also yielded a concurrent measure of (implicit) priming, based on the speed of responding to old versus new sequences. Recognition scores were computed as the difference between the mean judgment for old sequences minus the mean judgment for new sequences. Priming scores were computed as the difference between the mean reaction time elicited by the third element of old sequences minus the mean reaction time elicited by the third element of new sequences. Evidence of a dissociation between explicit recognition and implicit priming (i.e., poor recognition, but faster reaction times, for segments of the old sequences) was considered as supporting evidence of implicit learning during the training task.

The reliability of the probabilistic SRT task in the present study, using split-halves with the Spearman-Brown correction, was .44. This is a low reliability index when compared to reliability indices of measures of explicit learning, which are usually greater than .70, but it is similar to the indexes reported in other studies of implicit learning. Kaufman et al. (2010), from whom the SRT task was adapted, also reported a reliability of .44 and considered it standard for probabilistic SRT tasks, on the basis of the reliability of implicit learning previously reported in the literature (Reber et al., 1991; Dienes, 1992). Reber et al. (1991) and Robinson's (1996) replication study reported split-half reliabilities of .51 and .52, respectively, also

using the Spearman-Brown correction. According to Reber et al., "a Cronbach above

.4 or .5 is taken as reasonable support for the internal reliability of a test" (p. 893).

The less reliable is a measure, the lower its possible observed correlation with another

variable can be, regardless of the true correlation, given that lower reliability leads to

greater attenuation of correlation coefficients as a result of the amount of noise in the

measure. However, Kaufman et al. (2010) reported significant correlations between

implicit learning on their probabilistic SRT task and processing speed. These

correlations were in the middle third of effect sizes reported in psychology ($r = .2$ to

.3; Hemphill, 2003). Other studies have also shown correlations between implicit

learning and complex cognition (i.e., school grades in Math and English) (Gebauer &

Mackintosh, 2012; Pretz, Totz, & Kaufman, 2010).

4.3.5 General Intelligence Test

A Spanish version of the General Ability Measure for Adults (GAMA[9]) test was

used as a measure of general intellectual ability. GAMA is a commercially available

non-verbal test of intelligence published by Pearson that uses abstract designs,

shapes, and colors (i.e., non-verbal stimuli) to minimize the effects of confounding

variables such as language knowledge, verbal expression, and verbal comprehension

on test scores. It is a self-administered (booklets were used), 25-minute, timed test

with four subtests (66 items with response sets of five options) that require the

application of reasoning and logic to solve problems: Matching, Analogies,

---

[9] GAMA is considered a culture-fair test because its non-verbal nature allows evaluation of intellectual ability without substantial influence from linguistic, educational, and cultural factors. The Raven's Progressive Matrices (Raven, 1938) is also a non-verbal test of fluid intelligence, but it has been criticized for its use of matrix structures, considered a cultural construct (Greenfield, 1998).

Sequences, and Construction. The type of non-verbal reasoning measured corresponds roughly to fluid intelligence.

The Matching subtest requires examinees to determine which one of the six options is identical to the stimulus in color, shape, and configuration (see Figure 3).
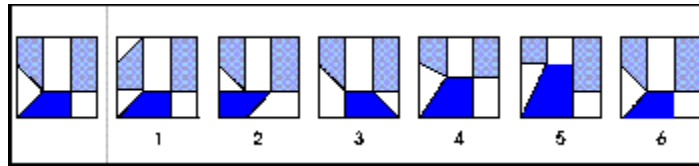
Figure 3. Sample matching item: Which answer is the same as the first picture?

The Analogies subtest requires examinees to recognize the relationship between two abstract figures and then identify the option that has a different pair of figures with the same conceptual relationship (see Figure 4).
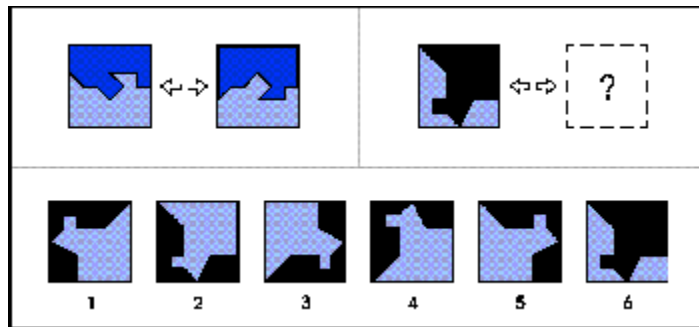
Figure 4. Sample analogies item: Which answer goes on the question mark?

The Sequences subtest requires examinees to recognize the pattern of change in a geometric design and choose the option that fits the pattern (see Figure 5).
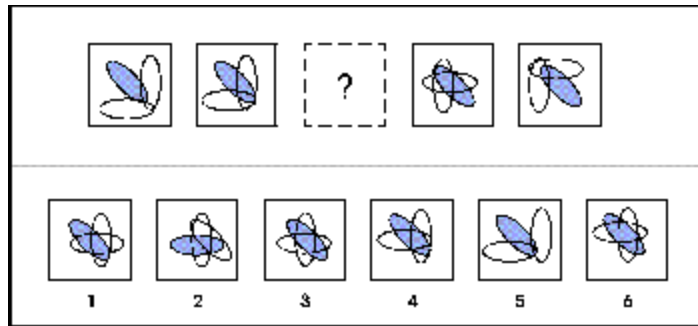
Figure 5. Sample sequences item: Which answer goes on the question mark to complete the pattern?

The Construction subtest requires examinees to determine how several shapes can be combined to produce one of the designs provided as options. The items require the examinee to analyze and synthesize the spatial characteristics of the shapes to mentally construct designs (see Figure 6).
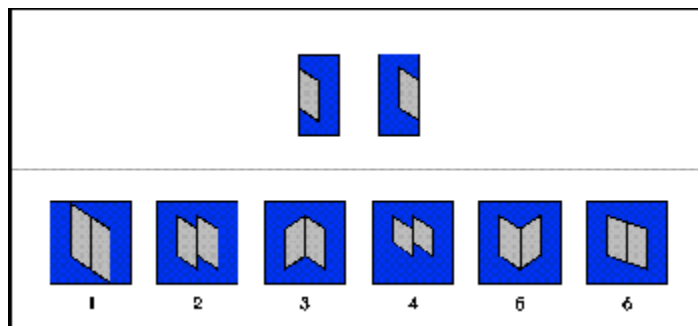
Figure 6. Sample construction item: Which answer can be made with the shapes in the top box?

The GAMA was normed on a sample of 2,360 individuals aged 18 to 96 (Naglieri & Bardos, 1997). Internal consistency using split-halves with Spearman-Brown correction ranged from 0.79 to 0.94 across normative age groups, with an average of 0.90 (reliability based on a linear composite). Average reliability coefficients for each of the four subtests across normative groups were .66, .81, .79, and .65 for the

Matching, Analogies, Sequences, and Construction subtests, respectively. The test-

retest reliability was 0.67 over a two- to six-week interval for a sample of 86 people.

In terms of validity (concurrent validity), the correlations between GAMA and other

general ability tests, the WAIS-R (Wechsler, 1981) and the K-BIT (Kaufman Brief

Intelligence Test, Kaufman & Kaufman, 1990) were .75 ($p < .001$) and .70 ($p < .001$),

respectively, for a sample of 194 participants. Skinner et al. (1996) looked at the

relationship between the GAMA and reading achievement among college students.

The results suggested that the GAMA non-verbal scores are significantly related to

reading achievement ($r = .39$, $p < .01$).

4.4 Procedure

Participants were tested individually by the researcher and all the tasks were

administered on a laptop computer. Data were collected in a seminar room of the

Education Department at the Universidad Complutense in Madrid. Upon their arrival,

participants were provided with a Spanish translation of the consent form[10] and had

the opportunity to ask questions before signing it (5 minutes). Two of the language

tests, the word monitoring task and the metalinguistic knowledge test, were

administered in a fixed order. The word monitoring task was always administered

first (20 minutes), and the metalinguistic knowledge test was administered last (20

minutes). The rationale behind this order was to reserve the tests that allow controlled

use of L2 knowledge and that encourage the highest degree of awareness to the end

and the most implicit L2 measure as pure as possible (for a similar order of

administration, see Ellis, 2005). The four remaining language tests were administered

---

[10] IRB (Institutional Review Board) Protocol #08-0138, approval date February 22, 2011.

between the word monitoring task and the metalinguistic knowledge test following a balanced Latin square design. The two timed GJTs took approximately 10 minutes each, whereas the untimed auditory and untimed visual GJT took approximately 20 minutes each. Every language test was followed by a cognitive test selected at random. Each cognitive test (the LLAMA, GAMA, and probabilistic SRT task) took 25 minutes. Overall testing time ranged between three to four hours. The six language tests included a 5-minute break halfway. In addition, participants were allowed to take rests between tests and/or as needed. Participants earned 50 euros (approximately $65) for their participation in the study and were provided with drinks and snacks.

4.5 Target Structures

In order to maximize variability among early L2 learners, the study included a combination of six grammatical structures that are early and late acquired items in L1 Spanish. Grammatical agreement, in languages that mark for agreement, is acquired by age 3 and with few errors (Slobin, 1985; Meisel, 1990), but it is among the most difficult grammatical structures for L2 learners. Later acquired items for L1 speakers include the conditional and subjunctive moods, subordinate clauses, passives, and tense and aspect. These are structures that are not mastered with 100% accuracy until age 7 and later, whereas grammatical agreement is acquired with almost 100% accuracy by age 3. Late acquisitions in Spanish such as the subjunctive and the passive depend on cognitive development in children and are more influenced by explicit instruction at school and by literacy skills.

The six target structures included in the present study were: (1) Noun-adjective gender agreement, (2) Subject-verb number agreement, (3) Noun-adjective number agreement, (4) Subjunctive mood, (5) Perfective/imperfective aspect contrasts, and (6) Passives with *ser*/*estar*. These structures are known to be difficult for NSs of a non-Romance language (Bruhn de Garavito & Valenzuela, 2008; Collentine, 1995; Jiang, Novokshanova, Masuda, & Wang, 2011; Johnston, 1995; Montrul, 2004; Smith, 1980; Terrell, Baycroft, & Perrone, 1987). Gender agreement, subject-verb agreement, and number agreement (agreement structures) are acquired by age 3 by L1 speakers of Spanish, while the subjunctive, aspectual contrasts, and passives (non-agreement structures) are acquired close to age 7 (Montrul, 2004). Grammatical structures acquired early in the L1, such as gender agreement, show greater variability in L2 acquisition than structures that are acquired late. For example, in the study by Granena and Long (2010), the variability (as indicated by standard deviations) among child L2 learners with ages of onset between 3 and 6 ($n = 20$) was greater for grammatical structures that are typically acquired before age 3 by L1 speakers (e.g., gender agreement, $M = 10.45$, $SD = 2.63$) than for structures that are acquired later (e.g., the subjunctive, $M = 21.00$, $SD = 1.45$). An example of each target structure is displayed in Table 5.

Table 5. *Target structures*

Early L1 acquisitions	Late L1 acquisitions
Agreement structures	Non-agreement structures
Noun-adjective gender agreement	**Subjunctive mood**
Cualquier corriente de aire puede resultar <u>molesto</u> (molesta) para la practica del esquí	*Es importante que los estudiantes de español <u>practican</u> (practiquen) el idioma todos los días*
"Any airstream can become <u>annoying</u> for skiing"	"It is important for Spanish learners to <u>practice</u> the language every day"
Subject-verb number agreement	**Perfective/imperfective aspect**
En el periódico se <u>publicó</u> (publicaron) todos los artículos escritos por Miguel Delibes	*En la edad de piedra, los seres humanos <u>aprendían</u> (aprendieron) a utilizar la rueda*
"All the articles written by Miguel Delibes <u>were</u> published in the paper"	"In the Stone Age, human beings <u>learned</u> how to use the wheel"
Noun-adjective number agreement	**Passives with *ser*/*estar***
Los votos de los que dispone el candidato son <u>mucho</u> (muchos) más de los que tiene la oposición	*El jefe informó de que el trabajo que <u>fuese</u> (estuviese) acabado para el viernes se pagaría doble*
"The votes the candidate has are <u>many</u> more than the opposition has"	"The boss announced that the work that <u>was finished</u> by Friday would be double-paid"

A pool of 360 target items was created and items randomly assigned to each of the six language measures. Each language test had a total of 60 items (10 per target structure), an equal number of which were grammatical and ungrammatical (see Appendix B for item pool).

Chapter 5: Results

This dissertation predicted that aptitudes that are more relevant for implicit language learning and processing would moderate L2 learners' attainment on tasks that require more automatic use of L2 knowledge. This prediction was made both for early L2 learners who are sequential bilinguals and for adult learners, since adults were still expected to be able to learn implicitly, but not for NSs, whose ultimate attainment is characterized by inter-individual homogeneity and, therefore, predicted to be independent of aptitude. Aptitude for implicit language learning should also moderate early L2 learners' attainment on tasks that require controlled use of L2 knowledge, since early L2 learners were expected to use the same type of knowledge regardless of language task. The nature of this knowledge is hypothesized to be implicit, like NSs' knowledge. Early L2 learners' ultimate attainment, however, is characterized by greater inter-individual variability than NSs and, therefore, was expected to be moderated by language aptitude.

On the other hand, aptitudes that are more relevant for explicit language learning were only expected to moderate adult L2 learners' attainment on tasks that allow controlled use of L2 knowledge. These tasks increase available test time and decrease processing demands; therefore, they provide an opportunity to utilize problem-solving and analytic skills. On these tasks, adult learners can rely on explicit L2 knowledge and compensate for their limited implicit competence. Adult learners with a higher aptitude for explicit language learning were expected to do better as a result of their greater analytic, metalinguistic abilities.

Finally, regarding general intelligence, it was hypothesized that, in ultimate L2 attainment, relationships between explicit aptitude and general intelligence and learning outcomes would pattern in the same way and would be different from effects of implicit aptitude on outcomes. This hypothesis was based on studies of artificial grammar learning, in which fluid intelligence correlates with learning when participants are instructed to look for patterns in the training materials, but not under more incidental learning conditions

This chapter presents the results of the study. The results for each of the cognitive and linguistic measures are presented first. Next, the results of the role of cognitive variables on language outcomes are reported. Overall performance on grammatical and ungrammatical items is reported first, and, then, follow-up detailed analyses are reported for ungrammatical items and for type of target structure (agreement structures –early L1 acquisitions- and non-agreement structures –late L1 acquisitions).

5.1 Cognitive Aptitudes

In this section, overall performance for each speaker group on the six cognitive tests (LLAMA B, D, E, and F, GAMA, and the probabilistic SRT task) is presented. Next, the results of a PCA, an exploratory factor analytic technique, are reported. PCA was conducted to reduce the dimensionality of the dataset by determining whether cognitive variables could be combined into different aptitude components as equally weighted composite scores.

5.1.1 The LLAMA Test

Table 6 shows the descriptive statistics for each of the four LLAMA subtests: LLAMA B (vocabulary learning), LLAMA D (sound recognition), LLAMA E (sound-symbol correspondence), and LLAMA F (grammatical inferencing). The maximum possible test score for each test was 100.

Table 6. *Descriptives of the LLAMA Language Aptitude Test*

Group	LLAMA B	LLAMA D	LLAMA E	LLAMA F
	M	*M*	*M*	*M*
NS Controls	56.50	33.50	79.00	61.50
(*n* = 20)	(18.36)	(16.39)	(19.17)	(22.31)
	30-95	0-60	30-100	10-100
Early AO	63.40	37.40	84.00	64.60
(*n* = 50)	(18.03)	(13.71)	(16.78)	(20.52)
	30-100	10-65	40-100	10-90
Late AO	50.20	28.60	70.40	55.80
(*n* = 50)	(17.35)	(12.90)	(25.55)	(24.00)
	5-80	0-55	0-100	0-90

Note. Standard deviations appear between parentheses

The four subtests were normally distributed, according to one-sample Kolmogorov-Smirnov (K-S) tests, in each of the groups: NS controls (p = .629, p = .284, p = .734, and p = .850), early AO (p = .354, p = .347, p = .556, and p = .932),

and late AO (p = .376, p = .266, p = .436, and p = .434). There were no extreme

outliers (i.e., +/- 3 SDs) from each group's mean.

Early L2 learners scored the highest on each of the four subtests: LLAMA B

(63.40%), LLAMA E (84%), LLAMA F (64.60%), and LLAMA D (37.40%),

whereas late L2 learners scored the lowest: LLAMA B (50.20%), LLAMA E

(70.40%), LLAMA F (55.80%), and LLAMA D (28.60%). Finally, the scores in the

NS group were 56.50% (LLAMA B), 79% (LLAMA E), 61.50% (LLAMA F), and

33.50% (LLAMA D).

NSs were not significantly different from either early or late L2 learners on any of

the LLAMA subtests, according to Scheffé posthoc tests: LLAMA B (p = .345 and p

= .412), LLAMA E (p = .674 and p = .314), LLAMA F (p = .871 and p = .629), and

LLAMA D (p = .569 and p = .412). Early and late L2 learners, however, were

significantly different on all the tests, except on LLAMA F (p = .148): LLAMA B (p

= .002), LLAMA E (p = .007), and LLAMA D (p = .008). These differences could be

due to the positive cognitive consequences that early bilingualism is claimed to have

on executive processes (e.g., Bialystok, 1999), age at time of testing (since late L2

learners were significantly older than early L2 learners), sampling bias (i.e., a biased

representation of early bilinguals who succeeded with L2 Spanish), or, perhaps, a

combination of two or more of these factors.

5.1.2 The GAMA Test

Table 7 shows the descriptive statistics for the GAMA general intelligence test.

Table 7. *Descriptives of the GAMA General Intelligence Test*

Group	GAMA	
	M	*Range*
Control (*n* = 20)	44.30 (7.01)	30-56
Early AO (*n* = 50)	45.88 (5.45)	28-56
Late AO (*n* = 50)	39.88 (7.44)	18-53

Note. Standard deviations appear between parentheses.

The maximum possible test score was 66. Tests scores were normally distributed, according to one-sample Kolmogorov-Smirnov (*K-S*) tests, in each of the groups: NS controls (*p* = .882), early L2 learners (*p* = .452), and late L2 learners (*p* = .465). Late L2 learners scored significantly lower than NS controls (*p* = .044) and early L2 learners (*p* < .001). There was no significant difference between NS controls and early L2 learners (*p* = .665). The fact that late L2 learners had a significantly lower IQ than NS controls was due to the effects of an outlier with a score of 18 out of 66 in the late L2-learner group (this was the participant with the oldest age at testing in the sample). When this participant was removed, late L2 learners were no longer significantly different from NSs (*p* = .064) and they only differed from early L2 learners (*p* < .001). The difference in intelligence between early and late learners could be due to late L2 learners' significantly older age at testing with respect to early L2 learners, or to other factors that could have contributed to early L2 learners' higher IQ scores (e.g., cognitive advantages of early bilingualism, sampling bias).

5.1.3 Probabilistic Serial Reaction Time (SRT) Task

To assess learning on the SRT task, the average response time on probable trials was subtracted from the average response time on improbable trials. The resulting difference was used as an index of sequence learning; greater differences indicated a greater degree of sequence learning. Error responses were discarded (0.90% of trials), as well as outlier responses that were +/- 3 standard deviations from the mean (1.68% of trials), computed individually for each block and participant.

A repeated-measures analysis of variance (ANOVA) with block (blocks 1 to 8) and type of trial (training vs. control) as within-subjects factors was conducted on the reaction time measures. The results showed a significant effect for block ($F(7, 112) = 12.552$, $p < .001$, $\eta_p^2 = .440$, $\lambda = .560$), and type of trial ($F(1, 118) = 108.842$, $p < .001$, $\eta_p^2 = .480$, $\lambda = .520$), as well as a significant interaction for block x type of trial ($F(7, 112) = 9.982$, $p < .001$, $\eta_p^2 = .384$, $\lambda = .616$), suggesting that learning of the training sequence occurred. Figure 7 shows the SRT learning performance for probable trials (i.e., congruent with the target sequence) and non-probable trials (i.e., incongruent with the target sequence) for the entire set of participants.
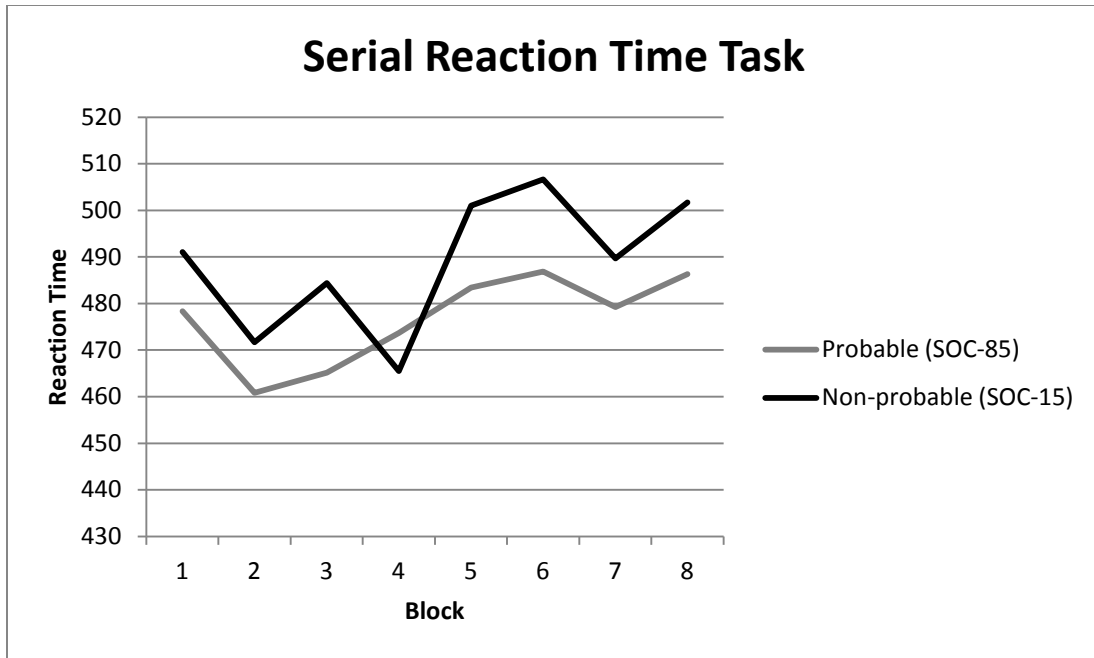
Figure 7. SRT learning performance

Reaction times for probable trials (SOC-85) were always faster than reaction times for non-probable trials in each of the blocks, except for block 4, where responses to non-probable trials, surprisingly, were faster ($t(119) = 2.193$, $p = .030$). The explanation for this pattern of results, which was observed in the sample as a whole, as well as in each of the groups separately, seems to be the percentage of non-probable trials in block 4. This block had the lowest percentage of non-probable trials (10%) and this seems to have decreased the amount of interference effects. Reaction times in block 6 support this interpretation, since this was the block with the largest percentage of non-probable trials (21.67%) and, perhaps for that reason, also the block with the largest difference between the average time to respond to probable trials and the average time to respond to improbable trials. Given that participants' responses showed high sensitivity to changes in probability levels, having kept the

.85/.15 probabilities for probable and improbable trials in each of the blocks throughout the task could have yielded more similar results across blocks.

As shown in Figure 7, blocks did not follow a linear trend. Reaction times for probable trials were faster at the beginning of the task and, then, became increasingly slower. This is a common effect in probabilistic versions of the SRT task (see, for example, Kaufman et al., 2010, page 331) and can be interpreted as an effect of the increasing control that takes place when participants learn the target sequence, but also realize that it does not always follow the same order (Jiménez, p.c., 01/13/2012).

The average reaction times for probable and improbable trials were 476.83 (SD = 75.39) and 489.08 (SD = 71.78), respectively. The resulting reaction time difference (i.e., index of sequence learning) was 12.25 ($t(119)$ = -10.564, $p < .001$). This difference was statistically significant in each of the groups: NS controls ($t(19)$ = -4.046, $p = .001$), early L2 learners ($t(49)$ = -8.352, $p < .001$), and late L2 learners ($t(49)$ = -5.645, $p < .001$), indicating a significant amount of learning in all the groups.

Table 8 shows the index of sequence learning in each of the groups. Sequence learning was normally distributed in every group, according to one-sample Kolmogorov-Smirnov (K-S) tests: NS controls ($p = .697$), early L2 learners ($p = .932$), and late L2 learners ($p = .983$). The three groups exhibited comparable amounts of sequence learning and did not differ significantly from one another, according to Scheffé posthoc tests, between late L2 learners and controls ($p = .495$), late and early L2 learners ($p = .120$), or controls and early L2 learners ($p = .930$).

Table 8. *Descriptives of the Probabilistic SRT Task*

Group	Probabilistic SRT Task	
	M	*Range*
Control (*n* = 20)	13.37 (14.78)	-17.41-44.91
Early AO (*n* = 50)	14.63 (12.39)	-10.51-49.26
Late AO (*n* = 50)	9.42 (11.79)	-13.00-38.64

Note. Standard deviations appear between parentheses.

The possible influence of explicit knowledge on participants' learning performance (i.e., conscious access to sequence knowledge) was assessed via a recognition test with an objective and a subjective component. Participants' confidence ratings given to old and new sequences (i.e., triads) on a six-point scale were compared. Low ratings indicated greater confidence in the sequence being old. If participants are unable to discriminate old from new, this may be evidence of implicit learning.[11] In addition, the reaction times elicited by the third element of the same old and novel triads were also compared. Response speed provides a direct index of the possible influence of unconsciously applied perceptual-motor programs. Tables 9 and 10 show the descriptive statistics for each of the groups. Table 9 displays confidence ratings, while Table 10 presents reaction times on new and old sequences.

---

[11] Shanks and Johnstone (1999) point out that, if discrimination of old and new sequences is possible, this could be due to the unconscious misattribution of fluency to oldness. In other words, participants may become aware of the fact that some of their responses are faster and judge those sequences as more familiar. That is why, if discrimination is possible, Shanks and Johnstone suggest comparing reaction times of sequences judged old and new, independently of actual old-new status, to test whether there is a contribution of explicit sequence memory to recognition performance over and above the fluency factor.

Table 9. *Mean Confidence Ratings for Old and New Triads*

Group	Old Sequences ($n = 12$)		New Sequences ($n = 12$)	
	M	*Range*	*M*	*Range*
Control	2.65 (0.50)	1.67-3.42	2.70 (0.53)	1.36-3.42
($n = 20$)				
Early AO	2.41 (0.60)	1.08-3.33	2.53 (0.69)	1.08-3.83
($n = 50$)				
Late AO	2.36 (0.64)	1.00-3.67	2.45 (0.69)	1.00-3.75
($n = 50$)				

Note. Standard deviations appear between parentheses.

Table 10. *Mean Reaction Times for Old and New Triads*

Group	Old Sequences ($n = 12$)		New Sequences ($n = 12$)	
	M	*Range*	*M*	*Range*
Control	560.98	390.36-	581.94	421.92-
($n = 20$)	(109.12)	810.92	(112.79)	763.50
Early AO	479.44	316.82-	501.95	351.73-
($n = 50$)	(92.49)	706.92	(82.39)	757.25
Late AO	554.29	369.75-	570.70	413.42-
($n = 50$)	(102.78)	822.42	(109.51)	845.42

Note. Standard deviations appear between parentheses.

As noted in Chapter 4, confidence ratings were indicated on a six-point scale, from 1 ("I am sure that this sequence was part of the test") to 6 ("I am sure that this sequence was <u>not</u> part of the test"). As can be seen in Table 9, slightly lower mean ratings were assigned to old sequences (low ratings indicate greater confidence in the sequence being old), but there was considerable overlap between the two distributions of ratings in all the groups. Participants' ratings were mostly located on the first half of the scale (points 1 to 3), from "I am sure that this sequence was part of the test" to "I think that this sequence was part of the test", regardless of old-new status. The mean score on new sequences exceeded the mean score on old sequences by only 0.05 of a scale unit in the NS control group, 0.12 in the early L2 learner group, and 0.09 in the late L2 learner group. This suggests that participants judged old and new sequences as being equally familiar and that they were unable to discriminate between them.

A repeated-measures ANOVA was conducted with rating (old vs. new) as a within-subjects factor and group as a between-subjects factor. Box's test was not significant ($p = .580$), suggesting that participants maintained their relative standing in the two treatment conditions. The equality of variances assumption was met, according to Levene's test, for both ratings to old sequences ($p = .424$) and new sequences ($p = .326$), indicating equal variability across groups. The repeated-measures ANOVA yielded a non-significant main effect for rating ($F(1, 117) = 3.340$, $p = .070$, $\eta_p^2 = .028$, $\lambda = .972$) and a non-significant interaction between rating and group ($F(2, 117) = .311$, $p = .733$, $\eta_p^2 = .005$, $\lambda = .995$), suggesting no significant differences between ratings to old and new sequences and a comparable effect in the

three groups of participants. This suggested that participants did not have explicit knowledge of familiar sequences.

Regarding reaction times, the third element of old sequences elicited faster responses in the three groups of participants. Given that the two first locations in the triads were the same in old and new sequences and that they only differed in their second-order conditional information (e.g., 1-2 was followed by 1 as an old sequence, but by 4 as a new sequence), the increased fluency observed for old sequences can be attributed to the oldness of the sequence and automatic retrieval of sequence knowledge.

A repeated-measures ANOVA was conducted with reaction times (old vs. new) as a within-subjects factor and group as a between-subjects factor. The assumption of equality of covariance matrices according to Box's Test was met ($p = .203$). The equality of variances assumption was met, according to Levene's test, for reaction times on old sequences ($p = .512$), but not for reaction times on new sequences ($p = .046$)[12], indicating unequal variability across groups. However, the largest standard deviation was less than three times the smallest standard deviation and Levene's test was approaching the .05 value. Therefore, ANOVA was considered robust. In addition, reaction times on old and new triads were normally distributed in the control group ($p = .845$ and $p = .895$), early AO group ($p = .553$ and $p = .282$), and late AO group ($p = .890$ and $p = .549$), according to K-S tests. The repeated-measures

---

[12] ANOVA is considered reasonably robust to moderate departures from the homogeneity assumption, if sample size is larger than 20, but the departure needs to stay smaller when the sample sizes are very different (largest to smallest > 1.5) (Keppel & Wickens, 2004). In addition, Levene's test is sensitive to Type I errors and, with a large sample size, it will tend to indicate a significant difference between variances when the real difference may not be that large. As a rule of thumb, if the largest standard deviation is three or four times larger than the smallest standard deviation, it is likely that the assumption has been violated (Houser, 2008). An alternative to transforming the data is to test at a more stringent alpha level, such as .01.

ANOVA yielded a significant main effect for reaction time ($F(1, 117) = 9.784$, $p =$ .002, $\eta_p^2 = .084$, $\lambda = .916$) and a non-significant interaction between reaction time and group ($F(2, 117) = .125$, $p = .883$, $\eta_p^2 = .002$, $\lambda = .998$). Reaction times were significantly faster on the third element of old (i.e., more familiar) sequences, and this effect was comparable in the three groups of participants.

5.1.4 Cognitive Aptitudes for Implicit and Explicit Learning

It was hypothesized that the LLAMA subtests B, E, and F were measures of explicit cognitive processes relevant for explicit language learning, whereas LLAMA D and the probabilistic SRT task were measures of implicit cognitive processes relevant for implicit language learning. These claims were based on the results of an exploratory factor analysis (Granena, 2011b, to appear), which showed that LLAMA subtests B, E, and F loaded together on one component, interpreted as analytic ability, whereas LLAMA D loaded on a separate component, interpreted as sequence learning ability. LLAMA B, E, and F have in common that all include a study phase prior to testing, allow time to think and use problem-solving strategies, and involve working out relations in a data set. LLAMA D, on the other hand, includes no study phase, does not allow time to rehearse, and involves recognition of phonological sequences.

To further validate the hypothesized distribution of cognitive aptitudes, a PCA was conducted ($n = 120$) on the scores of the four LLAMA subtests and on the amount of sequence learning in the probabilistic SRT task, as measured by the difference in reaction time between probable and improbable trials. An orthogonal rotation method (Varimax[13]) was used. The analysis yielded two principal

---

[13] Orthogonal rotation methods (e.g., Varimax, Equamax, Quartimax) result in uncorrelated factors,

components with eigenvalues greater than 1.0 that explained 59.23% of the total variance. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was greater than .600 (.712), and the Bartlett's test of sphericity was significant ($p < .05$), indicating that the correlation matrix differed significantly from zero. The first component had an eigenvalue of 1.861 and accounted for 37.22% of the variance. The second component had an eigenvalue of 1.101 and accounted for an additional 22.02% of the variance. The rotated component matrix showed that three tests loaded on the first component with loadings greater than .4: LLAMA B, $\lambda = .791$, LLAMA F, $\lambda = .790$, and LLAMA E, $\lambda = .639$. LLAMA D and the SRT task had loadings smaller than .4: LLAMA D, $\lambda = .351$, and SRT task, $\lambda = .219$. On the other hand, two tests loaded strongly on the second component: SRT task, $\lambda = .787$, and LLAMA D, $\lambda = .647$. LLAMA B, E, and F had loadings smaller than .4: LLAMA E, $\lambda = .295$, LLAMA B, $\lambda = -.030$, and LLAMA F, $\lambda = -.086$. The same pattern of results was obtained after applying a non-orthogonal rotation method (Direct Oblimin). This method, which allows factors to be correlated, showed that the correlation between the two components was -.093, indicating that there was no significant association between the two components and, therefore, that participants could have high ability in one component, but low ability in another, and vice-versa.

In addition to the four LLAMA subtests and the probabilistic SRT task, the study also included a general intelligence measure (GAMA) as part of the battery of tests. Although conventional general ability measures probably tap both explicit and implicit cognitive processes, several studies have found that they are highly correlated with attention-driven working memory measures (e.g., Engle et al., 1999; Kaufman et

whereas oblique rotation methods (e.g., Direct Oblimin, Promax) allow factors to be correlated.

al. 2010; Kyllonen, 1996; Kyllonen & Christal, 1990) and uncorrelated with measures such as probabilistic SRT tasks, claimed to tap implicit cognitive processes (e.g., Kaufman et al., 2010), and implicit memory measures of priming (e.g., Woltz, 1990, 1999). In order to uncover the underlying structure of all the cognitive measures in the present study and to see whether GAMA scores could be included as part of a composite measuring aptitudes for explicit or implicit learning, another principal components analysis was performed, including all cognitive measures (LLAMA B, LLAMA E, LLAMA F, LLAMA D, GAMA, and the SRT task). The analysis, conducted with Varimax rotation, yielded two principal components with eigenvalues greater than 1.0 that explained 57.34% of the total variance. The first component had an eigenvalue of 2.304 and accounted for 38.41% of the variance. The second component had an eigenvalue of 1.136 and accounted for an additional 18.94% of the variance. The rotated component matrix showed that the GAMA test loaded more strongly on the first component ($\lambda = .788$), together with LLAMA F ($\lambda = .749$), LLAMA B ($\lambda = .720$), and LLAMA E ($\lambda = .655$). Its loading on the second component, where LLAMA D and the SRT task kept loading more strongly ($\lambda = .660$ and $\lambda = .784$), was -.117, suggesting a negative association between intelligence and the second component. The same pattern of results was obtained via a non-orthogonal rotation, which further showed that the two components correlated at -.072.

On the basis of these results, and since GAMA scores correlated more strongly with LLAMA B, E, and F ($r = .56$, $p < .001$) than with LLAMA D and the SRT task ($r = .19$, $p = .04$), two equally weighted composite scores were created, one with GAMA, LLAMA B, LLAMA E, and LLAMA F scores, and one with LLAMA D

scores and sequence learning in the probabilistic SRT task (see Table 11 for the proposed underlying structure of cognitive aptitudes in this study). The two composite scores were created by converting each of the individual variables to z-scores in each of the groups. These scores were added and divided by the number of variables in the composite. The decision to create composite scores for each group separately was motivated by the fact that early and late L2 learners did not have comparable cognitive abilities. Early L2 learners performed significantly better in most cognitive measures (see Sections 5.1.1 and 5.1.2), which would have made the distribution of scores unbalanced across groups (i.e., a large number of participants would have been high-aptitude in the early AO group, but low-aptitude in the late AO group). The resulting composite scores were normally distributed, according to K-S tests: NS controls ($p = .886$ and $p = .833$), early L2 learners ($p = .629$ and $p = .956$), and late L2 learners ($p = .092$ and $p = .974$). The correlation between the two in each group was .06 ($p = .722$) in the control group, .17 ($p = .233$) in the early AO group, and .22 ($p = .119$) in the late AO group.

Table 11. *Cognitive Aptitudes*

Measures of Explicit Cognitive Processes	Measures of Implicit Cognitive Processes
LLAMA B (Vocabulary Learning)	LLAMA D (Sound Recognition)
LLAMA E (Sound-symbol Correspondence)	Probabilistic SRT Task (Implicit Learning)
LLAMA F (Grammatical Inferencing)	
GAMA (General Intelligence Test)	

Z-scores, which indicate the number of standard deviations above or below the mean, were used to divide participants in each of the groups into high-, mid-, and low-aptitude: *High* = *z*-scores >.5, *mid* = -.5 < *z*-scores < .5, and *low* = *z*-scores < -.5. Tables 12 and 13 display the average scores for each aptitude group. Participants in each of the aptitude levels were not the same across the two aptitude types. In the early AO group, six learners were high both in implicit and explicit language aptitude, eight were high only in implicit aptitude, and 12 were high in explicit aptitude. In the late AO group, six learners were high both in implicit and explicit language aptitude, 12 were high only in implicit aptitude, and seven were high in explicit aptitude. Therefore, there were more L2 learners who were high in implicit aptitude in the late AO group, perhaps indicating a sampling selection bias affecting proficient adult L2 learners who succeed in an immersion language learning context.

Table 12. *High-, Mid-, and Low-explicit Language Aptitude Groups (z-scores)*

	Control			Early AO			Late AO		
	High	Mid	Low	High	Mid	Low	High	Mid	Low
	n = 7	*n* = 8	*n* = 5	*n* = 18	*n* = 18	*n* = 14	*n* = 13	*n* = 24	*n* = 13
Explicit	.93	.01	-1.28	.99	.04	-1.32	1.11	.15	-1.40
Language	(.20)	(.32)	(.97)	(.38)	(.31)	(.46)	(.36)	(.26)	(.56)
Aptitude									

Note. Standard deviations appear between parentheses.

Table 13. *High-, Mid-, and Low-Implicit Language Aptitude Groups (z-scores)*

	Control			Early AO			Late AO		
	High	Mid	Low	High	Mid	Low	High	Mid	Low
	$n = 6$	$n = 6$	$n = 8$	$n = 14$	$n = 20$	$n = 16$	$n = 18$	$n = 16$	$n = 16$
Implicit	1.57	-.09	-1.11	1.17	.09	-1.14	1.47	-.05	-1.48
Language	(.63)	(.27)	(.76)	(.55)	(.32)	(.45)	(.85)	(.26)	(.75)
Aptitude									

Note. Standard deviations appear between parentheses.

5.2 Language Attainment

Overall performance for each speaker group on the six language tests (timed visual GJT, untimed visual GJT, timed auditory GJT, untimed auditory GJT, metalinguistic knowledge test, and word monitoring task) is presented in this section. Results are reported for grammatical and ungrammatical items first and, then, for ungrammatical items only, except in the case of the word monitoring task, which does not provide an interpretable measure for ungrammatical items.

5.2.1 Grammaticality Judgment Tests

Tables 14 and 15 display the groups' scores on the timed and untimed visual (Table 11) and auditory GJTs (Table 12). All dependent variables were normally distributed in each group, according to K-S tests ($p > .05$). There were no extreme outliers with values +/- 3 standard deviations from each group's mean. The three groups were significantly different from one another on the four GJTs, according to Bonferroni-adjusted comparisons. NS controls scored significantly higher than early

and late L2 learners on each of the GJTs ($p < .001$ and $p < .001$, respectively) and

early L2 learners scored significantly higher than late L2 learners ($p < .001$).

Table 14. *Group Mean Percentage Scores on Timed and Untimed Visual GJTs*

Group	Timed Visual GJT			Untimed Visual GJT		
	M	*SD*	*Range*	*M*	*SD*	*Range*
Control	84.21	7.65	67.31-	90.25	4.30	81.67-
($n = 20$)			98.21			98.33
Early AO	72.03	10.18	50.00-	76.77	9.13	53.33-
($n = 50$)			90.00			96.67
Late AO	57.88	8.39	41.51-	61.27	9.05	46.67-
($n = 50$)			71.74			91.67

Table 15. *Group Mean Percentage Scores on Timed and Untimed Auditory GJTs*

Group	Timed Auditory GJT			Untimed Auditory GJT		
	M	*SD*	*Range*	*M*	*SD*	*Range*
Control	92.04	5.13	78.57-	93.25	6.34	76.67-
($n = 20$)			98.33			100.00
Early AO	76.24	10.07	58.33-	79.39	9.73	56.67-
($n = 50$)			94.92			98.33
Late AO	57.63	7.90	38.33-	60.47	9.28	38.33-
($n = 50$)			78.38			85.00

In terms of test modality (auditory/visual), GJT scores in the group of NS controls were higher in the auditory than visual modality, and the same pattern was observed in the early AO group, whereas, in the late AO group, scores were almost the same in the two modalities. In terms of time pressure (timed/untimed), the pattern of results was the same in the three groups: scores were higher on untimed than timed GJTs, with a larger difference in the visual than auditory modalities. Figure 8 provides a visual comparison of the three groups of participants.
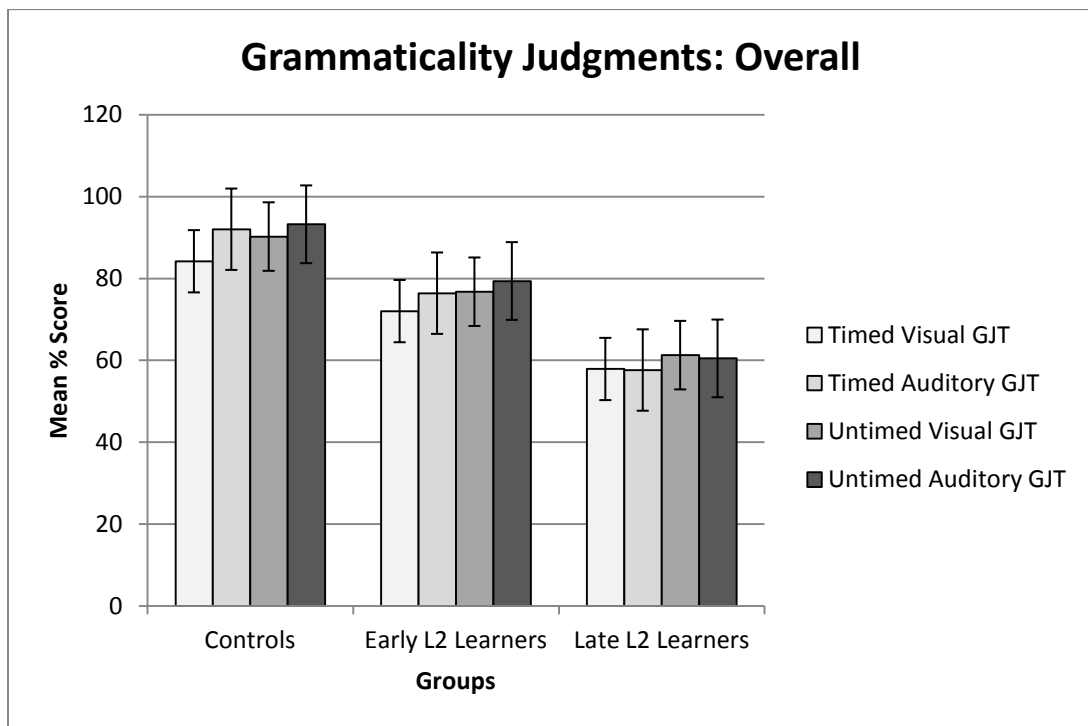
Figure 8. Group mean percentage GJT scores

A repeated-measures ANOVA with Modality and Time as within-subjects factors and Group as a between-subjects factor confirmed the descriptive results. Box's test did not reach the .050 level ($p = .046$), indicating a violation of the equality of covariances assumption, a common violation in behavioral studies that compare groups with very different variances (e.g., NSs vs. L2 learners). This suggested that

participants did not maintain their relative standing in the different treatment

conditions. Equality of variances, according to Levene's test, was met for the timed

visual GJT ($p = .093$) and the untimed auditory GJT ($p = .077$), but not for the timed

auditory ($p = .032$) and untimed visual GJTs ($p = .006$), indicating that scores did not

show equal variability across groups. However, the largest standard deviation was

less than three times the smallest standard deviation and, therefore, ANOVA was

considered robust (see Footnote 12).

The analysis yielded a significant main effect for Time ($F(1, 117) = 34.130$, $p <$

.001, $\eta_p^2 = .227$, $\lambda = .773$) and Modality ($F(1, 117) = 19.181$, $p < .001$, $\eta_p^2 = .142$, $\lambda =$

.858), as well as a significant interaction between Group and Modality ($F(2, 117) =$

8.509, $p < .001$, $\eta_p^2 = .128$, $\lambda = .872$), but not between Group and Time ($F(2, 117) =$

.360, $p = .698$, $\eta_p^2 = .006$, $\lambda = .994$). These results indicated that the entire set of

participants scored significantly higher on untimed tests, keeping modality constant,

and higher on auditory tests, keeping time pressure constant. However, modality was

further qualified by an interaction with group. Early L2 learners and NSs scored

higher on auditory tests than on visual tests, whereas late L2 learners scored higher on

visual than auditory tests.

When only correct responses to ungrammatical items[14] were considered (see

Tables 16 and 17, as well as Figure 9 for a visual depiction of the results), the three

groups also scored significantly different from one another ($p < .001$). Score

differences according to GJT modality and time increased, especially in the late AO

---

[14] The error in an ungrammatical item is the most likely reason for rejection of an item (DeKeyser, 2000), whereas acceptance of a grammatical item can respond to a variety of reasons. Although the margin of error becomes smaller when only ungrammatical items are considered, there is also a loss of power, resulting from the smaller sample of items considered.

group. While NS controls and early L2 learners scored higher in the auditory modality, late L2 learners scored higher in the visual modality. Also, the difference between timed and untimed test scores became larger in the late AO group. Late L2 learners' scores on the ungrammatical items of the untimed visual and untimed auditory GJTs were around 20% higher than on the timed versions of the tests. Therefore, time pressure had detrimental effects on late L2 learners' performance, regardless of the modality of the measure.

Table 16. *Group Mean Percentage Scores on Timed and Untimed Visual GJTs (Ungrammatical Items)*

Group	Timed Visual GJT			Untimed Visual GJT		
	M	*SD*	*Range*	*M*	*SD*	*Range*
Control	73.07	15.63	38.46-	86.33	7.79	70.00-
n = 20			100.00			96.67
Early AO	52.24	18.28	4.55-	64.07	17.28	10.00-
n = 50			83.33			93.33
Late AO	35.31	16.85	10.00-	55.40	19.12	16.67-
n = 50			88.89			90.00

Table 17. *Group Mean Percentage Scores on Timed and Untimed Auditory GJTs*

(Ungrammatical Items)

Group	Timed Auditory GJT			Untimed Auditory GJT		
	M	*SD*	*Range*	*M*	*SD*	*Range*
Control	86.10	9.70	62.96-	91.00	10.60	63.33-
n = 20			100.00			100.00
Early AO	57.10	19.12	20.00-	65.99	17.70	20.00-
n = 50			89.66			96.67
Late AO	28.25	15.43	00.00-	45.80	14.46	10.00-
n = 50			69.23			83.33

Figure 9. Group mean percentage GJT scores (ungrammatical items)

The repeated-measures ANOVA for ungrammatical items also violated the assumption of equality of covariance matrices ($p = .002$). Error variances were equal, according to Levene's tests, in the untimed auditory GJT ($p = .055$), the timed visual GJT ($p = .406$), and the timed auditory GJT ($p = .050$), but unequal in the untimed visual GJT ($p = .001$). However, the largest standard deviation was less than three times the smallest standard deviation and, therefore, ANOVA was considered robust. The analysis yielded a significant two-way interaction between Modality and Group ($F(2, 117) = 25.803$, $p < .001$, $\eta_p^2 = .308$, $\lambda = .692$) and also between Time and Group ($F(2, 117) = 6.684$, $p = .002$, $\eta_p^2 = .103$, $\lambda = .897$). Modality and Time, however, did not interact, either overall ($F(2, 117) = 3.096$, $p = .081$, $\eta_p^2 = .026$, $\lambda = .974$) or in any of the groups (i.e., there was no three-way interaction, $F(2, 117) = .174$, $p = .840$, $\eta_p^2 = .003$, $\lambda = .997$), indicating that the two modalities were similarly affected by time pressure. Figures 10 and 11 illustrate the two two-way interactions between Modality and Group (Figure 10) and Time and Group (Figure 11).

These results showed that, unlike NSs and early L2 learners, late L2 learners obtained higher scores on the visual GJT modalities. They also scored proportionally higher on untimed GJTs, when their performance is compared against NSs and early L2 learners, for whom the difference between scores on timed and untimed GJTs was not that large. Time pressure and modality had an effect on performance as separate factors, but not as combined factors, as the lack of an interaction between modality and time indicated. The effect of time pressure on performance was comparable across test modalities, and vice-versa, in all the groups. Therefore, performance on auditory and visual formats was similarly affected by whether the test was timed or

untimed, and performance on timed and untimed formats was similarly affected by

whether the test was auditory or visual.
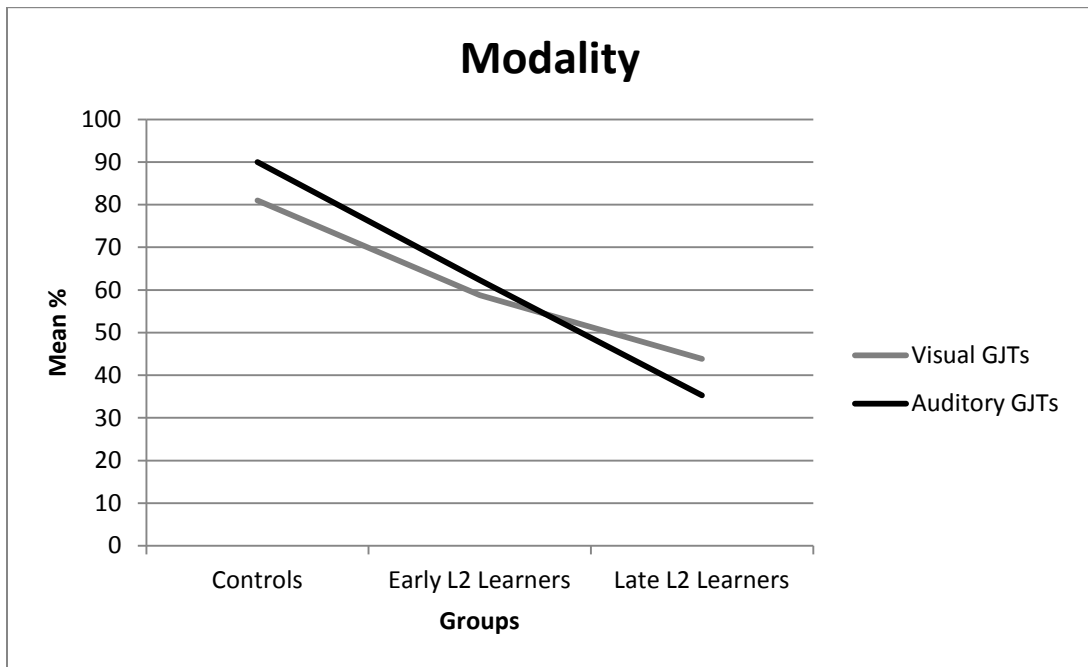
Figure 10. Modality x Group interaction
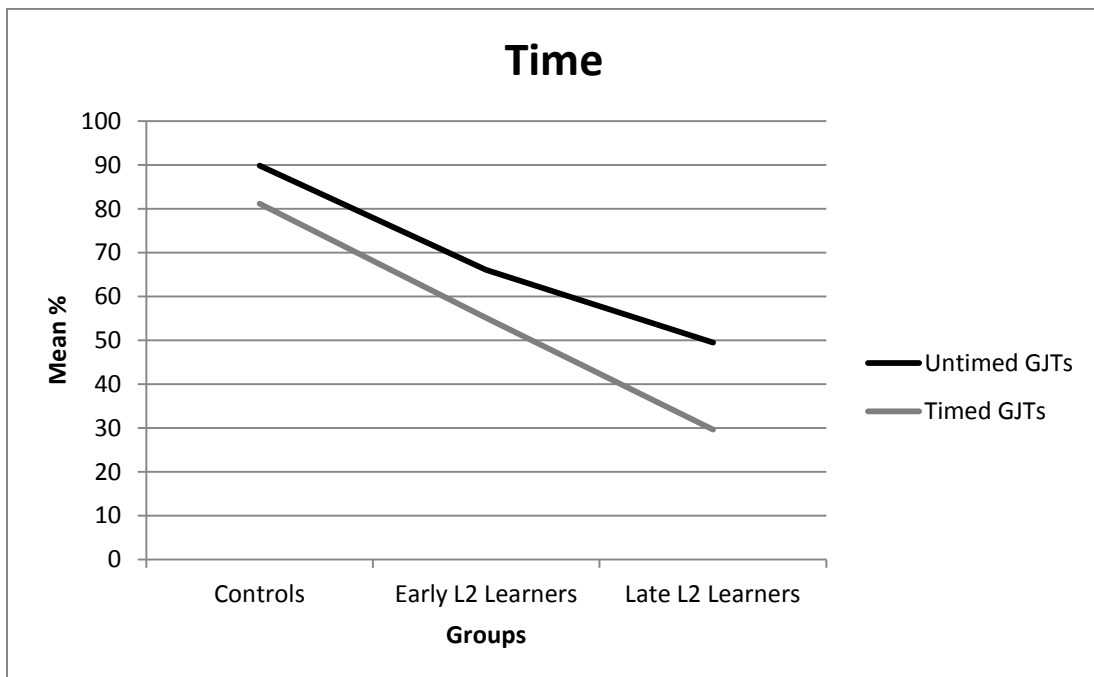
Figure 11. Time x Group interaction

5.2.2 Metalinguistic Knowledge Test

Table 18 shows the descriptive statistics for all the items in the metalinguistic test and Table 19 for ungrammatical items only (i.e., items that were successfully corrected and errors that were correctly explained). All dependent variables in each of the groups were normally distributed, according to K-S tests ($p > .05$). There were no extreme outliers with values +/- 3 standard deviations from each group's mean. The three groups were significantly different from one another, according to Bonferroni-adjusted pairwise comparisons. NS controls scored significantly higher than early and late L2 learners overall and on ungrammatical items only ($p < .001$) and early L2 learners scored significantly higher than late L2 learners ($p < .001$).

Table 18. *Group Mean Percentage Scores on the Metalinguistic Knowledge Test*

Group	Metalinguistic Test	
	M	*Range*
Control ($n = 20$)	89.75 (5.75)	76.67-96.67
Early AO ($n = 50$)	77.20 (9.84)	56.67-91.67
Late AO ($n = 50$)	63.50 (10.65)	43.33-88.33

Note. Standard deviations appear between parentheses.

Table 19. *Group Mean Percentage Scores on the Metalinguistic Test (Ungrammatical Items)*

Group	Metalinguistic Knowledge Test			
	Error Correction		Error Explanation	
	M	*Range*	*M*	*Range*
Control (*n* = 20)	80.67 (10.63)	53.33-93.33	75.35 (20.23)	31.25-100.00
Early AO (*n* = 50)	57.20 (18.85)	13.33-83.33	64.60 (21.32)	0.00-100.00
Late AO (*n* = 50)	36.33 (20.15)	3.33-80.00	73.46 (26.11)	0.00-100.00

Note. Standard deviations appear between parentheses.

Metalinguistic test scores (overall and on ungrammatical items) are displayed on Figures 12 and 13. Figure 14 further shows the average proportion of explained errors. As can be seen, the three groups were able to correct more errors than they could explain. NSs and late L2 learners were the groups that, proportionally, could explain a larger number of errors, although group differences were not significant, according to Scheffé posthoc tests, between late L2 learners and controls ($p = .954$), late and early L2 learners ($p = .168$), or controls and early L2 learners ($p = .223$). Since group differences were not significant, the fact that NSs had the largest percentage of correct grammatical explanations was interpreted as being due to chance. There were no features in the profile of the NSs that could account for their higher metalinguistic test scores. They were all studying or had studied university

degrees in a variety of disciplines (science, education, business, etc.), but were not linguistically-trained.
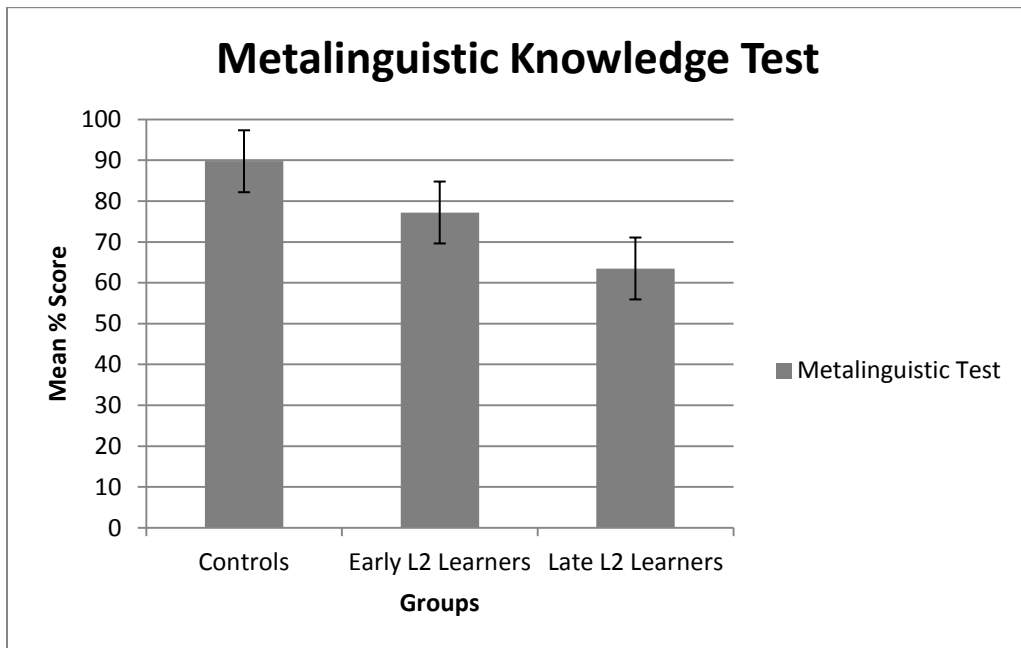
Figure 12. Group mean percentage scores on the metalinguistic knowledge test
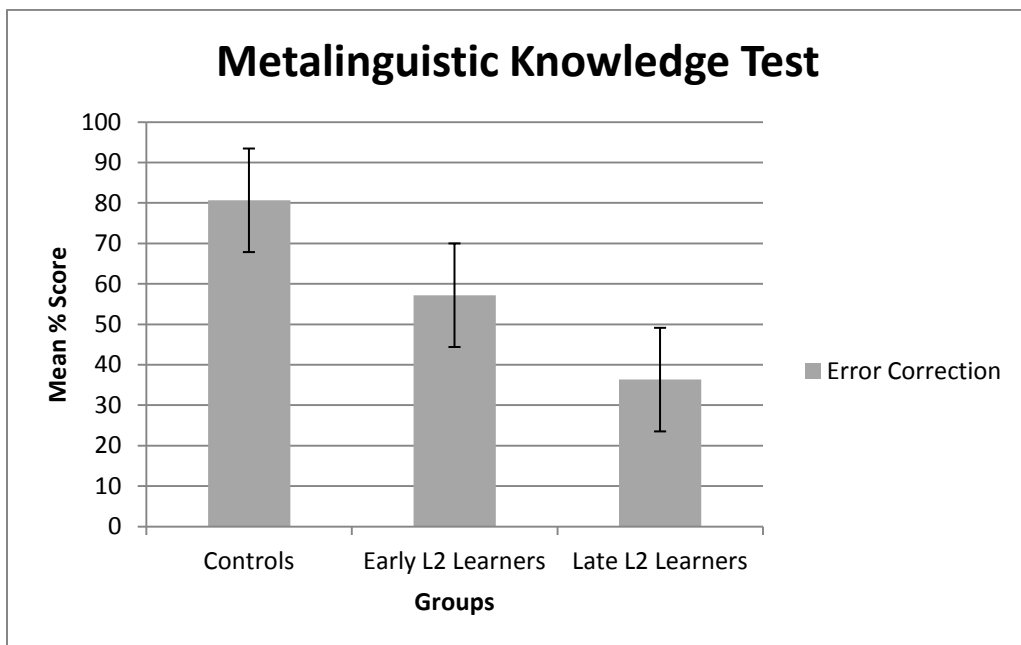
Figure 13. Group mean percentage scores on the metalinguistic knowledge test (correction of ungrammatical items)
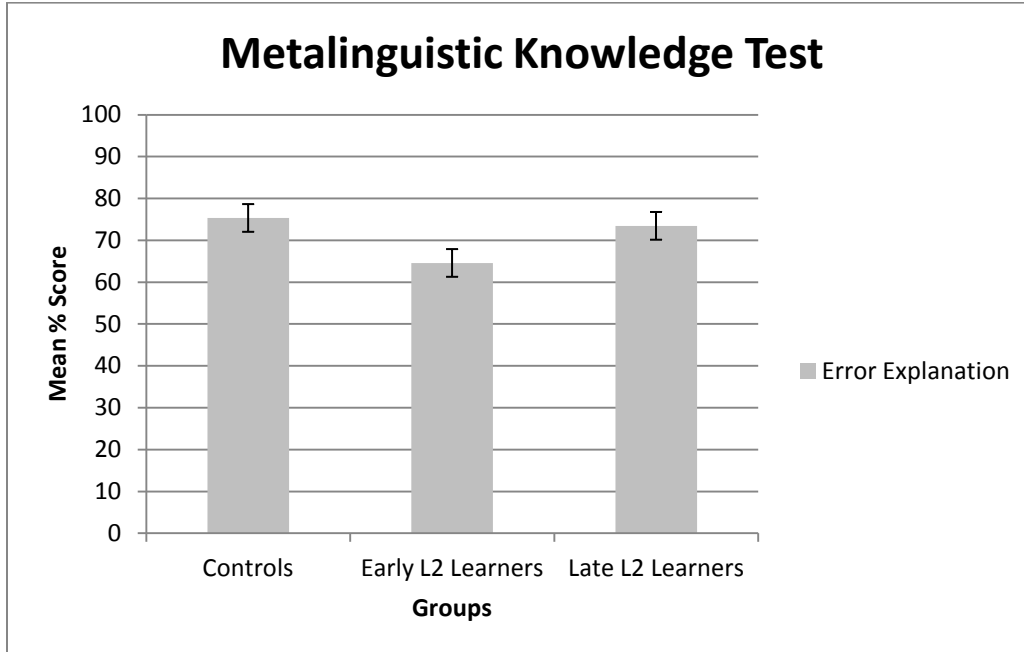
Figure 14. Group mean percentage scores on the metalinguistic knowledge test (explanation of ungrammatical items)

Finally, Figure 15 compares participants' performance on the metalinguistic knowledge test and the four GJTs. As reported in section 5.2.1, the three speaker groups were significantly different from one another on the four GJTs, and they were also significantly different on the metalinguistic knowledge test, according to a MANCOVA analysis ($F(10, 225) = 21.726$, $p < .001$, $\eta_p^2 = .492$, $\lambda = .258$). The NS control group scored significantly higher than the early and late AO groups, and the early AO group scored significantly higher than the late AO group. All pairwise Bonferroni-adjusted comparisons were $p < .001$.

Language Attainment: Overall
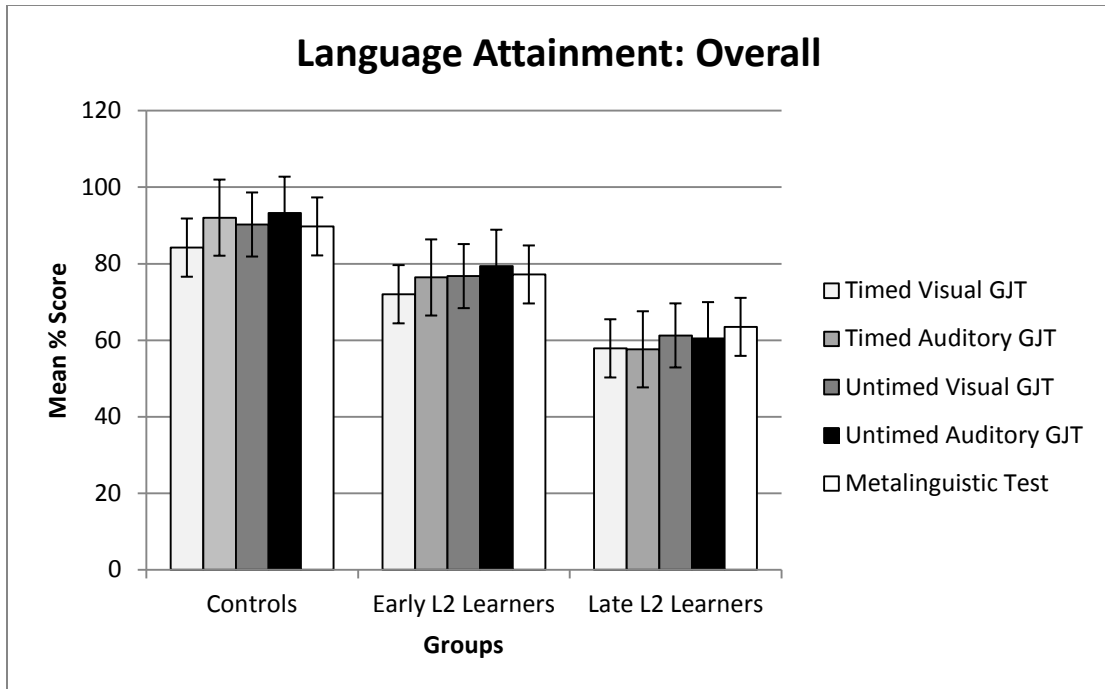
Figure 15. Group mean percentage scores on the four GJTs and the metalinguistic knowledge test.

Although, quantitatively speaking, NS controls and early L2 learners were significantly different (NSs scored between 12% and 14% higher on average on each of the tests), they shared the same pattern of scores, qualitatively speaking. Thus, both groups scored the highest on the untimed auditory GJT and the lowest on the timed visual GJT. Scores on the timed visual GJT were significantly lower than scores on the other tests within each of the two groups ($p < .05$), according to repeated-measures ANOVAs. In addition, NSs scored higher on the untimed auditory GJT than on the metalinguistic test ($p = .007$) and this comparison was marginally significant within the early AO group ($p = .073$). Unlike NSs and early L2 learners, late L2 learners scored the highest on the metalinguistic knowledge test and the lowest on the

two timed GJTs (visual and auditory). Their scores on the metalinguistic test were significantly higher than on the timed visual ($p = .014$) and auditory ($p = .001$) GJTs.

5.2.3 Word monitoring Task

Table 20 shows the descriptive statistics for word monitoring latencies on grammatical and ungrammatical critical items. A grammatical sensitivity index was created by subtracting latencies on ungrammatical items from latencies on grammatical items (see Table 21). Latencies were normally distributed in the NS control group ($p = .189$ and $p = .186$) according to K-S tests, but not in the early L2-learner group ($p = .003$ and $p = .001$) or late L2-learner group ($p = .011$ and $p = .029$). Visual inspection of boxplots showing the distribution of latencies in each group indicated the presence of five outliers in the early AO group and three outliers in the late AO group (see Figures 16 and 17).
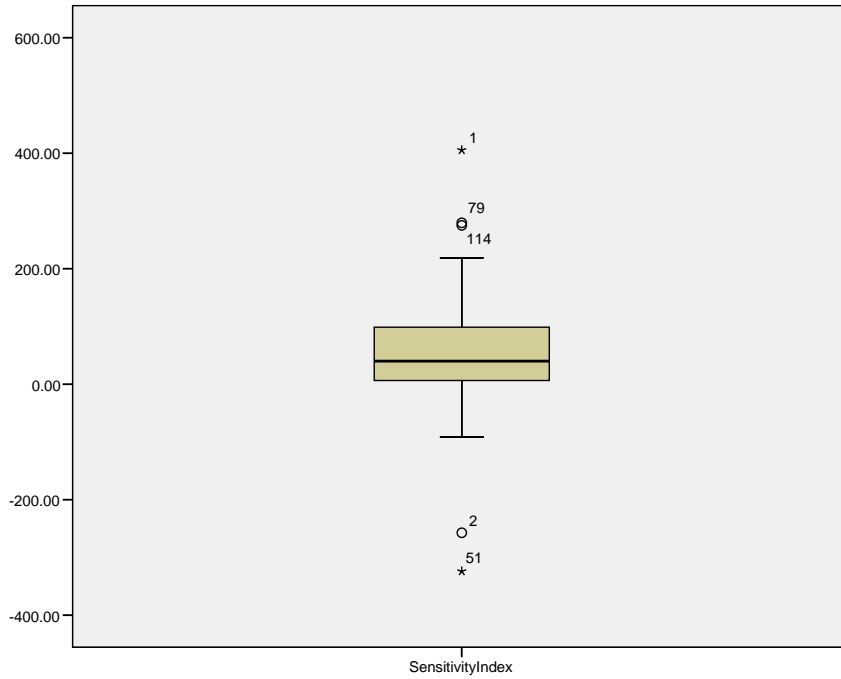
Figure 16. Distribution of overall word monitoring latencies in the early AO group
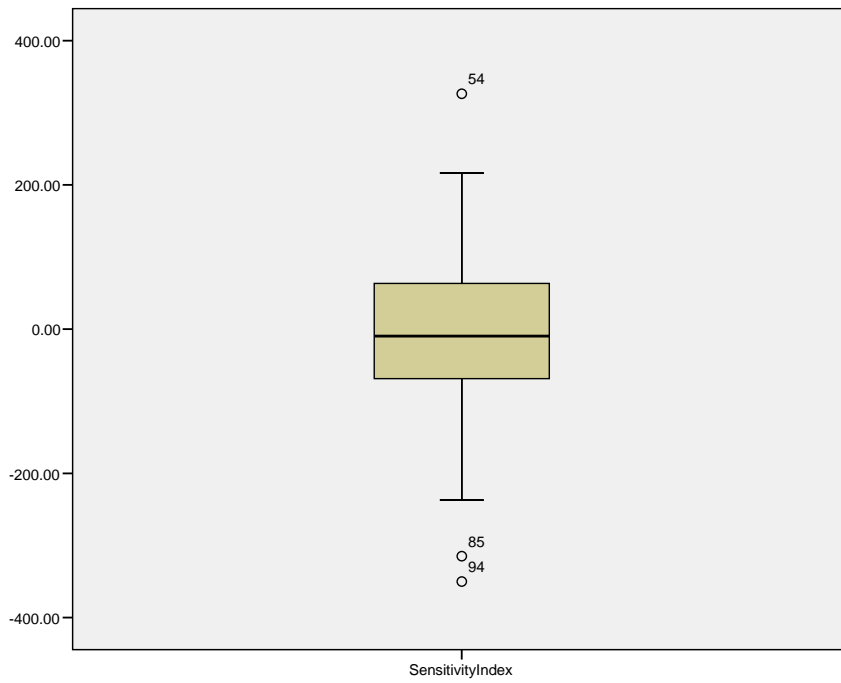
Figure 17. Distribution of overall word monitoring latencies in the late AO group

After removing these cases, normality was met for grammatical and ungrammatical items in the early AO group ($p = .103$ and $p = .237$), as well as for ungrammatical items in the late AO group ($p = .146$). Grammatical items in the late AO group approached normality ($p = .047$). The index computed as a measure of grammatical sensitivity was normally distributed in each of the groups: NS controls ($p = .998$), early L2 learners ($p = .195$) and late L2 learners ($p = .787$).

Table 20. *Word Monitoring Mean Latencies*

Group	Ungrammatical Items		Grammatical Items	
	M	*Range*	*M*	*Range*
Control	1105.66	677.41-	1043.47	646.67-
($n = 20$)	(418.29)	2629.14	(451.27)	2782.33
Early AO	920.92	743.07-	875.09	693.41-
($n = 45$)	(164.47)	1712.83	(177.24)	1799.60
Late AO	1704.41	796.72-	1711.95	827.70-
($n = 47$)	(755.58)	3202.62	(766.98)	3341.86

Note. Standard deviations appear between parentheses.

Table 21. *Grammatical Sensitivity Index (GSI)*

Group	GSI	
	M	*Range*
Control (*n* = 20)	62.19 (96.21)	-153.20-234.55
Early AO (*n* = 45)	45.83 (69.11)	-91.26-218.93
Late AO (*n* = 47)	-7.54 (101.08)	-237.03-216.46

Note. Standard deviations appear between parentheses.

A total comprehension score was computed on the basis of correct responses to the set of randomly distributed yes/no questions included in the word monitoring task. In order to ensure that participants had been focusing their attention on meaning while performing the task, a minimum of 75% response accuracy was required from each participant to be included in the analysis (see section 4.3.1 for rationale). No participant had an error rate higher than 25%. In the NS control group, the mean percentage response accuracy was 95.42% (*SD* = 3.0) (4.58% error rate), in the early L2-learner group it was 92.23% (*SD* = 4.2) (7.77% error rate), and, in the late L2-learner group 85.76% (*SD* = 7.06) (14.24% error rate).

Word monitoring latencies in the NS and early L2 learner groups were higher for ungrammatical items, indicating a delay in participants' word monitoring when the sentence included a grammatical error. In the late L2 learner group, mean latencies for grammatical and ungrammatical items were very similar and, even slightly higher for grammatical items, which yielded a negative GSI in this group. There was, however, considerable individual variation in GSIs among late L2 learners, as shown by the maximum and minimum GSI values.

Group monitoring latencies were compared in a 2x3 mixed factorial ANOVA. The model included a repeated factor with two levels (grammatical and ungrammatical) and a between-subjects factor with three levels (controls, early L2 learners, and late L2 learners). The assumptions of equality of covariance matrices and error variances were not met ($p < .001$). As a remedial measure (see footnote 12), and given that the largest standard deviation was more than four times the smallest standard deviation, a more stringent .01 alpha was adopted. Results[15] revealed that grammaticality was a significant factor ($F(1,110) = 13.777$, $p < .001$, $\eta_p^2 = .111$, $\lambda = .889$), suggesting overall differential sensitivity according to the grammaticality of the item. The average reaction time difference between grammatical and ungrammatical items was 33.49 milliseconds. The interaction between grammaticality and group was also statistically significant ($F(2,110) = 6.216$, $p = .003$, $\eta_p^2 = .102$, $\lambda = .898$): NSs and early L2 learners' reaction times were higher on ungrammatical items ($p = .009$ and $p = .005$, respectively), indicating group sensitivity to grammatical violations, whereas late L2 learners' reaction times were higher on grammatical items (almost overlapping with ungrammatical items), suggesting same sensitivity to both types of items as a group ($p = .677$) (see Figure 16).

---

[15] The results including outliers also showed that grammaticality was a significant main factor ($F(1,117) = 9.149$, $p = .003$, $\eta_p^2 = .073$, $\lambda = .927$) and that the two-way interaction with group was significant ($F(2,117) = 4.048$, $p = .020$, $\eta_p^2 = .065$, $\lambda = .935$).

Word Monitoring Task
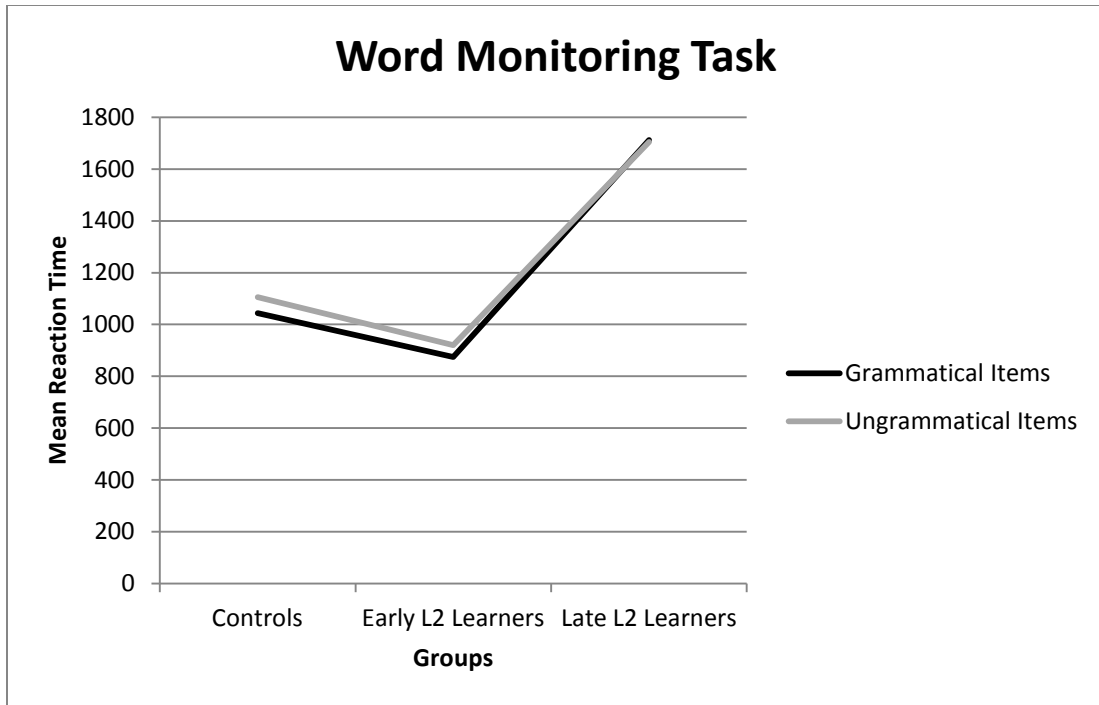
Figure 18. Group word monitoring latencies for grammatical and ungrammatical items

A similar pattern of results was found when the data were separated into agreement structures (gender agreement, number agreement, and person agreement) and non-agreement structures (aspect contrasts, the passive, and the subjunctive). Latencies for grammatical and ungrammatical items were again normally distributed in the NS control group ($p = .051$ and $p = .477$ for agreement structures, and $p = .071$ and $p = .230$ for non-agreement structures), but not normally distributed in the early L2 learner group ($p = .001$ and $p < .001$ for agreement structures, and $p = .001$ and $p = .016$ for non-agreement structures) or late L2-learner group ($p = .025$ and $p = .032$ for agreement structures, and $p = .018$ and $p = .020$ for non-agreement structures). Since non-normality may be caused by the presence of one or more outliers, the distribution of the data was visually inspected.

Visual inspection of boxplots showing the distribution of latencies for grammatical and ungrammatical items in each group indicated the presence of five outliers for agreement structures and four outliers for non-agreement structures in the early AO group (see Figures 19 and 20).
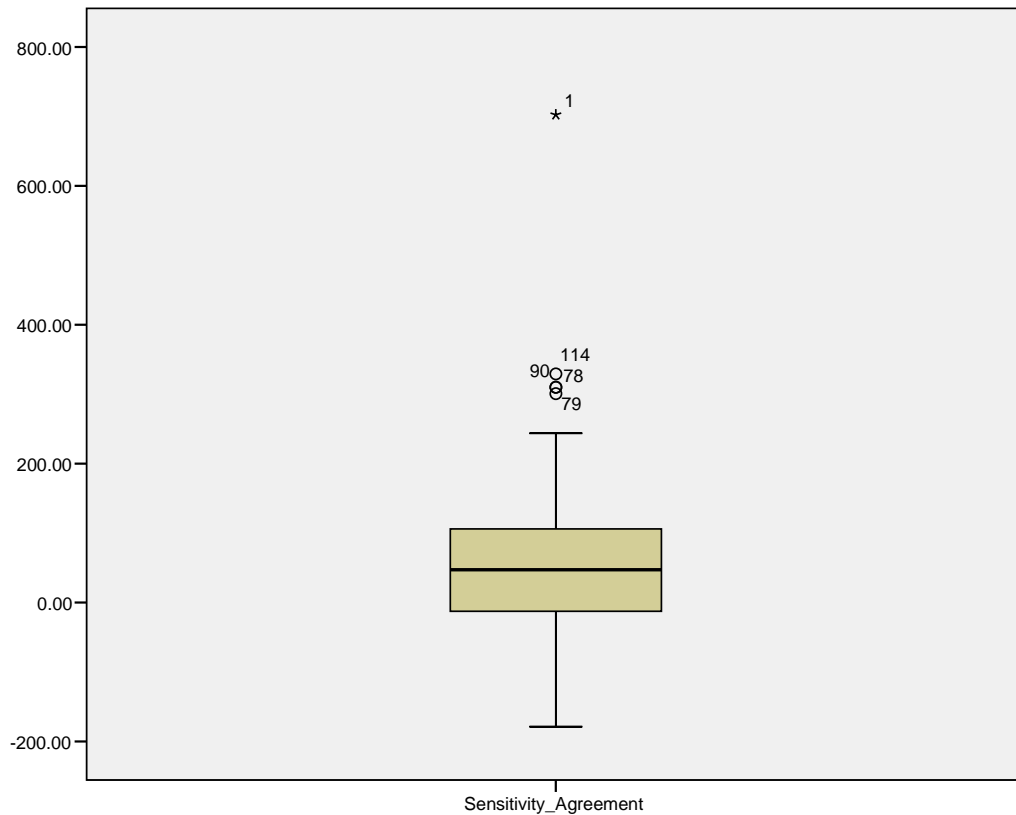
Figure 19. Distribution of word monitoring latencies for agreement items in the early AO group
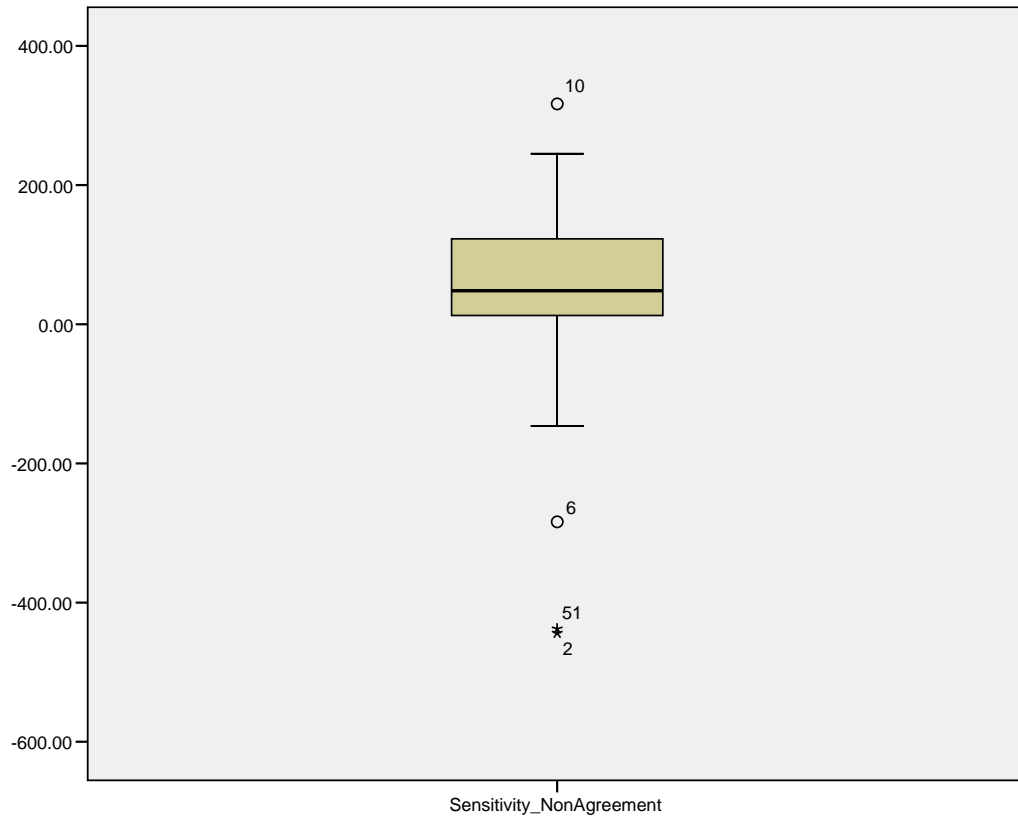
Figure 20. Distribution of word monitoring latencies for non-agreement items in the early AO group

After removing these cases, normality was met for grammatical and ungrammatical agreement items ($p = .981$ and $p = .826$) and non-agreement items ($p = .263$ and $p = .065$). In the late AO group, there were also five outliers for agreement structures and three for non-agreement structures (see Figures 21 and 22).
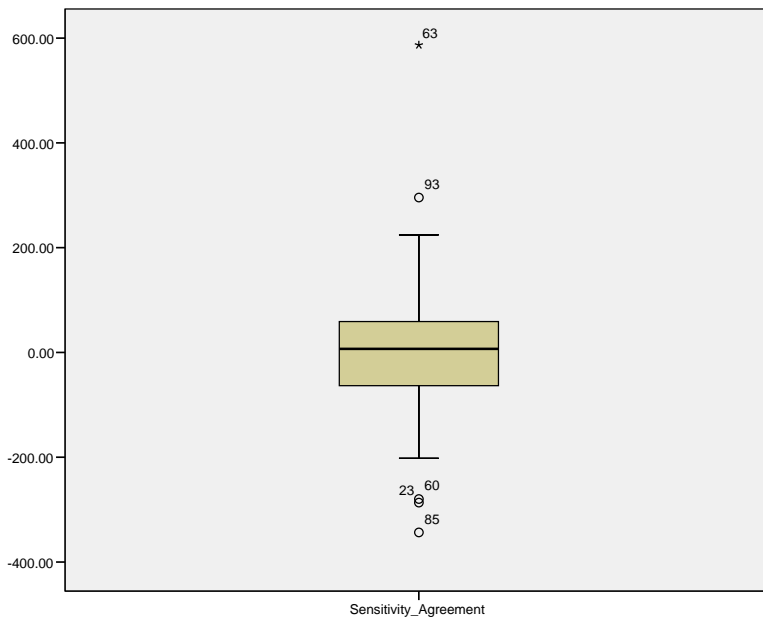
Figure 21. Distribution of word monitoring latencies for agreement items in the late
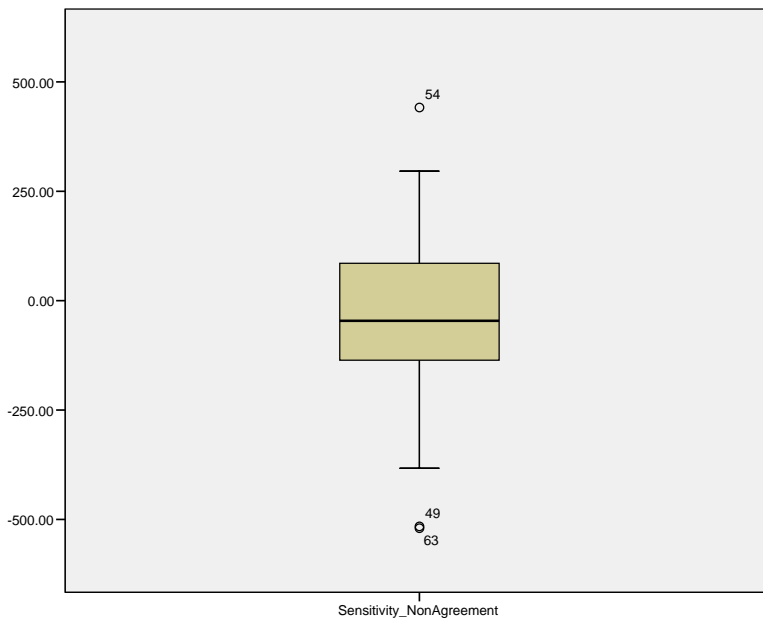
AO group

Figure 22. Distribution of word monitoring latencies for non-agreement items in the

late AO group

When they were removed, normality was met for ungrammatical items, both agreement and non-agreement ($p = .068$ and $p = .053$). Normality could only be approached for grammatical agreement and non-agreement items ($p = .043$ and $p = .042$), but ANOVA is considered robust to mild violations of the normality assumption.

The resulting GSIs for agreement and non-agreement items were all normally distributed in each of the groups: NS controls ($p = .798$ and $p = .801$), early L2 learners ($p = .103$ and $p = .280$), and late L2 learners ($p = .394$ and $p = .842$).

Table 22 shows the descriptive statistics for word monitoring latencies on grammatical and ungrammatical agreement items, and Table 23 the resulting GSIs.

Table 22. *Word monitoring Mean Latencies (Agreement Structures)*

Group	Ungrammatical Items		Grammatical Items	
	M	*Range*	*M*	*Range*
Control	1113.23	685.17-	1032.02	718.33-
($n = 20$)	(256.65)	1625.73	(247.44)	1653.20
Early AO	989.45	776.07-	958.81	695.90-
($n = 45$)	(304.68)	2708.47	(326.44)	2779.20
Late AO	1679.06	836.72-	1673.33	876.50-
($n = 45$)	(747.28)	3315.47	(745.25)	3208.52

Note. Standard deviations appear between parentheses.

Table 23. *Grammatical Sensitivity Index Agreement Structures*

	GSI Agreement	
	M	*Range*
Control (*n* = 20)	81.21 (114.40)	-108.13-267.35
Early AO (*n* = 45)	30.64 (84.41)	-178.73-243.67
Late AO (*n* = 45)	5.73 (104.07)	-201.93-224.35

Note. Standard deviations appear between parentheses.

A 2x3 mixed factorial ANOVA showed that grammaticality was a significant factor ($F(1,107) = 14.442$, $p < .001$, $\eta_p^2 = .120$, $\lambda = .880$). The average reaction time difference between grammatical and ungrammatical agreement items was 39.19 milliseconds. The interaction between grammaticality and group was also significant ($F(2,107) = 3.822$, $p = .025$, $\eta_p^2 = .067$, $\lambda = .933$) (see Figure 23). Differences in word monitoring latencies between grammatical and ungrammatical items were statistically significant in the NS group ($t(19) = 2.327$, $p = .032$) and early L2-learner group ($t(44) = 2.462$, $p = .018$), but not in the late L2 learner group ($t(44) = .370$, $p = .713$). This indicated that NSs and early L2 learners experienced involuntary delays in their responses when sentences included errors of gender, person, or number agreement. On the other hand, late L2 learners did not show the same grammatical sensitivity as a group. Their word monitoring latencies for grammatical items overlapped with ungrammatical items, indicating that grammatical violations involving agreement relations did not affect their reaction times.
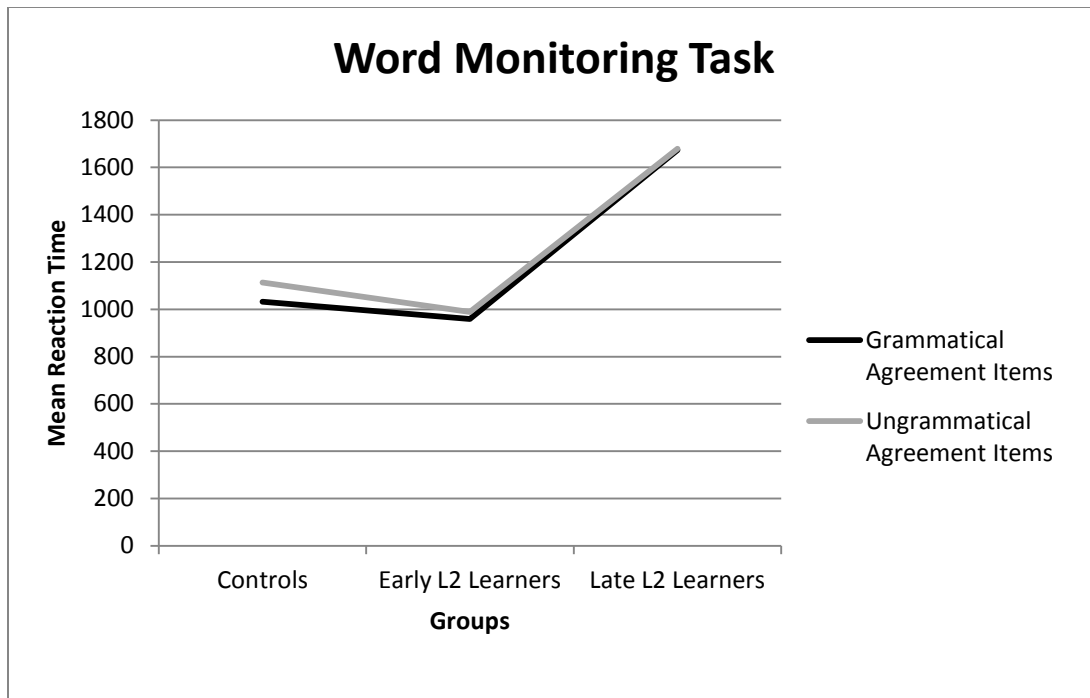
Figure 23. Group word monitoring latencies for grammatical and ungrammatical

items testing agreement structures (gender, person, and number agreement)

Table 24 shows the descriptive statistics for word monitoring latencies on

grammatical and ungrammatical non-agreement items, and Table 25 the resulting

GSIs.

Table 24. *Word monitoring Mean Latencies (Non-Agreement Structures)*

Group	Ungrammatical Items		Grammatical Items	
	M	*Range*	*M*	*Range*
Control	956.21 (198.13)	664.80-1404.47	874.33 (169.56)	575.00-1203.97
(*n* = 20)				
Early AO	908.43 (201.39)	705.20-1690.52	848.99 (172.30)	653.22-1533.07
(*n* = 46)				
Late AO	1664.31(749.89)	759.13-3131.60	1687.93(768.56)	755.47-3514.55
(*n* = 47)				

Note. Standard deviations appear between parentheses.

Table 25. *Grammatical Sensitivity Index (Non-Agreement Structures)*

	GSI Non-agreement	
	M	*Range*
Control (*n* = 20)	81.88 (96.26)	-109.77-256.15
Early AO (*n* = 46)	59.44 (86.78)	-146.00-245.07
Late AO (*n* = 47)	-23.62 (153.66)	-382.95-296.00

Note. Standard deviations appear between parentheses.

A 2x3 mixed factorial ANOVA showed that grammaticality was a significant factor ($F(1,110) = 9.909$, $p = .002$, $\eta_p^2 = .083$, $\lambda = .917$). The average reaction time difference between grammatical and ungrammatical non-agreement items was 39.23 milliseconds. The interaction between grammaticality and group was also significant

$(F(2,110) = 7.781, p = .001, \eta_p^2 = .124, \lambda = .876)$: NSs and early L2 learners' word

monitoring latencies were higher on ungrammatical items ($t(19) = 2.789, p = .012$ and

$t(45) = 4.645, p < .001$, respectively), whereas late L2 learners' latencies on

grammatical and ungrammatical items were not significantly different ($t(46) = -1.065$,

$p = .292$) (see Figure 24). This indicated sensitivity to errors involving the

subjunctive, the passive, and aspect contrasts in the NS control and early AO groups,

but lack of sensitivity in the late AO group.
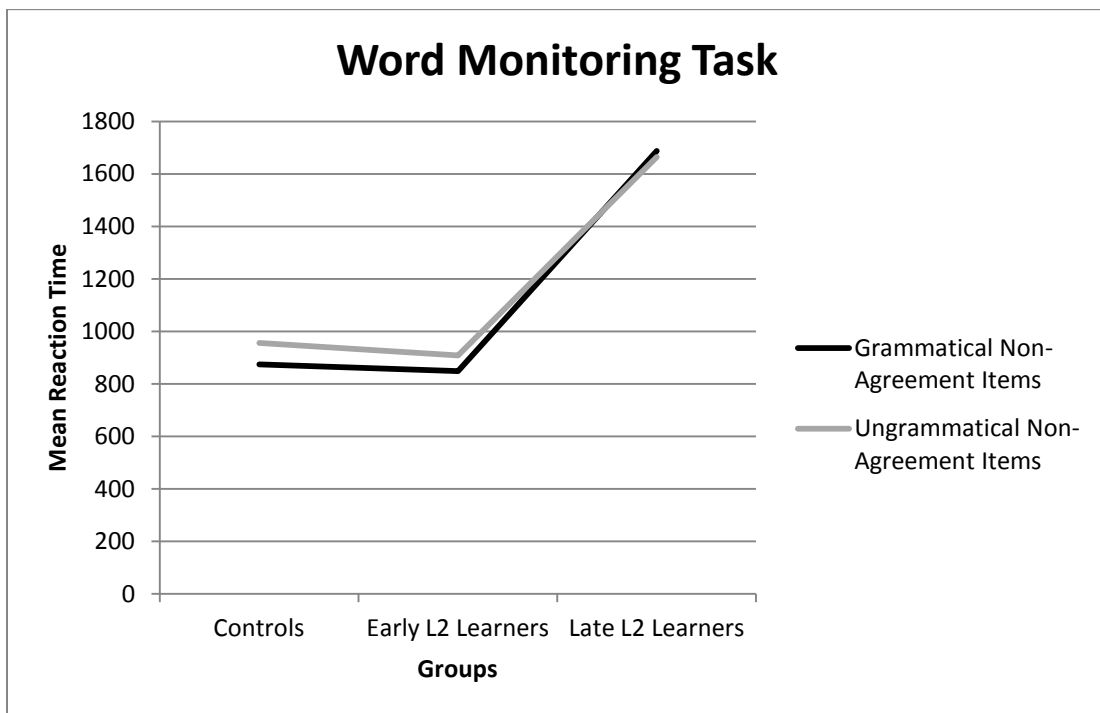
Figure 24. Group word monitoring latencies for grammatical and ungrammatical

items testing non-agreement structures (aspect, the subjunctive, and the passive)

Sensitivity to agreement and non-agreement structures was comparable in each of

the groups, as indicated by non-significant differences between the two GSIs in the

control group ($t(19) = -.339, p = .739$), early AO group ($t(40) = -.896, p = .376$), and

late AO group ($t(41) = .580, p = .565$). A between-subjects analysis (ANOVA)

further revealed that group was a significant factor in both the GSI for agreement structures ($F(2,107) = 4.697$, $p = .011$, $\eta_p^2 = .077$) and non-agreement structures ($F(2,110) = 7.781$, $p = .001$, $\eta_p^2 = .124$). Bonferroni-adjusted comparisons further indicated that NS controls were not significantly different from early L2 learners on either GSI for agreement ($p = .195$) or non-agreement ($p = .795$), but significantly different from late L2 learners on both ($p = .022$ and $p = .007$, respectively). Early and late L2 learners' sensitivity to non-agreement structures was also significantly different ($p = .005$), but their sensitivity to agreement structures was comparable ($p = .706$). Sensitivity to agreement structures could not, therefore, discriminate between early and late L2 learners, indicating that it is a feature that early and late acquisition may have in common, even though early L2 learners did not differ from NSs, either[16].

Finally, group was also a significant factor in overall grammatical sensitivity ($F(2,110) = 6.216$, $p = .003$, $\eta_p^2 = .102$). Multiple comparisons showed that NS controls were not significantly different from early L2 learners ($p = .791$), but significantly different from late L2 learners ($p = .012$). Early and late L2 learners were also significantly different ($p = .014$). These results at the between-subjects level confirmed the patterns observed at a within-subjects level. NSs and early L2 learners were highly sensitive to grammatical errors while monitoring words in a comprehension task. They also displayed comparable amounts of sensitivity. On the

---

[16] Interestingly, early L2 learners' performance on agreement structures also resembled late L2 learners' performance in the untimed visual, timed visual and metalinguistic test, according to Scheffé posthoc tests. In the untimed visual, the two groups of learners scored comparably on gender agreement ($p = .526$) and subject-verb (person) agreement ($p = .431$). In the metalinguistic test, they scored comparably on gender ($p = .110$) and subject-verb agreement ($p = .853$), and approached non-significance on number agreement ($p = .040$). Finally, in the timed visual, they scored comparably on gender agreement ($p = .683$). All these analyses yielded significant differences between NSs and early L2 learners ($p > .05$). The structures that did not yield any significant differences between NSs and early L2 learners were the subjunctive, in most tests, aspect, and the passive (all late L1 acquisitions).

other hand, late L2 learners did not show sensitivity to errors as a group and their sensitivity was significantly lower than NSs' and early L2 learners'.

5.2.4 Summary of Language Attainment

Table 26 shows the correlation matrix for the L2 learners' scores ($n = 100$) on the six language measures. This study hypothesized that the timed auditory GJT, the timed visual GJT, and the word monitoring task are language measures that require automatic use of L2 knowledge, whereas the untimed auditory GJT, the untimed visual GJT, and the metalinguistic test allow controlled use of L2 knowledge. The hypothesis was motivated by previous research, such as R. Ellis' (2005) psychometric study (recently replicated by Bowles, 2011), which showed that time pressure was a distinguishing factor between tasks that tap implicit and explicit L2 knowledge. As can be observed in the matrix, the six measures were positively correlated. The strongest relationships were between the metalinguistic test, the untimed visual GJT, and the untimed auditory GJT, suggesting that these three tests were measuring the same underlying construct, as hypothesized. However, the correlations between the language measures hypothesized to require automatic use of language knowledge were not so strong. Specifically, the correlations between the GSI, which was computed as an index of sensitivity to grammatical violations in the word monitoring task, and the other two measures hypothesized to require automatic use of language knowledge, the timed visual and timed auditory GJTs, were only moderately weak.

In fact, the two timed GJTs were more strongly correlated with the untimed GJTs and the metalinguistic test than with the GSI. This could be due to the nature of the tests, since the GJTs and the metalinguistic test shared the same format and scoring

120

procedure, whereas the GSI was a reaction time measure in milliseconds. The observed pattern of correlations could also suggest that the GSI is an index of a qualitatively different type of linguistic competence, the type of integrated language knowledge that word monitoring tasks have been hypothesized to measure, or, perhaps, an index of L2 processing capacity. Like the two timed GJTs, the word monitoring task involves performance in real time and minimizes controlled use of L2 knowledge. However, unlike the two timed GJTs and the other tests used in this study, the word monitoring task is carried out in a dual-task framework that focuses participants' attention on sentence meaning and on word monitoring, while all the other measures focus participants' attention on sentence correctness (i.e., language forms) and accuracy of grammaticality judgment.

Table 26. *Correlation Matrix for the Six Language Measures (L2 Learners)*

	Word Monitoring Task (GSI)	Timed Auditory (TA) GJT	Timed Visual (TV) GJT	Untimed Auditory (UA) GJT	Untimed Visual (UV) GJT	Metalinguistic Knowledge Test (MKT)
GSI	__	.28**	.27**	.27**	.26**	.33**
TA GJT		__	.70**	.80**	.79**	.76**
TV GJT			__	.70**	.66**	.66**
UA GJT				__	.84**	.82**
UV GJT					__	.86**
MKT						__

*p < .05

**p < .01

5.3 Cognitive Aptitudes and Language Attainment

In this section, the results of the role of cognitive variables on language outcomes are reported. The section is structured according to type of aptitude: aptitude for explicit learning, aptitude for implicit learning, and general intelligence. Each section is further subdivided into the two types of language outcome measure. These sections present the results of the effects of each type of aptitude on automatic and controlled outcome measures.

5.3.1 Aptitude for Explicit Learning and Language Attainment

This section presents the results of the role of aptitude for explicit learning (i.e., an equally weighted composite score combining LLAMA subtests B, E, and F and GAMA general intelligence scores) on language attainment as measured by tasks that allow controlled use of language knowledge (section 5.3.1.1) and measures that require automatic use of language knowledge (section 5.3.1.2). First, descriptive data are presented visually on scatterplots that show attainment scores as a function of age of onset with the aptitude for explicit learning dimension added. This visual display allows determining to what extent a high level of explicit aptitude is a necessary condition at an individual level in order to score within NS range. Next, multivariate analyses of covariance (MANCOVAs) are reported in order to determine the extent to which aptitude for explicit learning moderates language attainment in each of the groups. A MANCOVA was first conducted on overall test scores, grammatical and ungrammatical, and, then, re-run on ungrammatical items, agreement items, and non-agreement items in follow-up analyses.

5.3.1.1 Tasks that Allow Controlled Use of Language Knowledge

Figures 25, 26, and 27 display individual scores on the metalinguistic test, untimed visual GJT, and untimed auditory GJT, respectively, as a function of AO with the aptitude for explicit learning dimension added. The NS range is marked with a dotted line. The explicit aptitude groups (high, mid, and low) were created by establishing the following cutoffs on the aptitude for explicit learning composite score in every speaker group: *high* = z-scores >.5, *mid* = -.5 < z-scores < .5, and *low* = z-scores < -.5.

The highest scorers on the metalinguistic test in the early AO group were two learners with high explicit language aptitude (represented with black diamond markers). In the late AO group, six learners obtained scores as high as NSs. Three of them had high explicit aptitude (among them the highest scorer), two mid explicit aptitude (represented with dark gray circles), and one low explicit aptitude (represented with a light gray circle), suggesting that explicit language aptitude is advantageous, but not a necessary condition, to score within the NS range on the metalinguistic knowledge test. On the untimed visual GJT, the highest scorers in the early AO group were also two high explicit aptitude L2 learners, whereas in the late AO group, only one learner with mid explicit aptitude scored as high as NSs. Finally, on the untimed auditory GJT, the highest scorer in the early AO group was a high explicit aptitude L2 learner, while in the late AO group, two mid explicit aptitude L2 learners, but also one low explicit aptitude learner, scored within the NS range.
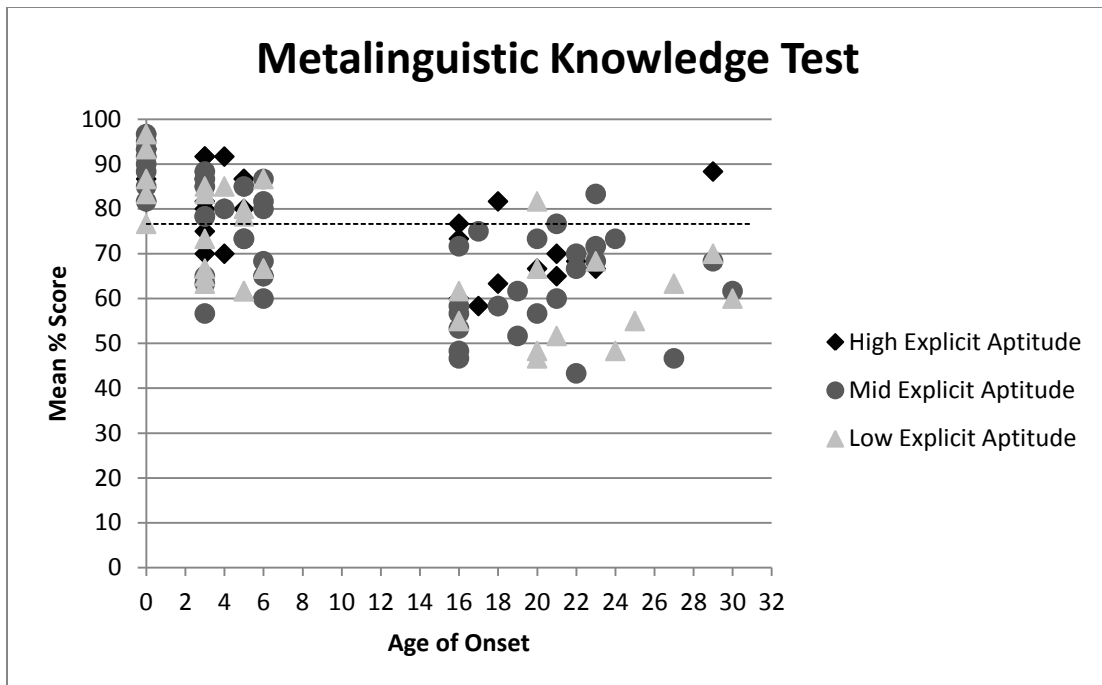
Figure 25. Metalinguistic knowledge test scores as a function of AO with the explicit

language aptitude dimension added
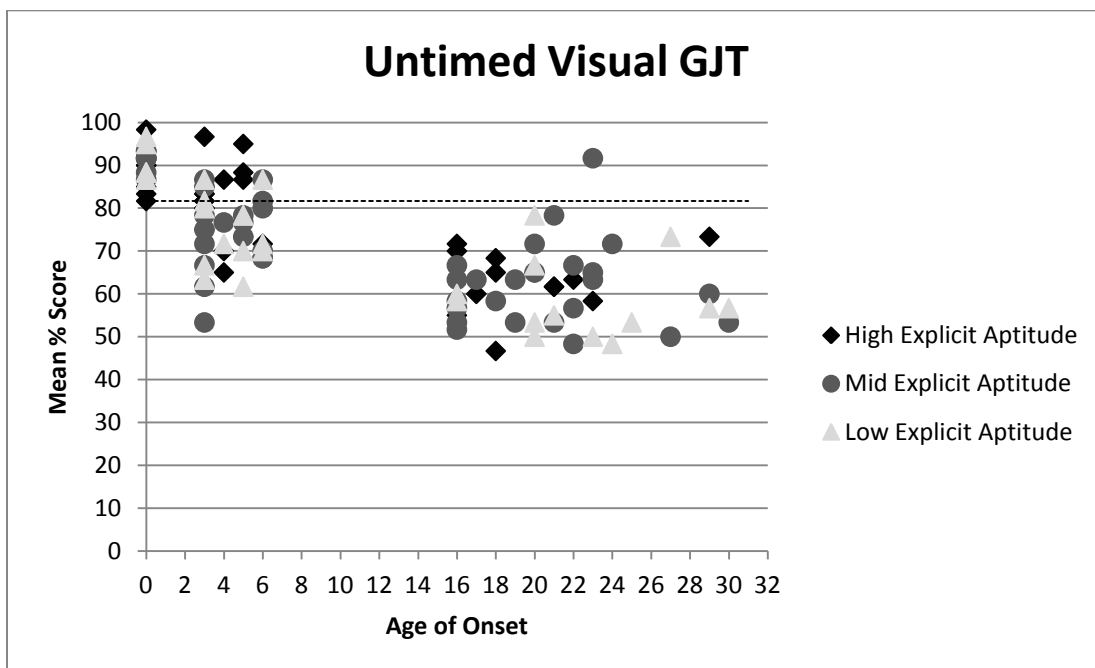
Figure 26. Untimed visual GJT test scores as a function of AO with the explicit

language aptitude dimension added

Untimed Auditory GJT
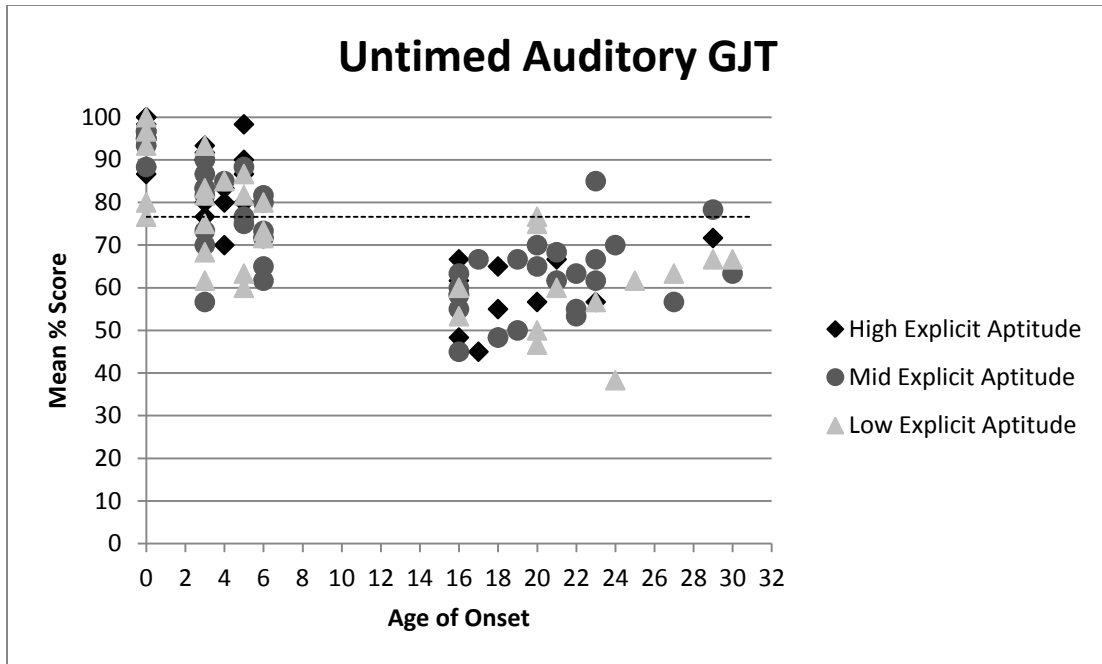
Figure 27. Untimed auditory GJT test scores as a function of AO with the explicit language aptitude dimension added

In order to investigate the role of aptitude for explicit learning in participants' language attainment as measured by tasks that allow controlled use of language knowledge, a MANCOVA was conducted with overall test scores on the untimed visual GJT, untimed auditory GJT, and metalinguistic test as dependent variables, group (NS controls, early L2 learners, and late L2 learners) as a fixed factor, and the composite aptitude score combining LLAMA B, E, F, and GAMA (i.e., aptitude for explicit learning) as a covariate. An interaction term was added, in addition to the group and covariate terms, to test for possible interactions between covariate and group as an independent factor. This is a necessary step to test for any aptitude-treatment interactions (ATIs)[17].

---

[17]Cronbach (1957) created ATI as a joint application of experimental and correlational methods, the two main approaches to psychological research at the time. This joint application examined

The assumption of equality of error variances (Levene's test) was met for the untimed auditory GJT ($p = .114$), but not for the untimed visual GJT ($p = .035$) or metalinguistic test ($p = .013$). The equality of covariances assumption (Box's test) was not met either ($p = .001$). Since the largest standard deviation was 9.1 in the untimed visual GJT data and 10.65 in the metalinguistic knowledge test data, and both were less than three times the smallest standard deviation (4.30 and 5.75, respectively), the MANCOVA was considered robust to the violation of the homogeneity assumption.

The analysis revealed a non-significant interaction between group and aptitude for explicit learning at the multivariate level ($F(6,224) = .950$, $p = .460$, $\eta_p^2 = .025$, $\lambda = .951$), indicating that the effect of aptitude was comparable across the groups and did not differ for any linear combination of test scores. The analysis also showed a significant multivariate effect of aptitude for explicit learning as a covariate ($F(3,112) = 3.581$, $p = .016$, $\eta_p^2 = .088$, $\lambda = .912$) on a linear combination of the three dependent measures (the metalinguistic test, the untimed visual GJT, and the untimed auditory GJT). The magnitude of this effect was medium. At the univariate level, the covariate was also significant for each of the three language measures separately, the untimed visual GJT ($F(1,114) = 5.308$, $p = .023$, $\eta_p^2 = .045$), the untimed auditory GJT

---

interactions between individual characteristics and treatment variables. The first term, "aptitude", refers to "any measurable person characteristic hypothesized to be propaedeutic to successful goal achievement in the treatment studied" (Snow, 1991:205). "Treatment" has a broad meaning of any experimental variable. "Interaction" is "the degree to which results for two or more treatments differ for persons who also differ on one or more aptitude variables" (Snow, 1991:206). ANCOVA models and covariate-adjusted means should be used when there is no significant interaction with the covariate. In an ATI model, the covariate shows a different relationship to an outcome variable in one treatment from the relationship it shows in another treatment. Factorial analyses are typically used to follow up ATI results. In this dissertation, where an ex-post-facto design was used, no treatment was delivered. ATI in the context of the present study means an interaction between a non-experimental independent variable (speaker group) and aptitude for a given dependent variable (L2 measure).

$(F(1,114) = 4.639, p = .033, \eta_p^2 = .039)$, both with a small effect size, and the

metalinguistic test $(F(1,114) = 10.814, p = .001, \eta_p^2 = .087)$, with a medium effect

size. Like the interaction at the multivariate level, the interactions between group and

covariate at the univariate level were not significant $(F(2,114) = 1.705, p = .186, \eta_p^2 = $

$.029, F(2,114) = 1.300, p = .277, \eta_p^2 = .022,$ and $F(2,114) = .760, p = .470, \eta_p^2 = .013,$

respectively). These results indicated that aptitude for explicit learning moderated

language attainment on the three measures of controlled use of knowledge and that

moderation was robust at the multivariate level, for a combination of the three

controlled measures, as well as at the univariate level, for each of the three measures

separately. These effects were comparable among NS controls, early L2 learners, and

late L2 learners.

The non-significant aptitude-treatment interaction suggested that the covariate

played a similar role in the two groups of L2 learners. As Figures 28, 29, and 30,

respectively, show, when untimed visual GJT, untimed auditory GJT, and

metalinguistic test scores were regressed on aptitude for explicit learning composite

scores, the slopes of the two L2-learner groups were similar, with the single exception

of the untimed auditory GJT. The slopes of the early and late AO groups were both

statistically significant for the untimed visual GJT ($p = .004$ and $p = .048$,

respectively) and metalinguistic test ($p = .002$ and $p = .004$, respectively), whereas the

slopes of the control group were not ($p = .927$ and $p = .241$). In the case of the

untimed auditory GJT, only the slope of the early AO group reached significance ($p = $

.011), but not the slope of the late AO or control group ($p = .397$ and $p = .565$,

respectively).
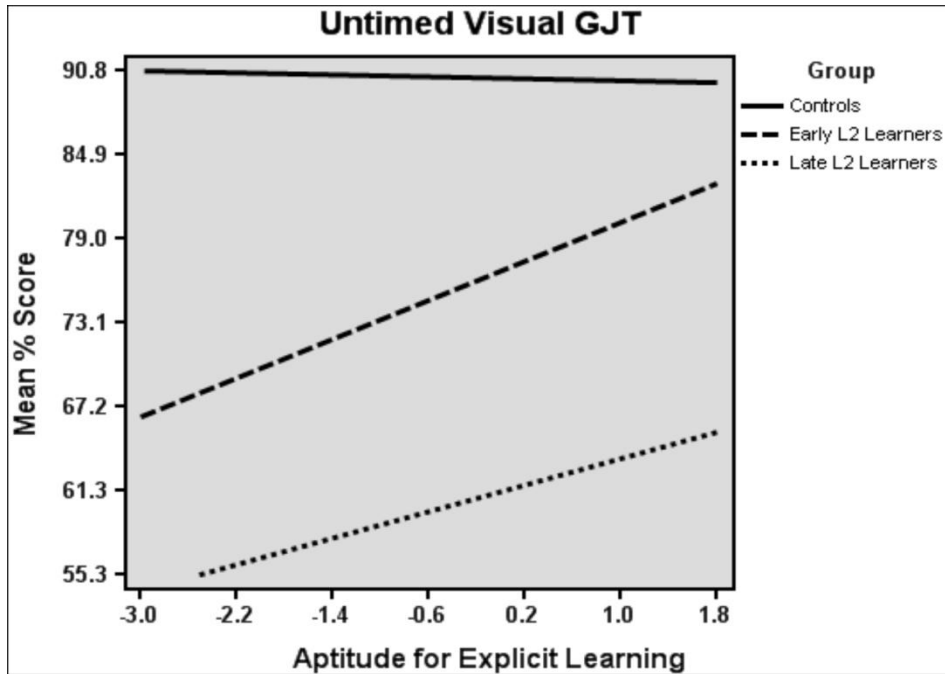
Figure 28. Regression of untimed visual GJT scores on aptitude for explicit learning composite scores at each group level
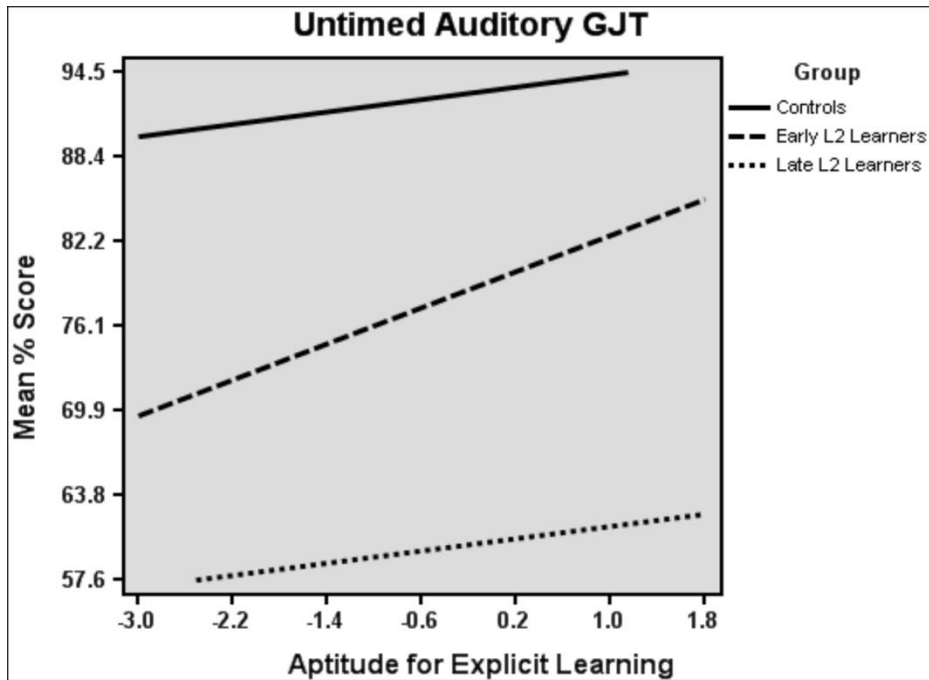
Figure 29. Regression of metalinguistic test scores on aptitude for explicit learning composite scores at each group level
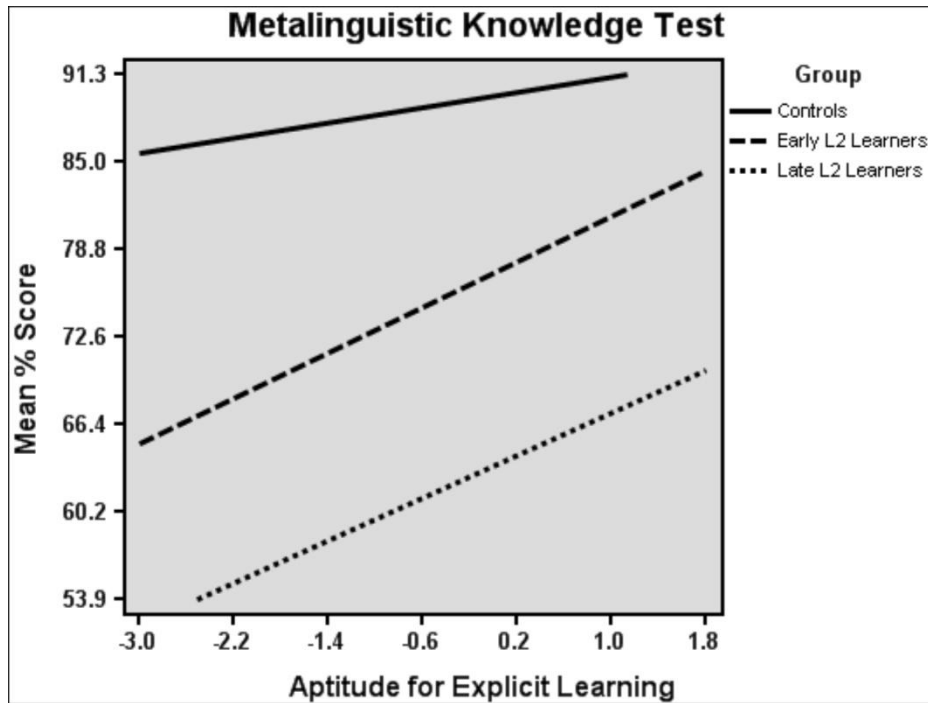
Figure 30. Regression of untimed auditory GJT scores on aptitude for explicit learning composite scores at each group level

Given that the non-significant aptitude-treatment interaction suggested that aptitude played a similar role in the two groups of L2 learners, but not in the NS group, a second MANCOVA model was run with NSs ($n = 20$) and L2 learners ($n = 100$). The results at the multivariate level remained robust. There was no significant interaction between group and aptitude for explicit learning ($F(3,114) = 1.631$, $p = .186$, $\eta_p^2 = .042$, $\lambda = .958$), but a significant multivariate effect of aptitude as a covariate ($F(3,114) = 2.741$, $p = .047$, $\eta_p^2 = .068$, $\lambda = .932$). At the univariate level, the effects of the covariate remained significant for each of the language tests, the untimed visual GJT ($F(1,116) = 4.150$, $p = .044$, $\eta_p^2 = .035$), the untimed auditory GJT ($F(1,116) = 4.978$, $p = .028$, $\eta_p^2 = .041$), and the metalinguistic test ($F(1,116) = 8.137$, $p = .005$, $\eta_p^2 = .066$), and they were further qualified by a significant two-way

129

interaction with group in the case of the untimed visual GJT ($F(1,116) = 4.752$, $p = .031$, $\eta_p^2 = .040$) and the metalinguistic test ($F(1,116) = 3.948$, $p = .049$, $\eta_p^2 = .034$). The two-way interaction between group and aptitude for the untimed auditory GJT approached significance ($F(1,116) = 3.722$, $p = .056$, $\eta_p^2 = .032$).

To further determine the effect of explicit aptitude on test scores using a factorial design, follow-up analyses were conducted in each of the groups (NS controls, early L2 learners, and late L2 learners) by comparing high and low explicit aptitude individuals in each group, according to a *z*-score distribution where *high* = *z*-scores >.5, *mid* = -.5 < *z*-scores < .5, and *low* = *z*-scores < -.5 (see Table 27 for a summary of descriptive statistics).

Table 27. *Summary of Overall Test Scores by Participants with High and Low Aptitude for Explicit Learning*

	Control		Early AO		Late AO	
	High	Low	High	Low	High	Low
	$n = 7$	$n = 5$	$n = 18$	$n = 14$	$n = 13$	$n = 13$
Untimed Visual GJT	88.81 (5.91)	91.00 (4.50)	81.67 (8.30)	73.57 (8.44)	63.08 (7.26)	58.46 (9.14)
Untimed Auditory GJT	94.52 (5.42)	89.33 (10.38)	84.80 (7.70)	76.07 (10.14)	58.33 (8.11)	59.62 (10.83)
Metalinguistic Test	90.95 (5.08)	87.33 (7.96)	82.31 (7.90)	73.10 (9.58)	68.97 (9.04)	59.74 (10.27)

Note. Standard deviations appear between parentheses.

High and low explicit aptitude controls ($n = 7$ and $n = 5$, respectively) were not significantly different on any of the three language tests: untimed visual GJT ($t(10) = -.694$, $p = .503$), untimed auditory GJT ($t(10) = 1.138$, $p = .282$), and metalinguistic test ($t(10) = .968$, $p = .356$). For this group ($n = 20$), the strongest correlation with aptitude for explicit learning corresponded to the metalinguistic test, and it had the same magnitude as the correlation in the late AO group, $r = .35$ ($p = .135$) (the disattenuated correlation[18] was .47).

In the early AO group, high and low explicit aptitude L2 learners ($n = 18$ and $n = 14$, respectively) differed significantly from each other on the three tests: untimed visual GJT ($t(30) = 2.716$, $p = .011$), untimed auditory GJT ($t(30) = 2.653$, $p = .014$), and metalinguistic test ($t(30) = 2.984$, $p = .006$). Correlations in this group ($n = 50$) were .38 ($p = .007$), .34 ($p = .016$) and .47 ($p = .001$) for the untimed visual GJT, untimed auditory GJT, and metalinguistic test, respectively (disattenuated correlations were .52, .46, and .63).

In the late AO group, high and low explicit aptitude L2 learners ($n = 13$ and $n = 13$, respectively) only differed on the metalinguistic knowledge test ($t(24) = 2.432$, $p = .023$), but not on the untimed visual or untimed auditory GJT ($t(24) = 1.426$, $p = .167$ and $t(24) = -.342$, $p = .736$, respectively). In this group, while the correlation between aptitude for explicit learning and performance on the metalinguistic test was significant ($r = .36$, $p = .010$), it did not reach significance for the untimed visual GJT

---

[18] Correlation coefficients disattenuated of measurement error were computed using the formula Rxy = rxy / sqrt (rxx ryy) (i.e., correlation coefficient divided by the square root of the product of the reliabilities of the two tests involved). The disattentuated coefficients suggest the upper bound of possible validity between the measures used.

($r = .26$, $p = 0.66$) and untimed auditory GJT ($r = .11$, $p = .472$) (disattenuated

correlations were .48, .36, and .15).

In order to further validate the results obtained for overall test scores (i.e., scores

on grammatical and ungrammatical items), MANCOVA analyses were re-run

including only scores on ungrammatical items, half of the items on each test ($k = 30$).

Multivariate and univariate analyses with group (NS controls, early L2 learners, and

late L2 learners) as a fixed factor yielded no significant interactions ($p > .05$). At the

multivariate level, aptitude for explicit learning remained a significant covariate for a

linear combination of the ungrammatical items on the three tests that allow use of

controlled language knowledge ($F(3,112) = 3.459$, $p = .019$, $\eta_p^2 = .085$, $\lambda = .915$),

and, at the univariate level, it remained significant for the metalinguistic test

($F(1,114) = 8.876$, $p = .004$, $\eta_p^2 = .073$). The effect size of these associations was

medium.

Figure 31 displays metalinguistic test scores on ungrammatical items as regressed

on aptitude for explicit learning scores in each of the groups. Simple slopes were

significant in the early and late L2 learner groups ($p = .004$ and $p = .014$,

respectively), but not in the control group ($p = .256$).
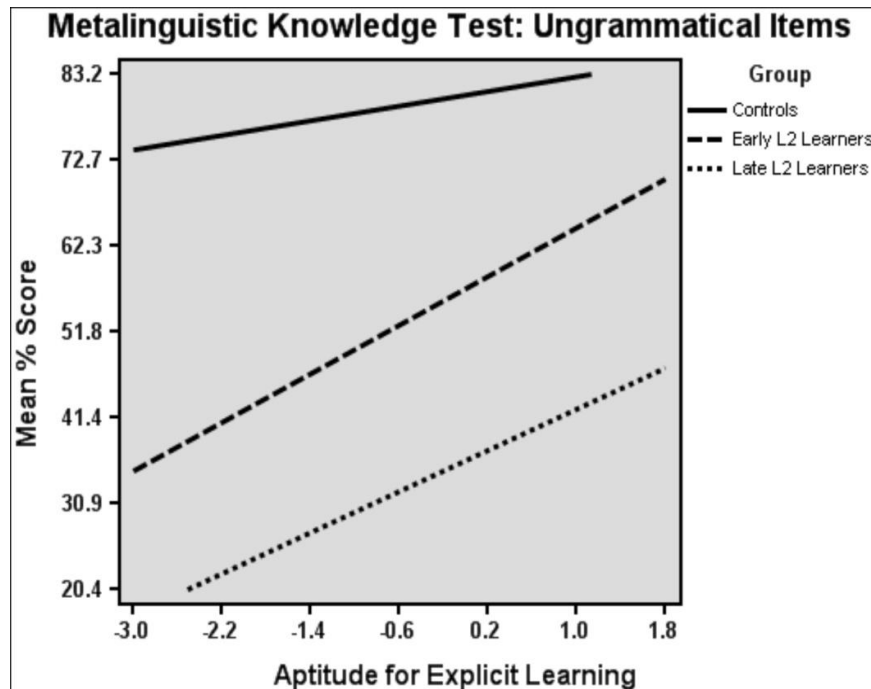
Figure 31. Regression of metalinguistic test scores on ungrammatical items on aptitude for explicit learning composite scores at each group level

There were no significant multivariate or univariate interactions, either, when early and late L2 learners were combined as a single group ($p > .05$), but aptitude for explicit learning remained a significant covariate with a medium effect size at the multivariate level ($F(3,114) = 3.054$, $p = .031$, $\eta_p^2 = .075$, $\lambda = .925$) and, at the univariate level, for the metalinguistic test ($F(1,116) = 5.911$, $p = .017$, $\eta_p^2 = .049$), although the interaction with group did not reach significance ($F(1,116) = 2.108$, $p = .149$, $\eta_p^2 = .018$).

Simple correlations between aptitude and metalinguistic test scores on ungrammatical items in each of the groups showed a significant relationship in the early and late AO groups ($r = .46$, $p = .001$ and $r = .32$, $p = .026$, respectively) (disattenuated correlations were .62 and .43). In the control group, the correlation had

133

practically the same magnitude as in the late AO group, but it did not reach

significance, probably due to the smaller sample size ($r = .34$, $p = .147$) (the

disattenuated correlation was .46). Differences between high and low explicit aptitude

individuals were significant in the early AO group ($M = 67.22$, $SD = 13.92$ and $M =$

49.76, $SD = 17.76$, respectively) ($t(30) = 3.121$, $p = .004$) and approached

significance in the late AO group ($M = 44.62$, $SD = 17.98$ and $M = 17.99$, $SD = 4.99$,

respectively) ($t(24) = 1.970$, $p = .060$), but were non-significant in the control group

($M = 82.38$, $SD = 9.17$ and $M = 76.67$, $SD = 15.28$, respectively) ($t(10) = .814$, $p =$

.435).

A last set of follow-up MANCOVA analyses was run distinguishing between

items testing agreement structures ($k = 30$) (gender, person, and number agreement)

and non-agreement structures ($k = 30$) (aspect contrasts, the subjunctive, and the

passive) on every language test. As for agreement items, multivariate and univariate

analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed

factor yielded no significant interactions ($p > .05$). Aptitude for explicit learning,

however, was a significant covariate with a medium effect size at the multivariate

level ($F(3,112) = 2.772$, $p = .045$, $\eta_p^2 = .071$, $\lambda = .929$) and, at the univariate level, for

agreement items on the metalinguistic test ($F(1,114) = 5.302$, $p = .023$, $\eta_p^2 = .046$). It

also approached significance for agreement items on the untimed auditory GJT

($F(1,114) = 3.882$, $p = .051$, $\eta_p^2 = .034$), but it was not significant for the untimed

visual GJT $F(1,114) = .163$, $p = .687$, $\eta_p^2 = .001$). When L2 learners were combined

as one group, interactions at the multivariate and univariate level remained non-

significant ($p > .05$). Aptitude for explicit learning was not a significant covariate at

the multivariate level, either, although it had a *p* value of .087 ($F(3,114) = 2.241$, *p* = .087, $\eta_p^2 = .056$, $\lambda = .944$). At the univariate level, aptitude was significant for the untimed auditory GJT ($F(1,116) = 3.990$, *p* = .048, $\eta_p^2 = .034$) and approached significance for the metalinguistic test ($F(1,116) = 3.776$, *p* = .054, $\eta_p^2 = .032$), in both cases with a small effect size, but it was not significant for the untimed visual GJT ($F(1,116) = .268$, *p* = .606, $\eta_p^2 = .002$).

Follow-up simple correlations revealed that the relationship between explicit aptitude and test performance for agreement items on the untimed auditory GJT and the metalinguistic test was not significant in the control group (*r* = .18, *p* = .453 and *r* = .16, *p* = .513, respectively) (disattenuated correlations were .24 and .22), but significant in the early AO group (*r* = .33, *p* = .023 and *r* = .35, *p* = .014) (disattenuated correlations were .44 and .47). In the late AO group, only the correlation with test scores for agreement items on the metalinguistic was significant (*r* = .39, *p* = .006), not the correlation with test scores for agreement items on the untimed auditory GJT scores (*r* = .17, *p* = .236) (disattenuated correlations were .53 and .23).[19] Differences between high and low explicit aptitude individuals in the early AO group were significant for the untimed auditory GJT ($t(30) = 2.775$, *p* = .010, mean difference of 12.34) and the metalinguistic test ($t(30) = 2.098$, *p* = .044, mean difference of 9.47). In the late AO group, only differences between high and low explicit aptitude individuals on the metalinguistic test were significant ($t(24) = 2.599$, *p* = .016, mean difference of 12.31), not on the untimed auditory GJT ($t(24) = .360$, *p*

---

[19] This stronger relationship between aptitude for explicit learning and performance on the untimed auditory GJT in the early AO group did not result in a significant group x covariate interaction ($F(6,224) = 1.437$, *p* = .202, $\eta_p^2 = .037$, $\lambda = .927$ with a three-level group factor –controls, early L2 learners, and late L2 learners- and $F(1,96) = 2.392$, *p* = .125, $\eta_p^2 = .025$ with a two-level group factor – early and late L2 learners).

= .722, mean difference of 1.54). In the control group, there were no differences between high and low explicit aptitude individuals for either the untimed auditory GJT ($t(10) = 1.033$, $p = .326$, mean difference of 4.86) or the metalinguistic test ($t(10) = .778$, $p = .454$, mean difference of 3.43).

Regarding non-agreement items, multivariate and univariate analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed factor and aptitude for explicit learning composite scores as a covariate yielded no significant interactions ($p > .05$). However, aptitude was a significant covariate with medium effect sizes at the multivariate level ($F(3,112) = 3.333$, $p = .022$, $\eta_p^2 = .083$, $\lambda = .917$), as well as for non-agreement items on the untimed visual GJT ($F(1,114) = 7.744$, $p = .006$, $\eta_p^2 = .064$) and the metalinguistic test ($F(1,114) = 7.473$, $p = .007$, $\eta_p^2 = .062$), but not for non-agreement items on the untimed auditory GJT ($F(1,114) = 1.387$, $p = .241$, $\eta_p^2 = .012$). When L2 learners were combined as a single group, aptitude for explicit learning remained a significant covariate at the multivariate level ($F(3,114) = 2.725$, $p = .048$, $\eta_p^2 = .067$, $\lambda = .933$), as well as at the univariate level for non-agreement items on the untimed visual GJT ($F(1,116) = 8.717$, $p = .004$, $\eta_p^2 = .071$) and metalinguistic test ($F(1,116) = 7.128$, $p = .009$, $\eta_p^2 = .058$), but not for non-agreement items on the untimed auditory GJT ($F(1,116) = 1.691$, $p = .196$, $\eta_p^2 = .015$).

In order to examine the effect of aptitude for explicit learning in each of the speaker groups separately, follow-up correlational and factorial analyses, were performed. Follow-up correlations showed a significant relationship between aptitude and performance on non-agreement structures on the untimed visual GJT in the early and late AO groups ($r = .40$, $p = .004$ and $r = .35$, $p = .013$, respectively)

(disattenuated correlations were .55 and .48), but not in the control group ($r = .17$, $p = $ .475) (the disattenuated correlation was .23). The correlation with performance on the metalinguistic test was only significant in the early group ($r = .38$, $p = .007$) (the disattenuated correlation was .51). In the late AO and control groups, the relationship was moderately weak[20] and positive, but non-significant ($r = .21$, $p = .151$ and $r = $ .28, $p = .224$, respectively) (disattenuated correlations were .28 and .38).

The results of the factorial analyses confirmed the correlational patterns. Differences between high and low explicit aptitude individuals were significant in the early AO group on the untimed visual GJT ($t(30) = 2.940$, $p = .006$, mean difference of 8.36) and metalinguistic test ($t(30) = 3.179$, $p = .003$, mean difference of 8.97). In the late AO group, only differences on the untimed visual GJT were significant ($t(24) = 2.551$, $p = .018$, mean difference of 8.21), but they did not reach significance on the metalinguistic test ($t(24) = 1.475$, $p = .153$, mean difference of 6.15). In the control group, there were no differences between high and low explicit aptitude individuals on either the untimed visual GJT ($t(10) = -.177$, $p = .863$, mean difference of -0.76) or the metalinguistic test ($t(10) = .865$, $p = .407$, mean difference of 3.81).

To summarize, aptitude for explicit learning moderated both early and late L2 learners' language attainment, as measured by untimed tests that focus participants' attention on language correctness and that allow controlled use of L2 knowledge, but it did not moderate the performance of NS controls. The effect of aptitude for explicit learning was observed at the multivariate level, in a combination of the three untimed measures, as well as at the univariate level, in each measure separately. Two of these

---

[20] Following Cohen (1988), the strength of a linear relationship can be weak ($0 < r < .20$), moderately weak ($.20 < r < .40$), moderate ($.41 < r < .60$), moderately strong ($.61 < r < .80$), and strong ($.81 < r < 1.0$).

measures were visual (the untimed visual GJT and the metalinguistic knowledge test) and one was auditory (the untimed auditory GJT). In the late AO group, only performance on visual tests was moderated by level of aptitude for explicit learning, while in the early AO group, performance on both untimed modalities, visual and auditory, showed a relationship with aptitude for explicit learning, although this difference did not yield any significant interactions between L2-learner group and covariate for the untimed auditory GJT. When only ungrammatical items were considered, aptitude effects remained robust for the metalinguistic test in the two L2 learner groups. This was the test that encouraged the greatest attention to language forms and the one for which the effect size of aptitude for explicit learning as a covariate was the largest.

The fact that aptitude for explicit learning moderated early, but not late, L2 learners' performance on the untimed auditory GJT seems to suggest that the auditory modality could have placed processing constraints on L2 learners that prevented them from making use of controlled L2 knowledge, even if the test was performed under untimed testing conditions. If this was the case, those late L2 learners with higher aptitude for explicit learning as measured by an auditory test should have been able to score higher on the untimed auditory GJT modality. The LLAMA aptitude subtest E, which was used as part of the composite of aptitude for explicit learning, requires test takers to work out relationships between sounds they hear and a writing system. The test gives participants time to freely navigate and work out those relationships by listening to the target sounds as many times as wished within the established testing time.

As a follow-up test to the results obtained for the untimed auditory GJT in the late AO group, the relationship between late L2 learners' aptitude for explicit learning and performance on the untimed auditory GJT was further investigated by examining LLAMA E test scores. Regarding overall test performance, while the correlation with aptitude for explicit learning was .11 ($p = .421$), it was .27 ($p = .061$) with LLAMA E scores (disattenuated correlations were .15 and .37). Similarly, when only ungrammatical items on the untimed auditory test were considered, the correlation increased from .01 ($p = .997$) to .28 ($p = .051$) (disattenuated correlations were .01 and .38). For agreement items testing gender, person, and number agreement, the increase was from .17 ($p = .236$) to .29 ($p = .051$) (disattenuated correlations were .23 and .40), and only for non-agreement items testing aspect contrasts, the subjunctive, and the passive, did the correlation not approach significance (from $r = .04$, $p = .780$ to $r = .15$, $p = .316$) (disattenuated correlations were .05 and .21).

5.3.1.2 Tasks that Require Automatic Use of Language Knowledge

Figures 32, 33, and 34 display individual scores on the timed visual GJT, timed auditory GJT, and word monitoring task as a function of AO with the aptitude for explicit learning dimension added. The NS range is marked with a dotted line. The explicit aptitude groups (high, mid, and low) were created by establishing the following cutoffs on the aptitude for explicit learning composite score in every speaker group: *high* = z-scores > .5, *mid* = -.5 < z-scores < .5, and *low* = z-scores < -.5.

The highest scorer on the timed visual GJT in the early AO group was a high explicit aptitude L2 learner, whereas, in the late AO group, a combination of high, mid, and low explicit aptitude L2 learners scored within NS range. On the timed

auditory GJT, a high explicit aptitude L2 learner obtained the highest score in the early AO group, while, in the late AO group, mid and low explicit aptitude learners overlapped within NS range. Finally, the highest grammatical sensitivity indices on the word monitoring task corresponded to high explicit aptitude L2 learners in the two learner groups.

The scatterplot for the word monitoring task further shows that practically all the L2 learners' sensitivity scores were within NS range. Due to the nature of reaction-time data, however, it is not possible to talk about ceiling effects. In addition, as the analyses of group GSIs in section 5.2.3 indicated, the performance of the three speaker groups on the task was not comparable. Late L2 learners' sensitivity scores were significantly lower than NSs' and early L2 learners' scores. Also, the difference between word monitoring latencies for grammatical and ungrammatical items on the task, which was used to compute GSIs, was non-significant in the late AO group, but significant in the NS control and early AO groups.
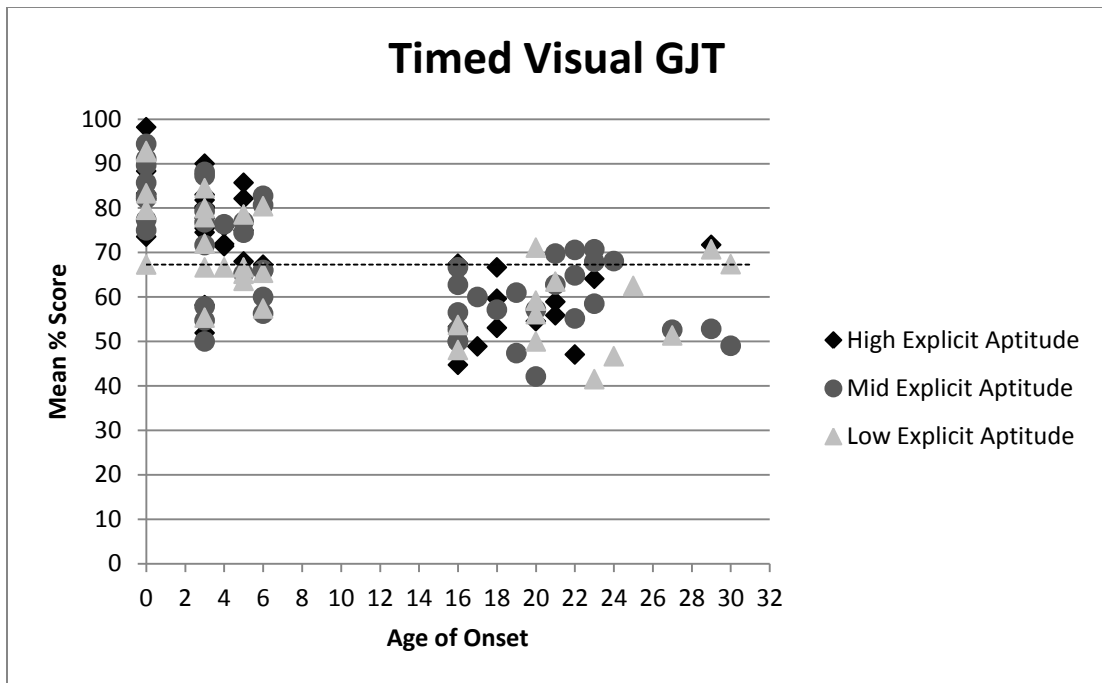
Figure 32. Timed visual GJT scores as a function of AO with the explicit language

aptitude dimension added
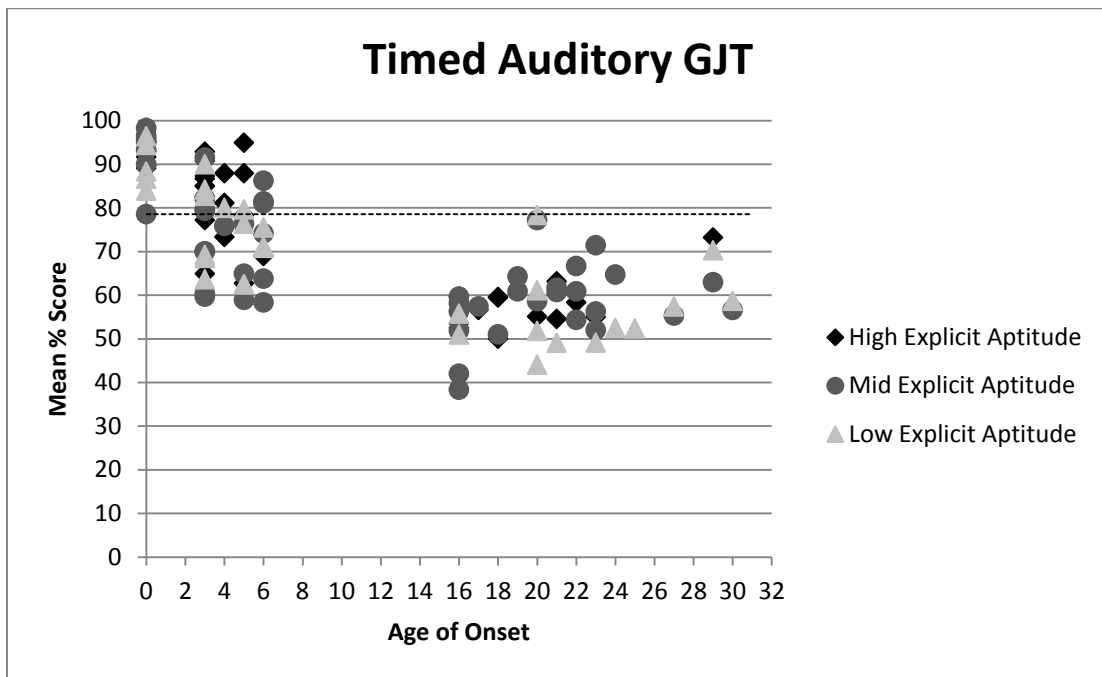
Figure 33. Timed auditory GJT scores as a function of AO with the explicit language

aptitude dimension added
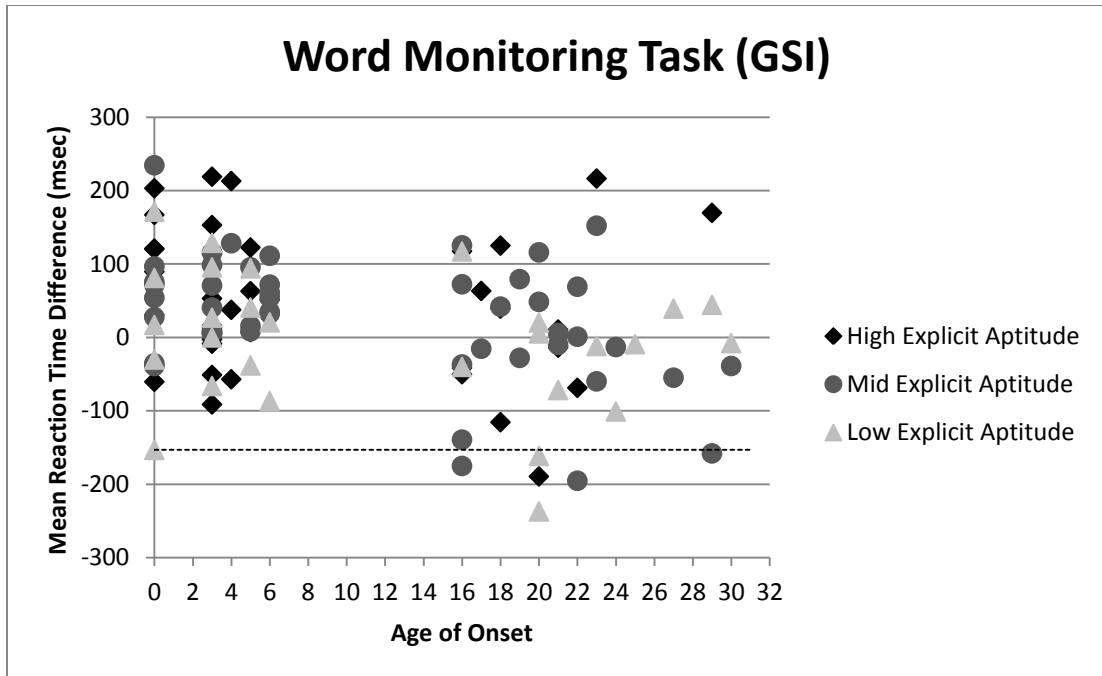
Figure 34. Word monitoring task scores (GSI) as a function of AO with the explicit language aptitude dimension added

In order to investigate the role of aptitude for explicit learning on participants' language attainment as measured by tasks hypothesized to require automatic use of language knowledge, a MANCOVA was conducted with overall test scores on the timed visual GJT, timed auditory GJT, and word monitoring task (i.e., GSI) as dependent variables, group (NS controls, early L2 learners, and late L2 learners) as fixed factor, and the composite aptitude score combining LLAMA B, E, F, and GAMA (i.e., aptitude for explicit learning) as a covariate. An interaction term was added, in addition to the group and covariate terms, to test for possible interactions between covariate and group as an independent factor. The assumption of equality of error variances was met for the timed visual GJT ($p = .224$), the timed auditory GJT

142

($p = .062$), and word monitoring task ($p = .637$). The equality of covariances

assumption was also met ($p = .462$).

At the multivariate level, the interaction between aptitude for explicit learning and

group was not significant ($F(6,210) = .381$, $p = .891$, $\eta_p^2 = .011$, $\lambda = .979$). Explicit

aptitude as a multivariate covariate did not reach significance, either ($F(3,105) =$

2.460, $p = .067$, $\eta_p^2 = .065$, $\lambda = .935$). At the univariate level, aptitude was significant

for the word monitoring task ($F(1,107) = 5.191$, $p = .025$, $\eta_p^2 = .046$), with a small

effect size, and it had a p value of .087 for the timed auditory GJT ($F(1,107) = 2.981$,

$p = .087$, $\eta_p^2 = .027$). It was not significant for the timed visual GJT ($F(1,107) =$

1.067, $p = .304$, $\eta_p^2 = .010$). Two-way interactions between group and covariate were

all non-significant: timed visual GJT ($F(2,107) = .546$, $p = .581$, $\eta_p^2 = .010$), timed

auditory GJT ($F(2,107) = .376$, $p = .687$, $\eta_p^2 = .007$), and word monitoring task

($F(2,107) = .481$, $p = .620$, $\eta_p^2 = .009$)[21].

When L2 learners were combined as a single group and compared against NS

controls, interactions at the multivariate and univariate level remained non-significant

($p > .05$). Aptitude for explicit learning remained a non-significant covariate at the

multivariate level, although it had a p value of .063 ($F(3,107) = 2.502$, $p = .063$, $\eta_p^2 =$

$.065$, $\lambda = .935$). It also remained non-significant at the univariate level for the timed

visual GJT ($F(1,109) = 1.987$, $p = .162$, $\eta_p^2 = .018$), and significant for the word

monitoring task ($F(1,109) = 4.282$, $p = .041$, $\eta_p^2 = .037$), as well as for the timed

---

[21] The analysis including outliers in the word monitoring task yielded similar results. At the
multivariate level, there was no interaction between group and covariate ($F(6,224) = .341$, $p = .915$, η_p^2
$= .009$, $\lambda = .982$) and aptitude for explicit learning approached significance as a covariate ($F(3,112) =$
2.653, $p = .052$, $\eta_p^2 = .066$, $\lambda = .934$). At the univariate level, aptitude was a significant covariate for
the timed auditory GJT ($F(1,114) = 4.122$, $p = .045$, $\eta_p^2 = .035$) and the word monitoring task ($F(1,114)$
$= 4.665$, $p = .033$, $\eta_p^2 = .039$), but not for the timed visual GJT ($F(1,114) = 1.327$, $p = .252$, $\eta_p^2 = .012$).
There were no significant interactions with group ($p > .05$).

auditory GJT ($F(1,109) = 4.575$, $p = .035$, $\eta_p^2 = .040$). Effect sizes were all small. The interactions between group and covariate were all non-significant, at the multivariate level ($F(3,107) = .521$, $p = .669$, $\eta_p^2 = .014$, $\lambda = .986$), and, at the univariate level, for the timed visual GJT ($F(1,109) = 1.379$, $p = .243$, $\eta_p^2 = .012$), timed auditory GJT ($F(1,109) = 1.223$, $p = .271$, $\eta_p^2 = .011$), and word monitoring task ($F(1,109) = .095$, $p = .759$, $\eta_p^2 = .001$).[22]

In order to examine the effect of aptitude for explicit learning in each of the speaker groups separately, follow-up correlational and factorial analyses were performed on the two tests that yielded significant results for the entire set of participants (the word monitoring task and the timed auditory GJT). Follow-up simple correlations between aptitude for explicit learning and performance on the word monitoring task and the timed auditory GJT were computed for each of the groups. Correlations had similar magnitudes across groups, but did not reach significance in any case. For the word monitoring task, the correlations in the control, early AO, and late AO groups were .29 ($p = .217$), .28 ($p = .062$), and .20 ($p = .182$), respectively (disattenuated correlations were .37, .36, and .26). For the timed auditory GJT, correlations were .21 ($p = .384$), .27 ($p = .056$), and .26 ($p = .075$), respectively (disattenuated correlations were .28, .36, and .34).

---

[22] The analysis including outliers in the word monitoring task yielded similar results. At the multivariate level, there was no interaction between group and covariate ($F(3,112) = .510$, $p = .676$, $\eta_p^2 = .013$, $\lambda = .987$) and aptitude for explicit learning approached significance as a covariate ($F(3,112) = 2.478$, $p = .065$, $\eta_p^2 = .061$, $\lambda = .939$). At the univariate level, aptitude was a significant covariate for the timed auditory GJT ($F(1,116) = 4.924$, $p = .028$, $\eta_p^2 = .041$) and approached significance for the word monitoring task ($F(1,116) = 3.726$, $p = .056$, $\eta_p^2 = .031$), but it was not significant for the timed visual GJT ($F(1,116) = 1.951$, $p = .165$, $\eta_p^2 = .017$). There were no significant interactions with group ($p > .05$).

Factorial analyses with high and low explicit aptitude individuals further showed no significant differences on the word monitoring task among NS controls ($t(10) = 1.308$, $p = .220$, mean difference of 78.71), early L2 learners ($t(30) = .798$, $p = .431$, mean difference of 41.24), or late L2 learners ($t(24) = 1.255$, $p = .221$, mean difference of 52.01). Score differences on the timed auditory GJT were not significant among controls, either ($t(10) = 1.285$, $p = .228$, mean difference of 3.16), or late L2 learners ($t(24) = .483$, $p = .634$, mean difference of 1.47), but high explicit aptitude early L2 learners outperformed their low-aptitude counterparts ($t(30) = 2.385$, $p = .024$, mean difference of 7.60). Although a significant effect of explicit aptitude was observed only in the early AO group, a MANCOVA conducted on the two L2-learner groups did not show a significant interaction between aptitude for explicit learning and early and late L2 learners' scores on the timed auditory GJT ($F(1,96) = 1.594$, $p = .210$, $\eta_p^2 = .016$). This indicated that the significant effect of explicit aptitude on early L2 learners' timed auditory GJT scores was not significantly different from the effect on late L2 learners' scores (as reported above, in the late AO group, the correlation between explicit aptitude and timed auditory GJT scores was .26, $p = .075$).

When only ungrammatical items on the timed visual and timed auditory GJTs were considered (the word monitoring task does not provide an interpretable measure for ungrammatical items only), multivariate and univariate analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed factor yielded no significant interactions ($p > .05$). Aptitude for explicit learning was not a significant covariate at the multivariate level, either ($p > .05$), but, at the univariate level, aptitude approached significance for the timed auditory GJT ($F(1,114) = 3.787$, $p =$

.054, η_p^2 = .032). When L2 learners were combined as a single group, explicit

aptitude remained a non-significant covariate at the multivariate level ($F(2,115)$ =

2.496, p = .087, η_p^2 = .042 λ = .958) and the interaction with group was not

significant, either ($F(2,115)$ = .344, p = .709, η_p^2 = .006, λ = .994). At the univariate

level, interactions were not significant ($p > .05$) and explicit aptitude was not a

significant covariate for the timed visual GJT ($F(1,116)$ = 1.225, p = .271, η_p^2 = .010).

The effect of the covariate on the timed auditory GJT, however, reached significance

($F(1,116)$ = 4.782, p = .031, η_p^2 = .040).

   In order to examine the effect of aptitude for explicit learning in each of the

speaker groups separately, follow-up correlational and factorial analyses were

performed on the test that yielded significant results for the entire set of participants

(the timed auditory GJT). Simple correlations in each of the groups revealed that the

relationship between aptitude scores and scores on ungrammatical items on the timed

auditory GJT was not significant in the control group (r = .18, p = .467) or late AO

group (r = -.03, p = .819), but significant in the early AO group (r = .30, p = .035)

(disattenuated correlations were .24, -.04, and .40). High explicit aptitude early L2

learners (n = 18) (M = 67.18, SD = 15.94) scored significantly higher than low

explicit aptitude early L2 learners (n = 14) (M = 52.40, SD = 16.93) ($t(30)$ = 2.533, p

= .017, mean difference of 14.78). High and low explicit aptitude participants in the

control and late AO groups were not significantly different (p = .210 and p = .947,

respectively). Although the effect of explicit aptitude on the timed auditory GJT was

only observed in the early AO group, a MANCOVA conducted on the two L2-learner

groups did not yield a significant interaction ($F(1,96)$ = .804, p = .372, η_p^2 = .008).

Therefore, explicit aptitude had a comparable effect in the two groups, even if it only reached significance in the early AO group.

A last set of follow-up analyses was run distinguishing between agreement items ($k = 30$) (gender, person, and number agreement) and non-agreement items ($k = 30$) (aspect contrasts, the subjunctive, and the passive) on every language test. As for agreement items, multivariate and univariate analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed factor yielded no significant interactions ($p > .05$). Aptitude for explicit learning was not a significant covariate at the multivariate level ($F(3,102) = 1.888$, $p = .136$, $\eta_p^2 = .051$, $\lambda = .949$) or, at the univariate level, for the timed auditory GJT ($F(1,104) = 1.438$, $p = .233$, $\eta_p^2 = .013$), or word monitoring task (i.e., GSI) ($F(1,104) = 1.534$, $p = .218$, $\eta_p^2 = .014$), but it was significant for the timed visual GJT ($F(1,104) = 4.239$, $p = .042$, $\eta_p^2 = .038$).[23]

When L2 learners were combined as one group, interactions at the multivariate and univariate level were all non-significant, as well ($p > .05$). As a covariate, aptitude for explicit learning was not significant for agreement items at the multivariate level ($F(3,104) = 1.897$, $p = .135$, $\eta_p^2 = .050$, $\lambda = .950$) or, at the univariate level, for the timed auditory GJT ($F(1,106) = 2.103$, $p = .150$, $\eta_p^2 = .018$) or word monitoring task ($F(1,106) = 2.246$, $p = .137$, $\eta_p^2 = .019$), but it approached significance for the timed visual GJT ($F(1,106) = 3.729$, $p = .056$, $\eta_p^2 = .031$)[24].

---

[23] The analysis including outliers in the word monitoring task also yielded non-significant results at the multivariate level. There was no interaction between group and covariate ($F(6,224) = .446$, $p = .847$, $\eta_p^2 = .012$, $\lambda = .977$) and aptitude for explicit learning was not a significant covariate ($F(3,112) = 2.274$, $p = .084$, $\eta_p^2 = .057$, $\lambda = .943$). At the univariate level, aptitude was not significant for the timed auditory GJT ($F(1,114) = .758$, $p = .386$, $\eta_p^2 = .007$) or word monitoring task ($F(1,114) = 1.518$, $p = .220$, $\eta_p^2 = .013$), but it was significant for the timed visual GJT ($F(1,114) = 5.526$, $p = .020$, $\eta_p^2 = .046$). There were no significant interactions with group ($p > .05$).

[24] The analysis including outliers in the word monitoring task also yielded non-significant results at the multivariate level. There was no interaction between group and covariate ($F(3,114) = .509$, $p = .677$,

In order to examine the effect of aptitude for explicit learning in each of the speaker groups separately, follow-up correlational and factorial analyses were performed on the test that yielded significant results for the entire set of participants (the timed visual GJT). Follow-up correlations in each of the groups revealed a significant correlation between aptitude and agreement items on the timed visual GJT only in the early AO group ($r = .36$, $p = .010$), but not in the NS control or late AO groups ($r = .13$, $p = .576$, and $r = .17$, $p = .242$, respectively) (disattenuated correlations were .48, .18, and .23). High explicit aptitude early L2 learners ($n = 18$) ($M = 75.10$, $SD = 8.08$) scored significantly higher on agreement items than their low-aptitude counterparts ($n = 14$) ($M = 65.20$, $SD = 8.79$) ($t(30) = 3.311$, $p = .002$, mean difference of 9.90).[25] High and low explicit aptitude participants in the control and late AO groups, on the other hand, were not significantly different ($p = .509$ and $p = .485$, respectively). The interaction between aptitude and L2-learner group was not significant ($F(1,96) = 2.165$, $p = .144$, $\eta_p^2 = .022$). Therefore, explicit aptitude had a comparable effect on agreement scores on the timed visual GJT in the two learner groups, even if it only reached significance in the early AO group.

Regarding non-agreement items, multivariate and univariate analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed factor and aptitude for explicit learning as a covariate yielded no significant interactions ($p > .05$). Aptitude was not a significant covariate at the multivariate level, either ($F(3,105) = 2.207$, $p =$

$\eta_p^2 = .013$, $\lambda = .987$) and aptitude for explicit learning was not a significant covariate ($F(3,114) = 2.079$, $p = .107$, $\eta_p^2 = .052$, $\lambda = .948$). At the univariate level, aptitude was not significant for the timed auditory GJT ($F(1,116) = 2.075$, $p = .152$, $\eta_p^2 = .018$) or word monitoring task ($F(1,116) = 2.291$, $p = .133$, $\eta_p^2 = .019$) but it was significant for the timed visual GJT ($F(1,116) = 4.340$, $p = .039$, $\eta_p^2 = .036$). There were no significant interactions with group ($p > .05$).

[25] The subtest in the aptitude composite that contributed the most to the significant relationship with timed visual GJT scores in the early AO group was LLAMA F, the grammatical inferencing subtest, with a correlation of .37 ($p = .009$).

.092, η_p^2 = .059, λ = .941). At the univariate level, aptitude for explicit learning was

not a significant covariate for the timed visual GJT ($F(1,107)$ = .013, p = .909, η_p^2 =

.000) or word monitoring task (i.e., GSI) ($F(1,107)$ = .218, p = .641, η_p^2 = .002), but it

was significant for the timed auditory GJT ($F(1,107)$ = 6.809, p = .010, η_p^2 = .057).[26]

This relationship remained significant when L2 learners were combined as one group

($F(1,109)$ = 5.468, p = .021, η_p^2 = .046). Interactions were all non-significant ($p > .05$)

and aptitude for explicit learning remained a non-significant covariate for the timed

visual GJT ($F(1,109)$ = .776, p = .380, η_p^2 = .007) and word monitoring task (i.e.,

GSI) ($F(1,109)$ = .754, p = .387, η_p^2 = .007).[27]

In order to examine the effect of aptitude for explicit learning in each of the

speaker groups separately, follow-up correlational and factorial analyses were

performed on the test that yielded significant results for the entire set of participants

(the timed auditory GJT). Follow-up correlations in each of the groups revealed a

significant correlation between explicit aptitude and non-agreement items on the

timed auditory GJT in the early AO group only (r = .31, p = .032) (the disattenuated

correlation was .41). High explicit aptitude early L2 learners (n = 18) (M = 84.44, SD

= 8.50) scored significantly higher than their low-aptitude counterparts (n = 14) (M =

---

[26] The analysis including outliers in the word monitoring task also yielded a non-significant interaction between group and covariate at the multivariate level ($F(6,224)$ = .227, p = .967, η_p^2 = .006, λ = .988), but a significant covariate effect ($F(3,112)$ = 2.732, p = .047, η_p^2 = .068, λ = .932). At the univariate level, aptitude was a significant covariate for non-agreement items on the timed auditory GJT ($F(1,114)$ = 5.057, p = .026, η_p^2 = .042) and it approached significance for the word monitoring task ($F(3,114)$ = 3.366, p = .069, η_p^2 = .029), but it was non-significant for the timed visual GJT ($F(1,114)$ = .020, p = .888, η_p^2 = .000). There were no significant interactions with group ($p > .05$).

[27] The analysis including outliers in the word monitoring task also yielded a significant effect of the covariate on non-agreement items on the timed auditory GJT ($F(1,116)$ = 7.247, p = .008, η_p^2 = .059) and a non-significant effect on the timed visual GJT ($F(1,116)$ = .704, p = .403, η_p^2 = .006) and word monitoring task ($F(1,116)$ = 2.370, p = .126, η_p^2 = .020).

74.93, SD = 8.67) (t(30) = 3.114, p = .004, mean difference of 9.52).[28] In the control group, the correlation between aptitude and non-agreement items on the timed auditory GJT had a slightly larger magnitude than in the early AO group, but did not reach significance (r = .32, p = .163), and, in the late AO group, the relationship was weak and non-significant (r = .19, p = .211) (disattenuated correlations were .42 and .25). High and low explicit aptitude NSs and high and low explicit aptitude late L2 learners were not significantly different (p = .168 and p = .692, respectively). The interaction between aptitude and L2-learner group was not significant (F(1,96) = 1.614, p = .207, η_p^2 = .017), indicating a comparable effect of aptitude after all.

To summarize, as expected, aptitude for explicit learning was not a significant covariate at the multivariate level for the language measures hypothesized to require automatic use of language knowledge (timed auditory GJT, timed visual GJT, and word monitoring task). There were no significant interactions with group, either, suggesting a comparable relationship between aptitude and language attainment in the three groups (NS controls, early L2 learners, and late L2 learners). Although there were no significant interactions between group and covariate, univariate tests and follow-up analyses showed an unexpected relationship in the early AO group between aptitude for explicit learning and performance on the timed auditory and timed visual GJTs that was not present in the late AO group. Early L2 learners with high explicit aptitude performed significantly better than their low-aptitude counterparts on the timed auditory GJT, both overall and when only ungrammatical items were considered. Specifically, there was a relationship between early L2 learners' aptitude

---

[28] The subtest in the aptitude composite that contributed the most to the significant relationship with timed auditory GJT scores in the early AO group was LLAMA E, the sound-symbol correspondence subtest, with a correlation of .30 (p = .036).

for explicit learning and scores on items testing non-agreement structures (aspect contrasts, the subjunctive, and the passive), but no effect on items testing agreement structures (gender, person, and number agreement) when the GJT was timed and auditory.[29] When the GJT was visual, however, early L2 learners with high explicit aptitude scored higher on agreement items than their low-aptitude counterparts.

The relationship between early L2 learners' explicit aptitude composite score and performance on the timed auditory GJT was mostly due to a significant correlation with LLAMA E, the aptitude subtest measuring sound-symbol correspondence ability, which was also found to be related to early and late L2 learners' performance on the untimed auditory GJT. On the other hand, the relationship between early L2 learners' explicit aptitude and performance on the timed visual GJT was due to a significant correlation with LLAMA F, the aptitude subtest measuring grammatical inferencing via pictures and written stimuli. Finally, as predicted, aptitude for explicit learning did not moderate participants' grammatical sensitivity as measured by the word monitoring task, the task hypothesized to be at the extreme of the continuum of tasks requiring automatic use of L2 knowledge.

---

[29] A trend towards dissociation was observed in the data between agreement and non-agreement items on the timed auditory GJT in the early AO group. Early L2 learners with high aptitude for *explicit* learning scored higher than their low-aptitude counterparts on non-agreement structures, while early L2 learners with high aptitude for *implicit* learning scored higher than their low-aptitude counterparts on agreement structures. The difference on timed auditory GJT scores for non-agreement items between early L2 learners with high and low aptitude for explicit learning was significant ($t(30) =$ 3.114, $p = .004$, mean difference of 9.52), but not the difference on scores for agreement items ($t(30) =$ 1.546, $p = .133$, mean difference of 6.38). On the other hand, the difference on timed auditory GJT scores for agreement items between early L2 learners with high and low aptitude for implicit learning approached significance ($t(28) = 1.779$, $p = .086$, mean difference of 7.30), but the difference on scores for non-agreement items was not significant ($t(28) = .122$, $p = .904$, mean difference of 0.48).

5.3.2 General Intelligence and Language Attainment

This section presents the results of the role of general intelligence on language attainment as measured by tasks that allow controlled use of language knowledge (section 5.3.2.1) and measures that require automatic use of language knowledge (section 5.3.2.2). First, descriptive data is presented visually on scatterplots that show attainment scores as a function of age of onset with the general intelligence dimension added. This visual display allows determining to what extent high intelligence is a necessary condition at an individual level in order to score within NS range. Next, multivariate analyses of covariance (MANCOVAs) are conducted in order to determine the extent to which intelligence moderates language attainment in each of the groups. A MANCOVA was first conducted on overall test scores, grammatical and ungrammatical, and, then, re-run on ungrammatical items, agreement items, and non-agreement items as follow-up analyses.

5.3.2.1 Tasks that Allow Controlled Use of Language Knowledge

Figures 35, 36, and 37 display individual scores on the metalinguistic knowledge test, untimed visual GJT, and untimed auditory GJT as a function of AO with the general intelligence dimension added. The NS range is marked with a dotted line. Like the aptitude groups, the general intelligence groups were created by converting GAMA raw scores into *z*-scores within each of the three speaker groups[30] and by establishing the following cutoffs reflecting distance from the mean in standard deviations: *high = z*-scores >.5, *mid = -.5 < z*-scores < .5, and *low = z*-scores < -.5.

---

[30] The decision to compute *z*-scores separately for each group was motivated by the fact that the three groups did not have comparable cognitive abilities (i.e., early L2 learners had significantly higher intelligence than late L2 learners).

The highest scorers on the three tests were either high- or mid-intelligence L2 learners in both the early and late AO groups, with the exception of the untimed auditory GJT where a low-intelligence late L2 learner also scored within NS range.
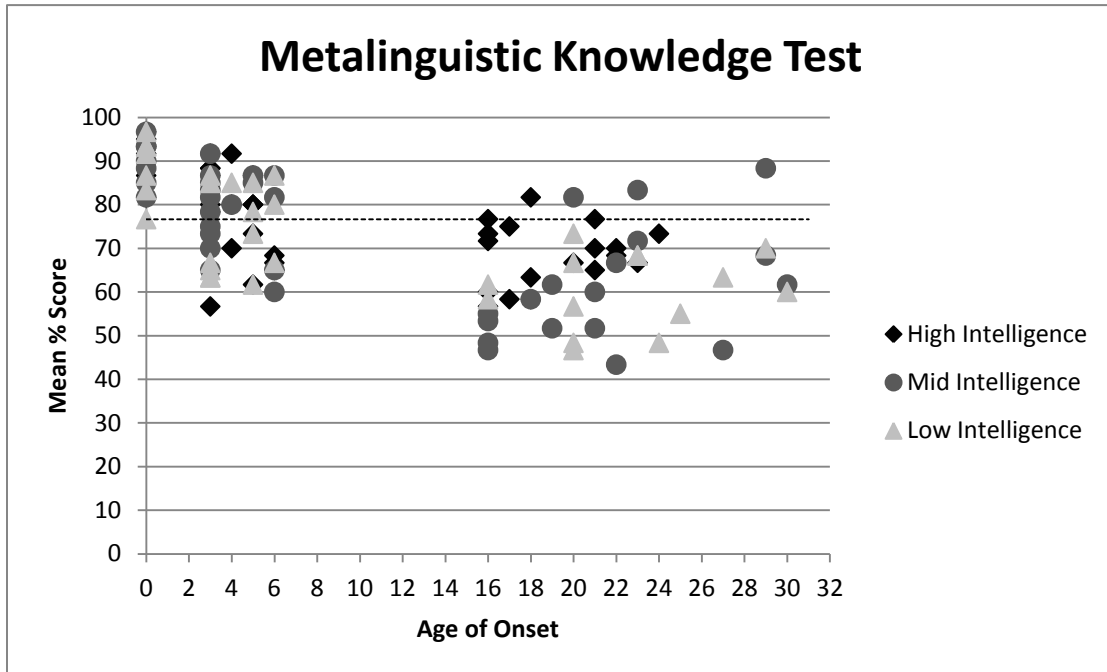
Figure 35. Metalinguistic knowledge test scores as a function of AO with the general intelligence dimension added
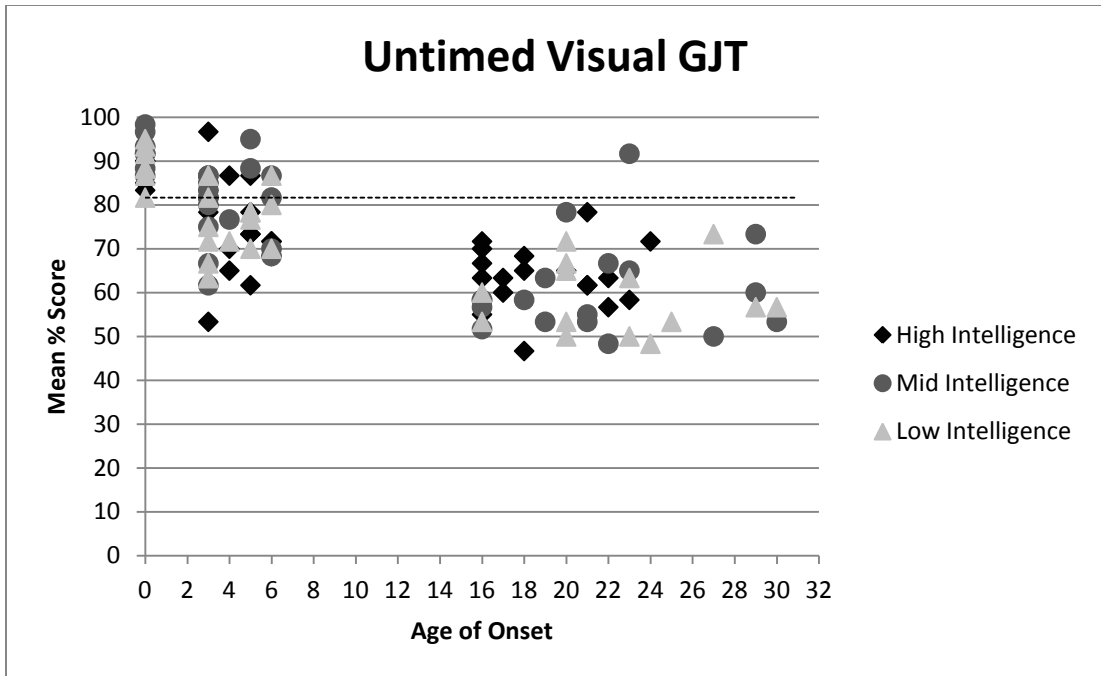
Figure 36. Untimed visual GJT scores as a function of AO with the general

intelligence dimension added
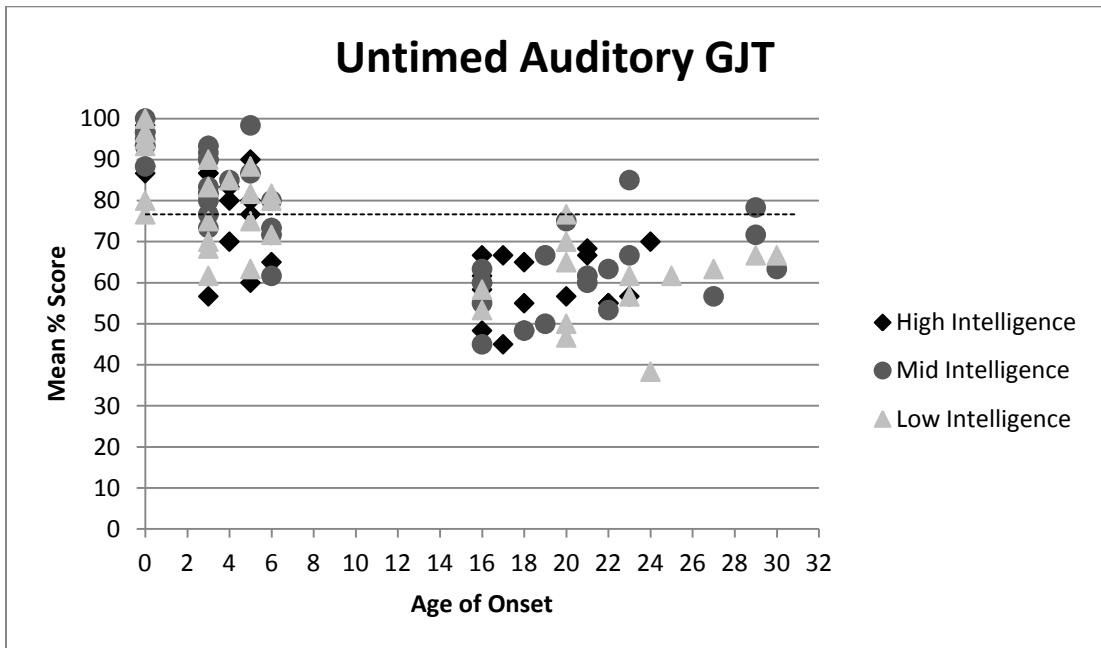
Figure 37. Untimed auditory GJT scores as a function of AO with the general

intelligence dimension added

A MANCOVA was conducted with overall test scores on the three tasks

hypothesized to allow controlled use of language knowledge (i.e., the untimed visual

GJT, untimed auditory GJT, and metalinguistic test), group as a fixed factor (NS

controls, early L2 learners, and late L2 learners), and general intelligence scores as a

covariate. An interaction term was added in a custom model to test for possible

interactions between covariate and group as an independent factor. The analysis

showed a non-significant interaction between group and intelligence at the

multivariate level ($F(6,224) = .720$, $p = .636$, $\eta_p^2 = .019$, $\lambda = .962$). Interactions

between the covariate and each of the tests at the univariate level were also non-

significant: Untimed visual ($F(2,114) = 2.107$, $p = .149$, $\eta_p^2 = .018$), untimed auditory

($F(2,114) = .326$, $p = .569$, $\eta_p^2 = .003$), and metalinguistic test ($F(2,114) = 3.140$, $p =$

$.079$, $\eta_p^2 = .027$).

Since a non-significant aptitude-treatment interaction could suggest a comparable

effect of the covariate in the two groups of L2 learners, a second MANCOVA was

performed with only two groups: NS controls ($n = 20$) and L2 learners ($n = 100$). The

analysis revealed that there was no significant interaction between group and general

intelligence at the multivariate level ($F(3,114) = 1.703$, $p = .170$, $\eta_p^2 = .043$, $\lambda = .957$).

General intelligence was not a significant covariate at the multivariate level, either

($F(3,114) = 1.402$, $p = .246$, $\eta_p^2 = .036$, $\lambda = .964$). However, univariate analyses

showed that intelligence was a significant covariate with a small effect size for the

untimed visual GJT ($F(1,116) = 3.972$, $p = .049$, $\eta_p^2 = .034$) and the metalinguistic

test ($F(1,116) = 4.200$, $p = .043$, $\eta_p^2 = .035$). These relationships were further

qualified by a group-by-covariate interaction, which was significant for the untimed

visual GJT ($F(1,116) = 5.253$, $p = .024$, $\eta_p^2 = .044$) and had a p value of .072 for the metalinguistic test ($F(1,116) = 3.301$, $p = .072$, $\eta_p^2 = .028$). General intelligence was not a significant covariate for the untimed auditory GJT ($F(1,116) = 2.649$, $p = .106$, $\eta_p^2 = .023$) and the interaction with group was not significant, either ($F(1,116) = 1.498$, $p = .223$, $\eta_p^2 = .013$). As shown in Figures 38, 39, and 40, with the exception of the untimed auditory GJT, where none of the slopes were significant ($p > .05$), general intelligence was more related to L2 learners' performance than NS controls' performance on the untimed visual GJT and metalinguistic test.
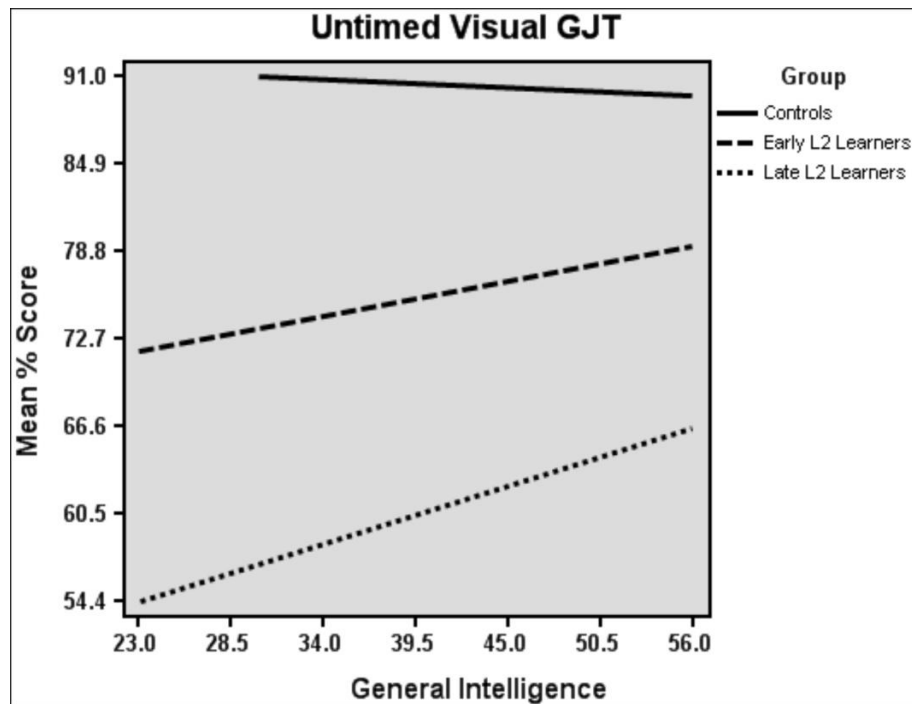
Figure 38. Regression of untimed visual GJT scores on general intelligence scores at each group level
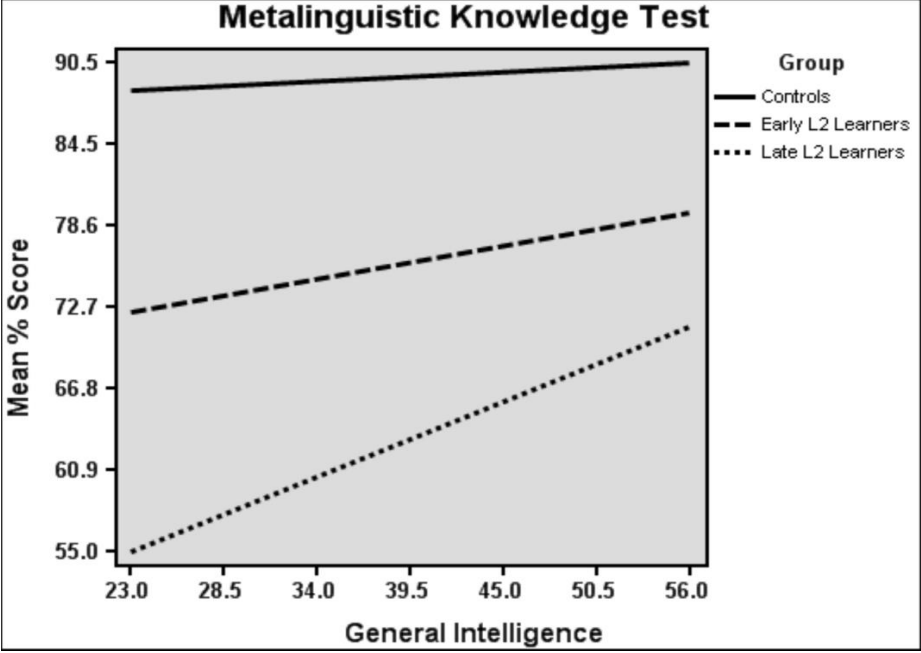
Figure 39. Regression of metalinguistic test scores on general intelligence scores at
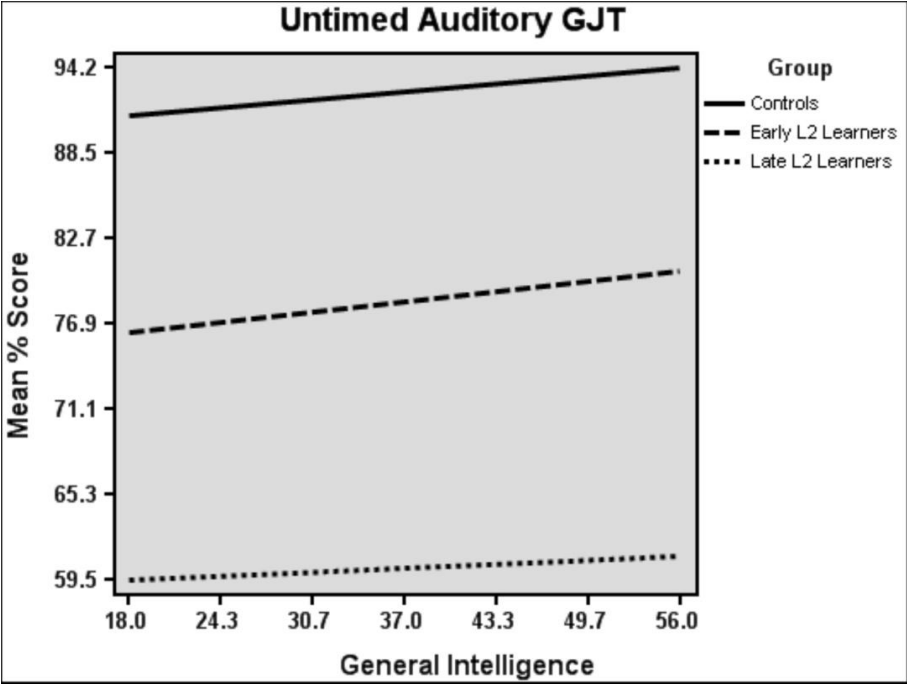
each group level

Figure 40. Regression of untimed auditory GJT scores on general intelligence scores

at each group level

There was also a difference between the slopes of the two groups of L2 learners. While the slope of the early AO group was not significant for either the untimed visual GJT or metalinguistic test ($p = .349$ and $p = .445$, respectively), the slope of the late AO group was significant in both cases ($p = .032$ and $p = .016$). Slope differences between early and late L2 learners, however, did not yield significant a significant interaction in either case ($F(1,96) = .007$, $p = .933$, $\eta_p^2 = .000$, and $F(1,96) = .168$, $p = .683$, $\eta_p^2 = .002$).

In order to further examine the effect of intelligence, follow-up factorial analyses were performed on the two tests that yielded significant results for the entire set of participants (the untimed visual GJT and the metalinguistic test) by comparing high- and low-intelligence individuals (i.e., $z > .5$ and $z < -.5$, respectively) within each group (NS controls, early L2 learners, and late L2 learners) (see Table 28 for a summary of descriptive statistics). High- and low-intelligence controls ($n = 6$ and $n = 7$, respectively) were not significantly different on either the untimed visual GJT or the metalinguistic test ($t(11) = -.169$, $p = .869$ and $t(11) = .567$, $p = .582$, respectively). There were no significant differences in the early AO group, either ($t(31) = .336$, $p = .739$ and $t(31) = .541$, $p = .593$, respectively). In the late AO group, differences in performance between high- and low-intelligence individuals were significant for the metalinguistic test ($t(30) = 2.894$, $p = .007$, mean difference of 8.07) and had a p value of .075 for the untimed visual GJT ($t(30) = 1.846$, $p = .075$, mean difference of 5.01).

Table 28. *Summary of Overall Test Scores on the Untimed Visual GJT and*

Metalinguistic Test by High- and Low-Intelligence Participants

	Control		Early AO		Late AO	
	High	Low	High	Low	High	Low
	$n = 6$	$n = 7$	$n = 18$	$n = 15$	$n = 18$	$n = 14$
Untimed	88.89	89.29	75.74	74.67	63.70	58.69
Visual GJT	(3.90)	(4.50)	(10.34)	(7.46)	(7.20)	(8.14)
Metalinguistic	90.83	88.81	77.41	75.44	68.43	60.36
Test	(5.65)	(6.99)	(10.70)	(9.99)	(7.17)	(8.60)

Note. Standard deviations appear between parentheses.

MANCOVA analyses were re-run including only scores on the ungrammatical
items in each language test ($k = 30$). Multivariate and univariate analyses with group
(NS controls, early L2 learners, and late L2 learners) as a fixed factor yielded no
significant interactions ($p > .05$). When early and late L2 learners were combined as a
single L2-learner group, there was no interaction between group and covariate at the
multivariate level, either ($F(3,114) = 1.338$, $p = .265$, $\eta_p^2 = .034$, $\lambda = .966$). The
interactions at the univariate level were also non-significant for all the tests, the
untimed visual GJT ($F(1,116) = 3.446$, $p = .066$, $\eta_p^2 = .029$), the untimed auditory
GJT ($F(1,116) = .769$, $p = .382$, $\eta_p^2 = .007$), and the metalinguistic test ($F(1,116) =$
2.598, $p = .110$, $\eta_p^2 = .022$). As a covariate, the effects of intelligence approached
significance for the metalinguistic test ($F(1,116) = 3.657$, $p = .058$, $\eta_p^2 = .031$), but

they were not significant for the untimed visual GJT ($F(1,116) = .363$, $p = .548$, $\eta_p^2 = .003$), or untimed auditory GJT ($F(1,116) = 1.820$, $p = .180$, $\eta_p^2 = .016$).

In order to examine the effect of intelligence in each of the speaker groups separately, follow-up correlational and factorial analyses were performed on the test that yielded significant results for the entire set of participants (the metalinguistic knowledge test). Simple correlations between intelligence and metalinguistic test scores on ungrammatical items in each of the groups showed a significant relationship in the late AO group ($r = .32$, $p = .022$) (the disattenuated correlation was .41). High-intelligence late L2 learners scored significantly higher ($M = 45.74$, $SD = 18.85$) than low-intelligence late L2 learners ($M = 29.05$, $SD = 15.49$) on ungrammatical items ($t(30) = 3.173$, $p = .003$). Correlations in the control and early AO group were weak and non-significant ($r = .06$, $p = .792$ and $r = .14$, $p = .343$, respectively) (disattenuated correlations were .03 and .18), thus replicating the results found for overall metalinguistic test scores. Unlike late L2 learners, high- and low-intelligence early L2 learners did not differ on their metalinguistic test scores for ungrammatical items ($M = 57.22$, $SD = 20.36$ and $M = 54.89$, $SD = 18.85$) ($t(31) = .339$, $p = .737$).

A last set of follow-up analyses was run, with general intelligence as a covariate, distinguishing between agreement items ($k = 30$) (gender, person, and number) and non-agreement items ($k = 30$) (aspect, the subjunctive, and the passive) on every test. For agreement items, multivariate and univariate analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed factor yielded no significant interactions ($p > .05$). As a covariate, general intelligence was significant at the univariate level for the metalinguistic test ($F(1,114) = 4.435$, $p = .037$, $\eta_p^2 = .038$).

The size of the effect was small. When L2 learners were combined and compared with controls, interactions at the multivariate and univariate level remained non-significant, and general intelligence had a *p* value of .064 as a covariate for the metalinguistic test ($F(1,116) = 3.496$, $p = .064$, $\eta_p^2 = .031$).

In order to examine the effect of intelligence in each of the speaker groups separately, follow-up correlational and factorial analyses were performed on the test that yielded significant results for the entire set of participants (the metalinguistic knowledge test). Simple correlations in each group showed that the relationship between intelligence and performance on agreement items on the metalinguistic test was not significant among early L2 learners ($r = .16$, $p = .257$) or controls ($r = .12$, $p = .607$), but significant among late L2 learners ($r = .38$, $p =.007$) (disattenuated correlations were .21, .16, and .49). High-intelligence late L2 learners ($n = 18$) scored significantly higher ($M = 74.26$, $SD = 11.07$) than low-intelligence late L2 learners ($n = 14$) ($M = 61.67$, $SD = 12.52$) on agreement test items ($t(30) = 3.014$, $p = .005$).

Regarding non-agreement items, multivariate and univariate analyses with group (NS controls, early L2 learners, and late L2 learners) as a fixed factor, and general intelligence as a covariate, yielded no significant interactions ($p > .05$). When L2 learners were combined as a group and compared with controls, results showed that intelligence was only a significant covariate at the univariate level for non-agreement items on the untimed visual GJT ($F(1,116) = 6.044$, $p = .015$, $\eta_p^2 = .051$). This relationship was further qualified by an interaction with group ($F(1,116) = 4.363$, $p = .039$, $\eta_p^2 = .037$). In both cases, the effect size was small.

As can be seen on Figure 41, the significant interaction was mostly due to the late AO group, which had a steeper slope than the early AO group ($p = .002$ and $p = .542$, respectively). Slope differences between early and late L2 learners did not yield a significant interaction ($F(1,96) = 1.107$, $p = .295$, $\eta_p^2 = .011$).
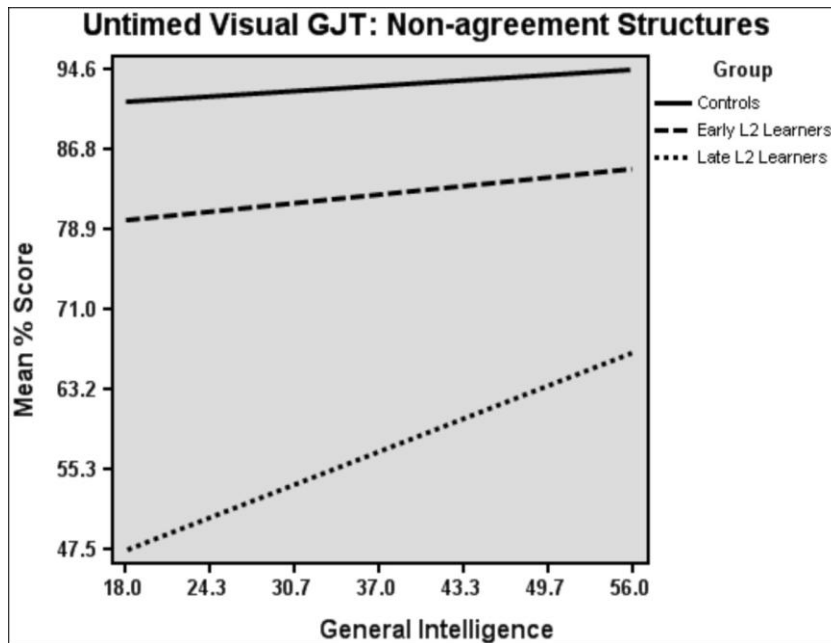
Figure 41. Regression of untimed visual GJT scores for non-agreement items on general intelligence scores at each group level

In order to further examine the effect of intelligence, follow-up factorial analyses were performed on the test that yielded significant results for the entire set of participants (the untimed visual GJT) by comparing high- and low-intelligence individuals (i.e., $z > .5$ and $z < -.5$, respectively) within each speaker group (NS controls, early L2 learners, and late L2 learners) (see Table 29 for a summary of descriptive statistics). High- and low-intelligence controls' performance on non-agreement items on the untimed visual GJT was not significantly different ($t(11) =$

.081, $p = .937$). There were no significant differences in the early AO group, either ($t(31) = -.053$, $p = .958$). In the late AO group, high-intelligence L2 learners scored significantly higher than low-intelligence L2 learners ($t(30) = 2.400$, $p = .023$, mean difference of 7.49).

Table 29. *Summary of Test Scores on the Untimed Visual GJT (Non-agreement Items) by High- and Low-Intelligence Participants*

	Control		Early AO		Late AO	
	High	Low	High	Low	High	Low
	$n = 6$	$n = 7$	$n = 18$	$n = 15$	$n = 18$	$n = 14$
Untimed	92.22	91.90	80.93	81.11	62.96	55.48
Visual GJT	(8.07)	(6.04)	(11.01)	(8.79)	(8.55)	(9.02)

Note. Standard deviations appear between parentheses.

To summarize, general intelligence did not moderate either NS controls or early L2 learners' language attainment on tasks that allow controlled use of L2 knowledge. It did not moderate performance at the multivariate level for a combination of those measures, either. Intelligence, however, moderated late L2 learners' attainment as measured by the metalinguistic test and the untimed visual GJT. This effect remained robust for ungrammatical items on the metalinguistic test. Late L2 learners' performance on agreement structures (gender, person, and number agreement) and non-agreement structures (aspect contrasts, the subjunctive, and the passive) was equally related to general intelligence, albeit on different tests. The only test where

intelligence did not show an effect in the late AO group was the untimed auditory GJT.

The fact that general intelligence did not moderate early L2 learners' attainment on measures that allow controlled use of language knowledge suggests that the intelligence factor did not contribute to the significant results reported for aptitude for explicit learning in the early AO group (despite a significant correlation between general intelligence and LLAMA aptitude subtests B, E, and F in this group of .30, p = .035). On the other hand, the fact that intelligence moderated late L2 learners' language attainment suggests that both general intelligence and LLAMA aptitude subtests B, E, and F contributed to the significant results reported for aptitude for explicit learning among late L2 learners. In this group, the correlation between intelligence and LLAMA B, E, and F was stronger than in the early AO group (r = .62, p < .001).

In order to tease apart the relationship between general intelligence (GAMA), explicit language aptitude (LLAMA B, E, and F), and ultimate L2 attainment in the two L2-learner groups, an analysis was conducted with the LLAMA B, E, and F composite as a covariate, group (early and late L2 learners) as a fixed factor and scores on the untimed visual GJT, untimed auditory GJT, and metalinguistic test. The assumptions of equality of covariances and error variances, Box's and Levene's tests, respectively, were all met (p > .05). At the multivariate level, there was no interaction between group and covariate, indicating a comparable effect of LLAMA B, E, and F on the two groups of learners ($F(3,94)$ = .954, p = .418, η_p^2 = .030, λ = .970). The LLAMA composite was, however, a significant covariate with a large effect size

164

($F(3,94) = 5.064$, $p = .003$, $\eta_p^2 = .140$, $\lambda = .860$). At the univariate level, interactions with group were also non-significant for the untimed visual GJT ($F(1,96) = 1.318$, $p = .254$, $\eta_p^2 = .014$), untimed auditory GJT ($F(1,96) = 2.467$, $p = .120$, $\eta_p^2 = .025$), and metalinguistic test ($F(1,96) = .561$, $p = .456$, $\eta_p^2 = .006$). The LLAMA composite was a significant covariate for each of the measures separately, the untimed visual GJT ($F(1,96) = 9.861$, $p = .002$, $\eta_p^2 = .094$), the untimed auditory GJT ($F(1,96) = 7.941$, $p = .006$, $\eta_p^2 = .077$), and the metalinguistic test ($F(1,96) = 15.494$, $p < .001$, $\eta_p^2 = .140$), for which the largest effect size was found ($\eta_p^2 = .140$).

Simple correlations between LLAMA B, E, and F and scores on the untimed visual GJT, untimed auditory GJT, and metalinguistic test were all significant in the early AO group: .37 ($p = .007$), .38 ($p = .007$), and .33 ($p = .018$) (disattenuated correlations were .54, .54, and .47). Early L2 learners with high and low LLAMA B, E, and F composite scores (i.e., $z > .5$ and $z < -.5$, respectively) were significantly different on the three language tests: Untimed visual GJT ($t(24) = 2.939$, $p = .007$, mean difference of 9.38), untimed auditory GJT ($t(24) = 2.876$, $p = .008$, mean difference of 10.30), and metalinguistic test ($t(24) = 3.313$, $p = .003$, mean difference of 11.06). In the late AO group, only the correlation between LLAMA B, E, and F and performance on the metalinguistic test was significant ($r = .43$, $p = .002$) (the disattenuated correlation was .61). Late L2 learners with a high LLAMA B, E, and F composite score performed significantly higher than late L2 learners with a low composite score ($t(26) = 2.289$, $p = .030$, mean difference of 9.28). The correlations between the LLAMA B, E, and F composite and late L2 learners' scores on the untimed visual GJT and untimed auditory GJT were .22 ($p = .117$) and .14 ($p = .320$),

respectively[31] (disattenuated correlations were .32 and .20) (but see section 5.3.1.1 for follow-up analyses on the relationship between LLAMA E and the untimed auditory GJT in the late AO group).

Similar results were obtained when only ungrammatical items were considered. The MANCOVA analysis yielded no significant interactions with group at any level ($p > .05$). As a covariate, the LLAMA B, E, and F composite was significant at the multivariate level with an effect size that was medium large ($F(3,94) = 4.454, p = .006, \eta_p^2 = .126, \lambda = .874$) and, at the univariate level, for the untimed visual GJT ($F(1,96) = 4.160, p = .044, \eta_p^2 = .042$) and the metalinguistic test ($F(1,96) = 12.507, p = .001, \eta_p^2 = .116$) with a small and a medium large effect size, respectively. The results for the untimed auditory GJT did not reach significance ($F(1,96) = 3.121, p = .080, \eta_p^2 = .032$).

Simple correlations between LLAMA B, E, and F and early L2 learners' performance on ungrammatical items on the untimed visual GJT, untimed auditory GJT, and metalinguistic test were .34 ($p = .018$), .28 ($p = .054$), and .42 ($p = .003$), respectively (disattenuated correlations were .49, .40, and .59). Early L2 learners with high and low LLAMA B, E, and F composite scores were significantly different on the untimed visual GJT ($t(24) = 2.156, p = .041$, mean difference of 12.78) and metalinguistic test ($t(24) = 3.116, p = .005$, mean difference of 19.60). The difference approached significance for the untimed auditory GJT ($t(24) = 2.055, p = .051$, mean difference of 12.34). In the late AO group, only the correlation between LLAMA B, E, and F and performance on ungrammatical items on the metalinguistic test

---

[31] The correlations between general intelligence and late L2 learners' scores on the untimed visual and untimed auditory GJTs did not reach significance, either ($r = .27, p = .062$ and $r = .04, p = .808$).

approached significance ($r = .27$, $p = .059$) (the disattenuated correlation was .38). The correlations with the untimed visual and auditory GJTs were .08 ($p = .572$) and .02 ($p = .883$)[32].

As for agreement ($k = 30$) (gender, person, and number) and non-agreement items ($k = 30$) (aspect contrasts, the subjunctive, and the passive), the MANCOVA analyses yielded no significant interactions at the multivariate or univariate level ($p > .05$). The LLAMA B, E, and F composite was a significant covariate for both agreement and non-agreement items at the multivariate level ($F(3,94) = 4.393$, $p = .006$, $\eta_p^2 = .124$, $\lambda = .874$, and $F(3,94) = 4.778$, $p = .004$, $\eta_p^2 = .134$, $\lambda = .866$, respectively), with medium large effect sizes in both cases. At the univariate level, it was also a significant covariate for agreement items on the untimed auditory GJT ($F(1,96) = 8.496$, $p = .004$, $\eta_p^2 = .082$) and metalinguistic test ($F(1,96) = 12.263$, $p = .001$, $\eta_p^2 = .114$), with a medium and a medium large effect size, respectively, and it approached significance for the untimed visual GJT ($F(1,96) = 3.949$, $p = .050$, $\eta_p^2 = .040$). In the case of non-agreement items, the LLAMA B, E, and F composite was significant for the untimed visual GJT ($F(1,96) = 12.144$, $p = .001$, $\eta_p^2 = .113$) and the metalinguistic test $F(1,96) = 10.442$, $p = .002$, $\eta_p^2 = .099$), with a medium large effect size, but non-significant for the untimed auditory GJT $F(1,96) = 2.816$, $p = .097$, $\eta_p^2 = .029$).

Simple correlations between the LLAMA B, E, and F composite and early L2 learners' performance on agreement items on the untimed visual GJT, untimed auditory GJT, and metalinguistic test were .29 ($p = .039$), .35 ($p = .015$), and .35 ($p = .012$). Early L2 learners with high and low LLAMA composite scores were

---

[32] The correlations between general intelligence and late L2 learners' scores on the untimed visual and untimed auditory GJTs were .25 ($p = .079$) and -.01 ($p = .942$), respectively.

significantly different on the three tests: untimed visual GJT ($t(24) = 2.709$, $p = .012$, mean difference of 10.36), untimed auditory GJT ($t(24) = 3.042$, $p = .006$, mean difference of 14.68), and metalinguistic test ($t(24) = 2.729$, $p = .012$, mean difference of 12.62). Correlations for non-agreement items were .39 ($p = .006$), .30 ($p = .034$), and .40 ($p = .004$), on the untimed visual GJT, untimed auditory GJT, and metalinguistic test, respectively (disattenuated correlations were .57, .42, and .57). Early L2 learners with high and low LLAMA B, E, and F composite scores were significantly different on the untimed visual GJT ($t(24) = 2.593$, $p = .016$, mean difference of 8.41) and the metalinguistic test ($t(24) = 2.781$, $p = .010$, mean difference of 9.48), but not on the untimed auditory GJT ($t(24) = 3.042$, $p = .006$, mean difference of 14.68).

In the late AO group, the only significant correlation was between the LLAMA composite and agreement items on the metalinguistic test ($r = .34$, $p = .016$) (the disattenuated correlation was .48). This was the only test that yielded a significant difference between learners with high and low scores on the composite ($t(26) = 2.103$, $p = .045$, mean difference of 10.85). The correlations between the LLAMA composite and agreement items on the untimed visual and untimed auditory GJTs were .09 ($p = .534$) and .18 ($p = .207$) (but see the follow-up analyses reported for the untimed auditory GJT and LLAMA E in section 5.3.1.1). As for non-agreement items, the only significant correlation in the late AO group was with the untimed visual GJT ($r = .29$, $p = .039$) (the disattenuated correlation was.42), although this test did not yield a significant difference between learners with high and low scores on the composite ($t(26) = 1.704$, $p = .100$, mean difference of 6.78). The correlations between the

LLAMA composite and non-agreement items on the untimed auditory GJT and metalinguistic test were .07 ($p = .622$) and .23 ($p = .110$) (disattenuated correlations were .10 and .33). Overall, in the late AO group, the correlations between language performance and the LLAMA B, E, and F composite score mirrored the correlations between language performance and intelligence scores. Like the LLAMA composite score, intelligence scores were significantly correlated with agreement items on the metalinguistic test ($r = .38$, $p = .007$) and with non-agreement items on the untimed visual GJT ($r = .36$, $p = .011$) (disattenuated correlations were .54 and .52).

To summarize, explicit language aptitude (LLAMA B, E, and F) moderated both early and late L2 learners' attainment on measures hypothesized to allow controlled use of L2 knowledge. General intelligence moderated attainment on the same language measures, but only among late L2 learners.

5.3.2.2 Tasks that Require Automatic Use of Language Knowledge

Figures 42, 43, and 44 display individual scores on the timed visual GJT, timed auditory GJT, and word monitoring task (GSI) as a function of AO, with the general intelligence dimension added. The NS range is marked with a dotted line. The general intelligence groups were created by converting GAMA raw scores into *z*-scores within each of the three speaker groups and by establishing the following cutoffs reflecting distance from the mean in standard deviations: *high* = *z*-scores >.5, *mid* = -.5 < *z*-scores < .5, and *low* = *z*-scores < -.5.

The highest scorer on the timed visual GJT in the early AO group was a high-intelligence L2 learner, whereas, in the late AO group, a combination of high-, mid-, and low-intelligence L2 learners scored within the NS range. On the timed auditory

GJT, a mid-intelligence L2 learner obtained the highest score in the early AO group, while, in the late AO group, two low-intelligence learners overlapped within the NS range. Finally, the highest grammatical sensitivity indices on the word monitoring task corresponded to high-intelligence L2 learners in both the early and late AO groups.
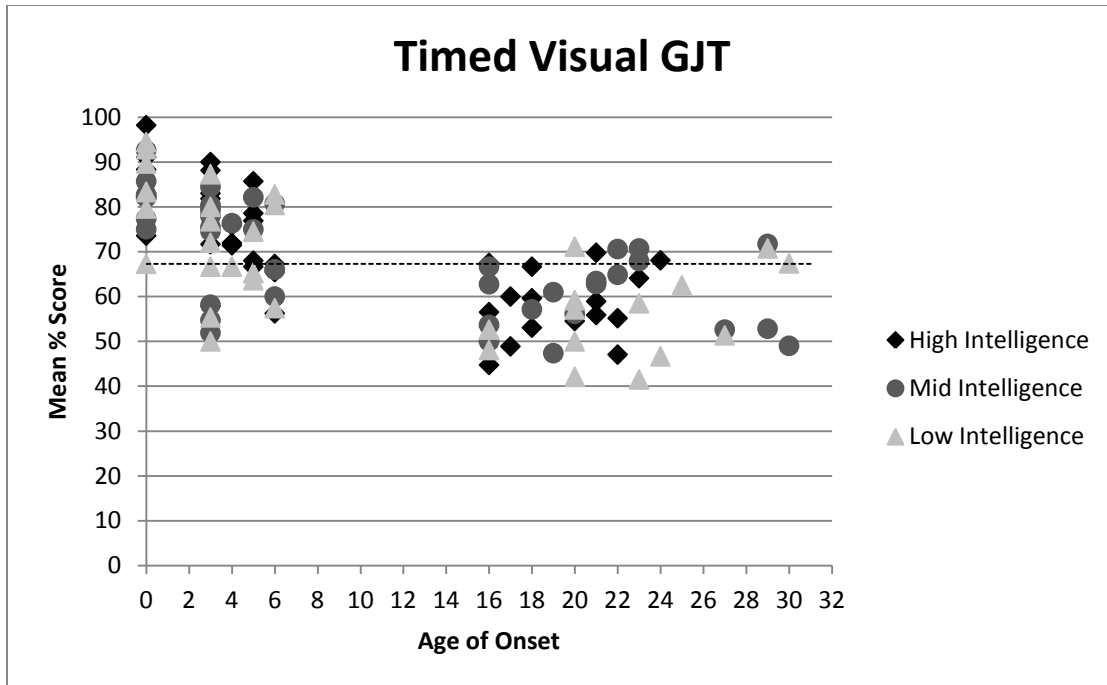
Figure 42. Timed visual GJT scores as a function of AO with the general intelligence dimension added
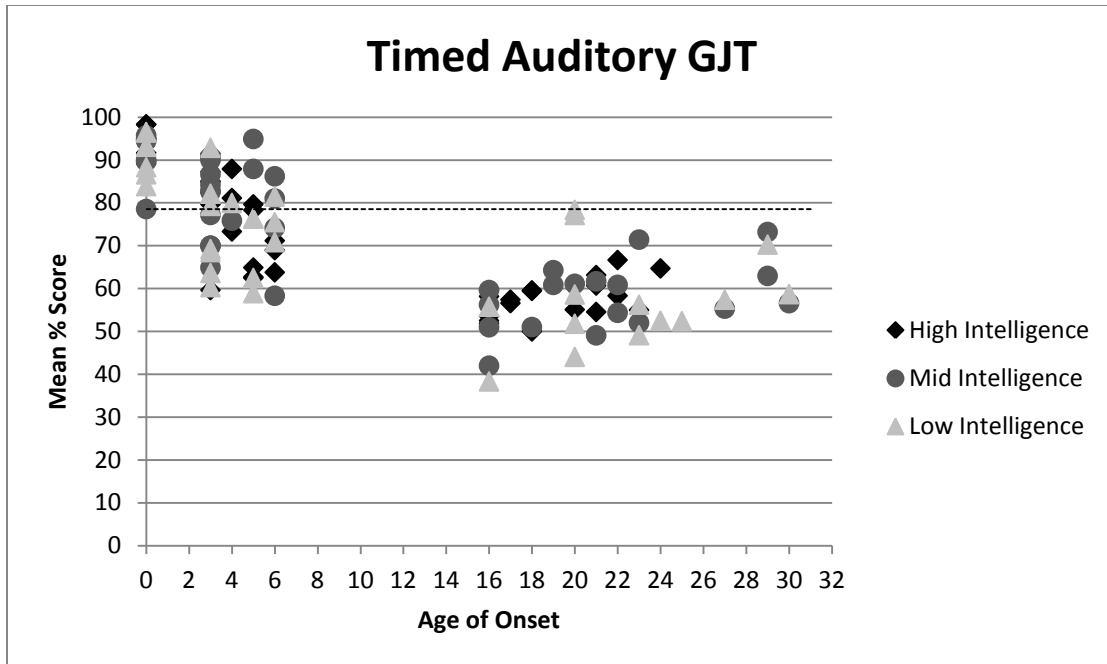
Figure 43. Timed auditory GJT scores as a function of AO with the general

intelligence dimension added
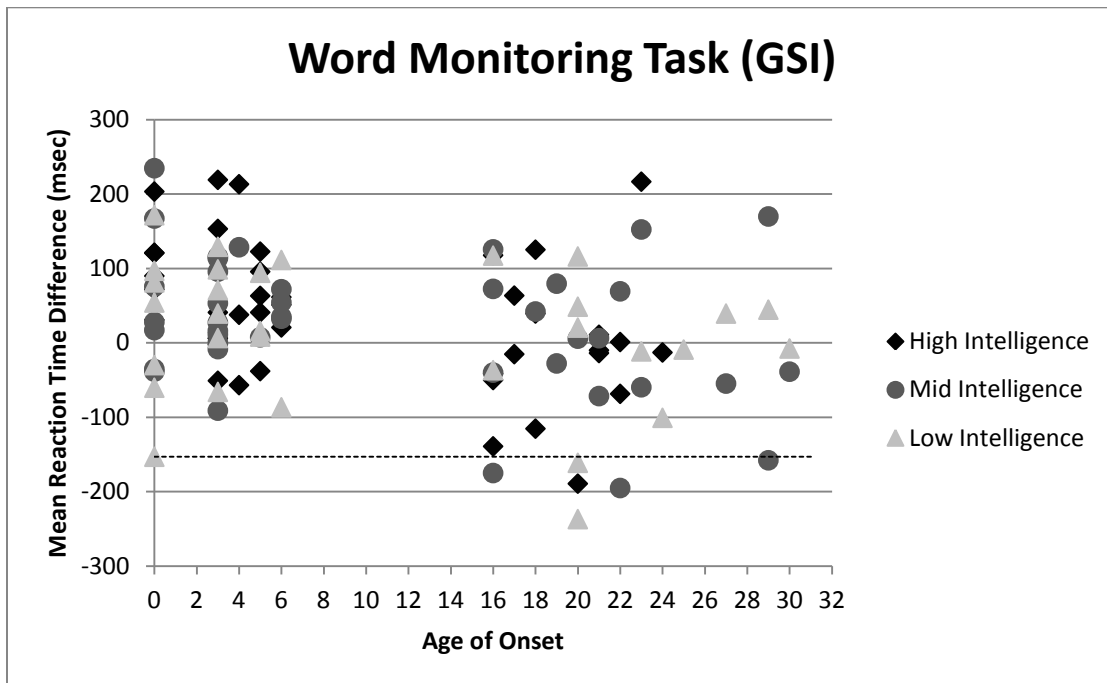
Figure 44. Word monitoring task scores (GSI) as a function of AO with the general

intelligence dimension added

In order to investigate the role of general intelligence on participants' language attainment as measured by tasks that require automatic use of language knowledge, a MANCOVA was conducted with overall test scores on the timed visual GJT, timed auditory GJT, and word monitoring task (i.e., GSI) as dependent variables, group (NS controls, early L2 learners, and late L2 learners) as fixed factor, and GAMA scores as covariate. An interaction term was added, in addition to the group and covariate terms, to test for possible interactions between covariate and group as an independent factor. The results revealed no significant interactions at the multivariate or univariate level ($p > .05$). Intelligence was not a significant covariate either, either at the multivariate level ($F(3,105) = .682$, $p = .664$, $\eta_p^2 = .018$, $\lambda = .964$), or univariate level for the timed visual GJT ($F(1,107) = .582$, $p = .447$, $\eta_p^2 = .005$), timed auditory GJT ($F(1,107) = 2.378$, $p = .126$, $\eta_p^2 = .020$), or word monitoring task ($F(1,107) = 1.011$, $p = .317$, $\eta_p^2 = .009$).[33] Results remained non-significant when L2 learners were combined as one group. Intelligence was not a significant covariate at the multivariate level ($F(3,107) = 1.960$, $p = .124$, $\eta_p^2 = .049$, $\lambda = .951$) or, at the univariate level, for the timed visual GJT ($F(1,109) = 1.663$, $p = .200$, $\eta_p^2 = .014$), timed auditory GJT ($F(1,109) = 2.486$, $p = .118$, $\eta_p^2 = .021$), or word monitoring task ($F(1,109) = 2.741$, $p = .101$, $\eta_p^2 = .023$). Interactions were all non-significant, as well ($p > .05$).[34]

---

[33] The analysis including outliers in the word monitoring task also yielded non-significant results: Intelligence was a non-significant covariate at the multivariate level ($F(3,112) = 1.376$, $p = .254$, $\eta_p^2 = .038$, $\lambda = .962$) and at the univariate level for the timed visual GJT ($F(1,114) = .392$, $p = .533$, $\eta_p^2 = .004$), timed auditory GJT ($F(1,114) = 1.434$, $p = .234$, $\eta_p^2 = .013$), and word monitoring task ($F(1,114) = 3.295$, $p = .072$, $\eta_p^2 = .030$). Interactions were all non-significant, as well ($p > .05$).

[34] The analysis including outliers also yielded non-significant results. Intelligence as a covariate was not significant at the multivariate level ($F(3,114) = 1.627$, $p = .187$ $\eta_p^2 = .041$, $\lambda = .959$) and, at the univariate level, for the timed visual GJT ($F(1,116) = 1.663$, $p = .200$ $\eta_p^2 = .014$), timed auditory GJT ($F(1,116) = 3.295$, $p = .072$ $\eta_p^2 = .028$), and word monitoring task ($F(1,116) = 2.741$, $p = .101$ $\eta_p^2 = .023$). Interactions were all non-significant, as well ($p > .05$).

The same results were found for ungrammatical items. The analysis with three groups (NS controls, early L2 learners, and late L2 learners) and the timed visual and timed auditory GJTs (the word monitoring task does not provide an interpretable measure for ungrammatical items only) yielded no significant interactions ($p > .05$). Intelligence was not a significant covariate at the multivariate level ($F(2,113) =$ 1.193, $p = .307$ $\eta_p^2 = .021$, $\lambda = .979$), or, univariate level, for the timed visual ($F(1,114) = .101$, $p = .752$, $\eta_p^2 = .001$) or timed auditory GJT ($F(1,114) = 1.941$, $p =$ .395, $\eta_p^2 = .016$). Combining L2 learners into one group made no difference to the results. Interactions remained non-significant ($p > .05$) and intelligence remained a non-significant covariate at the multivariate level ($F(2,115) = 1.322$, $p = .271$ $\eta_p^2 =$ .023, $\lambda = .977$) and at the univariate level for the timed visual ($F(1,116) = .254$, $p =$ .615, $\eta_p^2 = .002$) and timed auditory GJT ($F(1,116) = 2.182$, $p = .142$, $\eta_p^2 = .019$). Finally, the results for agreement and non-agreement target structures were all non-significant, as well. Intelligence was not a significant covariate and did not interact with group at any level ($p > .05$).

To summarize, general intelligence did not moderate any of the groups' language attainment on the three measures hypothesized to require automatic use of L2 knowledge.

5.3.3 Aptitude for Implicit Learning and Language Attainment

This section presents the results of the role of aptitude for implicit learning on language attainment as measured by tasks that allow controlled use of language knowledge (section 5.3.3.1) and measures that require automatic use of language knowledge (section 5.3.3.2). First, descriptive data is presented visually on

scatterplots that show attainment scores as a function of age of onset with the aptitude

for implicit learning dimension added. This visual display allows determining to what

extent a high level of implicit aptitude is a necessary condition at an individual level

in order to score within NS range. Next, multivariate analyses of covariance

(MANCOVAs) are conducted in order to determine the extent to which aptitude for

implicit learning moderates language attainment in each of the groups. A

MANCOVA was first conducted on overall test scores, grammatical and

ungrammatical, and then re-run on ungrammatical items, agreement items, and non-

agreement items in follow-up analyses.

5.3.3.1 Tasks that Allow Controlled Use of Language Knowledge

Figures 45, 46, and 47 display individual scores on the metalinguistic knowledge

test, untimed visual GJT, and untimed auditory GJT as a function of AO with the

aptitude for implicit learning dimension added. The NS range is marked with a dotted

line. The implicit aptitude groups (high, mid, and low) were created by establishing

the following cutoffs on the aptitude for implicit learning composite score in every

speaker group: *high* = z-scores >.5, *mid* = -.5 < z-scores < .5, and *low* = z-scores < -.5.

The highest scorers on the three tests in the two learner groups had either mid or

low implicit language aptitude. These are the same L2 learners that had either high or

mid aptitude for explicit language learning, except for a learner in the late AO group

who scored within the NS range on the metalinguistic test and who was high in both

types of aptitude.

174
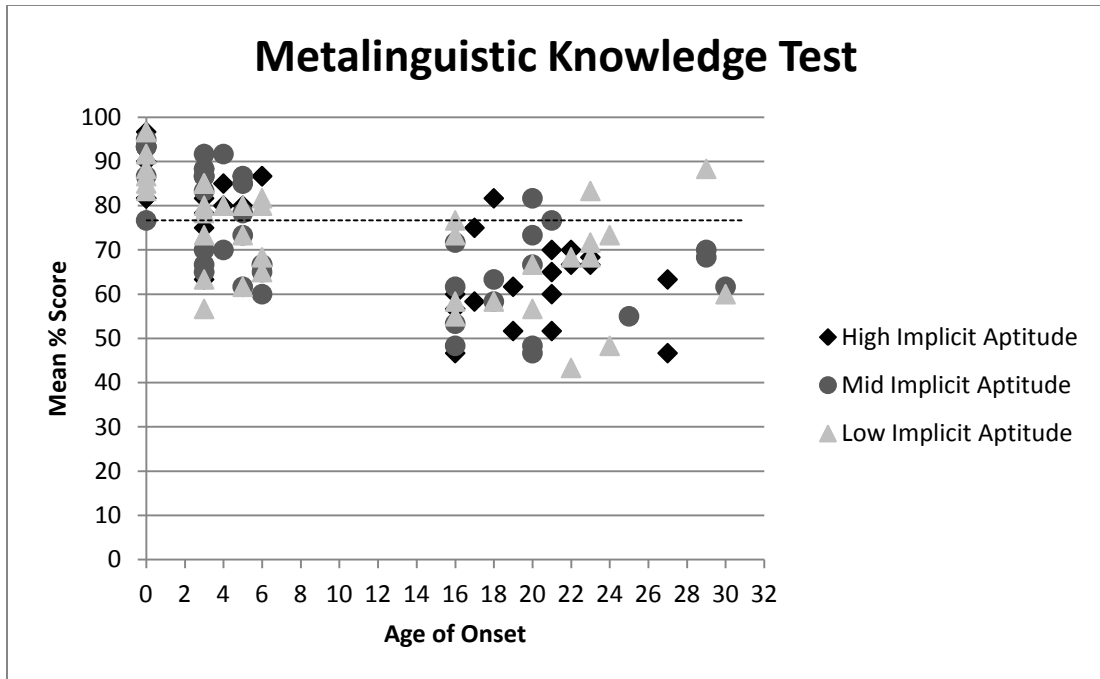
Figure 45. Metalinguistic knowledge test scores as a function of AO with the implicit

language aptitude dimension added
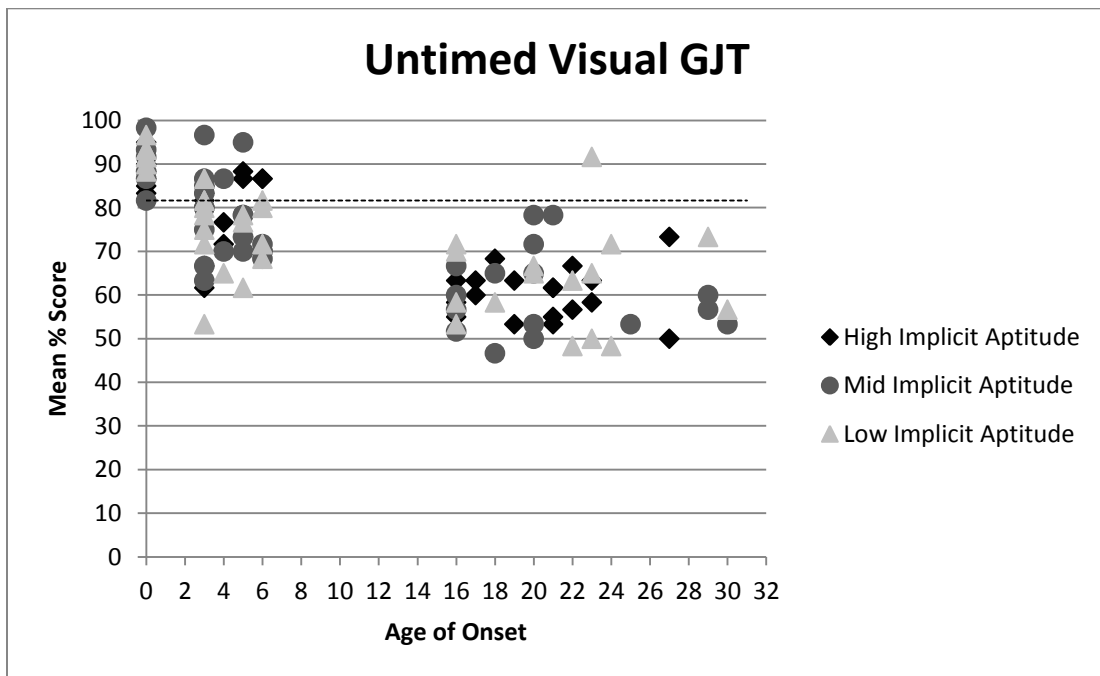
Figure 46. Untimed visual GJT scores as a function of AO with the implicit language

aptitude dimension added
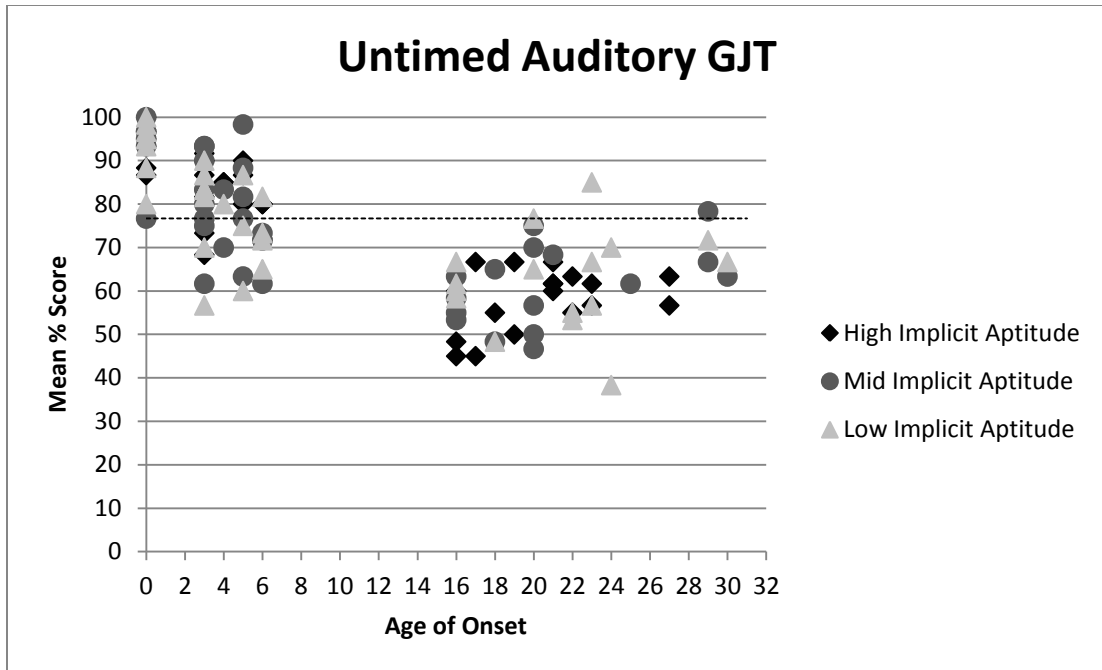
Figure 47. Untimed auditory GJT scores as a function of AO with the implicit

language aptitude dimension added

In order to investigate the role of aptitude for implicit learning in participants'

language attainment as measured by tasks that allow controlled use of language

knowledge, a MANCOVA was conducted with overall test scores on the untimed

visual GJT, untimed auditory GJT, and metalinguistic test as dependent variables,

group (NS controls, early L2 learners, and late L2 learners) as fixed factor, and the

composite aptitude score combining LLAMA D and the learning score on the

probabilistic SRT task (i.e., aptitude for implicit learning) as covariate. An interaction

term was added, in addition to the group and covariate terms, to test for possible

interactions between covariate and group as an independent factor.

The results revealed no significant interactions at the multivariate or univariate

level ($p > .05$). Aptitude for implicit learning was not a significant covariate, either,

either at the multivariate level ($F(3,112) = .723$, $p = .540$, $\eta_p^2 = .019$, $\lambda = .981$), or univariate level for the untimed visual GJT ($F(1,114) = .136$, $p = .713$, $\eta_p^2 = .001$), untimed auditory GJT ($F(1,114) = .113$, $p = .737$, $\eta_p^2 = .001$) or metalinguistic test ($F(1,114) = .379$, $p = .539$, $\eta_p^2 = .003$). Results remained non-significant when L2 learners were combined as a single group and compared with NS controls ($p > .05$). Implicit aptitude was not a significant covariate at the multivariate level ($F(3,114) = .281$, $p = .839$, $\eta_p^2 = .007$, $\lambda = .993$), or, at the univariate level, for the untimed visual GJT ($F(1,116) = .121$, $p = .729$, $\eta_p^2 = .001$), untimed auditory GJT ($F(1,116) = .592$, $p = .443$, $\eta_p^2 = .005$), or metalinguistic test ($F(1,116) = .151$, $p = .698$, $\eta_p^2 = .001$). Interactions were all non-significant, as well ($p > .05$).

The same results were found for ungrammatical items. The analysis with three groups (NS controls, early L2 learners, and late L2 learners) yielded no significant interactions ($p > .05$). Aptitude for implicit learning was not a significant covariate at the multivariate level ($F(6,112) = .345$, $p = .793$ $\eta_p^2 = .009$, $\lambda = .991$), or, univariate level, for the untimed visual GJT ($F(1,114) = .873$, $p = .352$, $\eta_p^2 = .000$), untimed auditory GJT ($F(1,114) = .124$, $p = .726$, $\eta_p^2 = .001$), or metalinguistic test ($F(1,114) = .602$, $p = .440$, $\eta_p^2 = .005$). Combining L2 learners as a single group made no difference to the results ($p > .05$).

Regarding agreement and non-agreement structures, the results for non-agreement items were all non-significant ($p > .05$). However, the MANOVA performed on agreement items yielded a significant two-way interaction between group and aptitude for implicit learning at the univariate level in two of the tests hypothesized to allow controlled use of language knowledge: the untimed auditory GJT ($F(2,114) =$

4.627, $p = .012$, $\eta_p^2 = .076$) and the metalinguistic test ($F(2,114) = 4.254$, $p = .017$, η_p^2 = .070), both associated with a medium effect size. The interaction in the case of the untimed visual GJT did not reach significance ($F(2,114) = 2.121$, $p = .125$, $\eta_p^2 = .036$). At the multivariate level, the interaction between group and covariate had a p value of .085 ($F(6,224) = 1.881$, $p = .085$, $\eta_p^2 = .048$, $\lambda = .906$). Finally, as a covariate at the multivariate and univariate level, aptitude for implicit learning was not significant ($p > .05$).

The existence of a significant two-way interaction between scores on agreement items and group indicated that the effects of aptitude for implicit learning were not comparable in the three groups of participants, as can be observed in Figures 48 and 49.
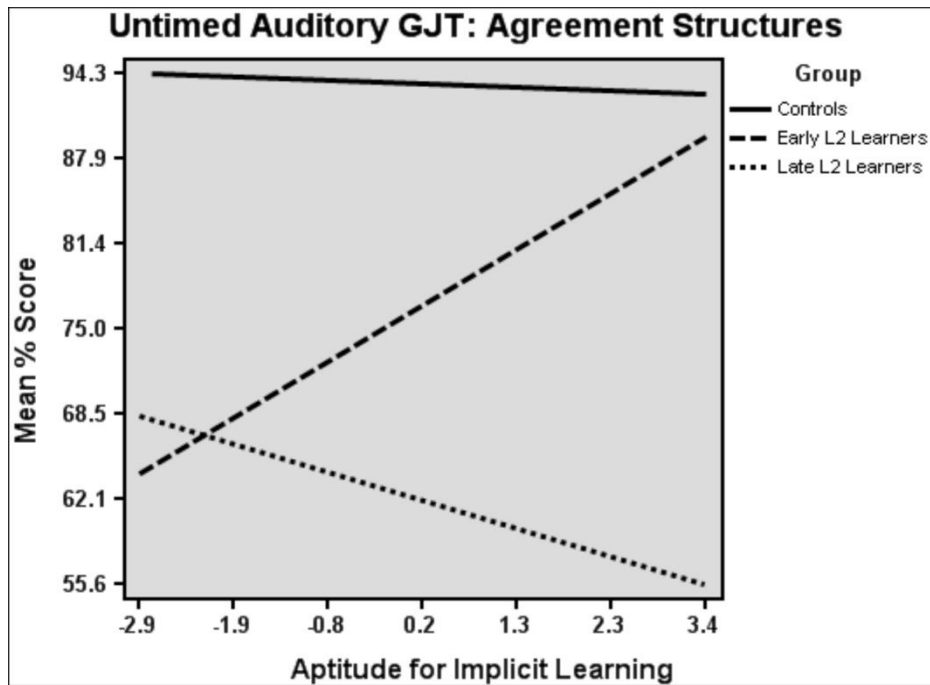
Figure 48. Two-way interaction between group and aptitude for implicit learning in the untimed auditory GJT (agreement structures)
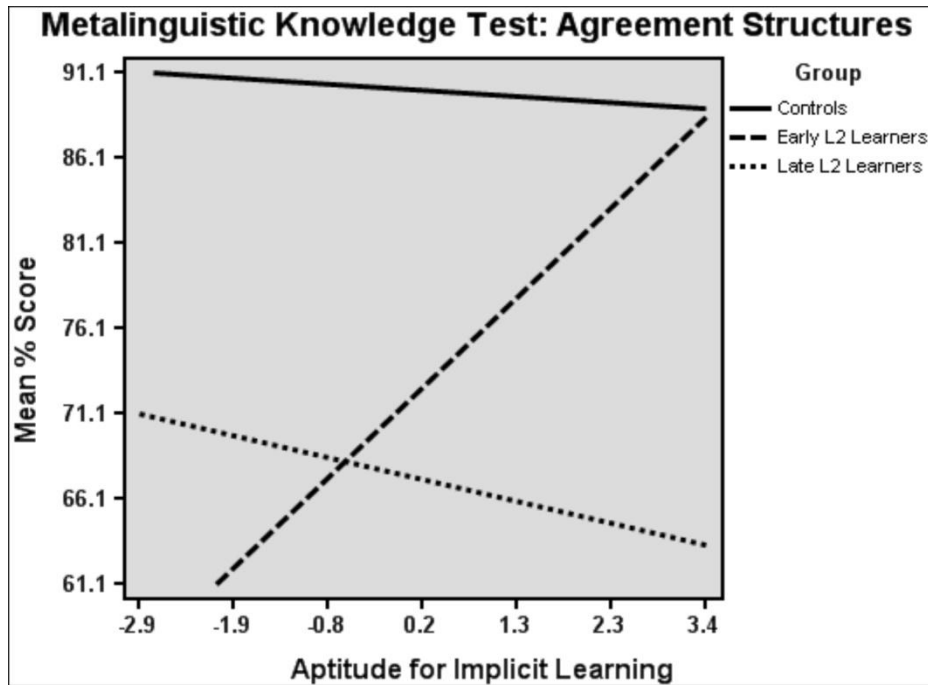
Figure 49. Two-way interaction between group and aptitude for implicit learning in the metalinguistic knowledge test (agreement structures)

Follow-up correlations in each of the groups showed a significant positive relationship between aptitude for implicit learning and agreement items on the untimed auditory GJT and metalinguistic test in the early L2 group only: .29 (p = .045) and .39 (p = .005), respectively (disattenuated correlations were .42 and .57). The correlation for agreement items on the untimed visual GJT was not significant (r = .19, p = .193). In the NS control and late AO groups, correlations did not reach significance and, unlike in the early AO group, they were all negative: -.37 (p = .110) and -.18 (p = .215) (untimed visual GJT), -.05 (p = .846) and -.27 (p = .056) (untimed auditory GJT), and -.06 (p = .792) and -.12 (p = .394) (metalinguistic test) (disattenuated correlations were -.55, -.27, -.07, -.40, -.09, and -.18).

Differences between high and low implicit aptitude individuals were only significant in the early AO group. High implicit aptitude early L2 learners ($n = 14$) scored significantly higher than their low-aptitude counterparts ($n = 16$) on the untimed auditory GJT ($M = 84.29$, $SD = 10.08$ and $M = 71.46$, $SD = 11.80$) ($t(28) = 3.177$, $p = .004$) and on the metalinguistic test ($M = 78.10$, $SD = 13.12$ and $M = 66.46$, $SD = 10.64$) ($t(28) = 2.681$, $p = .012$), but not on the untimed visual GJT ($p = .130$), even though high implicit aptitude early learners also scored higher on this test.

To summarize, individual differences in aptitude for implicit learning did not moderate language attainment in any of the groups on measures hypothesized to allow controlled use of L2 knowledge, when overall scores or scores on ungrammatical items were considered. Follow-up analyses on agreement structures, however, showed a positive effect of aptitude for implicit learning in the group of early L2 learners. This effect was present at the univariate level for agreement items on the untimed auditory GJT and the metalinguistic test. Early L2 learners with high implicit language aptitude scored higher on agreement items than early L2 learners with low implicit aptitude.

5.3.3.2 Tasks that Require Automatic Use of Language Knowledge

Figures 50, 51, and 52 display individual scores on the timed visual GJT, timed auditory GJT, and word monitoring task (GSI) as a function of AO with the implicit language aptitude dimension added. The NS range is marked with a dotted line. The implicit aptitude groups (high, mid, and low) were created by establishing the following cutoffs on the aptitude for implicit learning composite score in every speaker group: *high* = z-scores >.5, *mid* = -.5 < z-scores < .5, and *low* = z-scores < -.5.

The highest scorer on the timed visual GJT in the early AO group, a learner with high explicit language aptitude, was also high in terms of implicit language aptitude. In the late AO group, those learners who scored within the NS range and who had either low, mid, or high explicit language aptitude, had mostly low, or mid, implicit language aptitude. On the timed auditory GJT, the L2 learner who had high explicit language aptitude and who obtained the highest score in the early AO group also had high implicit language aptitude. In the late AO group, the two learners who overlapped within the NS range, and who had either low or mid explicit language aptitude, had both low implicit language aptitude. Finally, the highest grammatical sensitivity indices on the word monitoring task corresponded to L2 learners with high implicit language aptitude. These learners also had high explicit language aptitude.
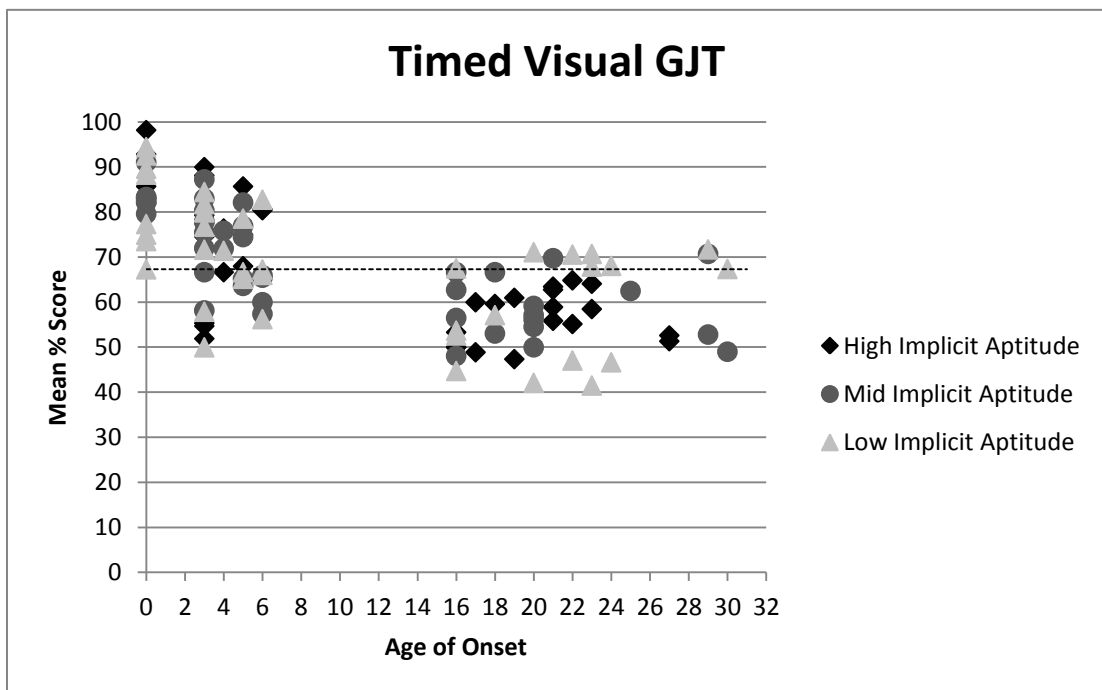
Figure 50. Timed visual GJT scores as a function of AO with the implicit language aptitude dimension added
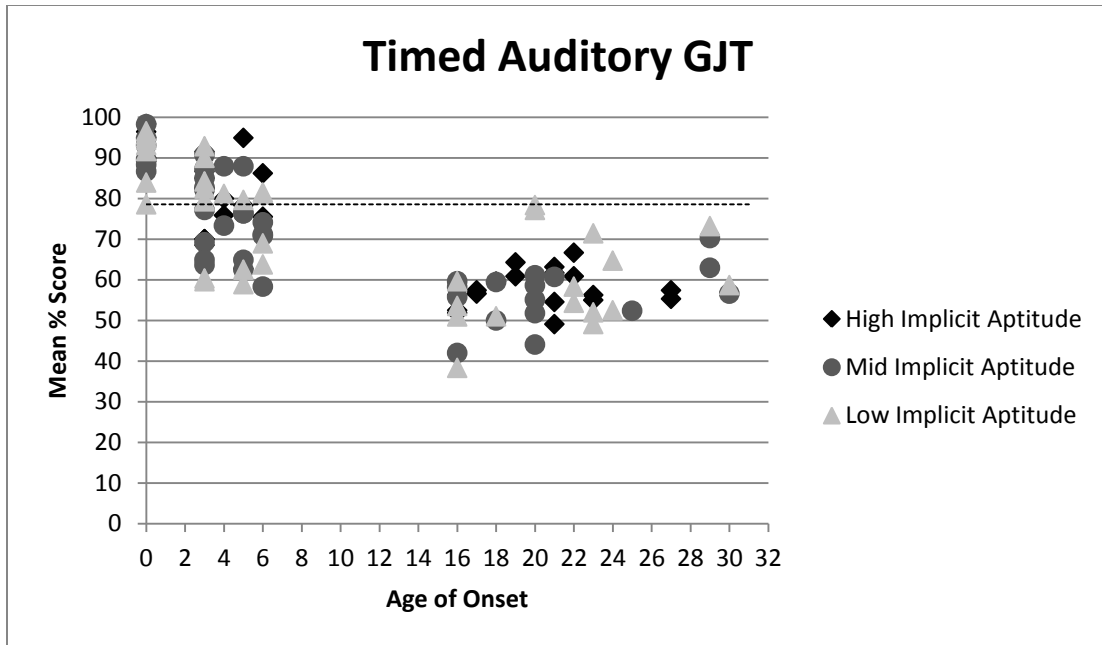
Figure 51. Timed auditory GJT scores as a function of AO with the implicit language

aptitude dimension added
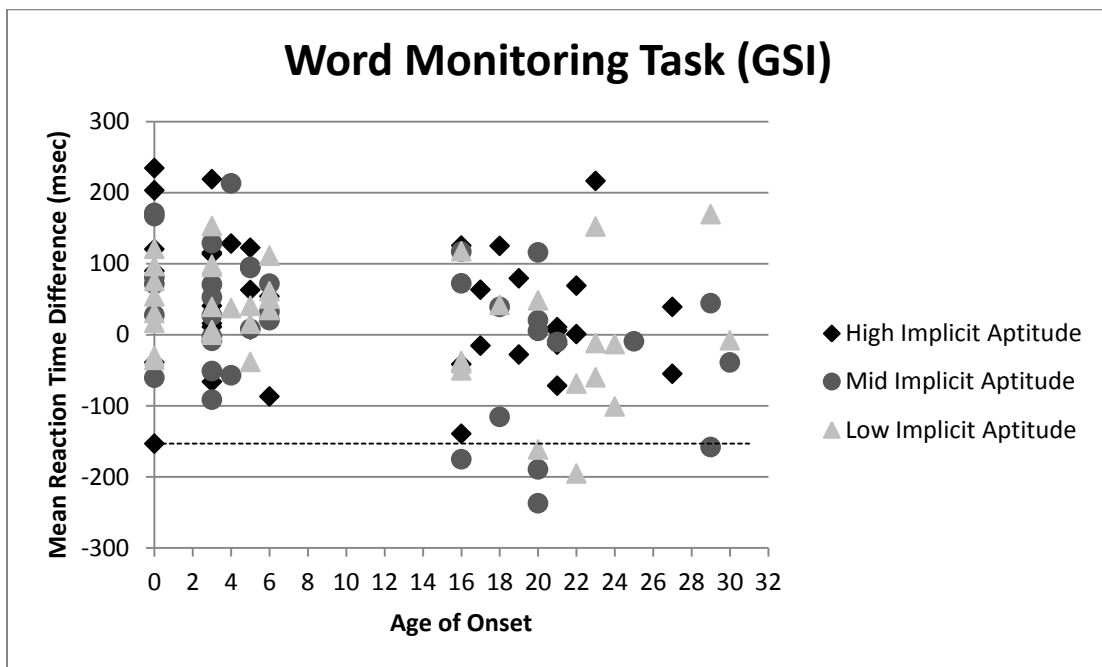
Figure 52. Word monitoring task scores (GSI) as a function of AO with the implicit

language aptitude dimension added

In order to investigate the role of aptitude for implicit learning in participants'
language attainment as measured by tasks hypothesized to require automatic use of
language knowledge, a MANCOVA was conducted with overall test scores on the
timed visual GJT, timed auditory GJT, and word monitoring task (i.e., GSI) as
dependent variables, group (NS controls, early L2 learners, and late L2 learners) as
fixed factor, and the composite aptitude score combining LLAMA D and the learning
score on the SRT task (i.e., aptitude for implicit learning) as a covariate. An
interaction term was added, in addition to the group and covariate terms, to test for
possible interactions between covariate and group as an independent factor.

The results revealed no significant interactions at the multivariate or univariate
level ($p > .05$). Aptitude for implicit learning did not reach significance, either, as a
covariate at the multivariate level ($F(3,105) = 2.216$, $p = .090$, $\eta_p^2 = .057$, $\lambda = .943$),
or univariate level for the timed visual GJT ($F(1,107) = 1.803$, $p = .182$, $\eta_p^2 = .016$),
timed auditory GJT ($F(1,107) = .450$, $p = .504$, $\eta_p^2 = .004$), or word monitoring task,
which had a p value of .082 ($F(1,107) = 3.079$, $p = .082$, $\eta_p^2 = .027$).[35] Results
remained non-significant when L2 learners were combined as one group and
compared with NS controls ($p > .05$). Aptitude was not a significant covariate at the
multivariate level ($F(3,107) = 1.948$, $p = .126$, $\eta_p^2 = .049$, $\lambda = .951$), or, at the
univariate level, for the timed visual GJT ($F(1,109) = .548$, $p = .461$, $\eta_p^2 = .005$),
timed auditory GJT ($F(1,109) = .666$, $p = .416$, $\eta_p^2 = .006$), or word monitoring task

---

[35] The analysis including outliers in the word monitoring task yielded similar results: aptitude for
implicit learning was not a significant covariate at the multivariate level ($F(3,112) = 1.805$, $p = .150$,
$\eta_p^2 = .046$, $\lambda = .954$), or, at the univariate level, for the timed visual GJT ($F(1,114) = 2.413$, $p = .123$,
$\eta_p^2 = .021$), timed auditory GJT ($F(1,114) = .201$, $p = .655$, $\eta_p^2 = .002$), or word monitoring task
($F(1,114) = 1.671$, $p = .199$, $\eta_p^2 = .014$). Interactions were all non-significant ($p > .05$).

($F(1,109) = 1.926$, $p = .168$, $\eta_p^2 = .016$). Interactions were all non-significant, as well ($p > .05$)[36].

The same results were found for ungrammatical items. The analysis with three groups (NS controls, early L2 learners, and late L2 learners) and the timed visual and timed auditory GJTs (the word monitoring task does not provide an interpretable measure for ungrammatical items only) yielded no significant interactions ($p > .05$). Aptitude for implicit learning was not a significant covariate at the multivariate level ($F(2,113) = 1.819$, $p = .167$ $\eta_p^2 = .031$, $\lambda = .969$), or, univariate level, for the timed visual ($F(1,114) = 1.742$, $p = .190$, $\eta_p^2 = .015$) or timed auditory GJT ($F(1,114) = .069$, $p = .793$, $\eta_p^2 = .001$). Combining L2 learners as a single group made no difference to the results. Interactions remained non-significant ($p > .05$) and intelligence remained a non-significant covariate at the multivariate level ($F(2,115) = 2.075$, $p = .130$ $\eta_p^2 = .035$, $\lambda = .965$) and at the univariate level for the timed visual ($F(1,116) = .998$, $p = .320$, $\eta_p^2 = .009$) and timed auditory GJT ($F(1,116) = .311$, $p = .578$, $\eta_p^2 = .003$).

As for agreement ($k = 30$) (gender, person, and number) and non-agreement items ($k = 30$) (aspect contrasts, the subjunctive, and the passive), the MANCOVA analyses yielded no significant interactions at the multivariate or univariate level ($p > .05$). However, aptitude for implicit learning was a significant covariate for agreement structures at the multivariate level with the three speaker groups as a fixed factor

---

[36] The analysis including outliers in the word monitoring task yielded similar results: aptitude for implicit learning was not a significant covariate at the multivariate level ($F(3,114) = 1.505$, $p = .217$, $\eta_p^2 = .038$, $\lambda = .962$), or, at the univariate level, for the timed visual GJT ($F(1,116) = .904$, $p = .344$, $\eta_p^2 = .008$), timed auditory GJT ($F(1,116) = .401$, $p = .528$, $\eta_p^2 = .003$), or word monitoring task ($F(1,116) = .679$, $p = .412$, $\eta_p^2 = .006$). Interactions were all non-significant ($p > .05$).

$(F(3,102) = 3.217, p = .026, \eta_p^2 = .086, \lambda = .914)$. The size of the effect was medium.

Univariate analyses further showed that this was mostly due to the significant effect

of the covariate on the word monitoring task[37] (i.e., GSI for agreement structures)

$(F(1,104) = 6.653, p = .011, \eta_p^2 = .060)$, since aptitude for implicit learning did not

moderate scores on the timed visual or auditory GJTs $(F(1,104) = 1.226, p = .271, \eta_p^2$

$= .012$ and $F(1,104) = .185, p = .668, \eta_p^2 = .002$, respectively).[38] The multivariate

effect of aptitude approached significance when L2 learners were combined and

compared with controls $(F(3,104) = 2.513, p = .063, \eta_p^2 = .068, \lambda = .932)$, as well as,

at the univariate level, for the word monitoring task $F(1,104) = 3.884, p = .051, \eta_p^2 =$

$.035)$, but not for the timed visual or auditory GJTs $(F(1,104) = .384, p = .537, \eta_p^2 =$

$.004$ and $F(1,104) = .750, p = .388, \eta_p^2 = .007$, respectively).[39] The interactions

between group and covariate were not significant for any of the language measures:

timed visual GJT $(F(2,104) = .382, p = .002, \eta_p^2 = .961)$, timed auditory GJT

$(F(2,104) = 1.115, p = .293, \eta_p^2 = .010)$, or word monitoring task $(F(2,104) = .268, p$

$= .606, \eta_p^2 = .003)$.

   The fact that aptitude for implicit learning was a significant covariate for

agreement structures in the word monitoring task and that there was no significant

---

[37] In addition to the word monitoring task, aptitude for implicit learning showed a trend towards significance in the early L2 group for the timed visual and timed auditory GJTs. Correlations in this group were .26 $(p = .074)$ and .23 $(p = .120)$, respectively. In addition, the difference between high- and low-aptitude early L2 learners had a p value of .086 $(t(28) = 1.779, p = .086)$ in the case of the timed auditory GJT.

[38] The analysis including outliers in the word monitoring task also yielded a significant multivariate effect of the covariate $(F(3,112) = 3.090, p = .030, \eta_p^2 = .076, \lambda = .924)$. Univariate tests were also significant for the word monitoring task $(F(1,114) = 6.010, p = .016, \eta_p^2 = .050)$, but not for the timed visual or auditory GJT $(F(1,114) = 2.153, p = .145, \eta_p^2 = .019$ and $F(1,114) = .177, p = .675, \eta_p^2 = .002$, respectively). Interactions were all non-significant $(p > .05)$.

[39] The analysis including outliers in the word monitoring task yielded non-significant results for the covariate at the multivariate level $(F(3,114) = 2.102, p = .104, \eta_p^2 = .052, \lambda = .948)$ and univariate level, for the timed visual GJT $(F(1,116) = .714, p = .400, \eta_p^2 = .006)$, timed auditory GJT $(F(1,116) = .777, p = .380, \eta_p^2 = .007)$, and word monitoring task $(F(1,116) = 2.593, p = .110, \eta_p^2 = .022)$. The interactions between group and covariate were all non-significant $(p > .05)$.

group-by-covariate interaction suggested a comparable effect of aptitude for implicit learning on grammatical sensitivity towards agreement structures in all the groups. As can be seen on Figure 53, the slopes of the three groups of participants were similar, although steeper in the case of the two L2-learner groups.
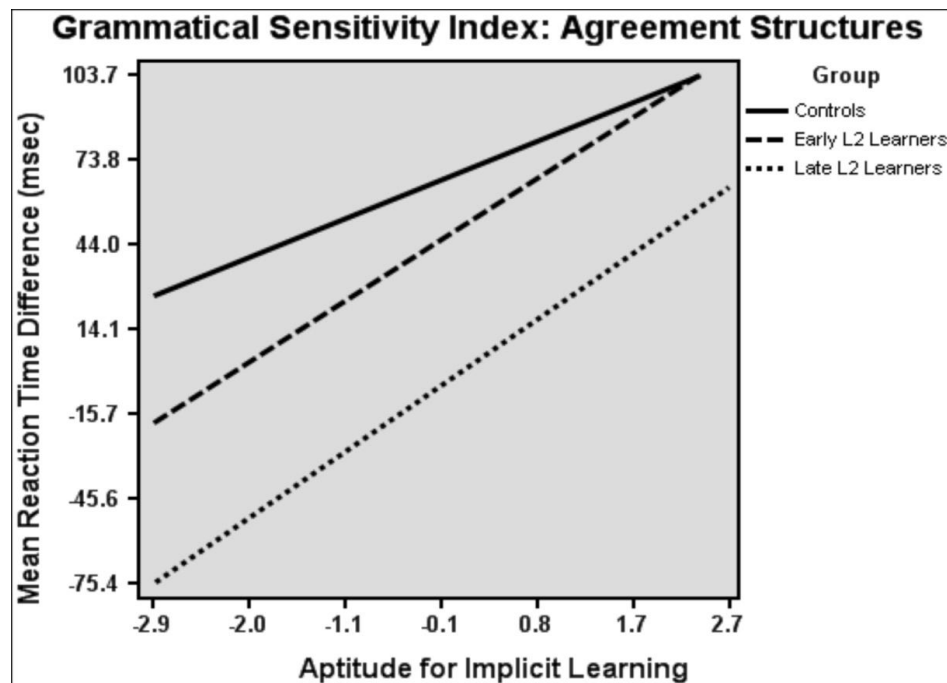
Figure 53. Regression of the grammatical sensitivity index for agreement items on aptitude for implicit learning at each group level

Follow-up simple correlations showed that there was a significant positive relationship between aptitude for implicit learning and sensitivity to agreement structures in the early and late AO groups ($r = .34$, $p = .021$, and $r = .31$, $p = .038$, respectively[40]), but not in the NS control group ($r = .15$, $p = .551$) (disattenuated

---

[40] The correlations including outliers were .28 ($p = .053$) and .40 ($p = .004$), in the early and late AO groups, respectively. The fact that the correlation in the late AO group increased in magnitude from .31 to .40 suggests that outliers are not always the result of task-irrelevant factors. They could be individuals with high aptitude that perform as outliers as a result of their cognitive ability. If so, perhaps they should not be eliminated, since they provide valuable information about the relationship

correlations were .47, .43, and .21). A comparison of the sensitivity indices displayed

by high and low implicit aptitude participants in each group (see Table 30) further

showed no significant differences in the NS control group ($t(12) = 1.009$, $p = .333$,

mean difference of 63.42), but significant differences in both the early and late AO

groups, where L2 learners with high aptitude for implicit learning showed higher

sensitivity than L2 learners with low implicit aptitude ($t(27) = 2.364$, $p = .026$, mean

difference of 69.86, and $t(29) = 3.048$, $p = .005$, mean difference of 97.16,

respectively).

Table 30. *Summary of GSIs for Agreement Structures on the Word monitoring Task*
by High and Low Implicit Aptitude Participants

	Control		Early AO		Late AO	
	High	Low	High	Low	High	Low
	$n = 6$	$n = 8$	$n = 13$	$n = 16$	$n = 15$	$n = 16$
GSI	86.92	23.50	90.32	20.47	55.68	-41.48
Agreement	(157.44)	(74.18)	(87.63)	(71.62)	(96.31)	(80.98)

Note. Standard deviations appear between parentheses.

Given the differential contribution of the LLAMA aptitude subtests B, E, and F,

and the GAMA intelligence scores on early and late L2 learners' language attainment,

a last follow-up analysis was conducted separating the effects of LLAMA D and SRT

learning scores on L2 learners' grammatical sensitivity for agreement structures. With

LLAMA D as a covariate and three groups as a fixed factor (NSs, early L2 learners,

---

between aptitude and test scores (Doughty, p.c., 4/10/2012).

and late L2 learners), multivariate effects were not significant. LLAMA D was not a significant covariate ($F(3,102) = 1.452$, $p = .232$, $\eta_p^2 = .039$, $\lambda = .961$) and the interaction with group was not significant, either ($F(6,204) = 1.130$, $p = .346$, $\eta_p^2 = .031$, $\lambda = .940$). At the univariate level, the effects of LLAMA D on grammatical sensitivity had a p value of .084 ($F(1,104) = 3.035$, $p = .084$, $\eta_p^2 = .027$), but there was no interaction between group and covariate ($F(2,104) = 1.401$, $p = .251$, $\eta_p^2 = .025$).

When learners were combined as a single group and compared with controls, results remained non-significant at the multivariate level ($p > .05$), but the interaction between group and grammatical sensitivity at the univariate level approached significance ($F(1,106) = 3.269$, $p = .073$, $\eta_p^2 = .029$). Simple correlations showed that the relationship between LLAMA D and sensitivity to agreement structures was significant in the early AO group ($r = .34$, $p = .021$) and approached significance in the late AO group ($r = .26$, $p = .079$), but was not significant in the control group ($r = -.07$, $p = .757$) (disattenuated correlations were .44, .34, and -.09). The effects of the SRT learning score as a covariate on sensitivity to agreement structures with three groups of participants as a fixed factor were not significant. The SRT score was not a significant covariate ($F(3,102) = .342$, $p = .795$, $\eta_p^2 = .009$, $\lambda = .991$) and it did not interact with group ($F(6,204) = 1.217$, $p = .307$, $\eta_p^2 = .031$, $\lambda = .969$) at the multivariate level. The results at the univariate level were also non-significant ($p > .05$). Combining L2 learners into a single group did not make any difference, and the SRT score remained non-significant covariate ($p > .05$). When simple correlations were computed, they were weak and non-significant in the control and early AO

groups ($r = .14$, $p = .553$ and $r = .16$, $p = .263$, respectively) (disattenuated correlations were .21 and .24). In the late AO group, the correlation had a slightly higher magnitude but remained non-significant ($r = .21$, $p = .141$) (the disattenuated correlation was .32). These results suggested that the significant relationship found between aptitude for implicit learning and early and late L2 learners' sensitivity to agreement structures was mostly due to the LLAMA D subtest in the early group and to a combination of LLAMA D and SRT scores in the late group.

Finally, the results for non-agreement items were all non-significant. Aptitude for implicit learning was not a significant covariate at any level ($p > .05$), and it did not interact with group either ($p > .05$). This indicated that individual differences in aptitude for implicit learning did not moderate L2 learners' grammatical sensitivity to non-agreement structures (aspect contrasts, the subjunctive, and the passive).

Given the significant relationship between aptitude for implicit learning and sensitivity to agreement violations in the late AO group, late L2 learners' performance on agreement items was examined across all the L2 measures in order to determine the extent to which late L2 learners displayed knowledge of grammatical agreement. Table 31 presents a breakdown of the average percentage scores obtained by the late L2 learners ($n = 50$) on agreement items.

Table 31. *Average Percentage Scores on Agreement Items in the Late AO Group*

	Agreement Structures
	M
Untimed Visual GJT	63.20 (11.39)
Untimed Auditory GJT	62.40 (10.41)
Metalinguistic Knowledge Test	67.47 (13.78)
Timed Visual GJT	57.62 (10.81)
Timed Auditory GJT	57.78 (10.08)

Note. Standard deviations appear between parentheses.

As can be seen, late L2 learners' scores on agreement items were higher on measures hypothesized to allow controlled use of L2 knowledge than on measures hypothesized to require automatic use of knowledge. The average score on the metalinguistic test, a measure hypothesized to allow controlled use of L2 knowledge, was significantly higher than the scores on all the other tests, according to Bonferroni-adjusted comparisons: untimed auditory GJT ($p = .018$), untimed visual GJT ($p = .044$), timed visual GJT ($p < .001$) and timed auditory GJT ($p < .001$). The average score on the timed visual GJT, a measure hypothesized to require automatic use of L2 knowledge, did not differ from the average score on the timed auditory GJT ($p = 1.000$), or from the scores on two of the measures hypothesized to allow controlled use of knowledge, the untimed visual GJT ($p = .067$) and the untimed auditory GJT ($p = .084$).

To summarize, aptitude for implicit learning moderated early and late L2 learners' scores on agreement structures. This effect was significant at the multivariate level for a combination of agreement scores on the three measures hypothesized to require automatic use of L2 knowledge (timed visual GJT, timed auditory GJT, and word monitoring task), and, at the univariate level, for the word monitoring task, at the automatic end of the L2 knowledge use continuum. In this task, individual differences in implicit aptitude were related to early and late L2 learners' degree of grammatical sensitivity to agreement violations.

5.3.4 Summary of Results: Cognitive Aptitudes and Language Attainment

The main findings regarding the relationship between cognitive aptitudes and language attainment among early L2 learners, late L2 learners, and NSs were the following:

- No significant interactions between L2-learner group and language aptitude in any of the analyses when overall test scores or scores on ungrammatical items were considered, suggesting comparable effects of cognitive aptitudes on language attainment among early and late L2 learners
    - Follow-up tests on agreement structures (gender, person, and number), however, yielded a significant interaction between group and aptitude for implicit learning: early L2 learners' scores on agreement items were moderated by aptitude for implicit learning in measures that allow controlled use of L2 knowledge and approached significance in measures that require automatic use of L2 knowledge
- Significant *multivariate* effects of aptitude for explicit learning as a covariate

191

on measures that allow controlled use of L2 knowledge, but not on measures that require automatic use of L2 knowledge

- o In the early AO group, the effect of aptitude for explicit learning was due to a combination of the three LLAMA aptitude subtests (B, F, and E), whereas in the late AO group it was due to a combination of the three LLAMA aptitude subtests plus general intelligence
- Significant *univariate* effects of aptitude for explicit learning as a covariate on measures that allow controlled use of L2 knowledge in both the early AO and late AO groups
  - o L2 learners with high aptitude for explicit learning outperformed L2 learners with low aptitude for explicit learning on 1) the metalinguistic knowledge test (in the early AO and late AO groups, but not in the control group), 2) the untimed visual GJT (in the early AO group and approaching significance in the late AO group, but not in the control group), and 3) the untimed auditory GJT (in the early AO group, but not in the late AO group or control group –though a trend was found in the late AO group for the LLAMA E subtest in the aptitude composite)
  - o For ungrammatical items (half of the items on each language test), results remained robust for the metalinguistic test: high- and low-aptitude early L2 learners' scores were significantly different, and high- and low-aptitude late L2 learners' scores approached significance
  - o Early and late L2 learners' performance on agreement and non-

agreement structures was equally moderated by aptitude for explicit learning on untimed visual tests (untimed visual GJT and metalinguistic test), but not on the untimed auditory GJT, which only showed a significant relationship with aptitude for explicit learning in the early AO group (this difference in modality between the two L2-learner groups did not yield a significant interaction)

- Significant *univariate* effects of aptitude for explicit learning as a covariate on measures that require automatic use of L2 knowledge in the early AO group only

  o Early L2 learners with high aptitude for explicit learning outperformed their low-aptitude counterparts on the timed auditory GJT and this difference remained significant for ungrammatical items

  o Early L2 learners with high aptitude for explicit learning also outperformed their low-aptitude counterparts on agreement items on the timed visual GJT

  o No significant effects of aptitude for explicit learning on the word monitoring task (i.e., grammatical sensitivity) in any of the groups

- Significant *multivariate* effects of aptitude for implicit learning on measures that require automatic use of L2 knowledge and significant *univariate* effects as a covariate on agreement items in the word monitoring task

  o L2 learners with high aptitude for implicit learning showed significantly greater sensitivity towards agreement violations than L2 learners with low aptitude for implicit learning, in both the early and

193

late AO groups, but not in the control group

- o In the early AO group, the effect of aptitude for implicit learning was mostly due to the LLAMA D aptitude subtest, whereas in the late AO group it was due to a combination of LLAMA D plus learning in the probabilistic SRT task

- No significant *multivariate* effects of general intelligence on either set of language measures, but significant interaction between L2-learner group and intelligence in the untimed visual GJT and metalinguistic test

- o 1) High- and low-intelligence late L2 learners were significantly different on overall metalinguistic test scores and approached significance on the untimed visual GJT, 2) Differences between high- and low-intelligence late L2 learners on the metalinguistic test remained significant for ungrammatical items, 3) Late L2 learners' performance on agreement and non-agreement structures was equally moderated by general intelligence on untimed visual tests (untimed visual GJT and metalinguistic test), 4) The only test where intelligence did not show an effect in the late AO group was the untimed auditory GJT

As a visual summary of the results, Table 32 shows the relationships observed in the data between aptitude for explicit learning, aptitude for implicit learning, and general intelligence on language attainment as measured by tests hypothesized to allow controlled use of knowledge or require automatic use of knowledge. A check indicates that at least one of the analyses conducted on either overall test scores,

ungrammatical items, agreement structures, or non-agreement structures was significant for the speaker group in question (early L2 learners and late L2 learners).

As can be observed, the two types of aptitude identified in the study played a role in early L2 learners' attainment, regardless of type of outcome measure. However, general intelligence did not play any role. The role of aptitude in late L2 learners' attainment was more specific and could only be observed in certain outcome measures. Like early L2 learners, late L2 learners' attainment on measures of controlled use of knowledge was moderated by aptitude for explicit learning. Like early L2 learners as well, late L2 learners' attainment on measures of automatic use of knowledge was moderated by aptitude for implicit learning. Unlike early L2 learners, however, late L2 learners' attainment on measures of controlled use of knowledge was also moderated by general intelligence. Therefore, general intelligence and aptitude for explicit learning had a similar effect among late L2 learners. No effects of cognitive aptitudes were observed in the NS control group.

Table 32. *Summary of Relationships between Types of Aptitude, General Intelligence, and L2 Attainment*

L2 Attainment	Explicit Aptitude		Implicit Aptitude		General Intelligence	
	Early	Late	Early	Late	Early	Late
Controlled Use of L2 Knowledge	✓	✓	✓	✗	✗	✓
Automatic Use of L2 Knowledge	✓	✗	✓	✓	✗	✗

Chapter 6: Discussion and Conclusions

This study set out to investigate the relationship between different cognitive aptitudes for L2 learning, including general intelligence, and ultimate level of language attainment by early (AOs 3-6) and late (AOs ≥ 16) L2 learners. Early bilinguals who start acquiring the L2 in an immersion context before age 6 were hypothesized not to be fundamentally different from NSs in terms of learning mechanisms (although they may still differ in ultimate success), whereas late bilinguals who start acquiring the L2 as adults (after age 16) should be fundamentally different from NSs in terms of learning mechanisms (and also different in ultimate success). Following DeKeyser's (2000) claim that any relationships between individual differences in language aptitude and learning outcomes constitute potential evidence for differences in learning processes, the present study examined whether individual differences in cognitive aptitudes hypothesized to play a role in either implicit or explicit learning relate to variation in L2 attainment in early and late L2 learners, as measured by tasks that allow controlled use of knowledge or that require more automatic use of knowledge.

A total of 120 participants took part in the study, 50 early L2 learners, 50 late L2 learners, all of them L1 Chinese-L2 Spanish bilinguals, and 20 NS controls. A set of six L2 attainment measures reflecting a continuum from automatic to controlled use of language knowledge was administered: four GJTs (timed visual, timed auditory, untimed visual, and untimed auditory), a metalinguistic knowledge test (at the controlled end of the L2 knowledge use continuum), and a word monitoring task (at the automatic end of the L2 knowledge use continuum). A battery of six cognitive

tests was also administered: four language aptitude subtests, a general intelligence test, and a probabilistic serial reaction time task.

6.1 Cognitive Aptitudes

Regarding cognitive aptitudes, this dissertation pointed out the heavy bias towards explicit cognitive processes that has characterized language aptitude constructs and measures in SLA. It also brought to attention the fact that SLA studies have neglected implicit cognitive processes as a source of potential aptitudes for language learning. One of the goals of this dissertation was to address this gap by including a cognitive task (a probabilistic SRT task) that could tap implicit learning. Implicit learning was further conceptualized to be an ability with meaningful individual differences that could be related to variation in L2 outcomes, in line with Kaufman et al. (2010) and Woltz (2003).

The results of an exploratory factor analysis, conducted using principal components analysis as the method of extraction and Varimax as the method of rotation, showed that the six cognitive tests administered (four language aptitude subtests, a general intelligence test, and a probabilistic SRT task) loaded on two different components. Four of the six tests (LLAMA B, E, F, and GAMA) loaded strongly on the first component, which accounted for the largest amount of variation and which was interpreted as "aptitude for explicit learning". On the other hand, the remaining two tests (LLAMA D and the probabilistic SRT task) loaded strongly on a second component that was interpreted as "aptitude for implicit learning". The interpretation of the underlying constructs was informed by the characteristics of the tests themselves, as well as by previous research findings.

The tests loading together on the "explicit aptitude" component had in common the fact that they involved explicit cognitive processes (i.e., attention-driven, conscious, and intentional processes). All involved working out relationships in either verbal or non-verbal datasets and allowed time to think and use problem-solving strategies. These skills can be broadly understood as analytic ability or explicit inductive learning ability, and they play a role in discovering patterns and rules (i.e., creating and testing hypotheses) on the basis of input data. This is one of the meanings that "inductive learning ability" has had in the SLA literature as a component of language aptitude. Moreover, this was one of the constructs that Carroll (1962) proposed as part of his four-factor model (phonetic coding, rote learning, grammatical sensitivity, and inductive learning), but that, nevertheless, was not represented in the MLAT battery, and that Skehan (1998) reconceptualized as language analytic ability together with grammatical sensitivity.

More recently, in a review of aptitude research, Skehan (2012) pointed out that the LLAMA aptitude test differed from the MLAT in that it "adds a receptive interpretation of inductive language ability" (p. 390). The results of this dissertation support Skehan's interpretation, but further qualify it by making a distinction between explicit and implicit inductive learning ability in the context of the LLAMA test battery and as broader aptitude components. Implicit inductive learning involves learning from input by analogy, not analysis (N. Ellis & Laporte, 1997). DeKeyser (1995) further hypothesized that implicit inductive learning is good for prototypicality (probabilistic) patterns (i.e., linguistic prototypes, such as number, case, or gender markings that are subject to allomorphy).

In the context of the LLAMA aptitude test battery, while LLAMA F (grammar inferencing) was the strongest loading on the component interpreted as "aptitude for explicit learning" and can be defined as a test measuring explicit inductive learning, LLAMA D (sound recognition) loaded on the component interpreted as "aptitude for implicit learning" and can be defined as a test measuring implicit inductive learning ability. LLAMA F requires test-takers to work out the grammar of an unknown language by means of pictures and short written sentences. LLAMA D, on the other hand, measures the ability to discriminate short stretches of spoken language by analogy.

As pointed out by Meara (2005), LLAMA D "owes something to Speciale (Speciale, N. Ellis, & Bywater, 2004)" who suggest "that a key skill in language ability is your ability to recognize patterns, particularly patterns in spoken language" (p. 8). Speciale et al.'s (2004) study included two cognitive factors as predictors of L2 vocabulary acquisition, a task of phonological sequence learning, measuring the ability to learn phonological regularities, and a nonword repetition task, measuring phonological short-term memory capacity. One of their findings was that phonological sequence learning ability constitutes a source of individual differences that can be dissociated from short-term store capacity. This line of research is based on a strand of cognitive psychology that investigates the implicit induction of phonological sequences (Saffran et al., 1996; Saffran, Johnson, Aslin, & Newport, 1999; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997). LLAMA D can, thus, be seen as an attempt to measure implicit induction learning ability.

Previous research has also shown a distinction between LLAMA B, E, and F, on the one hand, and LLAMA D, on the other, as well as the existence of aptitude profiles based on the LLAMA test (i.e., individuals with high scores on LLAMA F, but low scores on LLAMA D, and vice-versa, resulting in close-to-zero or weak correlations between both) (Granena, to appear). In addition, general intelligence, which, in this dissertation, loaded on the component interpreted as aptitude for explicit learning, has been consistently related to attention-driven working memory measures (e.g., Engle et al., 1999; Kyllonen, 1996; Kyllonen & Christal, 1990) and artificial grammar learning when participants are instructed to look for patterns in the training materials (e.g., Gebauer & Mackintosh, 2007; Reber et al., 1991; Robinson, 2002). However, it has exhibited low correlations with procedural skill performance beyond the early stages (e.g., Ackerman, 1987, 1988), indicating that, at least as represented by conventional tests, intelligence involves explicit cognitive processes similar to those that characterize some language aptitude components. Skehan (1998) further argued that the relationship between aptitude and intelligence was likely to be strongest for components of aptitude such as language analytic ability, but not for others such as phonetic coding ability.

Although language aptitude and intelligence overlap to some extent, as shown by either low to moderate or moderate to strong correlations (Gardner & Lambert, 1972; Sasaki, 1996; Skehan, 1982; Wesche, Edwards, & Wells, 1982), they still exhibit different correlations with L2 outcomes (e.g., Skehan, 1982). The results of this dissertation provided further support for the specificity of language aptitude. Intelligence loaded on the same component as three of the LLAMA subtests, and, in

combination with these subtests, it moderated L2 attainment as measured by tests that allow controlled use of knowledge in both early and late L2 learners. However, as an independent factor, it only showed a relationship among late L2 learners. This relationship only held when tests allowed controlled use of L2 knowledge.

Therefore, L2 attainment seems to be moderated by several factors (Carroll, 1983), specific factors when L2 learning starts at an early age, and both general and specific factors when L2 learning starts in adulthood. However, in this study, general factors did not moderate late L2 learners' attainment on tasks that required automatic use of L2 knowledge. In these tasks, only the type of language aptitude interpreted as being advantageous for implicit language learning, and which was unrelated to general intelligence, played a role. These findings suggest that there are abilities specific to language learning in post-critical period learning that do not overlap with general intellectual functioning.

6.2 Language Attainment

Regarding language attainment, this dissertation adopted a multiple-task design to measure participants' morphosyntactic attainment, following studies such as Abrahamsson and Hyltenstam (2009), where multiple tasks were used to sample different language domains. Notwithstanding the complexity of such designs, multiple assessment tasks are desirable in SLA research in order to provide a more comprehensive picture of learners' actual proficiency level (Chaudron, 2003). The multiple-task design used in this dissertation further aimed at addressing a gap in previous studies that have investigated language aptitude in single-task designs relying on L2 proficiency measures that have been biased towards explicit cognitive

processes (e.g., DeKeyser, 2000; DeKeyser et al., 2010; Abrahamsson & Hyltenstam, 2008).

A distinction was made between L2 measures that allow controlled use of L2 knowledge and measures that require automatic use of L2 knowledge. These measures were hypothesized to lie along a continuum of use of knowledge. The two control tasks at the extreme ends of the continuum were a metalinguistic knowledge test and a word monitoring task. In the metalinguistic test, participants' attention was directly focused on linguistic structure, correctness and grammatical rules (i.e., explicit declarative facts about language). It required language analysis rather than intuition about correctness. In the word monitoring task, participants' attention was focused on meaning. It required monitoring for a target word in a sentence and paying attention to sentence meaning, while the researcher measured sensitivity to grammatical violations. Four more tasks were administered that lay along the continuum: two timed GJTs (visual and auditory), hypothesized to require more automatic use of L2 knowledge, and two untimed GJTs (visual and auditory), hypothesized to allow controlled use of L2 knowledge.

The results showed that, as hypothesized, the metalinguistic knowledge test and the two untimed GJTs were strongly correlated with one another ($r > .80$). These results confirmed the findings of previous psychometric studies (Ellis, 2005 and Bowles, 2011), where metalinguistic knowledge tests loaded on the same factor as untimed GJTs. Contrary to Ellis' (2005) and Bowles' (2011) results, however, the correlations between language measures hypothesized to require automatic use of language knowledge were not strong. Specifically, the correlations between the GSI,

which was computed as an index of sensitivity to grammatical violations in the word

monitoring task, and the other two measures hypothesized to require automatic use of

language knowledge, the two timed GJTs, were only moderately weak ($r = .28$). In

fact, the two timed GJTs correlated more strongly with the measures hypothesized to

allow controlled use of language knowledge (magnitudes ranging between .66 and

.80). This could be due to the nature of the data, accuracy scores in the case of the

GJTs and the metalinguistic test, but reaction times in the case of the word

monitoring task. Alternatively, the word monitoring task could be measuring a

qualitatively different type of linguistic competence. This may be the type of

integrated language knowledge that the test has been claimed to measure (Kilborn &

Moss, 1996) and that several studies, mostly neurolinguistic studies investigating

language disorders, have provided evidence for (Karmiloff-Smith, Tyler, Voice,

Sims, Udwin, Howlin, & Davies, 1998; Kuperberg, McGuire, & David, 1998, 2000;

Marslen-Wilson & Tyler, 1980; Peelle, Cooke, Moore, Vesely, & Grossman, 2007).

GJTs, on the other hand, could be measuring controlled use of L2 knowledge to a

certain extent, regardless of the time pressure factor (Jiang, 2007).

Regarding early and late L2 learners' attainment, this dissertation found

significant differences between the two groups' overall scores in all the L2 measures

administered, in line with previous studies (Abrahamsson & Hyltenstam, 2009;

DeKeyser, 2000; DeKeyser et al., 2010; Granena & Long, 2010; Johnson & Newport,

1989). However, early L2 learners were also significantly different from NSs in all

the measures, except in the word monitoring task, where they showed the same

sensitivity to grammatical violations as NSs. When individual structures were

compared, early L2 learners did not differ from late L2 learners in two of the agreement structures, gender agreement and subject-verb agreement. These similarities were observed in the timed visual GJT, untimed visual GJT, and metalinguistic test. In addition, early and late L2 learners did not differ regarding their sensitivity to agreement structures in the word monitoring task, even though, in this case, early learners did not differ from NSs, either (only NSs and late learners did).

These results indicate that the acquisition of certain grammatical properties may be affected even when the L2 is acquired as early as age 3 or 4. These findings are partly similar to findings reported by Meisel (2009), who claimed that inflectional morphology is the domain in which child L2 acquisition can resemble adult L2 acquisition, and differ from L1 acquisition. He proposed a modified version of the Critical Period Hypothesis for certain domains of grammar. In this dissertation, an area that was especially affected was gender and subject-verb agreement, and the language pairing investigated Chinese-Spanish, two languages with very different inflectional paradigms (uniform vs. complex). Still, over half of the early L2 learners were able to score within NS-control range, several across the entire set of measures, whereas only a few late learners did, and none across the entire set, which suggests that native-like attainment remains possible for early L2 learners, but impossible for late L2 learners.

Meisel (2009) further hypothesized that language-specific learning mechanisms (processing and discovery mechanisms) may be also affected early and proposed applying the Fundamental Difference Hypothesis (Bley-Vroman, 1990) to child, as

well as adult L2 acquisition. According to Meisel, success in L2 acquisition depends on "a person's ability to inhibit the competing non-domain-specific cognitive resources" (p. 18), an explanation that seems compatible with the results reported in this dissertation, which are discussed in the next section, regarding the role of cognitive aptitudes not only in late L2 learners, but also in early L2 learners.

6.3 Cognitive Aptitudes and Language Attainment

This dissertation hypothesized that cognitive aptitudes that are more relevant for explicit language learning and processing would predict late, but not early, L2 learners' attainment on tasks that allow controlled use of L2 knowledge (Hypotheses 2a and 1a, respectively). These tasks increase available test time and decrease processing demands and, therefore, give L2 speakers an opportunity to rely on problem-solving and analytic skills. In these tasks, adult learners can rely on explicit L2 knowledge and compensate for their limited implicit competence. Adult learners with higher aptitude for explicit language learning were expected to do better as a result of their greater analytic, metalinguistic abilities.

Contrary to expectations, the results of the MANOVA analyses did not provide support for a differential role of aptitude for explicit learning in the two L2-learner groups. There was no evidence of a significant interaction between group and covariate and, therefore, the relationship between aptitude and attainment, as measured by tests that allow controlled use of knowledge, was comparable in the two groups of L2 learners. Moreover, the relationship between aptitude for explicit learning and tasks that allow controlled use of L2 knowledge was significant at the multivariate level for a linear combination of the three measures (i.e., untimed visual

GJT, untimed auditory GJT, and metalinguistic test), as well as, at the univariate level, for the untimed visual GJT and metalinguistic test in both the early and late AO groups, and for the untimed auditory GJT in the early group only. These results confirmed Hypothesis 2a and refuted Hypothesis 1a, since aptitude for explicit learning, unexpectedly, also played a role in the early AO group.

The fact that early L2 learners with high aptitude for explicit learning outperformed those with low aptitude on the untimed visual GJT, untimed auditory GJT, and metalinguistic test contradicts the findings in DeKeyser (2000) and DeKeyser et al. (2010), which showed a relationship between verbal analytic ability and scores on an untimed auditory GJT, like the one designed for this study, only in the late AO group. The relationship found between aptitude for explicit learning and morphosyntactic L2 attainment is, however, in line with the findings of Abrahamsson and Hyltenstam (2008), who concluded that aptitude played "not only a crucial role for adult learners but also a certain role for child learners" (p. 499).

Common to both studies (Abrahamsson & Hyltenstam, 2008, and the present study) was a larger *n* size of early L2 learners than in DeKeyser (2000) or DeKeyser et al. (2010). Sample sizes in Abrahamsson and Hyltenstam (2008) and in the current study were 31 and 50, respectively, whereas DeKeyser's (2000) early AO group included 15 participants and DeKeyser et al.'s (2010) included 20. As a result, the range of early L2 learners' test scores was more restricted. In fact, in DeKeyser (2000), all the early L2 learners scored above 90% on the GJT. In addition, the language aptitude test employed in DeKeyser (2000) and DeKeyser et al. (2010) was administered in the participants' L1. This could have further restricted the range of

scores, given that the test was measuring verbal analytic ability in the language that the early L2 learners might have felt less comfortable with and in which their literacy skills might have been the poorest. Therefore, the lack of a significant positive correlation between early L2 learners' language aptitude and GJT scores in DeKeyser (2000) and DeKeyser et al. (2010) could have been an artifact of the small variance (Long, 2007).

DeKeyser's (2000) explanation of the significant relationship between verbal analytic ability and morphosyntactic attainment only in the late, but not in the early, AO group was that adult learners relied on explicit learning mechanisms to compensate for increasingly inefficient implicit learning mechanisms. According to this explanation, the results in the present study would suggest that, not only late L2 learners, but also early L2 learners, rely on explicit, analytic, problem-solving capacities to reach higher levels of proficiency in morphosyntactic L2 attainment, as measured by certain tests. This claim would be in line with Paradis' (2009) position, according to which only children exposed to the L2 "before the age of 4 or 5 (and the younger the better) acquire the second language implicitly" (p. 110). According to him, the reason why some early L2 learners can still perform or be perceived as native-like is because of speeded-up controlled use of metalinguistic knowledge (i.e., conscious knowledge about form).

An alternative explanation could be that untimed L2 measures with a focus on language correctness allow L2 learners (both early and late) to control their performance consciously, inducing them to process language explicitly. As a result, untimed L2 measures would be partly measuring the same abilities as tests of aptitude

for explicit learning. This would explain the fact that the largest effect size observed in the data corresponded to the test at the most explicit end of the continuum from automatic to controlled use of L2 knowledge, the metalinguistic knowledge test. This test encouraged the highest degree of awareness and the greatest amount of attention to language forms by asking participants to correct grammatical errors and provide grammatical rules. Similar results were reported by Granena (2011), who found a positive effect for aptitude, measured by an average of the four LLAMA subtests, on an untimed visual GJT with a correction component in a group of 30 NSs of English, all of them adult L2 learners of Spanish and very advanced speakers.

The question remains whether early L2 learners who started learning the L2 as early as age 3, and who were hypothesized to have used the same (implicit only) language learning mechanisms as NSs, would rely on conscious knowledge about language form on untimed L2 measures that focus on language correctness. Like NSs, one would expect them, predominantly, to make use of feel judgments when responding to any language task, unless, as already argued, untimed L2 measures that focus on language correctness induce learners to approach the task analytically by placing a great deal of conscious (i.e., controlled) attention on sentence structure. L2 learners with higher analytic ability as measured by tests of aptitude for explicit learning could be more successful at detecting grammatical errors when a task requires focusing on language forms.

In Abrahamsson and Hyltenstam's (2008) study, early L2 learners' aptitude level was strongly related to scores on a GJT ($r = .70$, $p < .001$). This might be explained by the highly complex nature of the stimuli (i.e., very long, semantically complex

sentences) and/or by the fact that the GJT combined the results of an online and an offline version of the test. In other words, L2 learners with high aptitude in the domain of explicit, attention-driven memory processes could be more successful at processing and parsing sentence stimuli to identify grammatical errors.

If the relationship between aptitude for explicit learning and untimed L2 measures that focus on language correctness is due to the nature of the language test (i.e., test effects), rather than to reliance on explicit language knowledge, one would also expect a relationship between aptitude and performance on untimed L2 measures among NSs. However, this study did not find any significant relationships between NSs' language attainment and cognitive aptitudes on any of the attainment measures. One possibility is that the high inter-individual homogeneity that characterizes NSs' performance on language measures, in combination with the smaller sample size that typically characterizes NS control groups, precludes finding any significant results. In fact, Abrahamsson and Hyltenstam (2008) reported a correlation of .47 in their NS group, which did not reach significance ($p = .077$), probably due to the small size of the group ($n = 15$).[41] This would suggest that a relationship between NSs' cognitive aptitudes and their performance on certain types of GJTs cannot be discounted. The question is what the results from tasks that call for the use of analytic abilities and attention-driven memory processes can reveal about language competence in general, and to what extent they are similar to results from spontaneous language production tasks.

---

[41] In another study, Abrahamsson (p.c., 8/21/2011) did find a significant relationship between aptitude (as measured by the LLAMA aptitude test) and language performance on a GJT among native speakers, as well.

It was also predicted that cognitive aptitudes more relevant for explicit language learning and processing would not moderate either early or late L2 attainment on tasks that require automatic use of L2 knowledge (Hypotheses 4a and 5a, respectively), since these are online tasks that minimize the opportunities to plan responses or rely on problem-solving and analytic skills. The results confirmed Hypothesis 5a, but refuted Hypothesis 4a. As predicted, explicit language aptitude did not moderate late L2 learners' attainment on a timed visual GJT, a timed auditory GJT, and a word monitoring task. However, it did moderate early L2 learners' attainment on the two GJTs, the timed visual and the timed auditory. Therefore, explicit language aptitude moderated early L2 learners' performance on all the L2 measures administered, except for the word monitoring task, at the extreme end of the continuum of automatic use of L2 knowledge.

A feature that these measures have in common is the fact that they all focus test-takers' attention on language correctness and accuracy of grammaticality judgment. The word monitoring task, on the other hand, is carried out under a dual-task framework (e.g., Fodor, Ni, Crain, & Shankweiler, 1996; Furst & Hitch, 2000; Mullennix, Sawusch, & Garrison, 1992; Ransdell, Arecco, & Levy, 2001; Waters & Caplan, 1997; Wurm & Samuel, 1997) that focuses participants' attention on sentence meaning and word monitoring, while the researcher measures participants' sensitivity to linguistic violations (participants are never told about the presence of ungrammatical stimuli). One could argue that tests with a focus on language forms and language correctness call for, or may benefit from, test-takers' cognitive aptitudes for explicit language learning (i.e., analytic, metalinguistic abilities). According to

Jiang (2007), "a learner's performance in a GJT task (even a *timed* GJ task) can be a result of applying explicit knowledge rather than automatic competence" (p. 6, emphasis added). He further argued that psycholinguistic research paradigms, such as the one followed by the word monitoring task, are more likely to be informative about automatic activation of integrated linguistic knowledge, since participants react to grammatical errors without intending to do so.

If language tests that focus on language correctness call for analytic and/or metalinguistic abilities, the question remains as to why this study found a relationship between explicit language aptitude and attainment on the timed GJTs in the early AO group only, and not late AO group. The reason could be the time constraints imposed on the test and their effect on late L2 learners' performance. On timed language tests, test-takers are pressured to perform a task online under additional time constraints, in order to minimize controlled use of L2 knowledge. Performance typically declines when compared to untimed tasks (e.g., Bialystok, 1979; Bialystok & Miller, 1999; Murphy, 1997; Loewen, 2009), even among NSs.

In the present study, declines were significant regardless of test modality (visual or auditory), suggesting that time pressure creates a confounding factor at the level of processing above and beyond the possible confounding factor of phonological decoding typically associated with the aural presentation of stimuli. Time pressure made all participants' scores decline significantly, including NSs' scores. In the case of the late L2 learners, overall average raw scores (including missed items) on the timed visual and timed auditory GJTs were 26.68 (*SD* = 5.41) and 29.00 (*SD* = 5.25), out of a maximum of 60. Both raw averages were, therefore, below chance level. The

proportion of missed items in this group, with the corresponding loss of power it entails, was also considerable. The percentage scores taking into account total number of attempts were close to chance level, 57.88% and 57.63% for the timed visual and timed auditory GJTs, respectively.

Participants in the study by R. Ellis (2005) with 91 adult foreign language learners of mixed proficiency levels also scored close to chance level (54%) on the timed GJT, but well above chance on the untimed GJT (82%) and the oral narrative task (72%). These results indicate that the trade-off between processing demands and reliable use of (any type of) L2 knowledge may not be positive among adult L2 learners, when processing demands are considerable. Contrary to R. Ellis' (2005) suggestion to use timed testing formats, it seems that online tasks performed in real time, but not under time pressure, would make more reliable measures of automatic use of L2, since they would lie more comfortably within L2 learners' processing capacity (e.g., self-paced visual formats, auditory GJTs where sentences are played only once, and spontaneous language production tasks).

Regarding general intelligence, it was hypothesized that, in ultimate L2 attainment, relationships between explicit aptitude and general intelligence and learning outcomes would pattern in the same way and would be different from effects of implicit aptitude on outcomes. This hypothesis was based on studies of artificial grammar learning, in which fluid intelligence correlates with learning when participants are instructed to look for patterns in the training materials, but not under more incidental learning conditions. It was also based on studies in cognitive psychology that have shown psychometric intelligence to be more related to explicit

212

associative learning than to implicit learning. Hypotheses 1c and 4c predicted that intelligence would not moderate early L2 learners' language attainment, as measured by tasks that allow controlled use of knowledge (1c), or tasks that require automatic use of knowledge (4c). Hypotheses 2c and 5c posited that intelligence would moderate late L2 learners' attainment on tasks that allow controlled used of knowledge (2c), but not on tasks that require automatic use of knowledge (5c). All these hypotheses were supported by the findings. High-intelligence late L2 learners outperformed their low-intelligence counterparts on two measures of controlled L2 use (the metalinguistic test and the untimed visual GJT), but not on any other L2 outcome measures. Moreover, there were no effects of intelligence for early starters on any of the ultimate L2 attainment measures.

Follow-up analyses revealed that the intelligence factor did not contribute to the significant results reported for the composite of aptitudes for explicit learning in the early AO group (despite a significant correlation between general intelligence and LLAMA subtests B, E, and F, $r = .30$, $p = .035$). In the late AO group, on the other hand, general intelligence moderated L2 attainment on the same language measures that yielded a significant relationship with LLAMA B, E, and F (the metalinguistic test and the untimed visual GJT). Therefore, both general intelligence and language aptitude measures were relevant in late L2 learners' attainment, at least on tests that allow controlled use of L2 knowledge.

The main difference between the intelligence test used (the GAMA) and the LLAMA language aptitude subtests is the fact that the GAMA is a non-verbal (visual) test, whereas the LLAMA is a verbal (albeit language-independent) measure. This

may suggest that general learning mechanisms play a role in adult SLA (in combination with language-specific mechanisms), but no role in child SLA, in support of skill acquisition theory (Anderson 1983, 1993), and, as defended by DeKeyser (2001, 2003, 2007). However, the effect of general intelligence in the late AO group was only observed on tests that allow controlled use of L2 knowledge.

Similarly, studies comparing learning conditions have also found general intelligence to be more highly correlated with conditions where participants are explicitly instructed to look for underlying patterns than with incidental conditions in artificial grammar learning (e.g., Gebauer & Mackintosh, 2007; Reber et al., 1991; Robinson, 2002). Therefore, one cannot discount the possibility that the positive association between the two is due to the fact that they are measuring the same abilities. While being high- or low-intelligence did not make a difference for early L2 learners, late L2 learners needed the additional contribution of their general intellectual ability to perform on tasks that emphasize grammatical correctness and metalinguistic abilities. It seems that these tasks, then, would create a situation where late L2 learners may need to, and would be allowed to, resort to other cognitive resources, bringing into play all their verbal and non-verbal problem-solving capacities.

A factor that could have also contributed to the relationship between general intelligence and attainment on tasks of controlled use of L2 knowledge in the late AO group is formal instruction, since only 19 of the 50 late L2 learners had received either no instruction or instruction for a period less than one year (see Section 4.1). A comparison of late L2 learners with one year of instruction, or less, ($n = 19$) and late

L2 learners with more than two years of instruction ($n = 31$) revealed a significant correlation between general intelligence and metalinguistic knowledge test scores in the group with more than two years of instruction ($r = .39$, $p = .032$), but a close-to-zero negative correlation in the other group ($r = -.03$, $p = .900$). If late L2 learners who have received formal instruction potentially have more explicit language knowledge, these results would suggest a relationship between intelligence and stored, or use of stored, explicit language knowledge in adult learners[42].

Regarding cognitive aptitudes that are more relevant for implicit language learning and processing, it was predicted that these would moderate L2 learners' attainment on tasks that require more automatic use of L2 knowledge. This prediction was made for both early and late L2 learners (Hypotheses 4b and 5b), with the expectation that adult L2 learners can still learn implicitly, but not for NSs, whose ultimate attainment, characterized by inter-individual homogeneity, and mostly performance at ceiling, was considered independent of cognitive aptitudes. In addition, individual differences in aptitude for implicit language learning were predicted to be related to early, but not late, L2 learners' attainment on tasks that allow controlled use of L2 knowledge (Hypotheses 1b and 2b), since early L2 learners were expected to rely on the same type of knowledge, regardless of language task. Like NSs, this knowledge was hypothesized to be implicit. Unlike NSs, however, early L2 learners' ultimate attainment is characterized by greater inter-

---

[42] The subgroup of late L2 learners with more than two years of formal instruction obtained significantly higher scores on the metalinguistic knowledge test ($p = .001$) and results approached significance for the untimed visual GJT ($p = .064$). However, they were not significantly different from late L2 learners with one year of instruction, or less, on the rest of L2 measures ($p > .05$).

individual variability and was, therefore, expected to be moderated by cognitive aptitudes.

Results showed that, as predicted by Hypothesis 2b, implicit language aptitude did not moderate late L2 learners' attainment on tasks that allow controlled use of language knowledge. Contrary to expectations, implicit language aptitude did not moderate early L2 learners' attainment on such tasks either, at least when overall test scores or scores on ungrammatical items were considered. However, it moderated early L2 learners' attainment on agreement structures, thus, partially confirming Hypothesis 1b. Therefore, while only explicit language aptitude was a significant covariate for tasks that allow controlled use of L2 knowledge in the late AO group, both implicit and explicit language aptitude were significant covariates in the early AO group. The difference between the two was that aptitude for implicit learning only moderated early L2 learners' performance on agreement structures, but not on non-agreement ones, where only aptitude for explicit learning played a role.

An aptitude effect only occurring in early L2 learners could suggest a qualitative difference in the learning mechanisms of early and late L2 learners. However, because the effect belonged to a type of aptitude hypothesized to be relevant for implicit learning, and it was also present in late learners' scores on the word monitoring task, one could argue that it is indicative of early learners' advantage in implicit learning, or in the particular value of implicit learning for such features as [-interpretable] word-ending morphology. Early learners seem to have relied to some extent on this implicit knowledge, even when the L2 measure allowed controlled use

216

of language knowledge, whereas late learners largely relied on their analytic and/or metalinguistic abilities (aptitude for explicit learning and intelligence).

The relationship between aptitude for implicit learning and agreement structures in the early group was present to a greater or lesser extent in all the L2 measures. It was a significant relationship in measures that allow controlled use of L2 knowledge and on the word monitoring task, at the extreme end of the automatic use of knowledge continuum, and it showed a trend towards significance in the two timed GJTs. In the late AO group, on the other hand, aptitude for implicit learning only had an effect on agreement structures in the measure that drew late learners' attention away from language correctness, the word monitoring task. That was the task where late learners seem to have largely relied on implicit knowledge of grammatical agreement.

Hypotheses 4b and 5b, which predicted a relationship between implicit language aptitude and ultimate attainment on tasks that require automatic use of L2 knowledge, were partially confirmed. A significant relationship was found in the two groups of L2 learners, early and late, for the word monitoring task (at the extreme of the continuum of automatic use of L2 knowledge), but only for target structures involving grammatical agreement relations (gender, number, and person agreement). Both early and late L2 learners with high aptitude for implicit learning showed greater grammatical sensitivity towards agreement violations than L2 learners with low aptitude for implicit learning.

It is worth pointing out that the two aptitude composites patterned in the same way in the two groups of learners, as far as type of grammatical structure is

217

concerned. The aptitude composite hypothesized to be relevant for explicit learning moderated the two types of structures investigated (agreement and non-agreement). The aptitude composite hypothesized to be relevant for implicit learning, however, did not moderate participants' attainment on non-agreement structures, and it only played a role in structures involving grammatical agreement. This result may be relevant from the point of view of developmental patterns in acquisition and cognitive aptitudes, given the rationale behind the selection of structures in this dissertation, to which this discussion turns next.

The underlying rationale for the distinction between agreement and non-agreement structures was that L1 Spanish children acquire structures such as gender, number, and subject-verb agreement early (i.e., by age 3), whereas structures such as the subjunctive, the passive, and aspect contrasts are acquired later (i.e., at least age 7 or later) (Montrul, 2004). The late acquisition of the subjunctive, the passive, and aspect contrasts has to do with their linguistic complexity and children's cognitive developmental readiness. For example, in the case of the subjunctive (mood selection), children lack mental representations of "events that are independent or even incompatible with the reality of physical events" (Pérez-Leroux, 1998, p. 589). They are also structures at the syntax-semantics interface that make essential contributions to meaning and considered [+ interpretable] features (Tsimpli & Mastropavlou, 2007). However, their use is constrained to specific contexts. For example, past actions (aspect contrasts), topicalization (the passive), and negative commands (the subjunctive), among others. Finally, the passive and the subjunctive

are more frequently used in written language and formal registers and their acquisition is likely to be influenced by factors such as education and literacy level.

Agreement structures, on the other hand, are formal [- interpretable] non-salient features with a very high frequency of occurrence. Grammatical agreement is also characterized by the type of conditional (or transitional) probabilities that govern statistical learning, since it involves co-occurrence patterns within utterances and transitional probability, i.e., the probability of one event given the occurrence of another event (statistical regularity). For example, in the case of Spanish gender agreement, there are forward conditional probabilities of word-final phonemes *-a* and *–o*, given the feminine and masculine determiners *la* and *el* (.77 for word-final *-a* given *la*, and .56 for word-final *-o* given *el*) (Lindsey & Gerken, 2011).

Infants and young children are extremely sensitive and finely tuned to such distributional patterns in the input and learn them implicitly, as evidenced by the fact that Spanish children have acquired agreement structures with almost 100% accuracy by age 3. However, there is no consensus as to whether these learning mechanisms are still available to adults and, if so, under which circumstances they operate, and for what type of language features they can do so efficiently. The Fundamental Difference Hypothesis (Bley-Vroman, 1990) states that the implicit learning mechanisms that operate in child language learning are no longer efficient in adult language learning and that domain-general problem-solving mechanisms are used instead, a position supported by DeKeyser (2000), who also predicted that adults would need high verbal analytic ability to succeed in L2 learning. Meisel (2009) further claimed that the fundamental differences in learning mechanisms between

219

child and adult acquisition may already emerge in early childhood, earlier than the critical age range hypothesized by Bley-Vroman (1990) or DeKeyser (2000) (i.e., end of teens), and even Lenneberg (1967) (i.e., at puberty), and only for certain grammatical properties.

On the other hand, there is evidence from experimental settings that adults are sensitive to distributional patterns in non-linguistic input and that they can learn tone, noise, and visual sequences implicitly (Creel, Newport, & Aslin, 2004; Gebhart, Newport, & Aslin, 2009; Hunt & Aslin, 2010; Saffran et al., 1999). The results of the probabilistic serial reaction task used in the present study lend support to this body of findings, as well. They have led some researchers (Kaufman et al., 2010; Woltz, 2003) to conceptualize implicit learning as an *ability*, "the ability to automatically and implicitly detect complex and noisy regularities in the environment" (Kaufman et al., 2010: 321). This ability is characterized by automatic, associative, nonconscious, and unintentional learning processes. Contrary to Reber (1993), who views individual differences in implicit cognition as minimal, relative to individual differences in explicit cognition (due to the fact that implicit learning is evolutionarily older than explicit cognition), these researchers claim that implicit learning is a cognitive ability with meaningful individual differences. This implies that implicit learning can be significantly related to other cognitive abilities and/or language acquisition outcomes.

Adults are also sensitive to probabilities in linguistic input, as evidenced by the fact that they can compute how consistently sounds co-occur, and how frequently words occur online, and use this probabilistic information to acquire simple syntactic structure in miniature languages (Aslin & Newport, in press). The same learning

220

mechanisms could be at work in the acquisition of inflectional morphology (e.g., noun-adjective gender agreement or subject-verb agreement). This area of grammar is, in fact, a good candidate for implicit language learning, known to work through the slow accumulation of instances of input data (DeKeyser, 2003), especially in the type of immersion setting investigated, which is characterized by long-term exposure to large quantities of input.

If implicit learning mechanisms are affected in very early developmental phases, as suggested by Meisel (2009), and, as a result, become only partially available, an effect of language aptitude that could compensate for maturational changes should be observed not only in late, but also in early L2 learners, as found in this dissertation research. Some areas of grammar would be especially affected by this reduced capacity for implicit language learning, and it seems that inflectional morphology could be one of them, at least for language pairings with very different inflectional paradigms (e.g., Chinese-Spanish).

These are non-salient, [- interpretable] features, which are highly frequent in the input, but known to cause persistent difficulty in adult L2 acquisition. In the present study, even early L2 learners with AOs 3-6 performed significantly worse than NSs on gender and subject-verb agreement (whereas they did not differ from NSs on the subjunctive, which is a non-agreement structure). The fact that there may be a type of, apparently highly selective, cognitive aptitude that is advantageous for the acquisition of such non-salient features and that can compensate for partial loss of the implicit language learning capacity could explain why it is possible for some early and late L2 learners to attain higher levels of L2 competence than others (i.e., inter-individual

variation at a within-subjects level). Perhaps, this variation is a reflection of those L2 learners who were more able to keep relying on implicit learning mechanisms, despite other available mechanisms, in which case aptitude for implicit learning would mean the same as degree of implicit learning capacity.

The fact that aptitude for implicit learning predicted sensitivity towards grammatical agreement violations in both early and late L2 learners suggests some degree of similarity in early and late learners' language learning mechanisms. Following DeKeyser's (2000) hypothesis that relationships between individual differences in language aptitude and eventual learning outcomes potentially constitute evidence for differences in underlying learning processes, one could argue that those adults learning an L2 in a naturalistic (i.e., immersion) environment can also acquire certain features of the L2 implicitly, as indirectly indicated by the fact that those adults with higher implicit language aptitude showed greater sensitivity towards grammatical agreement violations. It should be noted that, despite any potential similarities between early and late L2 learners' learning mechanisms, success rate (i.e., the ability to perform in near-native like fashion) was still greater in the case of early L2 learners, suggesting that, if implicit language learning mechanisms remain partially available, they are less available to adult L2 learners or, alternatively, cognitive aptitudes cannot compensate for maturational effects equally effectively in adulthood as they can do in early childhood.

The potential role of implicit learning in eventual L2 outcomes by adult learners in an immersion setting is consistent with the findings of experimental studies that have focused on adult learners' implicit learning of semi-artificial grammars

(Rebuschat, 2008; Rebuschat & Williams, 2006, 2009; Williams, 1999, 2005). These studies typically show 65% accuracy in implicit learning groups, versus chance performance in control groups. A challenge that any study claiming implicit learning processes has to face, however, is the fact that evidence is based on learning outcomes (i.e., acquired knowledge) and such outcomes can be the result of implicit learning, explicit learning or a combination of both. Evidence of implicit learning can only be indirectly established by measuring the extent to which participants are aware of the acquired knowledge, in semi-artificial grammar learning studies, or by establishing a relationship between learning outcomes and cognitive aptitudes that are more relevant for either explicit or implicit learning, as suggested by DeKeyser (2000) and as investigated in the present study. Even the existence of verbalizable knowledge would not necessarily imply that learning did not happen implicitly, since implicitly acquired language knowledge (e.g., one's native language) can become verbalizable to a lesser or greater extent.

The last set of hypotheses in this study predicted no relationships between cognitive aptitudes and NSs' attainment on tasks that allow controlled use of language knowledge (Hypotheses 3a, 3b, and 3c) and tasks that require automatic use of language knowledge (Hypotheses 6a, 6b, and 6c). All these hypotheses were borne out by the data. Therefore, according to these results, NSs' linguistic competence can be considered independent of NSs' cognitive abilities. However, this study noted two factors that typically preclude finding such significant relationships in NS control groups: N size smaller than target groups, and inter-individual homogeneity, with performance usually close to ceiling. One cannot discount the possibility that larger N

sizes could show that cognitive aptitudes play a role in NSs' attainment on certain L2 measures, as some findings by Abrahamsson (p.c.) indicate. The type of language ability measured by the tests in question would have to be established. The prediction would be that no relationships would be observed in L2 measures that tap automatic use of language knowledge in tasks that do not carry any additional processing load.

The results of the present study can only speak to NSs' ultimate attainment. Cognitive aptitudes may still be a factor in rate of L1 acquisition, where inter-individual variation is probably greater than in ultimate attainment, as suggested by Skehan's (1990) findings in the Bristol Language Project (Wells, 1981, 1985). Skehan reported a number of significant correlations between language aptitude at age 13 and measures of acquisition derived from the children's speech when they were 42 months, and he argued that aptitude was a factor in the development of language competence in NSs. However, the significant relationships between aptitude and the biographical variables in the study make the role of environmental factors difficult to disentangle. Specifically, factors such as family background, parents' level of education, and parents' interest in literacy were significantly related to scores on aptitude measures, such as a verbal intelligence test and a grammatical sensitivity test, which, in turn, were related to linguistic indices, such as mean length of utterance and range of adjectives and determiners. Perhaps not surprisingly, only one of the aptitude measures, a sound discrimination test, was unrelated to biographical factors. This subcomponent of aptitude correlated with two of the comprehension indices in the study and with one of the vocabulary indices, suggesting a distinct dimension of aptitude in L1 acquisition.

6.4 Summary of Research Findings

As a summary of research findings, Table 33 displays the relationships that were predicted between aptitudes, general intelligence, and ultimate L2 attainment, as well as those relationships that were supported (✓) or unsupported (✗) by the data. Sixteen of the 18 expected relationships were either confirmed or partially confirmed. Partial confirmation should be understood as indicating that the predicted relationship was held at least in one of the analyses (either main or follow-up analyses). Thus, it could be a significant relationship for overall scores (grammatical and ungrammatical), for scores on ungrammatical items only, for scores on agreement structures, or on non-agreement structures.

Table 33. *Summary of the Study Predictions and Findings*

	Automatic L2 Use			Controlled L2 Use		
	Early AO	*Late AO*	*Control*	*Early AO*	*Late AO*	*Control*
General Intelligence	No✓	No✓	No✓	No✓	Yes✓	No✓
Explicit Aptitude	No ✗	No✓	No✓	No ✗	Yes✓	No✓
Implicit Aptitude	Yes✓	Yes✓	No✓	Yes✓	No✓	No✓

Note. A check mark (✓) stands for confirmed (or partially confirmed) and a cross mark (✗) stands for refuted.

6.5 Conclusions and Directions for Further Research

The current study adds to the current body of literature suggesting that different types of cognitive aptitudes have differential effects on long-term L2 outcomes. A broad distinction was made between explicit and implicit language aptitudes in an attempt to address the main limitation of conventional language aptitude measures, which have been heavily weighted in favor of explicit processes. Explicit language aptitude had an effect on L2 outcome measures that were untimed and that focused on language forms and language correctness. There was no evidence of any advantageous effects of this type of aptitude on language attainment, if the word monitoring task is taken as the most representative measure of implicit linguistic knowledge used in this study. The word monitoring task, however, is a psycholinguistic task that relies on reaction-time data, and this can be regarded as a limitation, since claims about integrated L2 knowledge are only indirectly established. Future research should investigate other L2 outcome measures to validate these findings, especially outcome measures that do not call for the use of the same analytic and/or metalinguistic abilities that also characterize explicit language aptitude measures.

Whereas explicit language aptitude had an effect on L2 outcome measures that were untimed and that focused on language forms and language correctness, implicit language aptitude had an effect on L2 learners' sensitivity to violations of grammatical agreement in the word monitoring task, which is online and has a meaning focus. The most relevant finding in this regard was the fact that implicit language aptitude moderated not only early L2 learners', but also adult learners',

226

sensitivity to gender, person, and number agreement. This finding is convergent with claims that implicit learning is crucial to language acquisition (e.g., N. Ellis, 1994) and with findings showing positive associations between measures of implicit learning and language acquisition (e.g., Gebauer & Mackintosh, 2012).

Further research should investigate other implicit language aptitudes, such as priming, in order to evaluate the possible effects of implicit induction (i.e., acquisition of patterns without awareness) on L2 outcomes. A limitation of the probabilistic SRT task used to measure implicit learning was its low reliability. Although the low reliability index was considered standard compared to previous studies in the literature (Dienes, 1992; Kaufman et al., 2010; Reber et al., 1991), it means that the assessment of implicit learning was less than optimal. Despite the noise in the data, there were significant relationships with L2 attainment. Previous studies have also shown significant correlations between implicit learning and complex cognition (Pretz et al., 2010). However, more reliable measures could show an even more prominent role of aptitude for implicit learning in acquisition.

Studies should also explore the extent to which aptitude for implicit learning is efficient and/or effective in instructed contexts that typically lack the massive input exposure that characterizes immersion settings, as well as the effects of aptitude for implicit learning on spontaneous language production tasks. As anecdotal evidence for this undertaking, five of the twelve adult learners classified as having high aptitude for implicit learning only (and either mid or low aptitude for explicit language learning) were also among the highest scorers on the oral interview used as an informal screening procedure for the study.

Finally, it would be very informative to investigate aptitude profiles in aptitude-treatment interaction studies. The following four aptitude profiles were observed in the present study: [high implicit, high explicit], [high implicit, low explicit], [low implicit, high explicit], and [low implicit, low explicit]. In the adult learner group ($n = 50$), 24% of the participants ($n = 12$) were high in implicit aptitude only. This percentage increased to 36% if adults high in implicit aptitude and high in explicit aptitude were considered ($n = 18$). On the other hand, only 14% were high in explicit aptitude only ($n = 7$) and 10% were low in both types of aptitude ($n = 5$). It would be interesting to investigate these different profiles in other populations of very advanced adult L2 learners in either immersion or instructed language contexts.

Appendix A

Biographical Questionnaire

CUESTIONARIO DE DATOS PERSONALES

1. Nombre y apellido: _____

2. Sexo: Hombre _____
   Mujer_____

3. Edad: _____

4. Correo electrónico: _____@_____

5. Teléfono de contacto: _____

6. Estudios realizados: _____

7. Profesión actual: _____

8. ¿Tienes alguna disminución o problema de tipo **visual** y/o **auditivo**? Por favor, especifica.

   _____

9. Lengua **dominante** (lengua en la que te sientes **más cómodo** hablando):
   _____

10. Otras lenguas por **orden de dominio** (de más a menos dominio):

+ dominio   **- dominio**
   _____ _____ _____ _____

11. ¿Es el español la lengua materna de tu padre o madre? Por favor, especifica.

   _____

12. ¿Sabrías decir qué lengua **se hablaba** en tu casa cuando **eras pequeño/a** (antes de que fueses a la guardería)? _____

   Si se hablaban varias lenguas, indica por favor un porcentaje:
   Lengua 1: _____ ____%
   Lengua 2: _____ ____%
   Lengua 3: _____ ____%

13. ¿Qué lengua **hablas actualmente** en tu casa? _____
    Si se hablan varias lenguas, indica por favor un porcentaje:

    Lengua 1: _____ ____%
    Lengua 2: _____ ____%
    Lengua 3: _____ ____%

14. ¿Puedes hacer un **ránking** de las lenguas que utilizas en un día normal de la
    que más utilizas a la que menos utilizas?
    Por favor, especifica un porcentaje aproximado de **uso diario**:

    Lengua 1: _____**%** aproximado de uso diario
    Lengua 2: _____**%** aproximado de uso diario
    Lengua 3: _____**%** aproximado de uso diario

15. ¿Cuál es tu nivel aproximado de Chino Mandarin (o lengua china hablada en
    tu casa)?

    ☐ Básico   ☐ Intermedio   ☐ Avanzado   ☐ Casi nativo   ☐ Nativo

    Sólo hablado   ☐
    Hablado y escrito   ☐

    Por favor, especifica en caso necesario:

    _____

16. Elige los **contextos** en los que normalmente **siempre** utilizas el **español**:

    Contextos Formales: En el trabajo   ☐
                                 En la universidad   ☐
                                 Para hacer gestiones   ☐
                                 Otros: _____

    Contextos Informales: En casa   ☐
                                 Con los amigos y conocidos ☐
                                 Con familiares   ☐
                                 En Internet   ☐
                                 Para ver la televisión   ☐
                                 Para escuchar la radio   ☐
                                 Para leer el periódico   ☐
                        Otros: _____

17. ¿Cuántas **horas por semana** utilizas el español para…?:

	1-2hrs	2-5hrs	6-10hrs	Más de 10hrs
Trabajar	_____	_____	_____	_____
Hablar con familiares	_____	_____	_____	_____
Hablar con amigos	_____	_____	_____	_____
Leer libros, periódicos	_____	_____	_____	_____
Ver TV, películas	_____	_____	_____	_____
Internet	_____	_____	_____	_____

18. ¿Hasta qué punto te identificas con la **cultura española**?
    Por favor, haz un círculo sobre el número correspondiente: **5** significa que **SÍ** te identificas totalmente (te sientes español) y **1** significa que **NO** te identificas con la cultura española en absoluto (no te sientes español):

+ Identificación   **-Identificación**

      5   4   3   2   1

19. ¿Convives o has convivido con hablantes nativos de español? ¿Durante cuánto tiempo?

    _____

20. ¿A qué edad llegaste a España por primera vez? (si has nacido en España escribe "nacido en España")

    _____

21. ¿Aprendiste español antes de llegar a España? (si has nacido en España, por favor ignora la pregunta? _____

22. ¿A qué edad comenzaste a aprender español?

    _____

23. ¿Dónde comenzaste a aprender el español?
    En un contexto de clase (curso de español) en país de origen ☐
    En un contexto de clase (curso de español) en España ☐
    De manera espontánea en España, hablando con los que me rodean ☐

24. ¿Cuántos años de cursos de idiomas de español has hecho?

    _____

25. ¿Has recibido educación en España (guardería, primaria, secundaria, estudios universitarios)? Por favor, especifica.

    _____

26. ¿Cuántos años llevas viviendo en España?

    _____

27. ¿Han sido años seguidos o has pasado temporadas en el extranjero?
    _____

28. ¿En qué poblaciones de España has vivido y cuánto tiempo en cada una de ellas?

    _____

29. ¿Has estado en algún otro país de habla española?
    País(es):   _____
    Tiempo de estancia: _____
    Edad de llegada: _____

30. ¿Crees que cuando hablas en español pareces nativo?

    Totalmente de acuerdo   ☐
    Bastante de acuerdo   ☐
    De vez en cuando pero no siempre ☐
    Totalmente en desacuerdo   ☐

31. ¿Estás satisfecho con tu pronunciación del español?

    Muy satisfecho   ☐
    Bastante satisfecho   ☐
    No muy satisfecho   ☐
    Totalmente insatisfecho ☐

32. ¿Es importante para ti pasar por hablante nativo de español?

    Totalmente de acuerdo   ☐
    Es importante pero no esencial para mi ☐
    No es muy importante   ☐
    No me importa   ☐

¡GRACIAS!

Appendix B

Item Pool

1. Noun-Adjective Gender Agreement

1. *La actriz que ganó el premio fue *aplaudido* calurosamente por el público
2. *Finalmente, la película no fue tan *aburrido* como pensábamos
3. *Los sistemas de iluminación en Europa son más *innovadoras* que en España
4. *Los terrenos que son demasiado *húmedas* tienen muchos inconvenientes
5. *Este libro resulta muy *apropiada* para lectores de todas las edades
6. *Esta criatura anda siempre *despitado* por culpa de sus compañeros
7. *Dicen que las fotos de mariposas son muy *complicados* de conseguir
8. *El reloj de la pared va *atrasada* siete minutos exactos
9. *En Méjico, la cerveza se ha de servir bien *frío* y con limón
10. *Dicen que las hijas de Miguel son muy *trabajadores* y serviciales
11. *Estoy de acuerdo que el piano de los abuelos es demasiado *antigua* para nosotros
12. *Según los expertos, la miel más saludable es *oscuro* de color y suave de textura
13. *La identidad del acusado permaneció *oculto* hasta el final del juicio
14. *Las manos de dedos largos son *delicados* y elegantes
15. *Mi compañera de piso está muy *nervioso* por los exámenes de mañana
16. *Mi madre se enfadó porque mi habitación estaba *sucio* y sin barrer
17. *La calle que lleva al centro estaba *abarrotado* de gente y coches por todos lados
18. *En este restaurante, el menú del día sale bastante cara pero vale la pena
19. *El cultivo del maíz es *apta* en cualquier superficie para la agricultura del planeta
20. *La torre de Pisa está cada vez más *inclinado* y con menos columnas
21. *Los edificios de la universidad están todos muy bien *equipadas* con tecnología punta
22. *El vuelo a Madrid fue muy *larga* pero agradable gracias a las atenciones de las azafatas
23. *Me gusta el suelo porque está bien *acabada* con materiales de alta calidad
24. *La llegada a Madrid fue mucho mas *agotador* de lo previsto por la organización
25. *Tus mensajes están *guardadas* en un archivo en el escritorio del ordenador
26. *La víctima del accidente *fue atendido* inmediatamente por los servicios de urgencias
27. *Las empleadas de esta empresa son más hábiles y *trabajadores* que en mi antigua empresa
28. *La habilidad y maestría del pintor son *asombrosos* e inigualables sin

duda alguna

29. *La guardería infantil está atendido eficazmente por su propietaria
30. Cada una de las empresas de nuestro sector está *dispuesta* a compartir información
31. Hay muchas personas que se levantan *cansadas* diariamente porque no pueden dormir bien
32. Las letras del alfabeto del castellano son *veintiuna* sin contar la *ll* y la *ñ*
33. La ropa sólo está medio *seca* porque hoy no ha hecho nada de sol
34. Cualquier chiste puede ser *aburrido* si no se cuenta con gracia y estilo
35. Marte es el planeta visible más *próximo* a la tierra
36. En un futuro, especies como el águila estarán más *protegidas* para evitar su extinción
37. La tensión entre los países implicados es demasiado *alta* para conseguir un acuerdo
38. Cualquier región del sur del Canadá es más *cálida* que Suecia
39. Cualquier volcán puede resultar *peligroso* cuando entra en erupción
40. Cualquier corriente de aire puede resultar *molesta* cuando se practica el esquí
41. En algunos países, los rostros de las mujeres quedan *ocultos* detrás de un velo
42. El crecimiento de la población española ha sido *contínuo* desde 1975
43. Las editoriales de libros antiguos se mantienen *ajenas* a la tecnología
44. Cualquier peaje de autopista debe ser *aprobado* unánimemente por el congreso
45. La superfície disponible para construir estará *regulada* este año por el gobierno
46. El paisaje del sur de España es mucho más *árido* que el del norte
47. La provisión de energía está *garantizada* mundialmente bajo cualquier ciscunstancia
48. En los últimos días el precio de la carne está más *caro* que el precio del pescado
49. La red de transporte público no es *satisfactorio* para los ciudadanos de Madrid
50. En este restaurante, la relación calidad-precio es de *las* mejores de la ciudad
51. La lección de piano de ayer no fue tan *buena* como otros días
52. Las cajas y bolsas que estén *vacías* servirán para la mudanza del viernes
53. Los armarios y sillas del abuelo están muy *nuevos* para tener tantos años
54. La moto esta recién *pintada* de azul metalizado con toques de color dorado
55. Los atardeceres en Grecia son muy *luminosos* y alegres
56. El mapa de pared parecía demasiado *pequeño* para nuestra habitación
57. Proporcionaremos toda la información que sea *necesaria* gradualmente y sin prisas
58. La soledad es más *llamativa* en los ancianos que viven solos
59. El motivo de la queja tiene que estar *relacionado* concretamente con el consumo de gas

60. Flores que sean así de *perfumadas* durante tanto tiempo no se encuentran fácilmente

1. Subject-Verb Number Agreement

   1. *Ayer por la noche dos ladrones le *intentó* robar el bolso a mi abuela
   2. *El estudiante pidió a los profesores que le *dejara* salir antes para ir al medico
   3. *Tu opinión y tu actitud *convenció* finalmente al director de la escuela
   4. *El fallo de esa empresa es que las decisiones las *toman* mucha gente
   5. *En la próxima reunión se *ampliará* con más detalles las causas de la crisis
   6. *Los actores se *dirigió* rápidamente al escenario para recoger su premio
   7. *El color de las flores *cambian* según la estación del año y el tiempo
   8. *El derecho de los trabajadores al descanso no lo *respeta* los empresarios
   9. *Los vendedores del mercado de mi pueblo *prefiere* recibir dinero en efectivo
   10. * Los jóvenes en los colegios de este país *sabe* muy poca geografía
   11. *Los árboles del parque *pierde* completamente sus hojas cuando llega el otoño
   12. *El chico en mi clase de matemáticas *interrumpen* constantemente a la profesora
   13. *Los bares cerca del campus *sirve* cervezas mejicanas y europeas
   14. *Los tíos de mi amiga *insistió* en pagarle el alquiler este mes
   15. *Los jugadores de fútbol americano *juega* los domingos y los lunes
   16. *A Manuel se le *cayó* todas las tarjetas de crédito al suelo
   17. *Los padres de Ramón le *hizo* soplar las velas el día de su cumpleaños
   18. *Los policías *fue* a buscar la pelota de fútbol que cayó a la calle
   19. *Los guardaespaldas no *deja* pasar a nadie que no lleve zapatos de vestir
   20. *Este año los Reyes Magos le *trajo* carbón a Óscar por su mal comportamiento
   21. *Los padres de Isabel la *puso* a dormir a las ocho de la noche como cada día
   22. *A los estudiantes que desafinan los profesores les *da* clases extras de canto
   23. *A Emilio siempre se le *acaba* las palomitas antes de empezar la película
   24. *A los asistentes se les *cayó* las lágrimas al oír el discurso del rey
   25. *Los ciudadanos se *queja* de las largas listas de espera en los hospitales públicos
   26. *Él y tú *conoces* los inconvenientes de viajar en avión con mascotas
   27. *Comer y correr a la vez *tienen* consecuencias negativas para el organismo
   28. *Al final los problemas de Miguel se *resolvió* a través de la justicia
   29. *Los entrenadores le *dió* la enhorabuena al equipo campeón de la final
   30. *Chema y tú *bailas* siempre hasta el amanecer cada fin de semana
   31. *Las patatas junto con la cebolla y el ajo picados se *ha* de freír durante una hora a fuego lento

32. El oxígeno y el hidrógeno los *proporciona* el medio ambiente en cantidades iguales
33. La niña y tú *cobraréis* mil euros de indemnización por el accidente
34. Se *permite* la entrada de camiones en horas de poco tráfico
35. El día de la inaguración, *vinieron* el alcalde y el regidor para celebrarlo
36. Mis viejos amigos me *reconocieron* inmediatamente nada más salir por la puerta
37. Finalmente, se *unieron* a la expedición alpinistas alemanes con muy poca experiencia
38. En mi clase, bastantes alumnos ya *saben* inglés y alemán de negocios
39. Mis padres me *han* comprado unos pantalones vaqueros y una bufanda negra
40. Muchas organizaciones se *especializan* en ayudar a las víctimas del terrorismo
41. Los jubilados de la plaza *observaban* atentamente las obras de restauración del ayuntamiento
42. En muchos transportes públicos, se *admiten* niños menores de tres años de forma gratuita
43. A los chicos no les *gustaron* los dibujos animados que daban por la tele
44. Los pasajeros del transatlántico *desembarcaron* ayer en el puerto principal de Atenas
45. Mis colegas del trabajo se *creen* más inteligentes que yo
46. Los vecinos del cuarto *dejaron* de saludarse después de la disputa por las obras
47. Los manifestantes *han* vuelto a ocupar las calles para reclamar justicia
48. Paco se marchó pero los demás *prefirieron* quedarse hasta el final del concierto
49. Los regalos te los *traerá* el cartero mañana por la mañana sin falta
50. A Juan le *gastaron* la broma sus nuevos compañeros de oficina
51. Los alumnos de cuarto curso *escriben* redacciones sobre temas de actualidad
52. El pastel se lo *comieron* con gusto los invitados a la cena de gala
53. Los ríos de España se *desbordan* contínuamente por la excesiva cantidad de agua que reciben
54. Los autores de esta obra *merecen* gratitud y reconocimiento por parte de todos
55. Muchas personas se *ven* afectadas por la gripe cada año por no querer vacunarse
56. Este jersey me lo *regalaron* compañeros de la facultad por mi cumpleaños
57. El equipo de búsqueda se *dispersó* por toda la zona del incendio para buscar a los desaparecidos
58. A mi siempre me *llamaron* la atención esos niños tan espabilados
59. Por esta razón no *son* recomendable baños de sauna para perder peso
60. *En el periódico se *publicaron* todos los artículos escritos por Miguel Delibes

2. Noun-Adjective Number Agreement

1. *Mis amigos pidieron botas *prestada* a sus vecinos para ir a esquiar
2. *Hay *mucho* más libros antiguos en esta biblioteca que en el museo
3. *Los niños que son así de *espabilado* siempre obtienen muy buenos resultados
4. *El vino se mete luego en unos barriles *similar* a los que se utilizan para el ron
5. *Todos los métodos son *válido* para atravesar el río y llegar al otro lado
6. * Los problemas de suministro están muy *unido* a la falta de recursos
7. *Debemos aprender a ser *mejor* cada día para poder alcanzar nuestros objetivos
8. *Un viento y una lluvia nunca *visto* antes afectaron toda la zona del norte
9. *Mi hermano les esperó *entusiasmados* para darles la bienvenida
10. *Los gatos de Pablo son todavía demasiado *pequeño* para ir de viaje
11. *La huelga general mantiene *paralizado* a bastantes transportes públicos
12. *Se ha presentado un número de pruebas bastante *elevados* en contra del acusado
13. *Siempre se viste con trajes *oscuro* perfectamente cortados a su medida
14. *Hay *bastante* más latinos en países como Estados Unidos que en España
15. *África tiene *poco* recursos naturales y mucha población necesitada
16. *Todo el mundo sabe que hay gatos que son *mejor* cazadores que otros
17. *La tormenta ha dejado *incomunicado* a varios pueblos del sur de España
18. *Las gaitas son instrumentos de viento *parecido* a las flautas y a las trompetas
19. *Le gusta hacer las cosas de una manera *distintas* al resto de los mortales
20. *En este momento ignoramos *cuál* van a ser las consecuencias de una crisis nuclear
21. *Tenemos *suficiente* candidatos en Europa para garantizar la continuidad
22. *El centro de mesa estaba hecho con manzanas recién *cogida* del árbol
23. *Pablo siempre lleva los pelos del bigote *enredado* y sin cuidar
24. *Ayer conocí a tus *futura* esposa y suegra casualmente en la calle
25. *Cada año *ciento* de pájaros emigran hacia el sur en busca del calor
26. *Determinados usuarios se pasan de *listo* intentando conseguir servicios gratis
27. *Los cambios en materia de educación han sido *mínimo* este año por culpa de la crisis
28. *Los coches *procedente* de Europa tendrán asientos mas amplios y cómodos
29. *Los servicios de transporte serán más *económico* respecto al año pasado
30. *En la subasta se vendieron artículos por importes muy *superior* a los mil euros
31. Juan y María estaban muy *felices* bebiendo champagne y brindando por su relación
32. Los cuatro gatos pasaban mucho tiempo *juntos* jugando y cazando ratones
33. Todos los invitados iban *vestidos* para la ocasión con chaqueta y corbata

34. La diferencia entre su estilo musical y *cualquier* otro estilo es el ritmo y la melodía
35. Las autoridades *soviéticas* no avisaron a los países europeos del peligro
36. Todo el festival costó seis *millones* de euros más de lo previsto por la organización
37. Martín explica que a los tres *días* de matrimonio Ángela lo dejó por otro
38. Los estudiantes de hoy en día están llenos de *deudas* a largo plazo
39. Andalucía cobra los precios de alquiler más *bajos* de toda España
40. Los aficionados volvieron a sus hogares *decepcionados* tras la victoria del equipo contrario
41. Los productos que son *originarios* de la India siempre tienen más demanda
42. Cada vez estamos mas *influídos* políticamente por los medios de comunicación
43. Necesitamos una ley que regule la exportación de *determinados* artículos de consumo al extranjero
44. La normativa de la universidad fue redactada por los *anteriores* consejeros hace más de diez años
45. Por regla general, los climas del sur son más *suaves* que los del norte
46. Han sido liberados todos los periodistas, *incluídos* los dos de nuestra agencia
47. Me compré unos pantalones *vaqueros* muy bonitos y una bufanda negra
48. El vecino tiene una sobrina y un sobrino *cariñosos* que le quieren mucho
49. La mayoría de los animales domésticos son muy *lentos* cuando se ven en peligro
50. España tiene el índice de accidentes más *elevado* de toda Europa
51. Las relaciones *hispano-alemanas* se han deteriorado mucho últimamente
52. Cristian tenía siempre las mejillas *rosadas* porque lo alimentaban muy bien
53. Después de los tres *primeros* días, Ana demostró que estaba en plena forma física
54. El nombre de Roma nos trae a la mente *imágenes* de antiguas civilizaciones y ruinas
55. El consumo de cigarillos es de veinte *millones* anuales en países como Colombia
56. Mis amigos no son *capaces* de ocultarme la verdad sobre lo sucedido
57. *Cualesquiera* que sean las causas del siniestro, la compañía de seguros está obligada a pagar
58. Tengo muchos amigos que estarían *dispuestos* a colaborar en el proyecto
59. Los votos de los que dispone el candidato son *muchos* más de los que tiene la oposición
60. Los excursionistas que fueron a escalar llevan *desaparecidos* más de tres días

3. Subjunctive Mood

1. *Jorge se irá a trabajar en cuanto lo *avisan* de la oficina
2. *Mi profesor de instituto siempre nos pedía que *lleguemos* pronto a clase
3. *La escuela exigirá que los alumnos de primer curso *hablan* el inglés
4. *Tu madre te pide que te *portas* bien durante la cena de nochevieja
5. *En el pasado, era imprescindible para los agricultores que *llueva* durante el verano
6. *A los niños siempre les prohibimos que *salen* solos a la calle
7. *Los expertos sugieren que los ancianos *toman* calcio y vitaminas cada día
8. *Nos gustó que todo *salga* bien el día de la boda de Carmen y Roberto
9. *Después de mirar toda la tienda no había nada que le *guste* a Pilar
10. *Nos impresionó que Silvia *apruebe* todos los exámenes de primer año
11. *Serviremos los aperitivos cuando *vienen* los invitados
12. *Estoy muy contento de que Miguel *sigue* trabajando para la compañía de seguros
13. *La vecina del quinto siempre nos invita a que *entramos* para tomar café
14. *No hay que dejar que los niños *comen* caramelos todos los días
15. *El muro impide que los prisioneros *pueden* escapar de forma fácil
16. *Nos quedaremos en la oficina hasta que el informe *está* listo para ser enviado
17. *Marcos renunciará a su puesto de trabajo cuando *consigue* algo mejor en otra empresa
18. *No me pareció que Antonio *tiene* buena pronunciación de los idiomas que habla
19. *El conserje siempre nos prohibía que *fumamos* cigarrillos en los pasillos del edificio
20. *Marta está harta de que el jefe la *hace* trabajar días festivos y fines de semana
21. *Todos nos sorprendimos mucho de que no *estabas* presente en la fiesta de Carlos
22. *Ayer pudimos llegar a la estación antes de que *salga* el tren a Madrid
23. *No saldré de casa mientras no *tengo* noticias de Rubén y sus amigos
24. *Los organizadores insisten que los asistentes *vuelven* mañana para devolverles el dinero
25. *Su ambición es que su hijo se *convierte* en presidente del país
26. *El equipo celebrará la victoria cuando *gana* la final de esta noche
27. *Ana irá de vacaciones cuando *aprueba* todas las asignaturas pendientes
28. *El nuevo jugador organizará una fiesta cuando *firma* oficialmente su contrato con el equipo
29. *Los periodistas darán la noticia cuando lo *permite* el gobierno
30. *Juan no cree que *puede* llegar a tiempo de ver el principio de la película
31. Mario no sube a un avión ni aunque le *paguen* una fortuna
32. Nos fuimos para casa antes de que *empezase* a llover
33. Ojalá que los bebés *durmiesen* así de bien durante toda la noche
34. Todo el mundo piensa que es bueno que te *cases* finalmente con el hijo del

alcalde
35. Los pacientes querían que el doctor los *atiendese* por orden alfabético
36. Mis hermanas dudan que yo *recuerde* sus cumpleaños
37. No tengo ningún amigo que *vaya* de vacaciones a Toledo este verano
38. Está muy bien que le *den* el premio a Sara por su papel en la obra de teatro
39. Nos podemos quitar los zapatos cuando *estemos* más cerca del detector de metales
40. Es fácil que Sergio se *olvide* hoy de llamarme por teléfono
41. Puede ser que *tengamos* que escalar la roca para cruzar el río
42. Es poco probable que mis padres *encuentren* hoy un sitio para aparcar
43. Damián no conoce a nadie que *haya* nacido en un país nórdico
44. Tan pronto como te *ajustes* el cinturón pondré el coche en marcha
45. Ustedes no pueden salir hasta que alguien *pague* la cuenta
46. Con tal que no te *hagas* daño puedes jugar en el parque con tus amigos
47. Los políticos siempre hablan como si lo *supiesen* todo sobre economía
48. El jurado duda que el acusado *diga* toda la verdad sobre lo occurrido
49. La agencia publicitaria busca a una chica que *tenga* aptitud para las lenguas extranjeras
50. Es indignante que la electricidad *sea* más cara en España que en cualquier otro país europeo
51. Es increíble que tantos españoles *perdiesen* familiares en la Guerra Civil
52. Es mejor que no *pidamos* pollo a la brasa en el nuevo restaurante
53. Es muy extraño que a Gloria le *guste* gastar tanto en zapatos de vestir
54. Es probable que Ángela *viaje* pronto a Méjico para conocer a su familia
55. Buscamos un apartamento que *esté* orientado al Este para aprovechar más el sol
56. Marcelo no va a recibir más ayuda del banco a no ser que *pague* la deuda
57. Los chicos no van a ver la televisión hasta que no *acaben* sus deberes
58. Cuando *termines* de limpiar tu cuarto, iremos al mercado a comprar fruta
59. Es importante que los estudiantes de español *practiquen* el idioma todos los días
60. Espero que *puedan* encontrar trabajo mejor pagado en otro sitio

4. Perfective/Imperfective Aspect

1. *En un momento el técnico *solucionaba* los problemas de conexión a internet
2. *Todos coincidían en que el recién nacido *tuvo* un cierto parecido con su padre
3. *Aquella mañana Alfonso *compraba* el periódico como cada mañana antes de ir a trabajar
4. *Nada más empezar a leer la carta ayer me *daba* cuenta de la gravedad del asunto
5. *Muy pocos *daban* con la solución al enigma de la semana pasada en el periódico
6. *De repente, me *acordaba* del regalo de cumpleaños para Adrián

7. *Por un instante, todos *pensábamos* que los dos coches iban a chocar
8. *Justo ahora *hicimos* palomitas para ver la película en la tele
9. *Hacía varios meses que no *comí* marisco de Galicia de esta calidad
10. *Por un segundo Sonia *creía* ilusionada que había ganado la lotería de Navidad
11. *Juan *pedía* tres días de permiso al encargado para ir a visitar a su familia
12. *Mi padre *conocía* a tu padre aquel día en la fiesta de cumpleaños de Rebeca
13. *Durante mi infancia, *iba* dos años a una escuela privada de monjas en Valladolid
14. *Conozco a una mujer que *estaba* mucho tiempo en Argentina antes de volver a España
15. *En las últimas vacaciones de verano *podía* descansar más de lo habitual en mi
16. *Durante varios años *estudiaba* inglés a distancia para mejorar mi currículum
17. *En mi antiguo trabajo *salí* puntualmente de mi oficina en el centro de Madrid a las cinco de la tarde
18. *Mis abuelos no fueron felices hasta que *vivían* cerca de nuestra casa
19. *A Pedro le *dolió* la cabeza hasta que se tomaba una aspirina
20. *Rodrigo *llevó* ocho días intentándolo antes de abandonar la competición
21. *Aquella tarde María *bailaba* rumbas con sus amigos durante horas
22. *Durante el fin de semana Jaime *estaba* más de cinco horas estudiando para el examen
23. *Aquel día Carlos *tuvo* pensado jugar durante dos horas en el patio
24. *Apenas el presidente *acababa* el discurso, alguien le disparó desde una terraza
25. *Durante toda esa mañana, el doctor López *visitaba* decenas de pacientes con gripe
26. *Los invitados *jugaban* a cartas hasta que dieron las doce de la noche
27. *A medida que Juan *habló* de sus problemas, Maite se ponía más nerviosa
28. *Javier y yo nos *conocimos* de haber estudiado juntos en la universidad
29. *Por lo menos ayer el abuelo *estaba* tranquilo por un rato
30. *Cada día Óscar *pensó* en su novia Carla y su familia
31. Me *entusiasmé* al conocer la noticia sobre el embarazo de Irene
32. Hasta los veinte años *viví* siempre en Cádiz con mis padres y mis abuelos
33. Contra todo pronóstico la lluvia *cayó* toda la tarde sin parar
34. Por aquel entonces siempre *cantabas* al ducharte por las mañanas
35. Carla fue mucho a la playa hasta que *tuvo* problemas en la piel
36. De niña *fui* a Andalucía tres veranos para visitar a mis abuelos maternos
37. He traído bombones para Ester porque la última vez le *gustaron* mucho
38. Durante las vacaciones, cada mañana Victor *compraba* pan y leche para desayunar
39. En Julio del 2000 *pasamos* dos semanas en el Caribe sin niños ni familiares
40. Pilar *conocía* a Nacho y a su familia desde hacía más dos años

41. Durante años Ramón *estuvo* estudiando la carrera equivocada en el extranjero
42. Mi hermano *corrió* dos veces esta mañana para prepararse para la maratón del viernes
43. Los viernes por la noche Juan siempre *miraba* películas de detectives
44. Todo el invierno *hizo* mucho frío en la zona de Cataluña
45. Ayer el tren directo de Barcelona a Bilbao *llegó* tarde por culpa de la nieve
46. De camino al trabajo, se me *ocurrió* cómo solucionar el problema
47. Por suerte para nuestros invitados, el bebé se *durmió* anoche antes de lo esperado
48. Durante ese rato *acabé* de preparar la cena de bienvenida
49. Los dinosaurios de hace 150 millones de años *comían* cualquier tipo de planta
50. Este profesor es el que *enseñaba* matemáticas los jueves en mi instituto
51. A los cinco años, Silvia se *quedaba* dormida en todas partes
52. Santiago me dijo que *salía* dentro de poco de su casa
53. En la foto de la entrada Maite *tenía* quince años recién cumplidos
54. Carlos *cumplió* catorce años el mismo día que Sonia
55. La reunión no *pudo* acabar a las dos como estaba previsto y se alargó más de una hora
56. En la edad de piedra, los seres humanos *aprendieron* a utilizar la rueda
57. Javier se *rompió* el brazo de niño a causa de un golpe durante un partido
58. Muchos años después Miguel *tuvo* noticias de su antigua novia
59. La familia *llegó* a la iglesia una hora antes de lo previsto
60. Por suerte cada día el bebé se *dormía* cinco minutos más pronto

5. Passives with Ser/Estar

1. *El nuevo museo de arte *estuvo* inaugurado oficialmente esta semana
2. *Madrid *es* situado estratégicamente en el centro de España
3. *En el siglo XV las iglesias *estuvieron* destruídas completamente en toda Europa
4. *El nuevo empleado *es* sobradamante cualificado para llevar la contabilidad
5. *Las duras condiciones en la mina *están* bien sabidas por todos
6. *Últimamente María *es* muy encariñada con mi madre
7. *El río de la zona afectada por el terremoto *es* contaminado indefinidamente
8. *Hasta el día de su inauguración, el museo podrá *estar* visitado gratuitamente
9. *Cazar linces *es* terminantemente prohibido en países como España y Portugal
10. *La biblioteca *estará* restaurada gracias a las donaciones de los ciudadanos
11. *El acusado *estuvo* declarado inocente de todos los delitos cometidos

12. *El niño desaparecido *estuvo* encontrado caminando tranquilamente cerca de un río
13. *Cada día cientos de delfines *están* rescatados de entre las redes de los pescadores
14. *La piscina del hotel *estará* vaciada temporalmente por motivos de limpieza
15. *Shakira *está* conocida mundialmente por sus ritmos y canciones de amor
16. *Los dos hermanos *eran* muy unidos hasta que discutieron por la herencia
17. *Después de varias horas de espera el vuelo *estuvo* cancelado definitivamente hasta nuevo aviso
18. *El proyecto de investigación ha *estado* aprobado finalmente por el Ministerio
19. *El perro de Sandra *estuvo* visto por última vez en una zona de bosque
20. *El cuadro de las Meninas *estuvo* pintado magistralmente por Diego de Velázquez
21. *El concierto de rock *estuvo* aplazado al próximo 10 de Junio por culpa de la lluvia
22. *Los ingredientes del pastel nupcial *estuvieron* seleccionados cuidadosamente por los mejores chefs del mundo
23. *Los terroristas han *estado* capturados huyendo en un coche robado de la policía
24. *Los ladrones han *estado* sorprendidos intentando abrir la caja fuerte de una joyería
25. *La constitución española *estuvo* aprobada unánimente en el 1978
26. *Con motivo de la boda real, las tiendas han *estado* cerradas a las 4 de la tarde
27. *La catedral vieja *estuvo* construida en el siglo XIII por el rey Jaime I el Conquistador
28. *Actualmente no *está* legal llevar animales domésticos a bordo de los aviones
29. *Muchas de las obras que *estuvieron* escritas por Cervantes se destruyeron en el siglo XVI
30. *Antes de ganar el premio, la novela ganadora ya había *estado* leída por miles de españoles
31. El equipo del Valencia dejó de *ser* invencible en la penúltima jornada de liga
32. La oposición se quejó por los temas que *estuvieron* ausentes en el discurso del presidente
33. La celebración de Carnaval de este año no va a *ser* olvidada fácilmente por los madrileños
34. La nueva ley *será* aprobada por el gobierno a pesar de los muchos votos en contra
35. Al salir al campo de juego, el jugador *fue* recibido cariñosamente por toda la afición
36. Los paquetes *fueron* entregados ayer a media tarde por el conserje del edificio

37. La península ibérica *está* bañada por el Atlántico y el Mediterráneo
38. Tras la amenaza de bomba, los pacientes *fueron* trasladados inmediatamente a hospitales cercanos
39. Mi último libro ha *sido* número uno en ventas en varios países europeos
40. Estas cestas *están* hechas a mano a base de material reciclado
41. El niño *estuvo* castigado sin salir de casa durante todo el fin de semana
42. La cantante Rosario *es* muy querida en todos los países de sudamérica
43. El gol de Messi *fue* muy celebrado por la afición del estadio
44. Los periodistas con más experiencia *fueron* destinados a zonas de conflicto
45. La canción *está* dedicada a las víctimas de atentados terroristas
46. La ensalada de espárragos de hoy *está* aliñada con aceite, limón y sal
47. Los delincuentes *fueron* condenados a dos meses de prisión incondicional
48. Más de mil personas *estuvieron* afectadas por los cortes de luz durante la tormenta
49. Varias personas *fueron* heridas a causa del atropello en el centro
50. Las pruebas *fueron* destruidas antes de que llegara la policía
51. El actor *fue* nombrado embajador de buena voluntad de las Naciones Unidas
52. El concierto de música clásica de ayer *fue* suspendido por la lluvia
53. La última película de Almodóvar *fue* premiada como la mejor película del festival
54. Las normas de juego han de *ser* cumplidas por todos los jugadores
55. Los montañeros desaparecidos *fueron* rescatados ayer por la noche
56. El mundo *está* actualmente gobernado por las grandes corporaciones
57. Cientos de árboles del Amazonas *son* talados cada año
58. Cada día cientos de animales abandonados *son* adoptados por familias españolas
59. Cuando hay un accidente, los coches *son* habitualmente desviados por rutas alternativas
60. El jefe informó de que el trabajo que *estuviese* acabado para el viernes se pagaría doble

Bibliography

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

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102,* 3–27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

Anderson, J. R. (1983). *The Architecture of Cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). Problem solving and learning. *American Psychologist*, *48*, 35–44.

Aslin, R. N., & Newport, E. L. (in press). Statistical learning: From learning items to generalizing rules. *Current Directions in Psychological Science*.

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning, 29*, 81–103.

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

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

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 235–249.

Bley-Vroman, R. (1988). The fundamental character of foreign language learning. In W. Rutherford & M. Sharwood Smith (Eds.), *Grammar and second language teaching: A book of readings* (pp. 133–159). Rowley, MA: Newbury House.

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis*, *20*, 3–49.

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning*, *29*, 81–103.

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.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *33*, 247–271.

Brooks, P. J., Kempe, V., & Sionov, A. (2006). The role of learner and input variables in learning inflectional morphology. *Applied Psycholinguistics*, *27*, 185–209.

Bruhn de Garavito, C., & Valenzuela, E. (2008). Eventive and stative passives in

  Spanish L2 acquisition: A matter of aspect. *Bilingualism: Language and

  Cognition*, *11*, 323–336.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2010). The role of language

  aptitude in first language attrition: The case of pre-pubescent attriters. *Applied

  Linguistics*, *31*, 443–464.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training.

  In R. Glaser (Ed.), *Training, research and education* (pp. 87–136). Pittsburgh, PA:

  University of Pittsburgh Press.

Carroll, J. B. (1964). *Language and thought*. Englewood Cliffs, NJ: Prentice Hall.

Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic

  theory for foreign language teaching. *International Journal of Psycholinguistics*, *2*,

  5–14.

Carroll, J. B. (1981). Twenty-five years of research in foreign language aptitude. In

  K. Diller (Ed.), *Individual differences and universals in language learning

  aptitude* (pp. 83–118). Rowley, MA: Newbury House.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*.

  Cambridge: Cambridge University Press.

Carroll, J. B., & Sapon, S. (1959). *Modern Language Aptitude Test: Form A*. New

  York: Psychological Corporation.

Chaudron, C. (2003). Data collection in SLA research. In C. Doughty & M. Long

  (Eds.), *The handbook of second language acquisition* (pp. 762-828). Oxford:

  Blackwell.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Collentine, J. (1995). The development of complex syntax and mood-selection abilities by intermediate-level learners of Spanish. *Hispania*, *78*, 122–135.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 1119–1130.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

De Graaff, R. (1997). The Esperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*, 249–276.

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*, 379– 410.

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

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). New York: Cambridge University Press.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 313–348). Oxford: Blackwell.

DeKeyser, R. M. (2007). The future of practice. In R. M. DeKeyser (Ed.), *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology* (pp. 287–304). New York: Cambridge University Press.

DeKeyser, R. M., Alfi-Shabtay, I., & Ravid, D. (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* (Vol. 2, pp. 395–406). London: Routledge.

Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic Bulletin & Review*, *8*, 343-350.

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science*, *16*, 41–79.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Mahwah: Lawrence Erlbaum.

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). Oxford: Blackwell.

Doughty, C., Bunting, M., Campbell, S., Bowles, A., & Haarmann, H. (2007). *Development of the High-level Language Aptitude Battery*. Technical Report: Center for Advanced Study of Language. University of Maryland, College Park.

Ellis, N. C. (Ed.). (1994). *Implicit and explicit learning of languages*. NewYork, NY: Academic Press.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition*, *18*, 91–126.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Ellis, N. C., & Laporte, N. (1997). Contexts of acquisition: Effects of formal instruction and naturalistic exposure on second language acquisition. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 53-83). Mahwah, NJ: Lawrence Erlbaum.

Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). Oxford: Blackwell.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. A psychometric study. *Studies in Second Language Acquisition*, *27*, 141–172.

Ehrman, M. E., & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 79*, 67–89.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General 128,* 309–331.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*, 147–171.

Fodor, J. D., Ni, W., Crain, S., & Shankweiler, D. (1996). Tasks and timing in the

    perception of linguistic anomaly. *Journal of Psycholinguistic Research*, *25*, 25–

    57.

Furst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological

    components of working memory in mental arithmetic. *Memory and Cognition*, *28*,

    774–782.

Gardner, R., & Lambert, W. E. (1972). *Attitudes and motivation in second language*

    learning. Rowley: Newbury House Publishers.

Gebauer, G. F., & Mackintosh, N. J. (2007). Psychometric intelligence dissociates

    implicit and explicit learning. *Journal of Experimental Psychology: Learning,*

    Memory, and Cognition, 33, 34–54.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent

    and non-adjacent dependencies among non-linguistic sounds. *Psychonomic*

    Bulletin & Review, *16*, 486–490.

Granena, G. (To appear). Reexamining the robustness of language aptitude in SLA. In

    G. Granena & M. H. Long (Eds.). *Sensitive periods, language aptitude, and*

    ultimate L2 attainment. To be published by John Benjamins in 2013.

Granena, G. (To appear). *Cognitive aptitudes for L2 learning and the LLAMA*

    aptitude test: What aptitude does LLAMA measure? In G. Granena & M. H. Long

    (Eds.). *Sensitive periods, language aptitude, and ultimate L2 attainment*. To be

    published by John Benjamins in 2013.

Granena, G. (2011a). *Reexamining the robustness of aptitude in naturalistic SLA*. Paper presented at the American Association for Applied Linguistics, Chicago, IL.

Granena, G. (2011b). *Cognitive aptitudes for L2 learning and the LLAMA aptitude test: What aptitude does LLAMA measure?* Paper presented at the EUROSLA Annual Conference, Stockholm University, Sweden.

Granena, G., & Long, M. H. (2010, October). *Age of onset, length of residence, aptitude and ultimate attainment in two linguistic domains*. Paper presented at the Second Language Research Forum Annual Conference, University of Maryland, College Park, MD.

Granfeldt, J., Schlyter, S., & Kihlstedt, M. (2007). French as cL2, 2L1 and L1 in pre-school children. *Petites Études Romanes de Lund*, *24*, 5–42.

Greenfield, P .M. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve* (pp. 81–124). Washington, DC: American Psychological Association.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition*, *19*, 379–400.

Harley, B., & Hart, D. (2002). Age, aptitude, and second language learning on a bilingual exchange. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 302–330). Amsterdam: Benjamins.

Hemphill, J. F. (2003). Interpreting the magnitudes of correlation coefficients. *American Psychologist*, *58*, 78–79.

Houser, R. (2008). *Counseling and educational research: Evaluation and application*. Thousand Oaks, CA: Sage.

Hunt, R. H., & Aslin, R. N. (2010). Category induction via distributional analysis: Evidence from a serial reaction time task. *Journal of Memory and Language*, *62*, 98-112.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Oxford: Blackwell.

Hyltenstam, K., Bylund, E., Abrahamsson, N. & Park, H.-S. (2009). Dominant language replacement: The case of international adoptees. *Bilingualism: Language and Cognition*, *12*, 121–140.

Ioup, G., Boustagui, E., El Tigi, M., & Moselle, M. (1994). Reexamining the critical period hypothesis: A case study of successful adult SLA in a naturalistic environment. *Studies in Second Language Acquisition*, *16*, 73–98.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603-634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*, 1–33.

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*, 940–967.

Jiménez, L., & Vázquez, G. (2005). Sequence learning under dual-task conditions: Alternatives to a resource-based account. *Psychological Research*, *69*, 352–368.

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.

Johnston, M. (1995). Stages of acquisition of Spanish as a second language. *Australian Studies in Language Acquisition*, *4*, 1–28.

Karmiloff-Smith, A. (1979). Micro- and macro-developmental changes in language acquisition and other representation systems. *Cognitive Science*, *3*, 91–118.

Karmiloff-Smith, A., Tyler, L. K., Voice, K., Sims, K., Udwin, O., Howlin, P., & Davies, M. (1998). Linguistic dissociations in Williams syndrome: Evaluating receptive syntax in on-line and off-line tasks. *Neuropsychologia*, *36*, 343–351.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Kaufman, A. S., & Kaufman, N. L. (1990). *K-BIT (Kaufman Brief Intelligence Test) manual*. Circle Pines, MN: American Guidance Service.

Kempe, V., & Brooks, P. J. (2008). Second language learning of complex inflectional systems. *Language Learning, 58*, 703–746.

Kempe, V., Brooks, P. J., & Kharkhurin, A. V. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*, 127–153.

Keppel, G. & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook.* Upper Saddle River, NJ: Pearson Prentice Hall.

Kilborn, K., & Moss, H. (1996). Word monitoring. *Language and Cognitive Processes*, *11*, 689-694.

Köpke, B., & Schmid, M. S. (2004). *Language attrition: The next phase*. In M. S.

   Schmid, B. Köpke, M. Keijzer, & L. Weilemar (Eds.), *First language attrition.*

   Interdisciplinary perspectives on methodological issues (pp. 1–43). Amsterdam:

   John Benjamins.

Kuperberg, G. R., McGuire, P. K., & David, A. S. (1998). Reduced sensitivity to

   linguistic context in schizophrenic thought disorder: Evidence from online

   monitoring for words in linguistically anomalous sentences. *Journal of Abnormal*

   Psychology, *107*, 423–434.

Kuperberg, G. R., McGuire, P. K., & David, A. S. (2000). Sensitivity to linguistic

   anomalies in spoken sentences: A case study approach to understanding thought

   disorder in schizophrenia. *Psychological Medicine*, *30*, 345–357.

Kyllonen, P. C. (1996). Is working memory capacity Spearman's *g*? In I. Dennis & P.

   Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49–75).

   Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than)

   working-memory capacity? *Intelligence, 14,* 389–433.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Lindsey, B. A., & Gerken, L. (2011). The role of morphophonological regularity in

   young Spanish-speaking children's production of gendered noun phrases. *Journal*

   of Child Language, 1–24.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit

   and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, &

H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Bristol, UK: Multilingual Matters.

Long, M. H. (2005). Problems with supposed counter-evidence to the critical period hypothesis. *IRAL*, *43*, 287–317.

Long, M. H. (2007). *Problems in SLA*. Mahwah, NJ: Erlbaum.

Marslen-Wilson, W. D., & Tyler, L. K. 1980. The temporal structure of spoken language processing. *Cognition, 8*, 1–71.

Meara, P. (2005). *LLAMA language aptitude tests*. Swansea, UK: Lognostics.

Meara, P., Milton, J., & Lorenzo-Dus, N. (2003). *Swansea language aptitude tests (LAT) v.2.0*. Swansea, UK: Lognostics.

Meisel, J. M. (1990). Inflection: Subjects and subject-verb agreement. In J.M. Meisel (Ed.), *Two first languages: Early grammatical development in bilingual children* (pp. 237–298). Dordrecht: Foris.

Meisel, J. M. (2009). Second language acquisition in early childhood. *Zeitschrift für Sprachwissenschaft*, *28*, 5–34.

Meisel, J. M. (2011). Bilingual language acquisition and theories of diachronic change: Bilingualism as cause and effect of grammatical change. *Bilingualism: Language and Cognition*, *14*, 121–145.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*, 302-331.

Miyake, A., & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne Jr.

(Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Montrul, S. (2004). Subject and object expression in Spanish heritage speakers. *Bilingualism: Language and Cognition*, *7*, 125–142.

Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. Amsterdam: John Benjamins.

Mullennix, J. W., Sawusch, J. R., & Garrison, L. F. (1992). Automaticity and the detection of speech. *Memory and Cognition*, *20*, 40–50.

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, *13*, 34–65.

Naglieri, J. A., & Bardos, A. N. (1997). *GAMA Manual*. Minneapolis, MN: Pearson.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology, 19*, 1–32.

Novoa, L. K., Fein, D., & Obler, L. (1988). Talent in foreign languages: A case study. In L. K. Obler & D. Fein (Eds.), *The exceptional brain: Neuropsychology of talent and special abilities* (pp. 294–302). New York: Guilford Press.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill, Inc.

Oller, J., & Perkins, K. (1978). A further comment on language proficiency as a source of variance in certain affective measures. *Language Learning*, *28*, 417–423.

Oyama, S. (1978). The sensitive period and comprehension of speech. *Working Papers on Bilingualism*, *16*, 1–17.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.

Peelle, J. E., Cooke, A., Moore, P., Vesely, L., & Grossman, M. (2007). Syntactic and thematic components of sentence processing in progressive nonfluent aphasia and nonaphasic frontotemporal dementia. *Journal of Neurolinguistics*, *20*, 482–494.

Pérez-Leroux, A. T. (1998). The acquisition of mood selection in Spanish relative clauses. *Journal of Child Language, 25*, 585–604.

Perruchet, P., & Amorim, M. A. (1992). Conscious knowledge and changes in performance in sequence learning: Evidence against dissociation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 785–800.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*, 233–238.

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

Pretz, J. E., Totz, K. S., & Kaufman, S. B. (2010). The effects of mood, cognitive style, and cognitive ability on implicit learning. *Learning and Individual Differences*, *20*, 215–219.

Ransdell, S., Arecco, M. R., & Levy, C. M. (2001). Bilingual long-term working memory: The effects of working memory loads on writing quality and fluency. *Applied Psycholinguistics*, *22*, 113–128.

Raven, J. C. (1938). *Progressive Matrices*. London: H. K. Lewis & Co., Ltd

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General, 118,* 219–235.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York: Oxford University Press.

Reber, A. S., Walkenfeld, F. F., & Hernstadt, R. (1991). Implicit and explicit learning: Individual differences and IQ. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 888–896.

Reber, A. S., & Allen, R. (2000). Individual differences in implicit learning: Implications for the evolution of consciousness. In R. G. Kunzendorf & B. Wallace (Eds.), *Individual differences in conscious experience* (pp. 227–247). Amsterdam: Benjamins.

Rebuschat, P. (2008). Implicit learning of natural language syntax. Unpublished Ph.D. Dissertation. University of Cambridge, Cambridge, UK.

Rebuschat, P. & Williams, J. N. (2006). Dissociating implicit and explicit learning of natural language syntax. In Sun, R. & Miyake, N. (Eds.) *Proceedings of the Annual Meeting of the Cognitive Science Society*, p. 2594. Mahwah, N.J.: Lawrence Erlbaum.

Rebuschat, P. & Williams, J. (2009). Implicit learning of word order. In N.A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th Annual Conference of the Cognitive Science Society* (p. 1031). Austin, TX: Cognitive Science Society.

Reves, T. (1983). *What makes a good language learner?* Unpublished Ph.D. dissertation, Hebrew University of Jerusalem, Israel.

Reed, J., & Johnson, P. (1994). Assessing implicit learning with indirect tests: Determining what is learned about sequence structure. *Journal of experimental Psychology: Learning, Memory, and Cognition*, *20*, 585–594.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*, 27–67.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning, 47*, 45–99.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes, and learning conditions in SLA. *Second Language Research, 17*, 368–392.

Robinson, P. (2002). Individual differences in intelligence, aptitude and working memory during adult incidental second language learning: A replication and extension of Reber, Walkenfeld, and Hernstadt (1991). In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 211–266). Amsterdam: Benjamins.

Roehr, K., & Gánem, A. (2009). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness*, *18*, 165–181.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606–621.

Saffran, J.R., Newport, E.L., Aslin, R.N., Tunick, R.A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*, 101–105.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence: Quantitative and qualitative analyses*. New York: Peter Lang.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences and instructional design. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 319–353). Cambridge: Cambridge University Press.

Schmid, M. S. (2006). Second language attrition. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (Vol. 11, pp. 74–81). Oxford: Elsevier.

Shanks, D. R., & Perruchet P. (2002). Dissociation between priming and recognition in the expression of sequential knowledge. *Psychonomic Bulletin & Review, 9*, 362–367.

Shanks, D. R., Wilkinson, L., & Channon, S. (2003). Relationship between priming and recognition in deterministic and probabilistic sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 248–261.

Sheen, Y. (2007). The effect of focused written corrective feedback and language aptitude on ESL learners' acquisition of articles. *TESOL Quarterly*, *41*, 255–283.

Skehan, P. (1982). Memory and motivation in language aptitude testing. Unpublished Ph.D. dissertation. University of London.

Skehan, P. (1989). *Individual differences in second language learning*. London: Arnold.

Skehan, P. (1990). The relationship between native and foreign language learning ability: Educational and linguistic factors. In H. Dechert (Ed.), *Current trends in European second language acquisition research* (pp.83–106). Clevedon: Multilingual Matters.

Skehan, P. (1998). *A cognitive approach to learning language*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam: Benjamins.

Skehan, P. (2012). Language aptitude. In S. M. Gass & A. Mackey (Eds.). *The Routledge handbook of second language acquisition* (pp. 381-395). New York: Routledge.

Skinner, C., Johnson, J., Bardos, A. N., & Rhee, S. (1996). *Brief measures of cognitive ability and their relationships with achievement*. Paper presented at the annual conference of the Colorado Society of School Psychologists, Vail, CO.

Slobin, Dan (1985). *Cross linguistic Study of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Smith, K. L. (1980). *Common errors in the compositions of students of Spanish as a second language*. Unpublished doctoral dissertation. University of Texas at Austin.

Sorace, A. (1993). Incomplete vs. divergent representations of unaccusativity in non-native grammars of Italian. *Second Language Research, 9,* 22–47.

Sparks, R. (1995). Examining the linguistic coding differences hypothesis to explain individual differences in foreign language learning. *Annals of Dyslexia*, *45*, 187–214.

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

Sparks, R., Ganschow, L., & Patton, J. (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. *Journal of Educational Psychology*, *87*, 638–655.

Speciale, G., Ellis, N. C., & Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Applied Psycholinguistics*, *25*, 293–321.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York: Cambridge University Press.

Tagarelli, K. M., Borges-Mota, M., & Rebuschat, P. (forthcoming). The role of working memory in implicit and explicit language learning.

Terrell, T. D., Baycroft, B., & Perrone, C. (1987). The subjunctive in Spanish interlanguage: Accuracy and comprehensibility. In B. Van Patten, T. Dvorak, J. Lee (Eds.), *Foreign language learning. A research perspective* (pp. 19–32). New York: Newbury House Publishers.

Tsimpli, I. M., & Mastropavlou, M. (2007). Feature interpretability in L2 acquisition and SLI: Greek clitics and determiners. In H. Goodluck, J. Liceras, & H. Zobl (Eds.). *The role of formal features in second language acquisition*, (pp.143-183). London: Routledge.

Waters, G. S., & Caplan, D. (1997). Working memory and on-line sentence comprehension in patients with Alzheimer's disease. *Journal of Psycholinguistic Research*, *26*, 377–400.

Wells, C. G. (1985). *Language development in the pre-school years*. Cambridge: Cambridge University Press.

Wechsler, D. (1981). *WAIS-R (Wechsler Adult Intelligence Scale-Revised) manual*. San Antonio, TX: The Psychological Corporation.

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K.C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 119–154). Rowley, MA: Newbury House.

Wesche, M. B., Edwards, H., & Wells, W. (1982). Foreign language aptitude and intelligence. *Applied Psycholinguistics*, *3*, 127–140.

Williams, J. N. (1999). Memory, attention, and inductive learning. *Studies in Second Language Acquisition, 21*, 1–48.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition, 27*, 269–304.

Willingham, D. B., Salidis, J., & Gabrieli, J. D. E. (2002). Direct comparison of neural systems mediating conscious and unconscious skill learning. *Journal of Neurophysiology*, *88*, 1451–1460.

Woltz, D. J. (1990). Repetition of semantic comparisons: Temporary and persistent priming effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 392–403.

Woltz, D. J. (1999). Individual differences in priming: The roles of implicit facilitation from prior processing. In P. L. Ackerman, P. C. Kyllonen,& R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 135–156). Washington, DC: American Psychological Association.

Woltz, D. J. (2003). Implicit cognitive processes as aptitudes for learning. *Educational Psychologist*, *38*, 95–104.

Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language*, *36*, 165–187.

Yilmaz, Y. (2010, October). *Relative effects of explicit correction and recasts: The role of working memory capacity and language analytic ability*. Paper presented at the Second Language Research Forum Annual Conference, University of Maryland, College Park, MD.

Yukawa, E. (1997). *L1 Japanese attrition and regaining: Three case studies of two early bilingual children*. Unpublished Ph.D. dissertation, Centre for Research on Bilingualism, Stockholm University.