Language Aptitude in Young Learners: The Elementary Modern Language Aptitude Test in Spanish and Catalan

Maria del Mar Suárez Vilagran

LANGUAGE APTITUDE IN YOUNG LEARNERS: THE ELEMENTARY MODERN LANGUAGE APTITUDE TEST IN SPANISH AND CATALAN

Tesi doctoral presentada per

Maria del Mar Suárez Vilagran

com a requeriment per a l'obtenció del títol de

Doctora en Filologia Anglesa

Programa de Doctorat *Lingüística Aplicada*

Bienni 2002-2004

Departament de Filologia Anglesa i Alemanya

Universitat de Barcelona

Barcelona, 2010

Directora: **Dra. Carmen Muñoz Lahoz**

CHAPTER 5: DISCUSSION

5.1. Discussion on the use of the MLAT-E in Spanish and Catalan in young learners

The elementary version of the MLAT is addressed to children between 8 and 12 years old. These children are between the age periods commonly referred to as middle childhood and pre-adolescence, as middle childhood starts at the age of six and finishes at twelve, while teenagers are considered to enter the adolescence period. During this period, children experience many changes, not only at the cognitive level, but also in other aspects. Lately, though, the age limits of middle childhood and onset of puberty are advancing with the times due to sociocultural changes. For instance, a common synonym for middle-childhood is the expression “school years”, which are preceded by the “pre-school period”, meaning that children before the age of six do not attend school. This is not the case in Spain, as most children start school before the age of six and they also attend extracurricular classes of all kinds, such as music, sports and foreign languages. Puberty is also said to begin earlier and earlier these days, at least in Western countries. In addition, the 6-year period that the *Manual* of the Spanish elementary version of the MLAT covers (from 8 to 13) is quite a long time span. Therefore, the developmental milestones which a child goes through during this period could probably reflect on the partial and total results on the MLAT-ES and the MLAT-EC. This is the topic addressed in the first research question of this study: To what extent are the MLAT-E in Spanish and Catalan suitable language aptitude measures for learners in grades 3 to 7?

The results presented in section 4.2 are peculiar for 3- and 7-graders for several reasons, not only because of the evolution in the scores, which show a clearly unsteady evolution from grades 3 to 6, but also because of the relative stability that scores reach between grades 6 and 7. These irregularities could be explained from the cognitive development perspective, especially from Piaget’s concept of constructivism, as well as from the children’s L1(s) developmental stages and literacy experience.

The participants in this study were in grades from 3 to 7. Following Piaget’s developmental stages, those in grades from 3 to 5-6 are in the concrete-operational stage and those in grades 6 and 7 are entering the formal operational stage. According to Piaget, all of them are capable of logical thinking at a concrete level, reasoning and problem solving, and those in grades 6 and 7 possess, in addition, abstract reasoning,

scientific thinking and consolidated problem-solving skills. Piaget's developmental stages might help us explain the change in the scores by 3-graders and 4-graders for the coincidence with the beginning of the concrete operational stage of development. They can also help us explain the absence of significant changes in the scores of grades 6 and 7 because 7-graders, having entered the formal stage, already master the cognitive operations of the previous stage, which seem to be the ones needed in order to answer most parts of the MLAT-E successfully. The advances which the formal operational stage supposes may not, consequently, affect significantly the scores of children in grade 7 because formal-operational activity may not be significantly tapped by some of the parts of the MLAT-E. However, two points should be mentioned regarding these stages. First, some studies have determined the age at which a given strategy is first used and the processes of its maturation, yet even the most systematic studies fail to make explicit the development of these strategies and the cognitive processes which allow passing from one problem-solving strategy to another (Berman, 1987). Therefore, it is unknown how important cognitive development is in this process and at which precise point in a person's lifetime the beginning of these developmental stages is to be established. Second, Piaget's cognitive stages have been questioned due to the rather strict division into slots, as many changes happen slowly and gradually and there does not seem to exist any period of cognitive stability. Moreover, no synchrony has been found across domains and no child develops in the same way as others although the same steps in development are found in homogeneous groups of children (Feldman, 1980).

While Piaget's developmental stages can help us explain the differences in the results of grades 3 and 4 and the suspicious similarities in the scores of grades 6 and 7, these results could also be interpreted from other cognitive developmental approaches, for instance, from an information-processing perspective. Also, the interpretation of scores of certain subtests may be simplistic if we take for granted that all the items in each subtest measure the construct they are meant to measure in each grade. As each MLAT-ES/EC part taps different aspects of aptitude, these will be discussed separately.

Parte 1 Palabras ocultas and *Part 1 Paraules ocultes* are meant to measure the ability to match sounds with the symbols they represent. Children in grade 3 show more difficulty in solving part 1, in contrast with children in grade 4. This difficulty could be due to several reasons. First of all, Grunwell (1986) suggests that phonological development across school years is closely related to the exposure to written language, which facilitates the arousal of metaphonological skills, and of all the

participants, those in grade 3 are the ones who have the least experience. In addition, children in grade 3 are only starting to read serif typeface, which is the font in which the tests are. The format in which the test is edited is a factor to be taken into account when analysing the significant increase between grade 3 and grade 4. The tests are written in 12-point size typeface Times New Roman, while children this age are used to reading bound writing, which is much rounder, in a larger font size. From an oral survey to primary education teachers and contrasting the information obtained in several textbooks, the author of this dissertation could observe that textbooks in grade 1 contain a large quantity of texts written in capital letters only. Both in grade 1 and grade 2, when lower-case letters start to be used as a general rule, they tend to be round and of the sans-serif type (e.g. Comic Sans, Lucida, Century Gothic) and of a 14- or 16-point size. It is towards the end of grade 2 that printed letters start being more widely used in most textbooks. Since for children in grade 3 Times New Roman is a rather new type of font, they could have got confused with some letters, such as in the confusion of capital <i> (I) with lower-case <l>. These look practically the same in the eyes of a child despite the fact that Times New Roman is a serif typeface, that is, a typeface which adds a curly cue to letters. However, this little end mark is so subtle in letters such as capital <I> and lower case <l> that the test takers may not have spotted the difference. Size and typeface are, therefore, an issue to consider when administering the test to 3-graders. In fact, standardised guidelines for item construction recommend using 14-point size letters, preferably of the serif typeface (especially Palatino) for children up to grade 6 (Osterlind, 1989).

In grade 3, children begin to understand the relationships between the speech sounds and the graphemes by which they are represented. Catalan children, who in principle have been taught to read following Teberosky's (e.g. 2001; Teberosky & Tolchinsky, 1995) constructivist model, hardly ever treat graphemes independently but within a larger structure, the syllable, and are aware of its structure (rime, nucleus, coda). Consequently, when trying to decipher the words in part 1, 3-grade children may have had more difficulty than older children because the syllabic structure of the hidden words does not correspond to what, for them, is either a syllable *stricto sensu* or a word written acceptably. Moreover, these words appear decontextualised, which makes the phoneme-grapheme recognition more difficult. It is also between grades 3 and 4 that children get used to deciphering unknown words (Frith, 1985), and from grade 4 until grade 6 children get to automatise reading. These milestones in the process of learning to read coincide with the large increase in the results obtained by 4-graders as

compared to 3-graders and with the relative stability in the results present in the upper grades.

In *Parte 1 Palabras ocultas / Part 1 Paraules ocultes,* spelling is altered within the syllable by deleting most vowels and taking advantage of the name of consonants to close the syllable. This makes this part rather more difficult for 3-graders, who are in the final stage of consolidating their ability to read and write, than for the upper grades. What children are expected to do in part 1 is mastered if they have been taught to read and write following the synthetic method, which is no longer taught in the Spanish/Catalan school system. This method wants the child to follow a bottom-up approach in deciphering words. It consists in reading and understanding a word by analysing the graphemes, deciphering their correspondence with the sound they represent and finally interpreting the syntax of the context where the word is and its meaning (Perfetti, 1984).

In order to answer the items in part 1, two steps have to be taken. First, the hidden word has to be decoded and second, it has to be matched with one of the four words presented as possible synonyms or short definitions of the misspelled word. Perhaps another variable that distinguishes the performance of lower grades as compared to the upper ones is the challenge of dealing with two concurrent tasks (Cowan, 1997), which is an ability regulated by the central executive system that also improves with age (Baddeley, 1986; Engle & Oransky, 1999). Having to address their efforts to two simultaneous tasks may overuse children's mental capacity to perform both tasks properly (Case, 1985). This could be exemplified by some of the wrong answers by mainly 3-graders to items 1, 2, 4, 8 and 12 on *Parte 1* of the MLAT-ES (see section 3.5.2.1 and appendix E) and to items 1, 2, 4 and 8 on *Part 1* of the MLAT-EC (see section 3.5.2.3 and appendix J), as some children matched the hidden word not with its synonym or definition but with a word that rhymed or whose spelling was very similar to the stem word. Therefore, it seems that these children failed to ignore the irrelevant stimuli in the task. This inhibition power improves noticeably between 6 and 9 years of age (Lin, Hsiao & Chen, 1999), which would help explain the better performance by children in the rest of grades.

In part 1 some items seem to be problematic especially for the lower grades and, for this population, they may be measuring something which they were not designed to test. This is the case of the items whose meaning young subjects, mainly those in grade 3 but also some in grade 4, ignore, such as those which involve abstract or not so common words. These are, for example, *exitoso* ("successful") written *eccitoso* in item 12 on the MLAT-ES or *excepcional* ("exceptional") written *esepzional*

on the MLAT-ES and *eccpcional* in item 19 on the MLAT-EC, with two confusing options *emocionante / emocionant* ("thrilling") and *único / únic* ("unique"). Both items were difficult for grades 3 to 5. Actually, items 12 and 19 in the MLAT-ES were the only very difficult items in this test for 3-graders (IF=-0.02 and IF=0.02 respectively). Item 4 in both tests could possibly be one more example of an item containing words that grade-3 children still do not have a clear concept of. These words are *círculo / cercle* (circle), *ángulo* (angle) and *circunferencia/circumferència* ("circumference") in item 4 of both tests. Therefore, in this part, besides sound-symbol association, it seems as if some aspects of verbal intelligence were also measured that make 3-graders different from 4-graders. More specifically, the IF of item 4 on the MLAT-EC almost doubled from grade 3 (IF=0.47) to grade 4 (IF=0.89). This performance by 3-graders could be due to the fact that semantic development spurts only in the first primary school years (Anglin, 1970). It should also be taken into account that some school children's semantic interpretations of some word sets do not fully coincide with the adults' interpretations (Asch & Nerlove, 1960; in McLaughlin, 1978). In addition, specialised vocabulary may only be at the older children's disposal, as they read more informative texts, where this kind of vocabulary appears (e.g. Duke, Bennet-Armistead & Roberts, 2002), than younger children do (Bernhardt et al., 1995).

Although 3-graders are the test takers who got the lowest mean, in this grade there were also children who got very high scores, the maximum being 28 in both versions of the test. There were also test takers in grades 4, 5, and 6 (not so much in grade 7) who got extremely low scores, the minimum scores being, following the grade order, 4, 8, 8 on the MLAT-ES; and 5, 8 and 4 on the MLAT-EC. Performing badly in this part (or in others) may have been caused by several factors. In this part, the LCDH (e.g. Sparks & Ganschow, 1991) could partly explain why decoding the hidden words may have been difficult for even some children in the upper grades, as some could present some handicap in decoding written language in their L1 and, consequently, transfer this disability to this part (Ganschow et al., 1991; Sparks et al., 2006, 2009).

The results on the MLAT-ES and the MLAT-EC may also have been affected by the participants' environment, including the type of instruction received at school, especially formal language instruction, even if it has been claimed that scores on the MLAT-E should not be affected by previous language training (Carroll, 1984) and that aptitude is mainly innate (e.g. Skehan, 1986c, 1989, 1990). This may be the case of, above all, students in grades 6 and 7, who, when presented *Parte 2 Palabras que se corresponden / Part 2 Paraules que es corresponen*, not only recognised that what they should do was identify the function of a given word in the sentence but they also

labeled the functions by making comments aloud such as “we are supposed to match the subject, the verb, the direct object or the adjective in the sentences”. That is, the children themselves put the test directions in those words because they had been taught this nomenclature in their curricular language subjects. Therefore, the curricular instruction these participants had received was possibly influencing their performance on this part of the test. It must be pointed out that this is true of the context in which the test was administered, as the *Currículum d’Educació Primària* (2009) contemplates the introduction of formal grammatical instruction in the second and third cycle of primary education, but it may not be the case of other countries in which grammatical instruction may not be taught explicitly so early. Actually, it has been found that not until children receive explicit instruction of certain structures (such as relative clauses or conditionals) do they master them (e.g. Guasti & Cardinaletti, 2003, in Montrul, 2008; Menyuk & Brisk, 2005). Nevertheless, there are subjects who escape this reasoning. These are the three grade-3 outliers in group 2 who obtained scores higher than the mean of 7-graders, as they scored 22, 25 and 29, the latter being the maximum possible score. Actually, the girl who scored 29 excelled in all the parts of the test, obtaining a final score of 115 out of 122 points whereas the other two test takers (a boy and a girl) obtained 78 and 79 points in total, which are still scores above the mean in grade 3 and much closer to the mean of 4-graders. It remains unknown whether these outliers also excelled in some acquisitional aspect of their L1s, which would be an example of the close relationship between L1 ability and FL aptitude (Skehan 1986b, 1989). Also, on the basis of their part scores, the children who scored 22 and 25 could be classified as “analytic learners”, while the girl who scored 122 could be considered a “balanced learner” (Skehan, 1998) with outstanding language talent for her age.

It has already been mentioned that children in grades 6 and 7 were possibly aware of what they had to do in part 2 and could explicitly state it using metalinguistic terms. Therefore, their results on this part may not exclusively be evidence of their sensitivity to grammatical structure, which would be implicit, but also of their effectiveness in retrieving information (rather a formal kind of grammatical information) from their declarative memory in a conscious and controlled way. This would clearly have implications as for which type of knowledge and learning, whether implicit, explicit or both, this part is meant to measure, since implicit and explicit knowledge are sometimes considered two separate types of knowledge. Nevertheless, the learning processes to acquire these two types of knowledge are not considered to follow two

different paths, rather they are considered to interact at the level of language performance (R. Ellis, 2004).

Despite the explicit formal grammar labeling by 6- and 7-graders, the range of scores in these grades is very similar to those of the rest of grades. Actually, both in grades 6 and 7 on the MLAT-ES there were several outliers, scoring as low as 6 and 3 respectively. However, the differences in the variance calculated by means of Levene's tests showed that there were no significant differences in the distribution of scores (Levene statistic .022, p=.882). Nevertheless, in both versions of the test there are significant differences in the means across certain grades. On the MLAT-ES, for instance, a significant difference is found between grades 3 and 4 as calculated by the Mann-Whitney U test (U=1475.00, p<.001, r=-.35) and between grades 5 and 6 (U=1127.50, p<.001, r=-.29), coinciding with the introduction of grammar instruction at school. Significant differences as calculated by the Mann-Whitney U test are also found on the MLAT-EC, but only between grades 3 and 4 (probably due to the outliers mentioned above), and between grades 4 and 5, coinciding with the introduction of grammatical formal instruction at school. It should be reminded that Carroll himself (1990) conceded that scores on this part may be susceptible to previous training in formal grammar and that, according to Ranta (1998, 2002), the way part 2 measures analytic ability needs previous development of metalinguistic skills.

From a cognitivist position, more specifically in the Piagetian line, language acquisition is dependent on the development of cognitive abilities, whose cognitive structures are mapped onto linguistic representations. Therefore, the stage at which the child is could determine their capability to answer part 2. It should also be taken into account that children in middle childhood are still acquiring some structures in their L1, which could also hamper their performance in this part. In addition, the choice of one word over another could have been prompted by its meaning rather than by its function, as 9-year-old children have been found to still have difficulty in disentangling syntax from semantics (C. Chomsky, 1969). Besides, other aspects of L1 acquisition are still in the process of being interiorised at the ages the MLAT-E can be administered and are mastered when children are in grade 6, such as passive structures and object relatives (Berman, 1993, 1997; Romaine, 1984; Zorriqueta, 1998, in Serra, 2000).

Some children, especially in the lower grades, were found to mark as a correct answer the word right below the word in capitals in the stem sentence without taking into account change in the canonical word order. That is, it seems as if they had interpreted grammatical functions depending on the word position. In addition, it has

been found that before the age of 10, children find it difficult to understand subordinate clauses in English where the canonical word order SVO has been inverted (Abrahamsen & Rigrodsky, 1984). It has also been found that it is not until children are 13 that they are able to interpret subordinate clauses with word order inversion focusing on their grammatical structure.

In part 3, children were asked to choose the word that rhymed with the stem. The aim of this part is to measure children's ability to hear speech sounds. For those in group 1 (who took the MLAT-ES first), this task proved to be very easy for all grades, but it was slightly more difficult for 3-graders (mean *p*-value=.65) than for 4-graders (mean *p*-value=.74), with a small increase in the means, 14%, which is exactly the same increase as between grades 4 and 5. The minor differences between grades despite the overall steady increase across grades can only be explained by the inner facility of the test as it was designed. In contrast, this part was much more difficult for 3-graders than for 4-graders in group 2 (those who took the MLAT-EC), as the increase in the means of these two grades is 60%, with a very large effect size (Cohen's *d*= 1.10). Apart from the differences in the design of this part of the test as compared to the Spanish version, this increase between grades 3 and 4 could be explained by the fact that phonological development is still going on after grade 3 (Carroll, 1971b; Menyuk, 1971; both in McLaughlin, 1978).

It is questionable whether this task actually taps the ability to hear speech sounds in Catalan and, above all, in Spanish, which are transparent languages as for sound-grapheme correspondences. To start with, pre-school children are already aware of the onset-rime structure of syllables (Lenel & Cantor, 1981; Maclean, Bryant, & Bradley, 1987) and are indeed very familiar with them thanks to the numerous activities related to this distinction and to the nursery rhymes they have been exposed to during their early childhood (Goswami & Bryant, 1990). Besides, phonemic awareness and reading have been found to develop at a faster rate for speakers who are learning languages with transparent orthographies, as both partially depend on the mapping phoneme-grapheme. For instance, comparing studies which used phoneme-counting tasks, English (Tunmer & Nesdale, 1985) and French (Demont & Gombert, 1996) grade-1 children lagged behind Turkish, Greek and German grade-1 children (e.g. Durgunoglu & Oney, 1999; Harris & Giannouli, 1999; Wimmer et al., 1991) due to almost one-to-one grapheme/phoneme correspondence of the latter languages as opposed to French and English. What is more, only 2 out of 38 items on the MLAT-ES (or 5 out of 38 for those who *sesean* or *cecean*) and 2 out of 32 items on the MLAT-EC aimed at finding consonant rhyme with no exact consonant grapheme/phoneme

correspondence. This means that test takers could have used the strategy of looking at how words were written in most of the words in this part of the test.

Nevertheless, a peculiar response pattern in this part was found only in the lower grades. It involved matching the stem word not with the word that rhymed with it but with the word which shared the onset and peak of the first syllable with the stem. It could be argued that this response pattern was due to the lower ability to maintain the focus of attention and to inhibit irrelevant information by 3- and 4-graders in contrast with the upper grades, who succeeded in ignoring this type of distractors, probably due to their higher level of cognitive development, which facilitated it (Bjorklund & Harnishfeger, 1990).

In Catalan, vocalic graphemes <e> and <o> and consonant phonemes such as <s> can represent more than one phoneme and vice versa (see section 3.4 and 3.5.3 for an account of the items containing these phenomena). This is perhaps one of the reasons why *Part 3 Paraules que rimen* was more difficult than *Parte 3 Palabras que riman*, especially for 3-graders, as the mean *p*-value of the MLAT–ES was .65 and that of the MLAT-EC was .55, while in the rest of grades it is much more similar. The reason why some participants may have had problems when identifying the rhyme in which sibilants were involved could be a case of interference from Spanish, as some Catalan speakers show some difficulty in pronouncing these consonants due to their Spanish-dominance. Regarding vowels, research shows that contrasts between /e/ and /ɛ/, and between /o/ and /ɔ/ are difficult for early Spanish-Catalan bilinguals to perceive (e.g. Sebastián-Gallés & Bosch, 2005), not only because of "single-category assimilation" patterns (Best & Tyler, 2007), which lead the speaker to hear only one single sound when faced with two weakly different sounds, but also because of the inconsistent pronunciation of these pairs of vowels by speakers of different Catalan dialects or even of the same dialect (e.g. Recasens & Espinosa, 2006). Moreover, the higher frequency of use of Spanish over Catalan has also been found to cause some difficulty in identifying these vowel contrasts (Mora, Keider & Flege, 2010). Actually, most of the items removed in this part dealt with these vowel distinctions, but no clear pattern was observed that showed a cause-effect relationship with the test takers' language preference, which was the information regarding the use of Catalan, Spanish or both collected in the biodata questionnaire (see section 3.5.3.1).

Top scores (38 out of 38) were achieved in part 3 by 4-graders already in both tests and significant differences in the means were only found between grades 3 and 4 and 4 and 5 on the MLAT-ES and between grades 3 and 4 on the MLAT-EC as calculated by the Mann-Whitney U test. It should be noticed that 35 out of 38 items in

the MLAT-ES (or all the items if test takers *sesean* or *cecean*) and 32 out of 38 items in the MLAT-EC attempted at identifying a consonant rhyme, which means that children had to pay attention not to phonemes but to whole syllables. It has often been argued that tasks that use the syllable (Fox & Routh, 1975; Leong & Haines, 1978; Liberman et al., 1974; Treiman & Baron, 1981; Treiman & Zukowski, 1991) or intrasyllabic units, i.e. onset/rime divisions (Treiman, 1992), as the minimum unit to analyse are easier than tasks that focus on phonemes, all other task's cognitive demands being equal.

Spanish and Catalan are languages with a C-V syllable structure, as opposed to English, in which most syllables end in a consonant. For Spanish and Catalan speakers, for whom onset-rime segmentation and phonemic segmentation are the same concept, the C-V syllable structure and the equivalence in segmentation might have favoured their excellent performance in this task too, as consonant rhymes are much easier to detect under these conditions. Moreover, this part is meant to measure the ability to hear speech sounds, which is an ability that does not need written support, yet it is measured using a paper-and-pencil test, which, in the case of transparent orthographies such as Catalan and, above all, Spanish, may not be the most appropriate test format. In addition, in phonologically transparent orthographies, almost perfect word reading accuracy is achieved after only a few months of formal reading instruction while it is reading speed what differentiates good from poor readers (Cossu, Giuliotta & Marshall, 1995; Wimmer & Hummer, 1990).

Two possible alternative ways to measure the ability to hear speech sounds is taking the test in an oral format and using pictures instead of written words. Both formats have been used in studies with children. For instance, in a study using a same-different judgment task in which participants had to listen to word pairs and choose those sharing the onset or the rime and those that rhymed, Treiman and Zukowski (1991) found that 7-year-olds got the highest score, reaching a ceiling effect, in syllable recognition. Nevertheless, preschoolers were also found to have an almost perfect command in the rhyme-matching task, in spite of their preliteracy. First-grade children were also superior to the other participants (preschoolers and kindergarteners) in the other conditions (onset/rime and phoneme recognition), as for them there was no significant difference among the three conditions. Pictures can also be used for phonological and phonemic awareness measuring purposes. For instance, children can be asked to group pictures corresponding to words that begin with the same phoneme (Bradley & Bryant, 1991) or to select the picture that rhymes with the target picture (e.g. Bryant, Maclean & Bradley, 1990; Stackhouse & Wells, 1993).

One aspect that distinguished some 3-graders from children in the rest of grades is that when taking both part 3 and part 1, as the author could observe during the data collection, they muttered the words they were reading while children in the other grades were able to read in silence. Therefore, children in grade 3 had to make use of strategies to answer the test that were not needed by children older than them. This trait that distinguished 3-graders from the upper grades may be connected to their language learning experience, to the stage at which their general cognitive abilities are (Ferguson & Macken, 1980, in Grunwell, 1986) as well as to their reading experience and development. Actually, not until children master writing can they keep their mouth closed when reading and recognise and understand the word written without having to decode it phoneme by phoneme (Montessori, 1950).

Parts 1, 2 and 3 were timed, but participants did not know how long each part took. They were only encouraged to answer as fast as they could and to leave blank any answer they did not know so as to be able to answer as many items as possible and get to finish the tests. A high number of participants in the lower grades (mainly in grade 3 but also in grade 4) did not get to finish these parts. For Case (1985), cognitive development depends on the increase in the capacity of processing information that develops through the use of effective strategies and also thanks to brain maturation, which contributes to increasing the speed of neural processes. Repeating tasks allows for automatisation, in such a way that less attention is required in any task situation, which permits evolving to more advanced and effective thinking as children grow and passing to the next developmental stage. Consequently, instruction at school and experience through intellectual tasks could favour the growth of children's effective thinking and, therefore, affect the number of items answered on the MLAT-ES and the MLAT-EC. That is, since the aptitude tests used are based mainly on the participants' L1, major experience with language, literacy and memorisation practices in this language, these could be exerting some kind of influence on the quantity of items answered.

Directions may also have affected the number of items answered in the lower grades, especially grades 3 and 4, in the parts consisting of more than one page, as at the end of each page the phrase "*Total de esta página",* aimed at helping the test corrector in the counting process, appears in bold letters, which could lead the children to think that this phrase marks the end of the subtest. Even though the test takers were explicitly told the number of pages each subtest contained, apparently very few of them (2.2%, 0.6%, 0.9% in parts 1, 2 and 3 on the MLAT-ES; and 2.9%, 3.3% and 1.3% in parts 1, 2 and 3 on the MLAT-EC) may not have turned over the page when they

reached the end, despite there being directions reminding them to do so. Consequently, these subjects may have thought that they had finished a subtest when, in fact, they had not. Besides, one convention among test developers is for instructions to be written in bold letters, as they make the instruction more salient. Although only very few children do not seem to have turned over the page, this hindrance could be solved easily just by deleting the phrase "*Total de esta página*" and leaving only the direction for the test taker "*Pasar a la próxima página*", which appears right underneath.

Regarding *Parte 4 Aprendamos números / Part 4 Aprenguem números,* we observe an increase in the means across grades, although this increase is not as noticeable as in the other parts. Actually, this part proved to be easy already from grade 3 (mean *p*-value 0.70 on the MLAT-ES and 0.62 on the MLAT-EC) and hardly any differences are observed in the means between grades 4 to 7. However, there were indeed differences in the means of grades 3 and 4 on the MLAT-EC (33% increase with a large effect size, Cohen's *d*=0.80). The different performance of 3-graders on this part of the test cannot be clearly justified, as numbers were called in the same way in both tests. The referents were, nevertheless, different, as the target language was Catalan in the MLAT-EC and Spanish in the MLAT-ES. Actually, WM appears to be dependent on the language tested (N. Ellis, 1992) and so could be this part of the MLAT-E. Besides, language dominance could affect the results somehow, as it has also been found to play some role on WM tests (Chincotta & Underwood, 1996), although this factor has also been found to be irrelevant (Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993). Consequently, more data are needed to see if this result is just a matter of chance or, on the contrary, explanations grounded on the literature can be given for it.

Alexiou (2005) believed that analytic skills improve after about the age of six while memory does not, but the results she obtained are not conclusive as far as memory is concerned. In a study involving children from 5 to 9 years old, 7-year-old children were those that seemed to have an advantage in the memory tasks, while 8-year-olds excelled in analytic skills, which are also relevant when it comes to organising the information to be stored in one's memory (Milton & Alexiou, 2006). This would have implications regarding FL learning, as young learners should not only be considered memorisers who learn FLs mainly implicitly and that they could also benefit from explicit learning. In our case, the scores on part 2 clearly increase across grades although they seem to reach a plateau between grades 6 and 7. In contrast, results in

part 4, leaving outliers aside, are so similar across grades that it cannot be determined whether there is an increase in rote memory from the ages of 8 to 14 or not.

Only minimal or even nonexistent differences are found in the means across grades 4 to 7 in part 4. Yet an increase in the means is observed overall except between grades 5 and 6 on the MLAT-EC, in which a negligible change of -4% is found. Several studies confirm that STM increases up to the age of 15-16 (Wingfield & Byrnes, 1981). This increase comes along with an increase in metamemorial knowledge and a more effective use of strategies to process the information to be stored (Chase & Ericsson, 1992; R. Kail, 1990). If we consider aspects such as memorisation strategies, children in grade 6 use rehearsal strategies spontaneously, whereas 3- and 4- graders have to be told explicitly to use them. Otherwise, they do not (Naus, Ornstein & Aivano, 1977; Ornstein, Naus & Liberty, 1975). Besides, children between 9 and 10 years old have been found to use a wider variety of memory strategies to memorise any input while younger children hardly ever use any memorization strategies (Bjorklund & Douglas, 1997) or use them with hardly any benefits in the task performance (Bjorklund, 1987; Bjorklund & Coyle, 1995; Kee, 1994) probably because they lack the conceptual ability to use any (Reese, 1962; in Bjorklund & Douglas, 1997). For instance, when memorising, it has been found that children between 9 and 10 years old usually repeat a list that consists of random numbers but do not repeat it or only repeat it once when the numbers on the list are in order. In contrast, younger children do not pay attention to the rationale behind a list of numbers before memorising it and so they have more difficulty in reproducing it later (McGilly & Siegler, 1990).

Children older than 11, besides being able to repeat items to be memorised and to organise them in a logical way, are able to elaborate on them, even if these items are not easy to classify. Although this strategy entails a great mental effort (Pressley et al., 1987) and needs prior memory training, it is found to be so effective that it becomes the most preferred when memorising (Schneider & Pressley, 1989). When answering part 4, some children in grade 3 may not have deduced anything having to do with the suffix *–ca* in the numbers when memorising the numbers that make the tens (10, 20 and 30), which would have allowed them to classify them as “tens” as opposed to the units, which did not finish in *–ca*. They may not have noticed the similarity between number “*vinca*” and “*vint*” (“twenty” in Catalan) or they may not have analysed that the stem *ras-* (3) plus the suffix *–ca* (10) made up number thirty, either. Therefore, they would be remembering each number without profiting from the relations between them (Bjorklund & Hock, 1982). Consequently, the input load to encode would have been

greater for some 3-graders than for the upper grades and this might have affected its later retrieval, as participants in the lower grades would have encoded all the numbers in a less elaborated way (Ceci, 1980), not having enough processing capacity left to memorise the numbers and, at the same time, to take advantage of their repeated pattern to alleviate the memory load needed (VanPatten, 1990).

Following Pascual-Leone's model of schemes and general-purpose operators, we could partially explain the increase in the means across grades, despite not being gradual. M-capacity could also partially explain the more effective learning of words if divided into meaningful segments. Numbers in the invented language consisted of 1 or 2 syllables if they were the ones taught first (*co, vein, ras*, *silca, vinca, rasca*) or of 3 syllables if they were compounds (e.g. *vinca-ras, silca-vein,* etc.). According to Morra and Camba (2009), following the M-capacity model, simple numbers plus the memorisation process result in two elements. Compound numbers plus memorising them result in three elements. To these two or three elements (depending on the number), in order to finish the memorisation process of the number we should add, as Morra and Camba (2009) suggest, a representation of the intended meaning and a unit binding process, which results in a total of four or five units. An M-capacity of four units is achieved at around 9 years of age. Therefore, combining numbers should have been beyond 3-graders' capacity. However, a large increase between the means of 3- and 4-graders is only observed in the MLAT-EC, not on the MLAT-ES.

In the upper grades there are almost no differences in the means and almost no test takers leave this part blank. From a developmental point of view, this could be explained by the increase in information processing capacity with age, be it the gradual increase in one's M-capacity defended by Pascual-Leone (1970), in the STM storing capacity (Case, 1992b) or to myelination in the brain (R. Kail, 2000). The increase in this capacity implies the possibility of focusing one's attention and storing a greater number of elements of a problem in memory, re-elaborating on the problem-solving strategies used in similar problems faced before. Moreover, it is not only capacity that increases, but also the speed at which information is processed, which will certainly allow for, first, more cognitive processes to be active while performing any task and using one's cognitive resources in higher-order processes like those present in formal operations (R. Kail, 1991; Kail & Bisanz, 1992) and, second, for one's more general and effective use of memory (Flavell, Miller & Miller, 1977), of problem-solving strategies (Case, 1981; Siegler, 1983) and of retrieval abilities (Howe et al*.,* 1985).

Nevertheless, if one of the four parts stands out for the amount of upper-grade outliers who got remarkably lower means than average, this is part 4, even though this

part came out to be the easiest of all. Several arguments could shed light on these somewhat surprisingly low results on behalf of some participants, who, nevertheless, performed within the average mean on the other parts. One of these reasons could be the test takers' lack of attentional control during task performance. Part 4 is the last part in the test, which takes almost one hour to complete, so test takers may have got tired or bored after having completed the first three parts or they could also have got intrinsically demotivated, especially if this test was not attention-demanding enough for them (Guttentag, 1995). Moreover, in both the MLAT-ES and the MLAT-EC the extreme facility of this part could have undermined the need for effortful remembering activity (Russo et al., 1995). Indeed, the difference in the means across grades is minimal and top scores are reached from grade 3 already. It has been demonstrated that the ability to control one's attention increases with age although this depends on the interest the task awakens. Perhaps the task was not meaningful enough or other external or internal stimuli were catchier so performance by test takers in the upper grades was not as good as one would have expected from their real capacity.

Carroll and Sapon (1967) based their original norming study for grades 3 to 6 while the norming study of the MLAT-ES covers grade 7 as well. Moreover, the MLAT-E was used with older children (in grades 7 and 8) and no significant differences were found between the means. What is more, boys in grade 8 scored lower (M=106.55) than boys and girls in grade 7 (M=110.35 and M=112.27 respectively), which seems to demonstrate that at least the MLAT-E can be administered in higher grades than those in the norming study. In both the MLAT-E and the MLAT-ES norming studies as well as in the data of this dissertation, very similar patterns in the means are found across grades. That is to say, most large or very large increases in the means in all parts of the tests are found between grades 3 and 4 and negligible changes between the upper grades, mainly between grades 6 and 7, as can be seen in all the tables in sections 4.2.3 and 4.2.4, and, on the whole, there is a steady increase in the means across grades. Cognitive maturation as well as language expertise, as defined by McLaughlin (1990), could have contributed to this evolution in the scores and show aptitude as a flexible trait (Sternberg, 1998).

A relative stability was found in the scores between grades 6 and 7, but the differences in the mean results between grades 5 and 7 were significant almost in all parts of the MLAT-EC and in all parts of the MLAT-ES. From these results only, it would be a little too daring to affirm that aptitude is not stable during middle childhood and that it seems to reach some kind of stability when children are 12 or 13, entering the pre-adolescence period. This is because several factors (mainly cognitive

development, L1 acquisition development, literacy and formal language experience) seem to affect in some way or another each part differently on both the MLAT-ES and the MLAT-EC across all grades, including grades 6 and 7. Moreover, data from 8-graders was not collected that could help to confirm Harper and Kieser's (1977) findings regarding children in grade 8. Therefore, in our case, Carroll's (1981:86) suggestion that "aptitude is *relatively* fixed over long periods of an individual's life span" could perhaps be reworded here as results on the MLAT-E, regardless of the language it is in, seem to reach a plateau when taken in the adolescence period or, in other words, when test takers have entered the formal operational stage.

5.2. The MLAT-ES and the MLAT-EC as aptitude measures for boys and girls

The aim of the second research question of this study was to see whether there were significant differences in the performance by boys and girls on the MLAT-ES and the MLAT-EC, as the tendency for females to be superior to males in verbal aptitude was found in the MLAT-E norming study, in which girls consistently scored higher than boys, although this superiority was not significant overall. That is, in most parts and grades negligible changes and small increases were found between the scores obtained by males and females. Only girls in grade 4 obtained a medium increase (24%) over boys in Part 2 Matching Words and a large increase (33%) in Part 4 Number Learning, both of them with medium effect sizes (Cohen's *d* 0.57 and 0.48 respectively). In the *MLAT-ES Manual* results neglected this variable, so no comparison is possible between the data of this study and that in the *MLAT-ES Manual*. Also, in the Harper and Kieser (1977) study, girls outperformed boys overall except for 7-grade girls in Part 3 Rhyming Words, but neither this nor any of the differences between the scores obtained by girls and boys was significant or had a large effect size.

In a study in which an adaptation of the Hungarian version of the MLAT-E was used, Kiss and Nikolov (2005) also found that grade-6 girls were significantly better than boys on this aptitude measure as well as on the FL proficiency measures they used in the study. In contrast with the consistent superiority of girls over boys in the *MLAT-E Manual* and with the previous study by Kiss and Nikolov, Kiss (2009) found that 2-grade girls scored only slightly better than boys in the pilot phase of the main

study, in which the results obtained by boys and girls were almost identical, as the boys scored only 2 decimals higher than the girls.

Contrary to the results on the MLAT-E norming study, results on the MLAT-ES and on the MLAT-EC are rather favourable to boys, who obtain higher means on most parts across grades except for some very specific cases. Focusing on the results of part 1, aimed at measuring not only vocabulary but also sound-symbol ability, boys perform better than girls, though not significantly, except in grade 3 on the MLAT-ES and in grades 5, 6 and 7 on the MLAT-EC. Vocabulary is actually one of those aspects in which research has not been able yet to conclude whether there are any significant differences between sex, as some studies favour females, for instance, those examining FL productive tasks related to vocabulary (Jiménez Catalán & Ojeda Alba, 2007), while some other favour males or find no significant differences between sexes, as in those in which FL receptive tasks (Grace, 2000), receptive vocabulary size tasks (Jiménez Catalán & Terrazas Gallego, 2005-2008) were used although girls' L1 lexicon has been found to be larger than boys' from an early age (Nelson, 1973). As for the sound-symbol ability task, this study's results on the MLAT-ES and especially on the MLAT-EC, which favour males over females, seem to challenge the conclusions research has found so far, as females do seem to be superior in L1 spelling abilities (Halpern & Wright, 1996) and actually, more males than females have been diagnosed with dyslexia or other reading disabilities (Sutaria, 1995).

Part 2 is the part on which girls do better than boys most times as compared to the other parts, although they still perform worse than boys overall. The two most salient increases are found in grades 3 and 7 on the MLAT-EC, the latter significantly at .030, while in the same grades; but on the MLAT-ES the differences are small and negligible respectively. In the norming study of the MLAT-E, girls were found to perform much better than boys in grade 4 on this same Part 2, while the differences in the means of this part in the rest of grades were small. Consequently, no clear conclusion can be drawn from the results obtained on this part in the present study, as no clear pattern can be observed regarding the sex variable.

The results obtained in part 3, which measures the ability to hear speech sounds, show that males are slightly better than girls across grades except in grade 7 on both the MLAT-ES and the MLAT-EC, and in grades 3 and 5 on the MLAT-ES, the latter being significant at 0.33. It is only on the MLAT-ES that grade-5 girls' mean scores experience a medium increase as compared to boys. The differences between the other means are, in contrast, negligible or very small overall. Therefore, hardly any differences seem to be found in this ability. In fact, no differences were found in this

part either in the MLAT-E norming study or in Harper's and Kieser's (1977). Therefore, the results of this study are in the same line as those in the larger studies mentioned, although in the present one boys are generally slightly, though not significantly, better than girls.

The references available as for the superiority of girls over males in rote memory are those available in the MLAT-E norming study and in Harper's and Kieser's study, as Kiss and Nikolov used another type of memory test. In the norming study and in the Harper and Kieser study, girls performed better than boys though not significantly. In our case, the differences across grades are mostly negligible except for the significant medium decrease in the means of 5-grade girls as compared to the results obtained by 5-grade boys on the MLAT-ES. Only in one case (in grade 7 on the MLAT-EC) are girls better than boys in this part of the test. Females have been found to have better LTM, especially for recalling past events (McGuinness, Olson & Chapman, 1990; Stumpf & Jackson, 1994) and in tasks where declarative memory is needed (Halpern & LaMay, 2000). As regards rote memory as measured by the MLAT-EC and the MLAT-ES, however, they do not seem to excel.

Total scores on the MLAT-ES and the MLAT-EC favour boys over girls except in grade 3 on the MLAT-ES and in grade 7 on the MLAT-EC. That is, the tendency is for girls to score lower than boys, getting to score 15% less than boys on the MLAT-ES in grade 5. These results are contrary to the results obtained in the MLAT-E norming study, in which girls outperformed boys in all tasks and parts, and also contrary to the study by Harper and Kieser, in which the girls' total scores were also higher than the boys'.

Studies on sex differences in language ability and performance, along with research into gender differences carried out so far, do not seem to reach a clear conclusion as to whether females are consistently superior to males in all or only some verbal tasks since, depending on the linguistic abilities — and the cognitive processes that underlie them—, one sex is found to outperform the other and vice versa. Moreover, the differences between sexes are sometimes not significant. The results of this study present more variability regarding this variable as compared to the studies which took the sex variable into account and tend to contradict the research done so far, since boys are found to perform generally better than girls in most test parts across grades. Nevertheless, the differences between means are most times minimal so, with the data available, it is not possible to conclude that the boys in this study have higher FL aptitude than girls.

5.3. Discussion on the use of the MLAT-E in Spanish and Catalan as concurrent predictors of FL proficiency

The third research question was worded in the following way: “Is there a relationship between language aptitude (as measured by the MLAT-ES and the MLAT-EC) and the subjects’ proficiency in English as a foreign language?”. In order to try to answer this question, several instruments were used. These were questionnaires administered to the teachers asking how they would grade their students regarding, first, their language-specific skills and, second, other skills of a more general scope such as their students’ general aptitude in FL, their general knowledge of the FL, and their aptitude in the academic context. They were also asked to write the marks they thought their students would obtain at the end of the same year in which the data of this study were collected. Students were also asked to grade themselves and to do some tasks in English. These varied depending on the grades in which they were. While both students from group 1 and group 2 answered the questions regarding their self-marks, only the marks provided by the teachers of the participants in group 1 could be used for statistical purposes. As a consequence, the discussion about the construct validity of the aptitude tests administered using the teacher criterion measures together with those supplied by the students only applies to the MLAT-ES, the aptitude test that group 1 took in the first place.

Using the marks assigned by the participants’ teachers was the method used to test the concurrent validity of the MLAT-E in the test norming study. Carroll and Sapon admit that checking the concurrent validity of the MLAT-E using these criterion measures is not the most reliable way to do so because instructors themselves have different criteria when grading their students and many factors other than aptitude come into play when deciding on a school grade. In spite of the limitations of the criterion measures used, many high validity coefficients appear in the MLAT-E norming study correlating the criterion measures with the total score, as 29 coefficients out of 40 were above .45 and 10 were even above .60. In the norming study of the MLAT-ES, the correlations found between the probable mark obtained by the student the same year of the data collection and the total scores are lower than those obtained in the MLAT-E norming study. The coefficients for the total score on the MLAT-ES range from .28 to .42, which is a narrower range of coefficients if compared to the range of the MLAT-E. When checking the correlation coefficients of this criterion variable with part

scores, the parts that correlate the strongest are *Parte 2* and *Parte 4* except in grade 7 (r_S=.19 *ns*). Lower coefficients are found between the probable mark and *Parte 1* in grade 3 and 7.

If high validity coefficients were not expected by Stansfield and Reed (2005) due to the supposed inaccuracy of the criterion measures used, they were not expected in this study either, especially as far as the marks the test takers assigned to themselves is concerned (see section 3.6.3). Regarding the marks given by the teachers, the correlations found between the probable FL mark and the total scores on the MLAT-ES are from low (considering all grades together) to moderate. All of them are statistically significant except in grade 3, in which none of the criterion measures correlates with the scores on the MLAT-ES, neither with the part, nor with the total scores. The part scores obtain lower coefficients across grades, especially *Parte 1* in all grades except grades 5 and 6, and *Parte 4* in grade 7, as it happened in the MLAT-ES norming study. Consequently, it is questionable if *Parte 1* is a valid predictor of concurrent proficiency as measured by criterion measures across all grades and so is *Parte 4* in the higher grades, especially in grade 7. Nevertheless, considering the total scores, the MLAT-ES does seem to be a valid predictor of FL proficiency, FL aptitude, general FL knowledge and even general aptitude in the eyes of the participants' teachers since, except when considering the whole cohort as one group, all the correlations obtained are significant and moderate. However, the validity of the teachers' criterion is somehow at stake when comparing the marks they considered their students deserved at the end of the year, as no significant difference was found between this measure and their ratings of general FL aptitude when performing a Mann-Whitney U test, taking into account that many factors other than just FL aptitude come into play when marking students.

Although significant differences were found between the marks the teachers estimated for their students and the marks the students gave to themselves, the correlations between these measures and the MLAT-ES are fairly similar. The only divergences are found, on the one hand, between the correlations with the total scores in grade 7 and, on the other hand, with the scores on *Parte 4* and these criterion variables in grade 6. As it happened with the marks assigned by the teachers, *Parte 1* does not correlate with the students' self-marks and no correlation is found either between the students' marks and the MLAT-ES in grade 3 in any of the parts or the total score.

The MLAT-ES norming study also provides the coefficients obtained when correlating the scores on the MLAT-ES with general aptitude, FL aptitude and the

marks the students had obtained the previous year provided by the students themselves. In the present study it was decided to remove this last variable due to the low reliability this data-collection method implied (see section 3.6.3). The index of the participants' FL aptitude in the MLAT-ES norming study was obtained by summing the language-specific skills rates provided by the teachers, while in our study it comprised just one estimate of achievement. The correlation coefficients obtained between the MLAT-ES total score and the FL aptitude variable in the norming study ranged from .30 (in grade 5) to .41 (in grade 4), while in our study, correlations were not significant and low in grade 3, and ranged from r_s=.479, p<.001 (in grade 6) to r_s=.706, p<.05 (in grade 4). Mind, though, that the coefficient was very low when considering the whole group (r_s=183, p<.05). All in all, the construct validity between the MLAT-ES and FL aptitude seems to be greater than the one obtained in the MLAT-ES norming study.

The only variable left to compare with the norming study is that of general aptitude. In the norming study, the correlation coefficients between this variable and the total score are the highest of all the correlations run with the general-scope criterion measures, ranging from r_s=.31 to r_s=.42. In this study, the correlations between this variable and the total scores are even higher, ranging from r_s=.556 to r_s=.674, both at p<.001, without taking into account, once more, grade 3 (r_s=274, *ns*) and the whole group (r_s=218, p<.001). All in all, general aptitude is not the variable that consistently correlates the strongest with the total score on the MLAT-ES.

In our case, the limited validity of the criterion measures regarding language-specific skills and the estimates of achievement grades at the end of the school year provided by the teachers was revealed by the Wilcoxon signed-rank tests, which showed that in grade 3 and grade 4 teachers did not make any distinction as regards the language-specific measures, that is, they seemed to consider their grade-3 and grade-4 students' FL skills unitarily. Besides, no significant differences were found either between the listening and the grammar estimates of achievement in any grade. In spite of these questionable criteria adopted by the teachers, the MLAT-ES proved to be a valid predictor of concurrent FL proficiency mostly in all grades (the significant positive coefficients ranging from r_s=.228 at p<.001 to r_s=.746 at p<.001 with the total score) except in grade 3. Actually, the scores on the MLAT-ES obtained by 3-graders are significantly lower than those obtained by 4-graders and it is only the scores on the MLAT-ES by 3-graders that do not correlate with the teachers' criterion measures.

Significant correlations do not mean that there is a cause-effect relationship between the measures correlated, only that there exists a relationship between the two variables. Consequently, it may be the case that 3-graders have either a different

(higher or lower) FL proficiency than that reflected by the proficiency measures used in this study or a different aptitude than that reflected by their scores on the MLAT-ES for the reasons explained in the discussion of research question 1 (see section 5.1). Consequently, perhaps the failure to show the relationship between 3-graders' FL aptitude and their FL proficiency as measured by the criterion variables lies in the design of the proficiency or the aptitude measure themselves. Another explanation could be grounded on Carroll's (1963) Model of School Learning, which defends that given optimal learning conditions, aptitude is taken over by them and its role is diminished. Therefore, aptitude in grade 3 may not be directly related to FL proficiency, but other factors such as motivation or general verbal intelligence may be more revealing at this age, although it has been repeatedly stated that they may overlap with aptitude despite their being autonomously different IDs (Skehan, 1991; see also section 1.6.6.1).

It has also been argued that aptitude is only relevant as a predictor of FL proficiency in formal contexts (Krashen, 1981, 1985; Reber 1989, 1993; Reber, Walkenfield & Hernstadt, 1991), as is ours, in which learning is supposed to take place mainly explicitly, although some have defended precisely the opposite view (Skehan, 1998) or even that aptitude is independent of the language instruction context (Carroll & Sapon, 1959; Reves, 1983). Actually, in two studies using the MLAT-E as an aptitude measure, it was found that children with high language aptitude (or high IQ) benefitted more from classes with a situational approach, i.e. similar to a naturalistic context (Hauptman, 1971). In the study by Harper and Kieser (1977), however, aptitude was significantly related to achievement marks in both an audiolingual context and a semi-formal context. Therefore, it is still not clear that any of the forms of the MLAT-E is relevant in any of these contexts.

The teachers of the youngest children in this study said that their teaching approach was mainly communicative, with hardly any writing activities and that they did not teach any grammar explicitly. Besides, younger children are supposed to learn mainly implicitly (Bley-Vroman, 1989, 1990; DeKeyser, 2000, 2003; DeKeyser & Larson-Hall, 2005). However, it would be too daring to say that aptitude as measured by the MLAT-ES may not be relevant for children in grade 3 but that it is for children in grade 4 as, although in this grade correlations are significant and moderate, only one year separates both groups. However, it is true that the correlation coefficients between the total score and the proficiency criterion measures are all optimal in the upper grades, while they are never so in grade 3.

The correlations between the language-specific skills and the partial scores have to be discussed on the basis of the construct each part is assumed to measure, as correlations between these measures one by one separately and the MLAT-ES were not published in the *Manual*. Instead, Stansfield and Reed (2005) used one variable that added up all the language-specific skills

The scores on *Parte 1, Parte 2* and the total score do not correlate with the speaking measure in grade 3 and considering group 1 as a whole. *Parte 1* and *Parte 4* also significantly correlate intermittently with this language skill and it is only in grade 4 that speaking correlates significantly with both part scores and the total scores. This is in accordance with the literature in which the MLAT for adults was found to have a lack of concurrent validity to predict speaking (Brecht, Davidson & Ginsberg, 1995; R. Ellis, 1986; Winke, 2005). Besides, Gardner et al. (1976) found aptitude as measured by the MLAT for adults to be more strongly bonded to class grades than to communicative skills. However, in our data the differences between the correlations of the MLAT-ES with the speaking measure and with the probable mark at the end of the course are minimal. In addition, in oral tests, an adaptation of the MLAT-E has been found to correlate moderately and significantly with the scores given by three raters to children performing oral proficiency tasks (Kiss, 2009).

Parte 1 is meant to measure vocabulary in the test takers' L1, so it is expected to correlate with the vocabulary criterion measure, but it only does so in grade 4 and grade 6. *Parte 4*, which in order to answer it the test taker has to learn some words, is supposed to be somehow related to vocabulary. It is significantly related to this variable only in grades 5 and 6. Therefore, according to the criterion measure used, these parts do not seem to tap the constructs they are meant to tap, although the adult version of this part has been found to be related to vocabulary (r=.50, p<.05) in previous research with adults (e.g. Winke, 2005b). Actually, an adaptation of this part was the only one that correlated significantly at a moderate level with three out of the four ratings she used provided by the participants' teachers and external raters (Kiss, 2009).

Parte 2 is meant to measure grammatical sensitivity, so this part would be expected to correlate significantly with the grammar criterion measure. This is precisely the relationship that this part has with the grammar criterion measure across grades except in grade 3. Actually, Words in Sentences has been found to predict FL proficiency, not only grammar knowledge in several previous studies (e.g. Alderson, Clapham & Steel, 1997; Elder et al., 1999; Ranta, 2002; Roehr, 2007). It also does so in the present study, as significant correlations with this part are also found with the rest of language-specific skills, except with vocabulary and listening in grade 4 and with

listening considering all grades together. The way *Parte 2* seems to relate to FL proficiency is the opposite of what Kiss (2009) found, as her adaptation of this task did not correlate significantly with any of the ratings supplied by the teachers in her study.

Regarding *Parte 3*, its design was somehow questioned in section 2.3.2.1 and in 2.3.4. However, it seems to be one of the parts which obtains the highest correlations with all the teachers' criterion measures and so does *Parte 2*, even if it is intended to tap grammatical sensitivity only. *Parte 3* was designed to detect one's ability to hear speech sounds and so, among the language-specific skills measures, it should be related to listening, in principle. It seems to be so, indeed, obtaining significant moderately-strong correlations, but it also obtains this kind of correlations with other language skills such as speaking and grammar in grade 4 or with vocabulary and grammar in grade 6. These results do not coincide with those obtained by Kiss (2009), as the adaptation she used of this part of the test did not correlate significantly with any of the ratings supplied by the teachers in her study.

Parte 4 loses its predictability power the higher the grade. Actually, Carroll (1990) himself admitted this part was too easy as the distribution of scores on this part always appeared to be negatively skewed in the data he handled. Consequently, this loss of predictability power could be due to the extreme facility of this part as compared to the rest, which shows in the disappearance of significant correlations in grade 7. However, it is positively and significantly related with all the skills in the rest of grades except with vocabulary in grade 4 and in grade 3, as mentioned above. This part is meant to measure auditory comprehension as well as memory, which is needed in several language tasks and is relevant for skills such as speaking and vocabulary, as it is also meant to tap the ability to learn a large number of associations of new words in a relatively short time. This part is a simplified version of Part 1 in the MLAT, which, in studies with adult populations, has been found to correlate, for example, with vocabulary and with speaking fluency (Winke, 2005b). An Arabic adaptation of this test by Reves (1983) also correlated with grammatical accuracy, oral fluency and course grades in both L1 Hebrew and English as a FL. Consequently, it seems as if this part was a useful measure to predict FL proficiency in general, and not just those tasks in which memory plays some kind of role. Actually, memory has been defended as perhaps the most important component of language aptitude (Skehan, 1982), although the type of memory measured by *Parte 4 Aprendamos números* is rote memory ability, which is of a rather limited scope as compared to working memory, considered to be the main underlying process of FL aptitude (e.g. Miyake, Friedman and Osaka, 1998).

The children in this study were also administered English proficiency measures in order to test the concurrent construct validity of the aptitude measures. Children in grades 3 and 4 were administered a listening test with a low reliability index and a cloze passage. The cloze passage was used because it is a measure considered to be integrative, measuring general proficiency in the FL. However, it could also be said that vocabulary and some analytic ability was required to answer it properly. The cloze passage correlated mostly with all parts except some specific cases (*Parte 1* in grade 3 – group 1; *Part 2* in grade 3 – group 2; and *Parte 4* and *Part 4* in grade 4 – groups 1 and 2). For its integrative nature, this test was expected to correlate with the total score, although it could also be expected to correlate with other parts due to its vocabulary focus (tapped in part 1 and part 4), and analysis component (tapped somehow in part 2). Certainly, across grades, this test correlated moderately and significantly with the total score, which confirms the construct validity of both the MLAT-ES and the MLAT-EC, and also with part 3, Rhyming Words, which in principle is not directly related to the skills needed to complete the cloze passage. Significant correlations with part 1 and part 4 were not found consistently in all grades, as mentioned above. Regarding the analytic component as measured by the Words in Sentences, which has been argued to be not so relevant in younger children, it seems to have some kind of relationship with proficiency as measured by the cloze, although the impact of the correlations is not homogeneous across grades depending on the aptitude measure used. Therefore, the measures that prove to be more powerful in grade 3 and grade 4 as a predictor of concurrent FL proficiency are the total score and Part 3, although there are no grounds in the literature to justify it.

5-, 6- and 7-graders were administered a cloze passage (slightly shorter for 5-graders), a listening test (longer than the one taken by 3- and 4-graders), and a dictation. The cloze passage, as said before, was considered an integrative measure. As such, it was expected to correlate significantly at least with the total scores on the aptitude measures. It does considering the grades separately and all together, except in grade 6 of group 1. In this group, this proficiency measure only correlates significantly with *Parte 2*, which is also related to the skills needed to answer the cloze passage. As for the correlations in the rest of grades, the cloze passage is found to correlate with all parts even with part 3, which is, in principle, unrelated to this proficiency measure in all grades. The exception is the lack of significant correlation with *Part 4* in grade 6.

The listening test was expected to correlate with, besides the total score, parts 3 and 4. However, this is not the case of grade 6 in group 1, in which no significant

relations are found between the listening test and any part scores except with *Parte 2.* It is not the case either of grade 5 in group 2, in which there is only a significant low correlation between this proficiency measure and *Part 3*. Besides, other no significant correlations are found between part scores and this proficiency measure in other grades in both groups. As for part 3, in both groups the significant correlations existing with this proficiency measure are from low to moderate. Something similar happens with part 4, as only significant moderate correlations are found between this part and the listening test in grade 5, and the whole group in group 1 and in grade 7 and the whole group in group 2. However, this proficiency measure does obtain significant moderate correlations with all parts when considering the three grades together. In front of these inconsistent results across grades, it could be suggested that perhaps the listening proficiency measure used should be fine-tuned for children in these grades, as while it fails to correlate with some parts and some total scores when examining each grade separately, it does not when considering the whole group.

While spelling mistakes were not taken into account when correcting the cloze passage, they were when grading the dictation test. Therefore, in order to do this task, apart from listening abilities and memory, spelling ability was needed. It would be up to a certain extent logical that this measure correlated, in addition to the total score, with part 1 as "(1) (…) spelling ability can be equally well measured by dictation tests or misspelling-recognition tests, other things being equal, (2) (...) spelling ability includes at least implicit knowledge of conventional spelling rules and phoneme-grapheme correspondences, but (3) (…) spelling ability does *not* involve immediate memory for visual forms of words." (Carroll, 1973:170). The dictation was also expected to correlate with part 4, as memory and "auditory alertness" (Carroll & Sapon, 1959) is needed to retain the phrases in order to write them correctly after having listened to them. All the expectations were met considering grades 5, 6 and 7 separately and all together, except with part 4 in grade 7 in group 1 and in grade 6 in group 2. Part 3, which has so far not been used for research purposes contrasted with dictation tests, also correlated significantly with this proficiency measure across grades. Therefore, the concurrent construct validity of both the MLAT-ES and the MLAT-EC in relation to this proficiency measure was, again, found in general terms, with the sporadic exception of part 4 just mentioned.

The only existing studies in which proficiency measures have been administered and a partial adaptation of the MLAT-E has been used are those by Kiss (2004, 2009) and Kiss and Nikolov (2005). They found that the total scores of the aptitude measure correlated significantly and moderately with language proficiency in

general (a composite of several listening, reading and writing tasks). Correlations with part scores are not reported in this study, so it is not possible to contrast them with ours. Kiss (2009) does report the correlations with part scores. In her study, Hidden Words correlates with three of the 5 proficiency criteria that she used in her study (marks estimated by raters related or not to the children in the study). Kiss (2009) also reports the correlations with the Number Learning task. These are not significant with any of the raters' estimates. What can be said of our study is that, as compared to Kiss' study, Part 1 Hidden Sounds along with Part 4 Number Learning seem to be the parts that, on the whole, correlate consistently the least with the proficiency measures used.

In conclusion, seeing the sporadic inconsistencies of the correlations between the aptitude scores and the proficiency measures used, it can be stated that the best predictor of concurrent proficiency is the total score on the aptitude measures. However, this does not mean that using other proficiency measures, perhaps more in the line of the tasks the children in the study are used to doing in their everyday school-life, stronger coefficients could be found between part scores and the proficiency measures.