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 4: RESULTS

4.1. Introduction

In this chapter, the results obtained from the statistical analyses for each part of the MLAT-ES and the MLAT-EC (depending on the group) will be presented. First of all, the differences between the mean scores across grades will be analysed in order to see if the aptitude tests follow a similar development across all grades or if, on the contrary, they vary in different degrees between certain grades. This will help to answer the first research question, which is whether the MLAT-ES and the MLAT-EC are equally suitable measures for all grades from 3 to 7. The results will also be contrasted with those available in the *Manuals* of both the MLAT-E and the MLAT-ES, to see if the MLAT-ES and the MLAT-EC conform to the data in the norming studies. As for the second research question, whether there are any differences in the aptitude results obtained by boys and girls, their means scores will be compared as well across grades and with those in the *MLAT-E Manual,* which are also presented taking this variable into account. Finally the third research question will analyse the construct validity of the MLAT-ES and the MLAT-EC administered in Catalonia by correlating the results obtained in these tests with those in the English proficiency measures administered to the participants as well as with the questionnaires filled out by the participants' teachers.

4.2. Research question 1: the MLAT-ES and the MLAT-EC in relation to the participants' age

The first research question was worded as follows:

To what extent are the MLAT-E in Spanish and Catalan suitable language aptitude measures for learners in grade 3 to 7?

All through section 3.5 of this dissertation, the differences in the means across grades have been a recurrent topic, particularly regarding the extreme grades in the sample, that is, grade 3 and grade 7. The descriptive results of both the MLAT-ES and the MLAT-EC administered in Catalonia show that while the increase in the mean

scores is more or less regular across grades 4, 5 and 6, it is not so between grades 3 and 4, where apparently there is a large increase as compared to grades 4, 5 and 6. Also, between grades 6 and 7, the results seem to reach a plateau and, therefore, the difference between the means of these grades cannot be significant.

In order to answer the first research question, previously, we needed to test for homogeneity of variance, as the data proved not to be normally distributed. However, it would be interesting to know if the existing variance is stable in all grades or if, on the contrary, the spread of scores around the means varies from one grade to another. The second set of analyses were attempted to explore whether the changes in the means across grades are significant and, consequently, threaten the validity of any of the tests for some grades, especially those at the extremes. Lastly, these results will be compared to the results obtained in the MLAT-E and the MLAT-ES norming studies. It seems sensible to do so because the MLAT-E is the test on which the MLAT-ES is based and the MLAT-ES norming study has been published as the official norm valid for all Spanish contexts. Moreover, the MLAT-EC was designed by adapting the MLAT-ES and translating it into Catalan.

4.2.1. The MLAT-ES in relation to the participants' age

As Wilcoxon tests showed that there was a task-training effect between the first sitting of the aptitude test and the second one, only the results of group 1 are reported in this section. Following are the boxplots that illustrate the distribution of group 1's scores on each part of the MLAT-ES, already described in section 3.5.2.3.

Figure 4.1. Boxplot of MLAT-ES *Parte 1 Palabras ocultas*

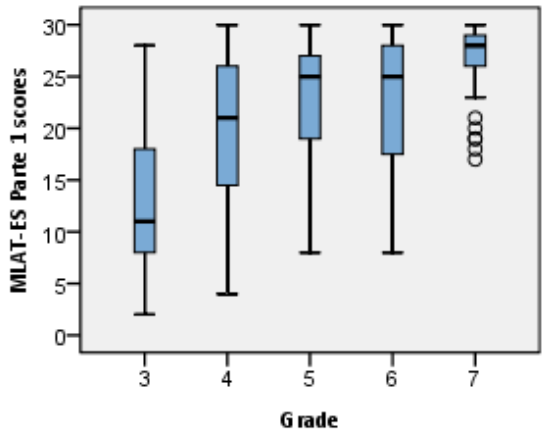


Figure 4.2. Boxplot of MLAT-ES *Parte 2 Palabras que se corresponden*

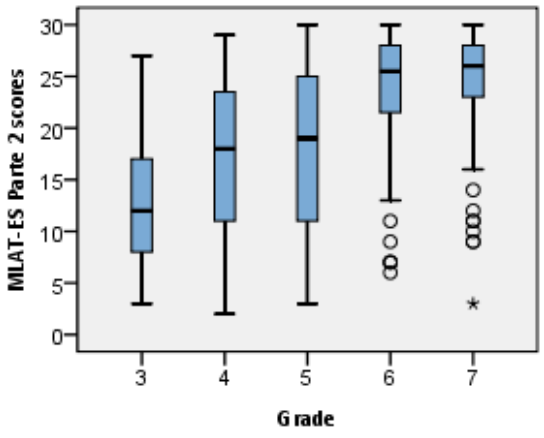


Figure 4.3. Boxplot of MLAT-ES *Parte 3 Palabras que riman*

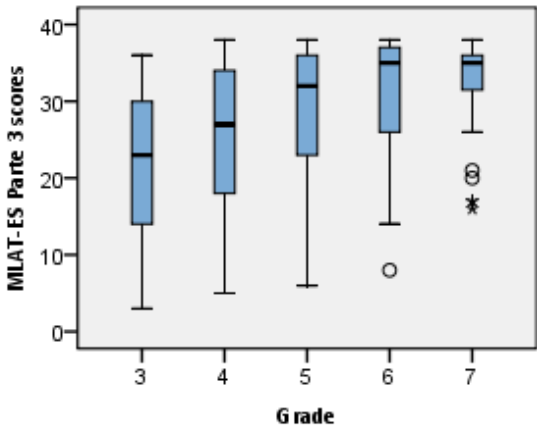


Figure 4.4. Boxplot of MLAT-ES *Parte 4 Aprendamos números*

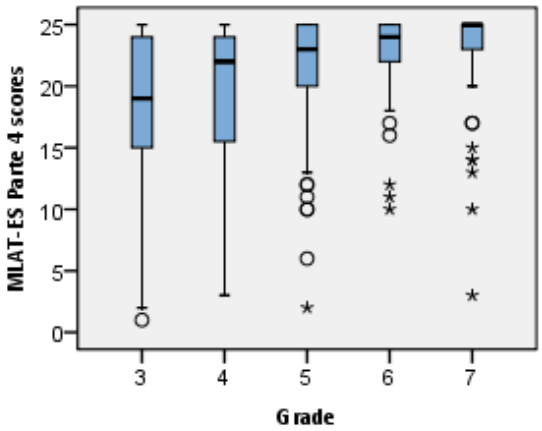
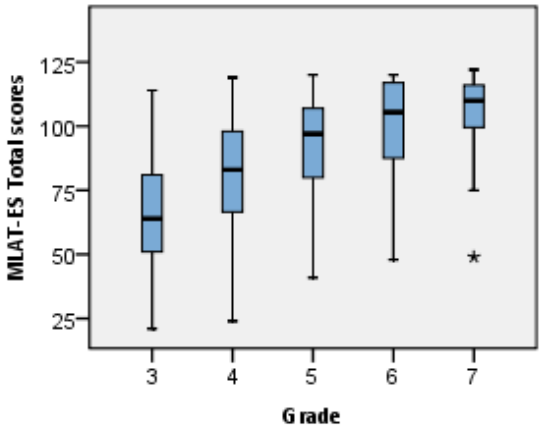


Figure 4.5. Boxplot of MLAT-ES Raw total score



While boxplots help us see graphically the distributions of scores depending on the grade, it is the Levene's test statistic, based on the mean, that determines whether the spread of the scores varies significantly from one grade to another. In Table 4.1, the significant levels are in bold. They reject the null hypothesis that there is no variance in the distribution of scores across grades.

Table 4.1. Levene's test for the MLAT-ES

Grades	Grades 3 & 4		Grades 4 & 5		Grades 5 & 6		Grades 6 & 7		All	
MLAT-ES	Levene Statistic	Sig.	Levene Statistic	Sig.	Levene Statistic	Sig.	Levene Statistic	Sig.	Levene Statistic	Sig.
Parte 1	5.188	**.024**	13.612	**.000**	.239	.626	49.173	**.000**	20.350	**.000**
Parte 2	.034	.854	.015	.902	9.568	**.002**	.022	.882	5.954	**.000**
Parte 3	.350	.555	3.898	**.050**	6.470	**.012**	15.869	**.000**	13.743	**.000**
Parte 4	12.162	**.001**	.004	.952	26.230	**.000**	.038	.846	10.649	**.000**
Total	.515	.474	1.772	.185	7.218	**.008**	12.767	**.000**	11.298	**.000**

On the basis of the descriptive results, it was expected that the group as a whole would present significant differences in the distribution of scores across grades. The variances in *Parte 1 Palabras ocultas* are significantly different between all grades, except for grades 5 and 6, $F(1, 165)=.239$, *ns*, as scores in grades 5 and 6 tend to concentrate below the median (see Figure 4.1 above). The opposite phenomenon is found between grades 3 and 4, in which 3-graders' scores tend to concentrate below the median while 4-graders' do not, $F(1, 180)=.024$, $p<.05$.

Scores on *Parte 2 Palabras que se corresponden* present a similar distribution from grades 3 to 5. Suddenly, this distribution changes between grades 5 and 6, $F(1, 165)=.002$, $p<.05$, in which scores are piled up near the median, as happens in grade 7. Consequently, between grades 6 and 7, Levene's test results in a non-significant value, $F(1, 189)=.882$, *ns*.

Scores on *Parte 3 Palabras que riman* have the tendency to be normally distributed except in the higher grades, in which they tend to concentrate below the median with a variance that has significant values across all grades but between grades 3 and 4, $F(1, 180)=.555$, *ns*.

Besides the evident lack of spread in the scores on *Parte 4 Aprendamos números* in the higher grades (see Figure 4.4 above), notice that there is no significant difference either in the distribution values between grades 4 and 5, $F(1, 175)=.004$, *ns.* This is probably due to the outliers in grade 5. The difference in the variance is not significant either between grades 6 and 7, since most scores are close to the maximum score in both grades. Therefore, it is expected that test takers in grade 7 or higher perform at the top marks of this test.

In order to see if the differences in the means across grades are significant, it is necessary to run a Mann-Whitney U test, which is the non-parametric alternative technique to the independent-samples t-test. The Mann-Whitney U test allows us to compare two groups with different participants when the resulting data of the condition are not normally distributed, as is our case. By default, SPSS provides the two-tailed significance value. If this value is less than .05, then the means of the two independent groups are significantly different. If it is predicted that one group will obtain a higher score than another one, then we need to calculate the one-tailed probability by dividing the two-tailed value by two. In this case, our prediction is that the higher the grade is, the higher the score will be, so it is necessary to obtain the one-tailed significance value. Both the two-tailed and the one-tailed values appear in bold when significant. The tables below provide the statistics of the differences between means across grades in each part of the MLAT-ES as well as in the total score. two-tailed and one-

tailed values as well as effect sizes are also provided. Significant values appear in bold and so do large effect sizes in this and all the tables in this section.

Table 4.2. Mann-Whitney U test between grades 3 and 4 on the MLAT-ES

Grades 3 and 4 N=141	MLAT-ES Parte 1	MLAT-ES Parte 2	MLAT-ES Parte 3	MLAT-ES Parte 4	MLAT-ES Total
Mann-Whitney U	1209.500	1475.000	1971.500	2082.000	1422.000
Z	-5.236	-4.136	-2.082	-1.631	-4.352
Asymp. Sig. (2-tailed)	**.000**	**.000**	**.037**	.103	**.000**
Asymp. Sig. (1-tailed)	**.000**	**.000**	**.019**	.052	**.000**
r (effect size)	**-.44 large**	-.35 medium	-.15 small	-.10 small	-.37 medium

The differences between the means of grades 3 and 4, as shown in Table 4.2 above are significant in all parts but 4. For 4-graders, the results were significantly higher in *Parte 1* (*Mdn*=21.00), obtaining a large effect size in *Parte 1*, *U*=1209.50, *p*<.001, *r*=-.44; a medium size effect in *Parte 2* (*Mdn*=18.00*, U*=1475.00, *p*<.001, *r*=-.35) and in the total score (*Mdn*=83.00*, U*=1422.00, *p*<.001, *r*=-.37); and a small size effect in *Parte 3* (*Mdn*=27.00*, U*=1971.50, *p*<.05, *r*=-.15). However, 4-graders (*Mdn*=19.00) did not seem to differ from 3-graders (*Mdn*=19.00) in their scores on *Parte 4*, *U*=2082.00, *p*<.001, *r*=-.10. On the whole, the test seems to work differently for these two grades, especially as for the first two parts, which inevitably affects the size of the differences in the mean total score when comparing grades 3 and 4.

Table 4.3. Mann-Whitney U test between grades 4 and 5 on the MLAT-ES

Grades 4 and 5 N=132	MLAT-ES Parte 1	MLAT-ES Parte 2	MLAT-ES Parte 3	MLAT-ES Parte 4	MLAT-ES Total
Mann-Whitney U	1508.000	1932.000	1654.000	1760.000	1608.500
Z	-2.897	-.945	-2.224	-1.751	-2.431
Asymp. Sig. (2-tailed)	**.004**	.345	**.026**	.080	**.015**
Asymp. Sig. (1-tailed)	**.002**	.173	**.013**	**.040**	**.008**
r (effect size)	-.25 small	-.08 small	-.19 small	-.10 small	-.15 small

Table 4.3 shows that significant differences in the means continue to exist between grades 4 and 5 in all parts but in *Parte 2 Palabras que se corresponden*, which means that 4- (*Mdn*=18.00) and 5-graders (*Mdn*=19.00) show a similar performance in this part, *U*=1932.00, *p*<.005, *r*=-.08. Notice, though, that the performance on *Parte 4* of both groups is significantly different only when looking at the one-tailed significance value, *U*=1760.00, *p*<.05, *r*=-.10. The effect sizes are small in all parts, no matter the significance value, only reaching a close to medium effect size in

Parte 1, in contrast with the effect sizes in the previous comparison involving grade 3 and grade 4.

Table 4.4. Mann-Whitney U test between grades 5 and 6 on the MLAT-ES

Grades 5 and 6 N=117	MLAT-ES Parte 1	MLAT-ES Parte 2	MLAT-ES Parte 3	MLAT-ES Parte 4	MLAT-ES Total
Mann-Whitney U	1626.000	1127.500	1357.000	1383.000	1263.000
Z	-.459	-3.183	-1.932	-1.831	-2.439
Asymp. Sig. (2-tailed)	.646	**.001**	.053	.067	**.015**
Asymp. Sig. (1-tailed)	.323	**.001**	**.027**	**.034**	**.008**
r (effect size)	-.04 small	-.29 medium	-.18 small	-.10 small	-.17 small

Looking at the two-tailed values in Table 4.4, *Parte 2* shows a significant difference between the test takers' performance in grade 5 (*Mdn*=19.00) and in grade 6 (*Mdn*=25.50), *U*=1127.50, *p*=001, *r*=-.29. This significant difference is also present in the total score, affected by the score on *Parte 2.* This means that the performance of both grades is very similar in all the parts except for *Parte 2.* The participants in grade 5 (*Mdn*=97.00) also obtain lower results in the total score than those in grade 6 (*Mdn*=105.50), *U*=1263.00, *p*<.05, *r*=-.17*.* The one-tailed value extends the significant differences to *Parte 3* and *Parte 4* with small size effects.

Table 4.5. Mann-Whitney U test between grades 6 and 7 on the MLAT-ES

Grades 6 and 7 N=127	MLAT-ES Parte 1	MLAT-ES Parte 2	MLAT-ES Parte 3	MLAT-ES Parte 4	MLAT-ES Total
Mann-Whitney U	1267.000	1844.500	1886.500	1793.500	1666.000
Z	-3.611	-.803	-.600	-1.106	-1.662
Asymp. Sig. (2-tailed)	**.000**	.422	.548	.269	.096
Asymp. Sig. (1-tailed)	**.000**	.211	.274	.135	**.048**
r (effect size)	-.32 medium	-.07 small	-.05 small	-.09 small	-.15 small

Between grades from 3 to 6 significant differences were found in the total scores. This is not the case of the difference in the results of grades 6 (*Mdn*=105.50) and 7 (*Mdn*=110.00) if we stick to the two-tailed value, the value that appears by default and that does not assume any difference in the results of the two groups (see Table 4.5). The one-tailed value is significant, yet very close to a non-significant value, *U*=1666.00, *p*<.05, *r*=-.15. *Parte 1* is the only part that carries a significant value in the difference between the means of grades 6 (*Mdn*=25.00) and 7 (*Mdn*=28.00), with a medium effect size, *U*=1267.00, *p*<.001, *r*=-.32. On the whole, it can be concluded that

grades 6 and 7 perform almost in the same way in this test and so do, up to a certain extent, grades 5 and 6, in contrast with what the results in the lower grades show.

Cohen's *d* is an index that informs us about the effect size of the difference between the means between independent samples. Its range of values goes from -0.15 to 1.45. Cohen's *d* can be calculated by knowing the N, the mean and the standard deviation of the two samples to be compared. This index, although not necessarily needed when having access to the raw data, is very useful when we need to compare our data with other published studies whose raw data is not at hand, as will be demonstrated later on. With the information available, it is also possible to estimate the relative size of percentage change. In this case, the two extremes of the percentage scale are <-75 (huge decrease) and >75 (huge increase). These values appear in Table 4.6 along with the numerical percentage change in the means and the effect size and direction this change has.

So far it has been shown that grade 3 seems to perform differently from the rest of grades, and that grades 6 and 7, and somehow grade 5, seem to perform very similarly. However, a deeper analysis of the differences in the performance on the different parts, especially the magnitude of the increase, can help us see whether the different parts of the MLAT-E in Spanish function in a different way depending on the grade or if, on the contrary, it is the whole test unitarily that determines the differences in the performance. In this way, it will also be possible to compare the performance of the bilingual Catalan/Spanish population to that of the norming study. The following table presents the differences between group 1's grades in each part of the MLAT-ES. Both large effect sizes and large or very large increases appear in bold.

Table 4.6. Differences in the means on the part and total scores of MLAT-ES across grades

MLAT-ES 1 *PALABRAS OCULTAS*	N	Cohen's *d*	Effect size	Percentage change	Direction
GRADE 3 – GRADE 4	141	**0.98**	**large**	**52**	**very large increase**
GRADE 4 – GRADE 5	132	0.56	medium	18	medium increase
GRADE 5 – GRADE 6	117	0.04	negligible	-1	negligible change
GRADE 6 – GRADE 7	127	**0.80**	**large**	18	medium increase
MLAT-ES 2 ***PALABRAS QUE SE CORRESPONDEN***	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	141	**0.76**	**large**	**40**	**large increase**
GRADE 4 – GRADE 5	132	0.14	negligible	6	small increase
GRADE 5 – GRADE 6	117	0.66	medium	25	medium increase
GRADE 6 – GRADE 7	127	0.11	negligible	3	negligible change
MLAT-ES 3 ***PALABRAS QUE RIMAN***	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	141	0.34	small	14	small increase
GRADE 4 – GRADE 5	132	0.38	small	14	small increase
GRADE 5 – GRADE 6	117	0.30	small	8	small increase
GRADE 6 – GRADE 7	127	0.29	small	6	small increase
MLAT-ES 4 ***APRENDAMOS NÚMEROS***	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	141	0.32	small	11	small increase
GRADE 4 – GRADE 5	132	0.47	medium	14	small increase
GRADE 5 – GRADE 6	117	0.43	medium	9	small increase
GRADE 6 – GRADE 7	127	0.02	negligible	0	negligible change
MLAT-ES RAW TOTAL SCORE	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	141	**0.78**	**large**	26	medium increase
GRADE 4 – GRADE 5	132	0.43	medium	11	small increase
GRADE 5 – GRADE 6	117	0.44	medium	10	small increase
GRADE 6 – GRADE 7	127	0.42	medium	7	small increase

Table 4.6 shows that, on the one hand, *Parte 1* and *Parte 2* seem to function differently across grades. Both parts present a very high Cohen's *d* value between grades 3 and 4, since there is a large increase in the means obtained in this part by 4-graders as compared to 3-graders (0.98 in *Parte 1* and 0.76 in *Parte 2*). *Parte 1* is very irregular as for the change in the means from grades 4 to 7 and so is *Parte 2.* In *Parte 1* the increase between grades 4 and 5 and between grades 6 and 7 is medium (18%) while there is a negligible decrease between grades 5 and 6 (-1%). This irregular pattern is found as well in *Parte 2*, though affecting other grades.

The only change that appears consistently and clearly in *Parte 1* and *Parte 2* is the large increase in the means between grades 3 and 4 (52% and 40% respectively). Probably more data are needed that can help us explain clearly the impact of the irregular increase across grades of *Parte 1* and *Parte 2*. It could be suggested that, as *Parte 1* depends very much on the spelling of words, the stage of literacy development

of 3-graders in contrast with that of 4-graders could explain the difference in the means on this part. As far as *Parte 2* is concerned, three large groups can be outlined on the basis of the test takers' performance. Grades 3 and 4 present a large increase with a large effect size in the means (Cohen's *d*=0.76) that stops between grades 4 and 5. It is between grades 5 and 6 that this difference becomes noticeable (25%) and with a medium effect size (Cohen's *d*=0.66) once more. Again, a negligible change is observed between grades 6 and 7 (3%). Grades 5 and 6 are, therefore, turning points as far as grammatical sensitivity is concerned. It is worth mentioning that the *Curriculum d'Educació Primària* (Curriculum of Primary Education, Catalonia, Servei d'Educació Curricular, 2009) contemplates the introduction of syntactic analysis in language subjects such as *Lengua Española* (Spanish language) and *Llengua Catalana* (Catalan language) between grades 5 and 6 and that children already learn the names for the grammatical categories (noun, verb, adjective, etc) in grade 4. This explicit linguistic training could be affecting the scores on *Parte 2* by participants in these grades.

On the other hand, parts 3 and 4 present from small to medium Cohen's *d* values consistently across grades. Also, the direction of the difference in the means is represented by a small increase or a negligible change across grades. Therefore, on the basis of the dimensions of this increase, these parts do not seem to work differently across grades or to be affected by any factor that has to do mainly with the test takers' age or linguistic training.

4.2.2. The MLAT-EC in relation to the participants' age

The descriptive results on the MLAT-ES are not normally distributed and neither are the results on the MLAT-EC obtained by group 2. The boxplots of the performance across grades in the different parts of the MLAT-EC clearly show that the variance in the score distribution is not stable, either. These represent graphically the descriptive results in section 3.5.3.3.

Figure 4.6. Boxplot of MLAT-EC *Part 1 Paraules ocultes*

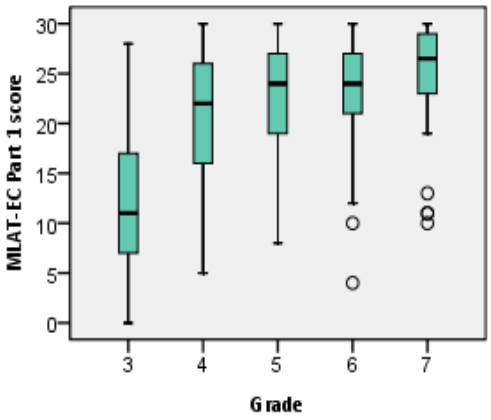


Figure 4.7. Boxplot of MLAT-EC *Part 2 Paraules que es corresponen*

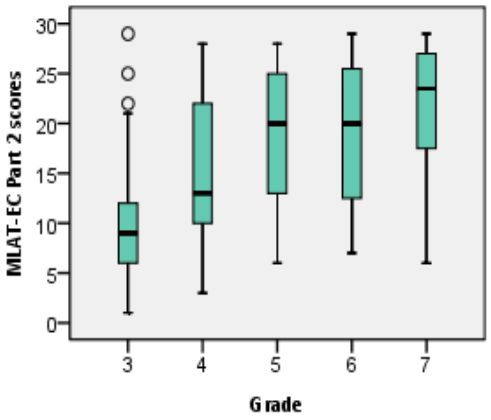


Figure 4.8. Boxplot of MLAT-EC *Part 3 Paraules que rimen*

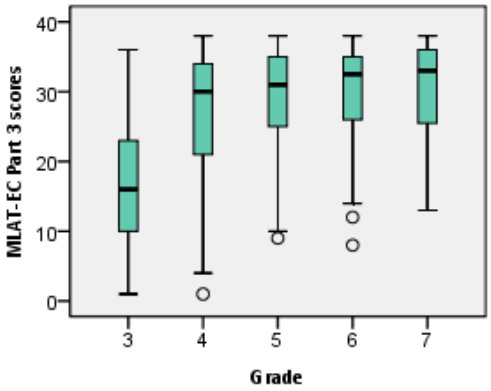


Figure 4.9. Boxplot of MLAT-EC *Part 4 Aprenguem números*

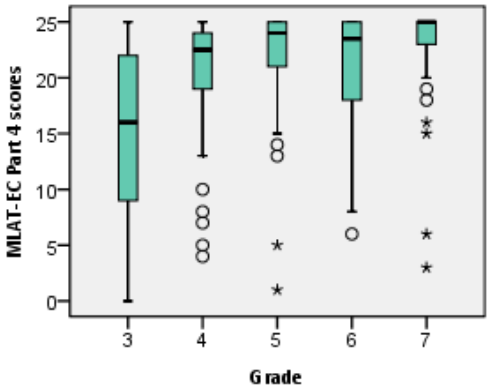
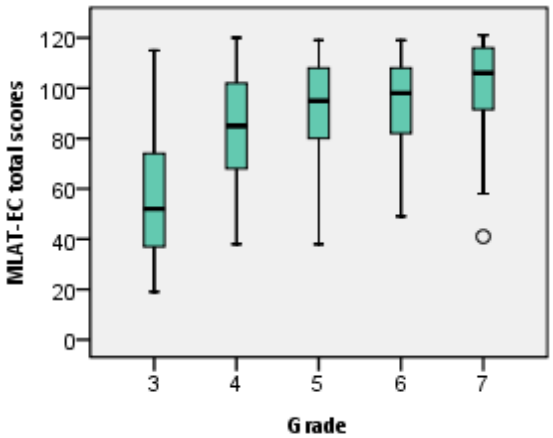


Figure 4.10. Boxplot of MLAT-EC Raw total score



On the basis of group 2's descriptive results (see section 3.5.3.3), it was also expected that the group as a whole would present significant differences in the distribution of scores across grades in all parts of the MLAT-EC. Table 4.7 shows the

Levene's statistics comparing the contiguous grades and the whole group. Significant values appear in bold.

Table 4.7. Levene's test for the MLAT-EC

Grades	Grades 3 & 4		Grades 4 & 5		Grades 5 & 6		Grades 6 & 7		All	
MLAT-EC	Levene Statistic	Sig.	Levene Statistic	Sig.	Levene Statistic	Sig.	Levene Statistic	Sig.	Levene Statistic	Sig.
Part 1	59.700	**.000**	12.302	**.001**	.1.864	.174	6.218	**.014**	24.774	**.000**
Part 2	2.042	.155	.436	.510	.000	.984	4.488	**.035**	2.5624	**.039**
Part 3	12.722	**.000**	5.636	**.019**	.612	.435	3.334	.069	8.897	**.000**
Part 4	25.691	**.000**	.120	.729	.758	**.**385	8.065	**.005**	13.468	**.000**
Total	21.052	**.000**	3.290	.071	.609	.436	3.190	.076	15.011	**.000**

The variances in *Part 1 Paraules ocultes* are significantly different between all grades except for grades 5 and 6, $F(1, 179)=.174$, *ns*, as scores in these two grades tend to concentrate below the median (see Figure 4.6 above).

In scores on *Part 2 Paraules que es corresponen*, in contrast with the MLAT-ES, the variance distribution is significantly different between grades 6 and 7, $F(1, 189)=.035$, $p<.05$, one course later than it happened in the MLAT-ES. Surprisingly, Levene's statistic is not significant between grades 3 and 4, $F(1, 192)=.155$, *ns*, although, according to the plot, the distribution of the scores in the box is fairly different. This is due to the fact that some outliers are found who scored much above the mean on this part.

The variance distribution of scores on *Part 3 Paraules que rimen* is significantly different between grades 3 and 4 and between grades 4 and 5, as the distribution of scores concentrates above the median in grade 3 but not in grade 4. It does in the rest of grades, but with different distribution of scores between grades 4 and 5. Hence the significance value $F(1, 178)=.019$ $p<.05$.

The spread of the scores on *Part 4 Aprenguem números* is larger in the MLAT-EC than it is in the MLAT-ES. Also, the significant values in the MLAT-EC do not totally coincide with those in the MLAT-ES, as they are found between grades 3 and 4, $F(1, 192)=.000$ $p<.001$, and between grades 6 and 7 $F(1, 192)=.005$ $p<.05$, while in the MLAT-ES they were between grades 3 and 4 but also between grades 5 and 6.

A Mann-Whitney U test was run in order to check if the differences in the means across grades are significant. Since it was clear from the descriptive results that the upper the grade, the higher the score, the one-tailed probability has been calculated, as it is used when the direction of the change in the means is known. This value is obtained by dividing the two-tailed value provided by SPSS by two. The tables below

show the differences between means across grades in each part of the MLAT-EC as well as in the total score.

Table 4.8. Mann-Whitney U test between grades 3 and 4 on the MLAT-EC

Grades 3 and 4 N=119	MLAT-EC Part 1	MLAT-EC Part 2	MLAT-EC Part 3	MLAT-EC Part 4	MLAT-EC Total
Mann-Whitney U	771.000	979.500	762.5500	1030.500	651.000
Z	-5.302	-4.200	-5.348	-3.934	-5.938
Asymp. Sig. (2-tailed)	**.000**	**.000**	**.000**	**.000**	**.000**
Asymp. Sig. (1-tailed)	**.000**	**.000**	**.000**	**.000**	**.000**
r (effect size)	-.49 medium / large	-.39 medium	**-.53 large**	-.41 medium	**-.55 large**

If the differences in the means on the MLAT-ES between grades 3 and 4 were significant in all parts but *Parte 4*, this tendency towards a significant difference in the values of these two grades is even more salient on the MLAT-EC, as shown in Table 4.8 above. Not only does grade 4 perform significantly higher than grade 3, but also the effect size of these higher scores is from medium to large in all parts, including, of course, the total score. Therefore, the MLAT-EC seems to function markedly differently in these two grades.

Table 4.9. Mann-Whitney U test between grade 4 and 5 on the MLAT-EC

Grades 4 and 5 N=123	MLAT-EC Part 1	MLAT-EC Part 2	MLAT-EC Part 3	MLAT-EC Part 4	MLAT-EC Total
Mann-Whitney U	1531.000	1327.500	1705.000	1505.500	1419.500
Z	-1.825	-2.855	-.943	-1.985	-2.386
Asymp. Sig. (2-tailed)	.068	**.004**	.343	**.047**	**.017**
Asymp. Sig. (1-tailed)	**.034**	**.002**	.172	**.024**	**.009**
r (effect size)	-.17 small	-.26 small	-.09 small	-.18 small	-.22 small

Significant differences are also found between the means of grades 4 and 5, although not in all parts (see Table 4.9). Grade 5 (*Mdn*=24.00) performs only better than grade 4 (*Mdn*=22.00) at the one-tailed significance value in *Part 1, U*=1531.00, *p*<.05, *r*=-.17. 5-graders (*Mdn*=20.00). They also perform much better in *Part 2* than do 4-graders (*Mdn*=13.00), *U*=1327.50, *p*<.05, *r*=-.26 and fairly better in *Part 4*, *U*=1505.50, *p*<.05, *r*=-.18. Naturally, grade 5 (*Mdn*=95.00) outperforms grade 4 (*Mdn*=85.00) in the total score, although the effect size of this increase is only small, *U*=1419.50, *p*<.05, *r*=-.22.

Table 4.10. Mann-Whitney U test between grades 5 and 6 on the MLAT-EC

Grades 5 and 6 N=121	MLAT-EC Part 1	MLAT-EC Part 2	MLAT-EC Part 3	MLAT-EC Part 4	MLAT-EC Total
Mann-Whitney U	1690.000	1816.000	1582.500	1653.500	1800.000
Z	-.728	-.073	-1.289	-.938	-.156
Asymp. Sig. (2-tailed)	.467	.942	.197	.348	.876
Asymp. Sig. (1-tailed)	.234	.471	.099	.174	.438
r (effect size)	-.07 small	-.001 small	-.12 small	-.09 small	-.01 small

Table 4.10 shows that the almost lack of difference that was subtly suggested in the performance by the test takers in grades 5 and 6 on the MLAT-ES appears more clearly in the Catalan version. While on the MLAT-ES significant values were only found at the one-tailed level in all parts and only at the two-tailed level in *Parte 2,* these significant differences between grades 5 and 6 disappear in the MLAT-EC. Therefore, it seems that grades 5 and 6 perform very similarly in this version of the test.

Table 4.11. Mann-Whitney U test between grades 6 and 7 on the MLAT-EC

Grades 6 and 7 N=124	MLAT-EC Part 1	MLAT-EC Part 2	MLAT-EC Part 3	MLAT-EC Part 4	MLAT-EC Total
Mann-Whitney U	1530.000	1586.000	1919.500	1305.000	1462.500
Z	-1.957	-1.673	-.003	-3.253	-2.289
Asymp. Sig. (2-tailed)	.050	.094	.998	**.001**	**.022**
Asymp. Sig. (1-tailed)	**.025**	**.047**	.499	**.001**	**.011**
r (effect size)	-.18 small	.15 small	.01 small	-.29 medium	-.21 small

Grades 6 and 7 also show a very similar performance on the MLAT-EC, as it is only in the one-tailed values that significant differences appear in all parts except for *Part 4* (see Table 4.11). This also makes the total score be significantly different at the two-tailed level, with a small-medium effect size, U=1462.50, p<.05, r=-.21. Actually, the significantly different value in *Part 4* is due to the fact that grade 6 (*Mdn*=23.50) performed worse than grade 5 (*Mdn*=24.00) on this part. Therefore, the significant differences in the results between grades 6 and 7, both in the total score and in *Part 4*, are not to be generalised. This is so because the general tendency is for grade 6 to perform better than grade 5 and, therefore, the distance in the means of grades 6 and 7 should normally not be so large.

In general terms, it can be concluded that grade 3 is the grade that performs most significantly differently in all parts of the MLAT-EC. It can also be concluded that significant differences disappear gradually across grades except for the peculiar

performance of 6-graders in *Part 4*, which makes the differences between grades 6 and 7 be larger than they were in the same part of the MLAT-ES.

Group 2's performance on the different parts of the MLAT-EC is fairly similar, though not exactly the same, to that of the MLAT-ES as for the size of change in the means and the grades where change happens, as can be seen in the following table.

Table 4.12. Differences in the means on the part and total scores of MLAT-EC across grades

MLAT-EC 1 *PARAULES OCULTES*	N	Cohen's *d*	Effect size	Percentage change	Direction
GRADE 3 – GRADE 4	119	**1.16**	**very large**	**68**	**very large increase**
GRADE 4 – GRADE 5	123	0.36	small	10	small increase
GRADE 5 – GRADE 6	121	0.13	negligible	3	negligible change
GRADE 6 – GRADE 7	124	0.33	small	7	small increase
MLAT-EC 2 *PARAULES QUE ES CORRESPONEN*	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	119	0.51	medium	**32**	**large increase**
GRADE 4 – GRADE 5	123	0.56	medium	25	medium increase
GRADE 5 – GRADE 6	121	0.02	negligible	-1	negligible change
GRADE 6 – GRADE 7	124	0.30	small	11	small increase
MLAT-EC 3 *PARAULES QUE RIMEN*	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	119	**1.10**	**very large**	**60**	**very large increase**
GRADE 4 – GRADE 5	123	0.24	small	7	small increase
GRADE 5 – GRADE 6	121	0.19	small	5	small increase
GRADE 6 – GRADE 7	124	0.01	negligible	0	negligible change
MLAT-EC 4 *APRENGUEM NÚMEROS*	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	119	**0.80**	**large**	**33**	**large increase**
GRADE 4 – GRADE 5	123	0.28	small	7	small increase
GRADE 5 – GRADE 6	121	0.18	small	-4	negligible change
GRADE 6 – GRADE 7	124	0.48	medium	10	small increase
MLAT-EC RAW TOTAL SCORE	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 – GRADE 4	119	**1.30**	**very large**	**52**	**very large increase**
GRADE 4 – GRADE 5	123	0.46	medium	11	small increase
GRADE 5 – GRADE 6	121	0.06	negligible	1	negligible change
GRADE 6 – GRADE 7	124	0.32	small	**6**	small increase

Grade 4 consistently obtains higher means than grade 3 with a large or very large increase in all parts. In the MLAT-ES this only happened in *Parte 1* and in *Parte 2*. The fact that 4-graders perform significantly better than 3-graders in *Part 3* (60%, Cohen's *d*=1.10) could be explained by the fact that rhymes in Catalan are more difficult to detect from written linguistic input than they are in Spanish, as the correspondence between letter and sound is not so exact as it is in Spanish, especially as for vowels, which were not an issue in the MLAT-ES. Other factors that could

explain the lower results of 3-graders could be the fact that they have less experience with written language than 4-graders. The cognitive developmental stage at which they are could also be affecting the results. As happened in the other parts, grade 4 also performs much better than does grade 3 in *Part 4 Aprenguem números*, while in the rest of grades, the increase in the means is minimum. This is somewhat surprising, as both *Part 4 Aprenguem números* and *Parte 4 Aprendamos números* are virtually identical and, group 1 showed a minimum increase in the means on this part across grades.

In *Parte 1* of the MLAT-ES, the evolution of scores was a bit irregular across grades, although with a clear tendency to increase. The increase was medium between grades 4 and 5, and between grades 6 and 7. In *Part 1* of the MLAT-EC the scores also increase, but this increase is not so large from grades 4 to 7 and does not have large effect sizes either.

Although with slight changes as compared to the MLAT-ES, *Part 2* also presents the cohort as two different groups. That is, 3-graders obtain much lower scores than 4-graders (32% increase for the 4-graders, Cohen's *d*=0.51). A medium increase with a medium effect size is found between grade 4 and 5 (25%, Cohen's *d*=0.56), followed by negligible and small changes with small and negligible effect sizes between grades 5 and 6 and grades 6 and 7. The different patterns regarding the increase as compared to the MLAT-ES are due to the fact that the means in grades 3 and 4 on the MLAT-ES are higher than they are on the MLAT-EC, while the mean of grade 6 is lower on the MLAT-ES than on the MLAT-EC. Broadly speaking, though, it can be said that the great changes in the means in part 2 of both tests coincide with the introduction of formal instruction of grammar in the L1s at school, as established in the *Curriculum d'Educació Primària* (Curriculum of Primary Education, Catalonia, Servei d'Educació Curricular, 2009).

The changes and effect sizes in the means on the raw total score confirm that the performance of 3-graders on the MLAT-EC is very different from the performance of the rest of grades. Notice that not only is the increase in the means of the total scores between grades 3 and 4 very large (52%), but its effect size is also the largest of all, as Cohen's *d* value is 1.30. A similar pattern is consistently found in all the parts separately.

4.2.3. The participants' age in the MLAT-E norming study and in the MLAT-ES and the MLAT-EC administered in Catalonia

As mentioned before, the question of whether the increase in the means across grades is significant or not can be measured with Cohen's *d*, even without having access to the raw data. This index will be useful to see whether the test takers of this study, who took the MLAT-ES and the MLAT-EC, performed in the same way as the test takers in the official norming studies of the MLAT-ES and the MLAT-E. In this way, it will also be possible to attach the contrast between the performances of certain grades to factors other than the language in which the test is written. The significance of the increase across grades is not provided in the *Manuals*, yet it is possible to estimate it without necessarily having access to the raw data, as Cohen's *d* can be calculated just by knowing the N, the mean and the standard deviation, which do appear in the *Manual*. With the information available, it is also possible to estimate the relative size of percentage change.

It was shown in the norms that appear in section 3.5.5 that in none of the tests do the scores increase gradually across all grades. Instead, the increase is sharper between grades 3 and 4 than between grades 4 and 5, and grades 5 and 6 for both boys and girls in the MLAT-E and for the whole population in the MLAT-ES and the MLAT-EC. Also, the results do not differ much in the upper grades either on the MLAT-ES or the MLAT-EC, especially between grades 6 and 7.

As for the MLAT-E, the *Manual* presents the data dividing the population according to the sex variable. The descriptive results and further analysis of the MLAT-E are run using the data based on the results from four schools, as the data of the designated percentiles (used in section 3.5.5) does not appear broken down per grades. In order to compare the population of this study with the one of the MLAT-E, it is also necessary to provide the data of the bilingual Catalan/Spanish participants according to this variable. The descriptive results of the MLAT-E as well as of the MLAT-ES and the MLAT-EC administered in Catalonia were already presented in sections 2.3.1.3, 3.5.2.3 and 3.5.3.3 respectively, but these did not include the distinction according to sex. The sex variable appears here only as a convention to distribute the data in the same way as it appears in the MLAT-E norming study and, therefore, will not be explored in depth in this section, but in section 4.3.

In Table 4.13, the statistical information of the MLAT-ES norming study is not provided, as it does not appear divided according to the sex variable in the *Manual*, but it will appear later on in this section. The data of the performance of grade 7 on the

MLAT-E is not provided either, as the *Manual* includes the data only for grades from 3 to 6.

Table 4.13. Descriptive raw scores of part 1 on the MLAT-E, MLAT-ES and MLAT-EC according to sex

		Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part		MLAT-E Part 1 Hidden words (30 items)			MLAT-ES *Parte 1 Palabras ocultas* (30 items)			MLAT-EC *Part 1 Paraules ocultes* (30 items)		
Grade	Gender	N	Mean	SD	N	Mean	SD	N	Mean	SD
3	Boys	112	14.7	6.3	26	12.65	6.216	27	12.48	7.753
	Girls	105	14.7	5.3	40	12.92	7.241	30	12.27	8.399
4	Boys	113	19.8	5.3	41	21.39	5.826	26	21.77	5.421
	Girls	88	21.3	4.5	34	17.12	7.256	36	19.83	7.061
5	Boys	112	22.6	4.6	25	24.24	4.994	33	22.76	5.268
	Girls	109	22.4	4.2	32	21.91	5.432	28	22.75	5.681
6	Boys	167	23.8	4.0	32	24.03	6.008	28	24.29	5.276
	Girls	174	24.7	3.8	28	21.11	7.104	32	22.75	5.623
7	Boys	-	-	-	36	27.06	2.735	31	24.61	4.958
	Girls	-	-	-	31	26.42	3.713	33	25.67	4.695

The descriptive results of part 1 consistently show small or negligible changes in the means of grades from 4 to 7, the smallest ones being no different at all in the case of girls in grades 5 and 6 who took the MLAT-EC. In some cases, instead of an increase across grades, negligible decreases can be found, for instance, in the girls from grades 5 to 6 of group 1 and in the 6- and 7-grade boys of group 2. There is, though, a noticeable increase in all tests from grades 3 to 4. This increase, as shown in Table 4.14, is large or very large and, except for the girls in group 1, the size effect as measured by Cohen's *d* is also considerable. Other subgroups also show large effect sizes and significant two-tailed values, although the increase these effect sizes are linked to are mostly small. The girls in grade 7 of group 1 do show a medium increase as compared to the results obtained by 6-grade girls, but this increase is probably a consequence of the 4% decrease of 6-grade girls as compared to 5-graders in the same group.

Table 4.14. Differences in the mean scores on part 1 on the MLAT-E, MLAT-ES and MLAT-EC across grades according to sex

Grades and test versions	S.	N	M-W. U	Z	Asymp. Sig (2-t.)	r (eff. s.)	C.'s *d*	Eff. size (Cohen)	% chnge	Direct.
GRADE 3 – GRADE 4 MLAT-E PART 1	**B**	225	-	-	-	-	**0.88**	**large**	**35**	**large increase**
	G	193	-	-	-	-	**1.34**	**very large**	**45**	**large increase**
GRADE 3 – GRADE 4 MLAT-ES *PARTE 1*	**B**	67	165.000	-4.742	**.000**	**-.58**	**1.48**	**huge**	**69**	**v. large increase**
	G	74	454.000	-2.455	**.014**	-.29	0.59	medium	**33**	**large increase**
GRADE 3 – GRADE 4 MLAT-EC *PART 1*	**B**	53	121.500	-4.089	**.000**	**-.56**	**1.41**	**very large**	**74**	**v. large increase**
	G	66	271.500	-3.462	**.001**	-.43	**1.00**	**large**	**62**	**v. large increase**
GRADE 4 – GRADE 5 MLAT-E PART 1	**B**	225	-	-	-	-	0.57	medium	14	small increase
	G	197	-	-	-	-	0.25	small	5	small increase
GRADE 4 – GRADE 5 MLAT-ES *PARTE 1*	**B**	66	362.000	-1.997	**.046**	-.24	0.52	medium	13	small increase
	G	66	333.500	-2.705	**.007**	-.33	**0.76**	**large**	28	medium increase
GRADE 4 – GRADE 5 MLAT-EC *PART 1*	**B**	59	375.500	-.819	.413	-.11	0.19	small	5	small increase
	G	64	388.000	-1.574	.116	-.20	0.46	medium	15	medium increase
GRADE 5 – GRADE 6 MLAT-E PART 1	**B**	279	-	-	-	-	0.28	small	5	small increase
	G	283	-	-	-	-	0.58	medium	10	small increase
GRADE 5 – GRADE 6 MLAT-ES *PARTE 1*	**B**	57	380.500	-.315	.753	-.04	0.04	negligible	-1	negligible change
	G	60	443.000	-.074	.941	-.01	0.13	negligible	-4	negligible change
GRADE 5 – GRADE 6 MLAT-EC *PART 1*	**B**	61	372.000	-1.307	.191	-.17	0.30	small	7	small increase
	G	60	434.000	-.208	.835	-.03	0	negligible	0	negligible change
GRADE 6 – GRADE 7 MLAT-ES *PARTE 1*	**B**	68	395.500	-2.235	**.025**	-.27	0.67	medium	13	small increase
	G	59	239.500	-2.978	**.003**	-.39	**0.97**	**large**	25	medium increase
GRADE 6 – GRADE 7 MLAT-EC *PART 1*	**B**	59	421.500	-.191	.849	-.03	0.06	negligible	1	negligible change
	G	65	333.000	-2.571	**.010**	-.32	0.57	medium	13	small change

S.: sex; B: boys; G: girls; M-W. U: Mann-Whitney U; Asymp. Sig (2-t.): asymptotic significance, 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d*: Cohen's *d*; % chnge: percentage change; Direct: direction

From these analyses, it is clear that no matter the language, part 1 of the elementary version of the MLAT has a different functioning for 3-graders as compared to the other grades. In addition, the performance by 5-, 6- and 7-graders is also very

similar regardless of the language of the test. Therefore, both the population of the MLAT-E study and the one of this piece of research perform very similarly on the aptitude tests analysed here.

The oscillations in the descriptive results of part 2 (see Table 4.15 below) are very similar to those of part 1 in all senses. As in part 1, scores in grade 3 are noticeably lower than in grade 4 and they hardly change in the upper grades.

Table 4.15. Descriptive raw scores of part 2 on the MLAT-E, MLAT-ES and MLAT-EC according to sex

		Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part		MLAT-E Part 2 Matching words (30 items)			MLAT-ES *Parte 2 Palabras que se corresponden* (30 items)			MLAT-EC *Part 2 Paraules que es corresponen* (29 items)		
Grade	Gender	N	Mean	SD	N	Mean	SD	N	Mean	SD
3	Boys	112	11.2	5.4	26	12.00	5.463	27	9.56	5.323
	Girls	105	11.4	5.5	40	12.95	6.164	30	11.17	6.581
4	Boys	113	13.5	5.8	41	17.88	6.750	26	16.35	7.456
	Girls	88	16.8	5.6	34	17.15	7.809	36	14.97	6.835
5	Boys	112	17.0	6.0	25	20.64	6.999	33	19.55	7.146
	Girls	109	18.2	6.2	32	17.06	8.277	28	19.32	6.864
6	Boys	167	19.2	5.6	32	22.41	7.116	28	19.36	7.098
	Girls	174	21.4	4.8	28	24.39	5.405	32	19.22	7.142
7	Boys	-	-	-	36	24.17	5.988	31	19.45	7.084
	Girls	-	-	-	31	23.77	6.386	33	23.06	6.139

Table 4.16 shows that, like in part 1, in part 2 large and very large increases are found between grades 3 and 4, with medium and large effect sizes. In contrast with part 1, the increase in the scores of grade 5 as compared to grade 4 is mostly medium while small and negligible changes are found only between grades 5 and 6 on the MLAT-EC and in all tests between grades 6 and 7. The increase in the means scores from girls in grade 6 is sharper than in the other tests due to the fact that girls in grade 5 scored lower than 4-grade girls.

Table 4.16. Differences in the mean scores on part 2 on the MLAT-E, MLAT-ES and MLAT-EC across grades according to sex

Grades and test versions	S.	N	M-W. U	Z	Asymp. Sig (2-t.)	r (eff. s.)	C.'s *d*	Eff. size (Cohen)	% chnge	Direct.
GRADE 3 – GRADE 4 MLAT-E PART 2	B	225	-	-	-	-	0.41	medium	21	medium increase
	G	193	-	-	-	-	**0.98**	**large**	**47**	**large increase**
GRADE 3 – GRADE 4 MLAT-ES *PARTE 2*	B	67	264.000	-3.466	**.001**	-.42	**0.95**	**large**	**49**	**large increase**
	G	74	463.000	-2.357	**.018**	-.27	0.61	medium	**32**	**large increase**
GRADE 3 – GRADE 4 MLAT-EC *PART 2*	B	53	160.500	-3.399	**.001**	-.47	**1.07**	**large**	**71**	**v.large increase**
	G	66	347.500	-2.484	**.013**	-.31	0.57	medium	**34**	**large increase**
GRADE 4 – GRADE 5 MLAT-E PART 2	B	225	-	-	-	-	0.60	medium	26	medium increase
	G	197	-	-	-	-	0.24	small	8	small increase
GRADE 4 – GRADE 5 MLAT-ES *PARTE 2*	B	66	394.500	-1.563	.118	-.19	0.41	medium	15	medium increase
	G	66	542.000	-.026	.979	-.003	0.01	negligible	-1	negligible change
GRADE 4 – GRADE 5 MLAT-EC *PART 2*	B	59	324.000	-1.606	.108	-.21	0.45	medium	20	medium increase
	G	64	324.500	-2.434	**.015**	-.30	0.65	medium	29	medium increase
GRADE 5 – GRADE 6 MLAT-E PART 2	B	279	-	-	-	-	0.38	small	13	small increase
	G	283	-	-	-	-	0.60	medium	18	medium increase
GRADE 5 – GRADE 6 MLAT-ES *PARTE 2*	B	57	351.500	-.782	.434	-.10	0.26	small	9	small change
	G	60	214.000	-3.475	**.001**	-.45	**1.05**	**large**	**43**	**large increase**
GRADE 5 – GRADE 6 MLAT-EC *PART 2*	B	61	450.500	-.167	.868	-.11	0.03	negligible	-1	negligible change
	G	60	447.000	-.015	.988	-.001	0.01	negligible	-1	negligible change
GRADE 6 – GRADE 7 MLAT-ES *PARTE 2*	B	68	491.000	-1.051	.293	-.13	0.27	small	8	small increase
	G	59	427.500	-.099	.921	-.01	0.11	negligible	-3	negligible change
GRADE 6 – GRADE 7 MLAT-EC *PART 2*	B	59	429.500	-.068	.945	-.01	0.01	negligible	0	negligible change
	G	65	356.000	-2.267	**.023**	-.28	0.59	medium	20	medium increase

S.: sex; B: boys; G: girls; M-W. U: Mann-Whitney U; Asymp. Sig (2-t.): asymptotic significance, 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d* : Cohen's *d*; % chnge: percentage change; Direct: direction, v. large: very large

But for the large increase in the scores of girls in group 1, generally speaking, it can be concluded that part 2 of these three versions of the test function in a very

similar way, no matter the language in which it is, in the sense that perhaps three groups could be outlined from the descriptive results and the percentage changes in this part. First of all, grade 3 consistently obtains much lower results than grade 4, as opposed to the smaller differences in the means on the other grades compared consecutively, that is, compared to their immediate upper grades. Secondly, medium increases are observed between grade 4 and grade 5, while these changes are small or even in the opposite direction between the rest of grades. At this point another line could be drawn between grades 4 and 5, while grades 5, 6 and 7 could form a group themselves, as the test takers' performances are very similar across these grades. Finally, it could also be pointed out that even though the languages of the tests differ in some grammatical aspects, these do not seem to affect the performance of the populations compared.

In the description of the tests in previous sections (see 2.3.1.1, 2.3.2.1 and 3.4), part 3 has been described as the part which is most language dependent. So far, part 1 and part 2 have shown similar score patterns in any of the languages at work (English, Spanish and Catalan), but this may not be the case for part 3, as the intrinsic differences between the tests may show as well in different score patterns across grades. In table 4.17, it can be seen that the means on Part 3 of the MLAT-E are higher because the English version of this test has seven items more than the Spanish and Catalan versions and because raw part scores, not percentages, are used in these tables. However, the Spanish and Catalan versions have the same number of items. Nevertheless, not only do group-2 test takers in grade 3 obtain lower scores than group-1 test takers in the same grade, but the difference between their scores and those obtained by 4-graders is larger than the difference in the scores of their counterparts.

Table 4.17. Descriptive raw scores of part 3 on the MLAT-E, MLAT-ES and MLAT-EC according to sex

		Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part		MLAT-E Part 3 Finding rhymes (45 items)			MLAT-ES *Parte 3 Palabras que riman* (38 items)			MLAT-EC *Part 3 Paraules que rimen* (38 items)		
Grade	Gender	N	Mean	SD	N	Mean	SD	N	Mean	SD
3	Boys	112	31.0	10.9	26	21.69	8.817	27	17.04	9.545
	Girls	105	31.9	10.8	40	22.40	9.831	30	16.77	9.457
4	Boys	113	36.0	9.7	41	27.07	8.577	26	28.00	7.715
	Girls	88	39.1	6.6	34	23.18	10.080	36	26.25	10.058
5	Boys	112	39.8	7.1	25	31.48	6.765	33	29.97	6.844
	Girls	109	41.1	5.3	32	26.59	9.725	28	27.79	7.904
6	Boys	167	41.2	5.5	32	31.28	7.900	28	30.54	6.438
	Girls	174	42.2	6.0	28	31.07	7.760	32	30.12	7.893
7	Boys	-	-	-	36	32.64	5.958	31	29.13	7.251
	Girls	-	-	-	31	33.65	4.550	33	31.36	6.772

The increase expressed in percentages in Table 4.18 confirms that while the Spanish and English versions of the MLAT-E are slightly more difficult for 3-graders than for 4-graders, the Catalan version appears to be much more difficult for 3-graders than it is for 4-graders, with large and very large effect sizes for boys and girls (1.28 and 0.98 respectively). The language version of the test could be playing a part in the differences of the scores of 3-graders, as this contrast is especially salient in group 2, who took the MLAT-EC, yet it does not seem to affect the results of the other grades in any of the versions of this part of the aptitude test, regardless of the test takers' sex. Therefore, what makes the performance of 3-graders different from that of the rest of grades is probably something else in addition to the language in which the test is. This could be, for instance, due to cognitive developmental factors or due to previous metalinguistic experience with this kind of exercises. Summarising, two well-defined groups can be outlined from the percentage changes and effect sizes on this part as grade 3 seems to react very differently from the other grades when facing this part of the test.

Table 4.18. Differences in the mean scores on part 3 on the MLAT-E, MLAT-ES and MLAT-EC across grades according to sex

Grades and test versions	S.	N	M-W U	Z	Asymp. Sig (2-t.)	*r* (eff. s.)	C.'s *d*	Eff. size (Cohen)	% chnge	Direct.
GRADE 3 – GRADE 4 MLAT-E PART 3	B	225	-	-	-	-	0.49	medium	16	medium increase
	G	193	-	-	-	-	**0.79**	**large**	23	medium increase
GRADE 3 – GRADE 4 MLAT-ES *PARTE 3*	B	67	345.500	-2.415	**.016**	-.29	0.63	medium	25	medium increase
	G	74	633.000	-.511	.610	-.06	0.08	negligible	3	negligible change
GRADE 3 – GRADE 4 MLAT-EC *PART 3*	B	53	127.000	-3.993	**.000**	**-.55**	**1.28**	**very large**	**64**	**v. large increase**
	G	66	262.000	-3.583	**.000**	-.44	**0.98**	**large**	**57**	**v. large increase**
GRADE 4 – GRADE 5 MLAT-E PART 3	B	225	-	-	-	-	0.45	medium	11	small increase
	G	197	-	-	-	-	0.34	small	5	small increase
GRADE 4 – GRADE 5 MLAT-ES *PARTE 3*	B	66	347.000	-2.193	**.028**	-.27	0.56	medium	16	medium increase
	G	66	434.500	-1.407	.160	-.17	0.35	small	15	small increase
GRADE 4 – GRADE 5 MLAT-EC *PART 3*	B	59	376.500	-.804	.422	-.10	0.28	small	7	small increase
	G	64	480.000	-.325	.745	-.04	0.17	small	6	small increase
GRADE 5 – GRADE 6 MLAT-E PART 3	B	279	-	-	-	-	0.23	small	4	negligible change
	G	283	-	-	-	-	0.19	small	3	negligible change
GRADE 5 – GRADE 6 MLAT-ES *PARTE 3*	B	57	380.500	-.315	.753	-.04	0.03	negligible	-1	negligible change
	G	60	302.000	-2.170	**.030**	-.28	0.51	medium	17	medium increase
GRADE 5 – GRADE 6 MLAT-EC *PART 3*	B	61	430.500	-.457	.648	-.06	0.09	negligible	2	negligible change
	G	60	358.500	-1.331	.183	-.17	0.30	small	8	small increase
GRADE 6 – GRADE 7 MLAT-ES *PARTE 3*	B	68	550.500	-.315	.752	-.04	0.20	small	4	negligible change
	G	59	398.000	-.550	.582	-.07	0.42	medium	8	small increase
GRADE 6 – GRADE 7 MLAT-EC *PART 3*	B	59	395.000	-.593	.553	-.08	0.21	small	-5	small decrease
	G	65	493.500	-.455	.649	-.06	0.17	small	4	negligible change

S.: sex; B: boys; G: girls; M-W U: Mann-Whitney U; Asymp. Sig (2-t.): asymptotic significance, 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d*: Cohen's *d*; % chnge: percentage change; Direct: direction, v. large: very large

In Table 4.19, it can be seen that the means in part 4 of the English version of the elementary MLAT are much lower for all grades and sexes than they are for MLAT-

ES and the MLAT-EC. As already commented on in previous sections, the similarity of the numbers created for this test with the numbers in the test takers' L1s and the fact that number 30 is called "*rasca*", which is a word with content both in Spanish and Catalan, could have rendered the Spanish and Catalan versions of the test easier than the original in English.

Table 4.19. Descriptive raw scores of part 4 on the MLAT-E, MLAT-ES and MLAT-EC according to sex

		Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part		MLAT-E Part 4 Number Learning (25 items)			MLAT-ES *Parte 4 Aprendamos números* (25 items)			MLAT-EC *Part 4 Aprenguem números* (25 items)		
Grade	Gender	N	Mean	SD	N	Mean	SD	N	Mean	SD
3	Boys	112	7.2	5.9	26	17.58	6.113	27	16.70	7.488
	Girls	105	7.9	6.4	40	17.58	7.267	30	14.33	7.572
4	Boys	113	10.7	7.3	41	19.66	5.803	26	20.88	5.046
	Girls	88	14.2	7.4	34	19.38	5.560	36	20.39	5.458
5	Boys	112	14.8	7.3	25	23.40	2.327	33	22.58	4.070
	Girls	109	15.2	7.2	32	18.81	6.433	28	21.25	5.254
6	Boys	167	16.7	7.0	32	23.19	3.063	28	21.36	5.079
	Girls	174	16.7	6.6	28	22.25	3.787	32	20.94	4.529
7	Boys	-	-	-	36	22.86	3.735	31	22.74	5.112
	Girls	-	-	-	31	22.81	4.629	33	23.64	2.655

In addition, the evolution of scores across groups is also different depending on the language the test is in. Rather than attach this disparity to the version of the test, other factors could be affecting the results, such as previous training when using the type of memory tapped in this test or the familiarity with this type of task. Also, while substantial changes are observed in the scores obtained by grades 3 and 4 and grades 4 and 5 on the MLAT-E, these changes are observed in the same grades but with a much lower impact and in a more irregular way depending on the test takers' sex (see Table 4.20). For instance, in group 1 there is a small increase in the means of 4-graders as compared to the 3-graders' in both the girls' and boys' group. Group-1 boys in grade 5 are the only ones who present a medium increase in their mean score (19%) with a large effect size (Cohen's d=0.79) as compared to the boys in grade 4 of the same group. In contrast, the girls in the same grade actually perform worse than 4-graders. On the other hand, 4-grade girls in group 2 obtained a much higher mean score with a large effect size (Cohen's *d*=.95) than 3-graders and, as it happens on the MLAT-ES, the mean score of 5-grader girls presents a negligible change, though it does not decrease as compared to that of 4-grade girls in group 2. In this same group

2, the boys' mean scores increase considerably in grade 5 as compared to the mean scores of male test takers in grade 4.

Table 4.20. Differences in the mean scores on part 4 of the MLAT-E, MLAT-ES and MLAT-EC across grades according to sex

Grades and test versions	S.	N	M-W U	Z	Asymp. Sig (2-t.)	*r* (eff. s.)	C.'s *d*	Eff. size (Cohen)	% chnge	Direct.
GRADE 3 – GRADE 4 MLAT-E PART 4	**B**	225	-	-	-	-	0.53	medium	**49**	**large increase**
	G	193	-	-	-	-	**0.96**	**large**	**80**	**huge increase**
GRADE 3 – GRADE 4 MLAT-ES *PARTE 4*	**B**	67	418.500	-1.480	.139	-.18	0.36	small	12	small increase
	G	74	602.500	-.846	.398	-.10	0.28	small	10	small increase
GRADE 3 – GRADE 4 MLAT-EC *PART 4*	**B**	53	222.500	-2.300	**.021**	-.32	0.66	medium	25	medium increase
	G	66	293.000	-3.193	**.001**	-.39	**0.95**	**large**	**42**	**large increase**
GRADE 4 – GRADE 5 MLAT-E PART 4	**B**	225	-	-	-	-	0.56	medium	**38**	**large increase**
	G	197	-	-	-	-	0.14	negligible	7	small increase
GRADE 4 – GRADE 5 MLAT-ES *PARTE 4*	**B**	66	268.500	-3.272	**.001**	-.40	**0.79**	**large**	19	medium increase
	G	66	522.000	-.284	.776	-.04	0.10	negligible	-3	negligible change
GRADE 4 – GRADE 5 MLAT-EC *PART 4*	**B**	59	336.000	-1.453	.146	-.19	0.38	small	8	small increase
	G	64	428.000	-1.041	.298	-.13	0.16	small	4	negligible change
GRADE 5 – GRADE 6 MLAT-E PART 4	**B**	279	-	-	-	-	0.27	small	13	small increase
	G	283	-	-	-	-	0.22	small	10	small increase
GRADE 5 – GRADE 6 MLAT-ES *PARTE 4*	**B**	57	390.000	-.169	.865	-.02	0.08	negligible	-1	negligible change
	G	60	303.000	-2.176	**.030**	-.28	0.65	medium	18	medium increase
GRADE 5 – GRADE 6 MLAT-EC *PART 4*	**B**	61	438.000	-.360	.719	-.05	0.27	small	-5	small decrease
	G	60	402.500	-.686	.493	-.09	0.06	negligible	-1	negligible change
GRADE 6 – GRADE 7 MLAT-ES *PARTE 4*	**B**	68	566.500	-.125	.901	-.02	0.10	negligible	-1	negligible change
	G	59	341.500	-1.474	.140	-.19	0.13	negligible	3	negligible change
GRADE 6 – GRADE 7 MLAT-EC *PART 4*	**B**	59	355.000	-1.280	.201	-.17	0.28	small	6	small increase
	G	65	294.500	-3.218	**.001**	-.40	0.74	medium	13	small increase

S.: sex; B: boys; G: girls; M-W U: Mann-Whitney U; Asymp. Sig (2-t.): asymptotic significance, 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d* : Cohen's *d*; % chnge: percentage change; Direct: direction, v. large: very large

Despite the different total number of items, the means of grade 3 of group 1 and of the MLAT-E norming study are very similar, as shown in Table 4.21. Therefore, 3-graders in group 1 performed better on the MLAT-ES than did 3-graders on the MLAT-E. Mean total raw scores are consistently lower in group 2, the one who took the MLAT-EC due to the inner IF of this test.

Table 4.21. Descriptive raw total scores on the MLAT-E, MLAT-ES and MLAT-EC according to sex

		Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part		MLAT-E (130 items)			MLAT-ES (123 items)			MLAT-EC (122 items)		
Grade	Gender	N	Mean	SD	N	Mean	SD	N	Mean	SD
3	Boys	112	64.0	22.8	26	63.92	17.955	27	55.78	20.885
	Girls	105	65.9	22.3	40	65.85	23.088	30	54.53	24.653
4	Boys	113	80.0	22.9	41	86.00	19.388	26	87.00	19.926
	Girls	88	91.3	18.5	34	76.82	24.453	36	81.44	22.999
5	Boys	112	94.3	20.2	32	99.76	15.975	33	94.85	17.096
	Girls	109	96.8	17.9	25	84.38	22.381	28	91.11	20.030
6	Boys	167	100.9	17.9	32	100.91	18.438	28	95.54	16.093
	Girls	174	105.0	16.6	28	98.82	19.982	32	93.03	18.499
7	Boys	-	-	-	36	106.72	10.416	31	95.94	19.112
	Girls	-	-	-	31	106.65	16.088	33	103.73	17.076

In spite of the differences in the mean raw total scores, it appears that, on the whole, grade 3 performs significantly different from the other grades (see Table 4.22), as the increase in the means of grade 4 is from medium to very large in the three versions of the tests and this increase also has considerable effect sizes. Grade 4 also seems to perform somehow in a different way from grades 5, 6 and 7, although the increase in the means is medium only for boys in the MLAT-E norming study and for boys in group 1. Total mean raw scores of 5-graders in the three groups and 5-grade boys in group 2 present only a small increase with small or medium effect sizes. This same pattern is found again in grade-6 girls in group 2, although this large increase in the mean score is, as pointed out before, due to the fact that 5-graders obtained lower scores than expected. On the whole, it could be stated that the MLAT-E, and the MLAT-ES and MLAT-EC administered in Catalonia present a different type of challenge for test takers in the lower grades, especially for those in grade 3.

Table 4.22. Differences in the mean scores on the total scores on the MLAT-E, MLAT-ES and MLAT-EC across grades according to sex

Grades and test versions	S.	N	M-W. U	Z	Asymp. Sig. (2-t.)	*r* (eff. s.)	C.'s *d*	Eff. size (Cohen)	% chnge	Direct.
GRADE 3 – GRADE 4 MLAT-E TOTAL	**B**	225	-	-	-	-	0.70	medium	25	medium increase
	G	193	-	-	-	-	**1.24**	**very large**	**39**	**large increase**
GRADE 3 – GRADE 4 MLAT-ES TOTAL	**B**	67	215.500	-4.087	**.000**	**-.50**	**1.19**	**very large**	**35**	**large increase**
	G	74	503.000	-1.920	.055	-.22	0.47	medium	17	medium increase
GRADE 3 – GRADE 4 MLAT-EC TOTAL	**B**	53	100.000	-4.467	**.000**	**-.61**	**1.56**	**huge**	**56**	**v. large increase**
	G	66	230.000	-3.994	**.000**	**-.49**	**1.15**	**very large**	**49**	**large increase**
GRADE 4 – GRADE 5 MLAT-E TOTAL	**B**	225	-	-	-	-	0.67	medium	18	medium increase
	G	197	-	-	-	-	0.3	small	6	small increase
GRADE 4 – GRADE 5 MLAT-ES TOTAL	**B**	66	294.000	-2.890	**.004**	**-.36**	**0.77**	**large**	16	medium increase
	G	66	457.500	-1.110	.267	-.14	0.33	small	10	small increase
GRADE 4 – GRADE 5 MLAT-EC TOTAL	**B**	59	330.500	-1.504	.132	-.20	0.43	medium	9	small increase
	G	64	379.000	-1.692	.091	-.21	0.45	medium	12	small increase
GRADE 5 – GRADE 6 MLAT-E TOTAL	**B**	279	-	-	-	-	0.35	small	7	small increase
	G	283	-	-	-	-	0.48	medium	8	small increase
GRADE 5 – GRADE 6 MLAT-ES TOTAL	**B**	57	369.500	-.491	.623	-.06	0.07	negligible	1	negligible change
	G	60	267.000	-2.684	**.007**	-.35	**1.07**	**large**	26	medium increase
GRADE 5 – GRADE 6 MLAT-EC TOTAL	**B**	61	459.000	-.043	.965	-.01	0.04	negligible	1	negligible change
	G	60	419.000	-.430	.667	-.06	0.1	negligible	2	negligible change
GRADE 6 – GRADE 7 MLAT-ES TOTAL	**B**	68	509.500	-.818	.413	-.10	0.08	negligible	1	negligible change
	G	59	332.000	-1.550	.121	-.20	0.44	medium	8	small increase
GRADE 6 – GRADE 7 MLAT-EC TOTAL	**B**	59	420.500	-.205	.838	-.03	0.02	negligible	0	negligible change
	G	65	308.500	-2.884	**.004**	-.36	0.61	medium	12	small increase

S.: sex; B: boys; G: girls; M-W. U: Mann-Whitney U; Asymp. Sig. (2-t.): asymptotic significance, 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d*: Cohen's *d*; eff. size: effect size; % chnge: percentage change; Direct: direction, v. large: very large

4.2.4. The participants' age in the MLAT-ES norming study and the MLAT-ES and the MLAT-EC administered in Catalonia

In contrast with the MLAT-E, the published norming study of the MLAT-ES includes the information of grade 7. However, it does not make any distinction between males and females. Consequently, in order to see if the evolution of scores across grades on the MLAT-ES and MLAT-EC administered in Catalonia is similar to the official norming study, it is necessary to repeat the same statistical analysis done with the MLAT-E but treating the cohorts without taking into account the sex variable. Language will only be a factor to take into account when comparing the performances on the MLAT-ES and the MLAT-EC while other factors should be taken into consideration when differences, if any, are found between the MLAT-ES norming study and the performance by group 1 in this study. The descriptive results and the independent-sample comparisons appear per parts. The mean total scores and the corresponding test comparisons appear in the last place. Thus, the tables that follow include, on the one hand, the means and the standard deviations of the partial and total scores of the norming study and of the tests administered in Catalonia. As for the statistical analysis to compare the independent samples, only the data regarding Cohen's *d* and percentage change are included, as the Mann-Whitney U tests and their significance appear at the beginning of this section already.

Mean scores on part 1 are consistently higher for the test takers of this study than for the population of the MLAT-ES norming study, as shown in Table 4.23. This is, though, not the only difference between these two populations. While the means of the norming study present a steady increase, this is not the case of the Catalan/Spanish cohort, as 3-graders obtain much lower mean scores than the next grade, and the differences between the means across the other grades are not so large in both groups 1 and 2.

Table 4.23. Descriptive raw scores on part 1 on the norming study of the MLAT-ES and on the MLAT-ES and the MLAT-EC administered in Catalonia

	Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part	**MLAT-ES *Parte 1* (30 items)**			**MLAT-ES *Parte 1* (30 items)**			**MLAT-EC *Part 1* (30 items)**		
Grade	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
3	207	11.0	7.7	66	12.82	6.81	57	12.37	8.03
4	206	13.0	7.8	75	19.45	6.81	62	20.65	6.45
5	289	16.3	7.6	57	22.93	5.33	61	22.75	5.46
6	306	19.1	7.3	60	22.67	6.65	60	23.47	5.47
7	178	22.6	7.0	67	26.76	3.22	64	25.16	4.82

Cohen's *d* and the percentage change confirm that the Catalan/Spanish cohort presents a different score pattern in part 1 from that of the MLAT-ES norming study (see Table 4.24). While the increase between grades 3 and 4 in the means of the MLAT-ES and the MLAT-EC is of 52% and 68% with large (Cohen's *d*=0.98) and very large (Cohen's *d*=1.16) effect sizes, the increase in the means of grades 3 and 4 of the norming study is medium and its effect size is small (Cohen's *d*=0.26).

Table 4.24. Differences in the mean scores on part 1 of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia across grades

	N	Cohen's *d*	Effect size (Cohen)	Percentage change	Direction
GRADE 3 – GRADE 4 MLAT-ES (*Manual*)	413	0.26	small	18	medium increase
GRADE 3 – GRADE 4 MLAT-ES (Group 1)	141	**0.98**	**large**	**52**	**very large increase**
GRADE 3 – GRADE 4 MLAT-EC (Group 2)	119	**1.16**	**very large**	**68**	**very large increase**
GRADE 4 – GRADE 5 MLAT-ES (*Manual*)	495	0.43	medium	25	medium increase
GRADE 4 – GRADE 5 MLAT-ES (Group 1)	132	0.56	medium	18	medium increase
GRADE 4 – GRADE 5 MLAT-EC (Group 2)	123	0.36	small	10	small increase
GRADE 5 – GRADE 6 MLAT-ES (*Manual*)	595	0.38	small	17	medium increase
GRADE 5 – GRADE 6 MLAT-ES (Group 1)	117	0.04	negligible	6	small increase
GRADE 5 – GRADE 6 MLAT-EC (Group 2)	121	0.13	negligible	3	negligible change
GRADE 6 – GRADE 7 MLAT-ES (*Manual*)	484	0.49	medium	18	medium increase
GRADE 6 – GRADE 7 MLAT-ES (Group 1)	127	0.80	large	18	medium increase
GRADE 6 – GRADE 7 MLAT-EC (Group 2)	124	0.33	small	7	small increase

As far as grades from 4 to 7 are concerned, both the population in the MLAT-ES norming study and the bilingual Catalan/Spanish population who took the MLAT-ES in the first place present a very similar evolution of mean total scores on part 1. That is, both groups present larger increases across grades than does group 2, who took the MLAT-EC in the first place. Again, it is clear that this part of the test presents some kind of inner difficulty for 3-graders that disappears progressively the higher the grade, but this difficulty is more perceptible in the cohort of this study.

The results in part 2 tend to be higher for group 1 as compared to the rest of groups, except for test takers in grade 5 when compared to their group-2 counterparts, as shown in Table 4.25.

Table 4.25. Descriptive raw scores on part 2 on the norming study of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia

	Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part	**MLAT-ES *Parte 2* (30 items)**			**MLAT-ES *Parte 2* (30 items)**			**MLAT-EC *Part 2* (29 items)**		
Grade	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
3	207	11.0	6.6	66	12.58	5.87	57	10.40	6.02
4	206	14.9	8.2	75	17.55	7.21	62	15.55	7.08
5	289	17.7	8.1	57	18.63	7.88	61	19.44	6.96
6	306	20.4	7.8	60	23.33	6.40	60	19.28	7.06
7	178	22.0	6.9	67	23.99	6.13	64	21.31	6.81

At least two groups can be identified as regards the direction and percentage change of part 2 mean scores and, up to a certain extent, as regards the effect size of these changes, too. On the one hand, grade 3 obtains significantly lower mean scores on part 2 both in this study and in the MLAT-ES norming study. On the other hand, the next increase in the means expands itself for 2 grades, as from grades 4 to 6 there is a medium increase in the means. What is clear is that almost no differences either in the percentages or in the effect sizes are found between grades 6 and 7 in this part of the test, regardless of the language it is in or any other features that make the cohorts distinct from each other, such as the social or the educational context.

Table 4.26. Differences in the mean scores on part 2 of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia across grades

	N	Cohen's *d*	Effect size (Cohen)	Percentage change	Direction
GRADE 3 – GRADE 4 MLAT-ES (*Manual*)	413	0.53	medium	**35**	**large increase**
GRADE 3 – GRADE 4 MLAT-ES (Group 1)	141	**0.76**	**large**	**40**	**large increase**
GRADE 3 – GRADE 4 MLAT-EC (Group 2)	119	0.51	medium	**32**	**large increase**
GRADE 4 – GRADE 5 MLAT-ES (*Manual*)	495	0.34	small	19	medium increase
GRADE 4 – GRADE 5 MLAT-ES (Group 1)	132	0.14	negligible	6	small increase
GRADE 4 – GRADE 5 MLAT-EC (Group 2)	123	0.56	medium	25	medium increase
GRADE 5 – GRADE 6 MLAT-ES (*Manual*)	595	0.34	small	15	medium increase
GRADE 5 – GRADE 6 MLAT-ES (Group 1)	117	0.66	medium	25	medium increase
GRADE 5 – GRADE 6 MLAT-EC (Group 2)	121	0.02	negligible	-1	negligible change
GRADE 6 – GRADE 7 MLAT-ES (*Manual*)	484	0.21	small	8	small increase
GRADE 6 – GRADE 7 MLAT-ES (Group 1)	127	0.11	negligible	3	negligible change
GRADE 6 – GRADE 7 MLAT-EC (Group 2)	124	0.30	small	11	small increase

Part 3 appears to be consistently more difficult for the cohort of the MLAT-ES norming study. Only 3-graders in group 2 obtain similar results to those of the norming study population, and it is from grade 4 that group 2 catches up group 1 (see Table 4.27). While it is a fact that this part presents a wider array of possible rhymes in Catalan than in Spanish, this does not seem to affect the results of group 2, who obtain very similar results to those obtained by group 1 from grades 4 to 7. Also, the Latin American variety of Spanish presents a wider array of possible consonant rhymes than the peninsular Spanish variety. This could be the reason why the mean scores of grade 3 of the norming study are markedly lower than those of group 1.

Table 4.27. Descriptive raw scores on part 3 on the norming study of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia

	Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part	**MLAT-ES *Parte 3* (38 items)**			**MLAT-ES *Parte 3* (38 items)**			**MLAT-EC *Part 3* (38 items)**		
Grade	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
3	207	16.9	10.7	66	22.12	9.38	57	16.89	9.41
4	206	23.0	10.3	75	25.31	9.43	62	26.98	9.12
5	289	24.3	9.8	57	28.74	8.83	61	28.97	7.37
6	306	27.7	9.1	60	31.18	7.77	60	30.32	7.19
7	178	28.9	8.5	67	33.10	5.34	64	30.28	7.04

Consequently, while the increase between grades 3 and 4 of group 1 is small (14%), it is large (36%) and very large (60%) for the norming study group and for group 2 (see Table 4.28). The rest of grades do not present any noticeable change in their mean scores on this part. Therefore, it seems that one possible factor to be influencing the scores of 3-graders in the norming study and in group 2 on this part is, to start with, the language in which this part of the test is written, among other possible factors.

Table 4.28. Differences in the mean scores on part 3 of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia across grades

	N	Cohen's *d*	Effect size (Cohen)	Percentage change	Direction
GRADE 3 – GRADE 4 MLAT-ES (*Manual*)	413	0.58	medium	36	**large increase**
GRADE 3 – GRADE 4 MLAT-ES (Group 1)	141	0.34	small	14	small increase
GRADE 3 – GRADE 4 MLAT-EC (Group 2)	119	**1.10**	**very large**	**60**	**very large increase**
GRADE 4 – GRADE 5 MLAT-ES (*Manual*)	495	0.13	negligible	6	small increase
GRADE 4 – GRADE 5 MLAT-ES (Group 1)	132	0.38	small	14	small increase
GRADE 4 – GRADE 5 MLAT-EC (Group 2)	123	0.24	small	7	small increase
GRADE 5 – GRADE 6 MLAT-ES (*Manual*)	595	0.36	small	14	small increase
GRADE 5 – GRADE 6 MLAT-ES (Group 1)	117	0.66	medium	25	medium increase
GRADE 5 – GRADE 6 MLAT-EC (Group 2)	121	0.19	small	5	small increase
GRADE 6 – GRADE 7 MLAT-ES (*Manual*)	484	0.14	negligible	4	negligible change
GRADE 6 – GRADE 7 MLAT-ES (Group 1)	127	0.29	small	6	small increase
GRADE 6 – GRADE 7 MLAT-EC (Group 2)	124	0.01	negligible	0	negligible change

As for part 4, besides the fact that the test takers of this study obtain higher results than those in the norming study, it is curious that 3-graders in group 2 obtain rather lower mean scores than 3-graders in group 1, even though this part is virtually the same in the MLAT-ES and in the MLAT-EC.

Table 4.29. Descriptive raw scores on part 4 on the norming study of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia

	Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part	**MLAT-ES *Parte 4* (25 items)**			**MLAT-ES *Parte 4* (25 items)**			**MLAT-EC *Part 4* (25 items)**		
Grade	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
3	207	12.4	8.15	66	17.58	6.79	57	15.46	7.56
4	206	15.1	8.0	75	19.53	5.66	62	20.60	5.25
5	289	17.3	7.5	57	20.82	5.52	61	21.97	4.66
6	306	19.4	6.2	60	22.75	3.42	60	21.13	4.76
7	178	20.5	5.1	67	22.84	4.14	64	23.20	4.03

As a consequence of the lower mean scores of 3-graders in group 2, the increase in the mean scores on this part for group 2 between grades 3 and 4 is large (33%) and it also has a large effect size (Cohen's *d*=0.80) (see Table 4.30). Also, the increase between grades 3 and 4 and between grades 4 and 5 of the norming study is medium with a small effect size (Cohen's *d*=0.15), while it never reaches this level for the rest of grades in the other groups. Consequently, except for this large increase between grades 3 and 4 in group 2, it could be affirmed that part 4 does not present any significant difference across grades as far as the population of this study is concerned, while medium increases can be observed in the norming study.

Table 4.30. Differences in the mean scores on part 4 of the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia across grades

	N	Cohen's *d*	Effect size (Cohen)	Percentage change	Direction
GRADE 3 – GRADE 4 MLAT-ES (*Manual*)	413	0.34	small	22	medium increase
GRADE 3 – GRADE 4 MLAT-ES (Group 1)	141	0.32	small	11	small increase
GRADE 3 – GRADE 4 MLAT-EC (Group 2)	119	**0.80**	**large**	**33**	**large increase**
GRADE 4 – GRADE 5 MLAT-ES (*Manual*)	495	0.29	small	15	medium increase
GRADE 4 – GRADE 5 MLAT-ES (Group 1)	132	0.47	medium	14	small increase
GRADE 4 – GRADE 5 MLAT-EC (Group 2)	123	0.28	small	7	small increase
GRADE 5 – GRADE 6 MLAT-ES (*Manual*)	595	0.31	small	12	small increase
GRADE 5 – GRADE 6 MLAT-ES (Group 1)	117	0.43	medium	9	small increase
GRADE 5 – GRADE 6 MLAT-EC (Group 2)	121	0.18	small	-4	negligible change
GRADE 6 – GRADE 7 MLAT-ES (*Manual*)	484	0.19	small	6	small increase
GRADE 6 – GRADE 7 MLAT-ES (Group 1)	127	0.02	negligible	0	negligible change
GRADE 6 – GRADE 7 MLAT-EC (Group 2)	124	0.48	medium	10	small increase

In spite of the clear differences in the total mean scores between the scores of the norming study and those of the participants in this study, the descriptive results present a similar pattern of scores and standard deviations across grades, regardless of the version of the aptitude test taken; yet there are quantitative differences in the pattern. 3-graders score approximately 15 points fewer than 4-graders on the MLAT-ES and 28.65 fewer on the MLAT-EC, while in the rest of grades the gap between the total mean scores diminishes considerably on all the tests, although with some specific differences between some grades and tests. For instance, the difference between the means on the MLAT-ES norming study and on the MLAT-EC fluctuates a bit between grades 5 and 7 with a clear tendency to diminish across grades, while it diminishes steadily in group 1.

Table 4.31. Descriptive raw total scores on the norming study of the MLAT-ES and on the MLAT-ES and the MLAT-EC administered in Catalonia

	Norming study (*Manual*)			Group 1 (Catalonia)			Group 2 (Catalonia)		
Part	**MLAT-ES *Total* (123 items)**			**MLAT-ES *Total* (123 items)**			**MLAT-EC *Total* (122 items)**		
Grade	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**
3	207	51.2	25.3	66	65.09	21.09	57	55.12	22.75
4	206	65.9	28.0	75	81.84	22.16	62	83.77	21.77
5	289	75.6	25.9	57	91.12	21.12	61	93.13	18.44
6	306	86.5	23.0	60	99.93	19.04	60	94.20	17.32
7	178	94.0	19.4	67	106.69	13.24	64	99.95	18.37

As observed in the comparison between the total mean scores on the MLAT-ES and the ones administered in Catalonia, it can be observed in Table 4.32 that grade 3 performs significantly differently from the way the other grades perform. While the increase in the means on the MLAT-ES administered in Catalonia is similar to that of the norming study (29% and 26% respectively), this increase is remarkably larger in the MLAT-EC (52%, Cohen's *d*=1.30). Therefore, the Catalan version of the MLAT-E seems to be particularly difficult for grade 3 as compared to the rest of grades and to the Spanish version of the test. The cause of this difference between the MLAT-ES and CAT can be explained by the language in which the test is, as the participants of group 1 and group 2 come from very similar contexts regarding their social, educational and linguistic characteristics. A medium increase in the means can still be found between grades 4 and 5 in the norming study with a small effect size. The percentage change of this group (15%) is very close to the percentage changes of the Catalan/Spanish groups (11%), considered to be small increases. As for the rest of grades, the increase is very similar in all groups, regardless of the version of the test.

Table 4.32. Differences in the mean raw total scores on the MLAT-ES, and on the MLAT-ES and the MLAT-EC administered in Catalonia across grades

	N	Cohen's *d*	Effect size (Cohen)	Percentage change	Direction
GRADE 3 – GRADE 4 MLAT-ES (*Manual*)	413	0.55	medium	29	medium increase
GRADE 3 – GRADE 4 MLAT-ES (Group 1)	141	**0.78**	**large**	26	medium increase
GRADE 3 – GRADE 4 MLAT-EC (Group 2)	119	**1.30**	**very large**	**52**	**very large increase**
GRADE 4 – GRADE 5 MLAT-ES (*Manual*)	495	0.36	small	15	medium increase
GRADE 4 – GRADE 5 MLAT-ES (Group 1)	132	0.43	medium	11	small increase
GRADE 4 – GRADE 5 MLAT-EC (Group 2)	123	0.46	medium	11	small increase
GRADE 5 – GRADE 6 MLAT-ES (*Manual*)	595	0.45	medium	14	small increase
GRADE 5 – GRADE 6 MLAT-ES (Group 1)	117	0.44	medium	10	small increase
GRADE 5 – GRADE 6 MLAT-EC (Group 2)	121	0.06	negligible	1	negligible change
GRADE 6 – GRADE 7 MLAT-ES (*Manual*)	484	0.35	small	9	small increase
GRADE 6 – GRADE 7 MLAT-ES (Group 1)	127	0.42	medium	7	small increase
GRADE 6 – GRADE 7 MLAT-EC (Group 2)	124	0.32	small	6	small increase

On the whole, it could be said that the patterns of scores of the MLAT-ES norming study and of group 1 are very similar across grades and that grade 3 performs consistently differently from the way the rest of grades do. At the beginning of this section two main issues were raised. The first one was whether the differences in the means between grades 3 and 4 that had been observed in the descriptive results were significant, as they were larger than the differences between the other grades. Indeed, these differences, as measured by the Mann-Whitney U tests and the percentage changes, are consistently significant for this grade. Therefore, there is no doubt that, on the whole, the constructs tapped by the elementary version of the MLAT are at a different stage of evolution when measured in grade 3. Regarding the MLAT-EC, it also presents the same patterns of scores as the MLAT-ES administered in Catalonia and the scores in the *Manual*. However, the overall performance, especially of 3-graders, is far worse than it is on the MLAT-ES, as compared to both the norming study and the performance by group 1. This lower performance is particularly lower in parts 1 and 3, which are the ones that depend the most on the language in which the test is. However, they are not so in part 2, which is the one that depends the least on the test

takers' L1, as it uses an invented language. However, it should also be pointed out that this part is markedly easier in the MLAT-ES and the MLAT-EC than in the MLAT-E.

The second issue raised was the fact that the decrease in the distance between means in the upper grades could somehow invalidate the test for the upper grades. In all the tables that include the means of the total scores on all the tests analyzed, it can be observed that the percentage of change in the means diminishes the upper the grade. Therefore, there could be a moment when the test does not have enough discrimination power due to the low IF of some parts, especially of part 4, as the variance from grades 5 to 7 is very small in all the versions of the test.

In order to determine if the results plateau in grades 5 or 6 and thus be able to suggest that aptitude as measured by the MLAT-ES and the MLAT-EC stabilises at a concrete point, it is necessary to run Mann-Whitney U tests between grades 5 and 7.

Table 4.33 Differences in the mean part and total scores on the MLAT-ES (norming study), and on the MLAT-ES and the MLAT-EC administered in Catalonia between grades 5 and 7

Grades and test versions	N	M-W. U	Z	Asymp. Sig.(2-t)	*r* (eff. s.)	C.'s *d*	Eff. size (Cohen)	% chnge	Direct
GRADE 5 – GRADE 7 MLAT-ES PARTE 1 (MANUAL)	367	-	-	-	-	**0.86**	**large**	**39**	**large**
GRADE 5 – GRADE 7 MLAT-ES PARTE 1 (CATALONIA)	124	986.000	-4.563	**.000**	-.29	**0.89**	**large**	17	medium
GRADE 5 – GRADE 7 MLAT-EC PART 1	125	1377.000	-2.851	**.004**	-.18	0.47	medium	11	small
GRADE 5 – GRADE 7 MLAT-ES PARTE 2 (MANUAL)	367	-	-	-	-	0.56	medium	24	medium
GRADE 5 – GRADE 7 MLAT-ES PARTE 2 (CATALONIA)	124	1133.500	-3.902	**.000**	-.25	**0.77**	**large**	29	medium
GRADE 5 – GRADE 7 MLAT-EC PART 2	125	1620.000	-1.644	.100	.10	0.27	small	10	small
GRADE 5 – GRADE 7 MLAT-ES PARTE 3 (MANUAL)	367	-	-	-	-	0.49	medium	19	medium
GRADE 5 – GRADE 7 MLAT-ES PARTE 3 (CATALONIA)	124	1360.000	-2.768	**.006**	-.18	0.61	medium	15	medium
GRADE 5 – GRADE 7 MLAT-EC PART 3	125	1683.500	-1.329	.184	.184	0.18	small	5	small
GRADE 5 – GRADE 7 MLAT-ES PARTE 4 (MANUAL)	367	-	-	-	-	0.48	medium	18	medium
GRADE 5 – GRADE 7 PARTE 4 MLAT-ES (CATALONIA)	124	1373.000	-2.795	**.005**	-.18	0.42	medium	10	small
GRADE 5 – GRADE 7 PART 4 MLAT-EC	125	1456.000	-2.603	**.009**	-.16	0.29	small	6	small
GRADE 5 – GRADE 7 MLAT-ES TOTAL (MANUAL)	367	-	-	**-**	-	**0.78**	**large**	24	medium
GRADE 5 – GRADE 7 MLAT-ES TOTAL (CATALONIA)	124	1027.000	-4.426	**.000**	-.28	**0.91**	**large**	17	medium
GRADE 5 – GRADE 7 MLAT-EC TOTAL	125	1452.500	-2.468	**.014**	-.16	0.37	small	7	small

Mann-Whitney U tests show that while there is some variability when considering the tests subparts on the MLAT-EC, the differences in the mean total scores of grade 5 and grade 7 are significant in both the MLAT-ES and in the MLAT-EC. However, there is no consistency either in the effect size of this increase or the percentage change, although consistent small increases or negligible changes were found in all the mean total scores between grades 6 and 7 when comparing them.

Therefore, it seems that aptitude as measured by the MLAT-ES and the MLAT-EC seems to stabilise at grade 6.

4.3. Research question 2: MLAT-E Spanish and Catalan in relation to the participants' sex

Sex is an individual difference that has been used to study several aspects related to FL learning. Results so far are not conclusive, as some studies favour one sex over the other and the opposite happens in some others (see section 1.6.4). In order to try to find out if this variable plays a role in the cohort of this study, the second research question is worded as follows:

RQ 2: Is there a relationship between language aptitude (as measured by the MLAT-E in Spanish and Catalan) and the subjects' sex?

In the norms for students in grades 3 to 6 on the MLAT-E provided in the *MLAT-E Manual*, the population appears divided according to the sex variable, while this division according to this variable disappears in the *MLAT-ES Manual*. In Table 2.1 in section 2.3.1.2, it can be seen that girls consistently obtain higher raw mean total scores than boys. This difference is not significantly strong, as calculated by means of Cohen's *d* and using the percentage change in the means as a reference value. These values have been calculated resorting to the statistical information available in the *MLAT-E Manual*.

Table 4.34. Cohen's *d* effect size and percentage change of the means on the MLAT-E norming study according to sex (calculated from the data in Carroll & Sapon, *MLAT-E Manual,* 2002:6)

Grades	Grade 3		Grade 4		Grade 5		Grade 6	
Sex	**BOYS**	**GIRLS**	**BOYS**	**GIRLS**	**BOYS**	**GIRLS**	**BOYS**	**GIRLS**
N	493	528	505	510	495	500	670	640
Mean	61.1	64.4	76.3	83.5	88.9	94.7	99.5	104.5
SD	24.7	23.6	23.3	22.8	20.6	19.2	20.4	18.1
Cohen's *d*	.14		.31		.29		.26	
Effect size	negligible effect		small effect		small effect		small effect	
Percentage change	5		9		7		5	
Direction	small increase		small increase		small increase		small increase	

Girls also appear to consistently obtain either the same or higher scores than boys on all parts of the MLAT-E in all grades, as shown in Table 4.35 based on the data from four schools.

Table 4.35. Part and total mean scores on the MLAT-E norming study (based on data from four schools) - (adapted from Carroll & Sapon, *MLAT-E Manual,* 2002:8-9)

Parts			Part 1 (30 items)		Part 2 (30 items)		Part 3 (45 items)		Part 4 (25 items)		Total (130 items)	
Grade	**Gender**	**N**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
3	**Boys**	112	14.7	6.3	11.2	5.4	31.0	10.9	7.2	5.9	64.0	22.8
	Girls	105	14.7	5.3	11.4	5.5	31.9	10.8	7.9	6.4	65.9	22.3
4	**Boys**	113	19.8	5.3	13.5	5.8	36.0	9.7	10.7	7.3	80.9	22.9
	Girls	88	21.3	4.5	16.8	5.6	39.1	6.6	14.2	7.4	91.3	18.5
5	**Boys**	112	22.6	4.6	17.0	6.0	39.8	7.1	14.8	7.3	94.3	20.2
	Girls	109	22.4	4.2	18.2	6.2	41.1	5.3	15.2	7.2	96.8	17.9
6	**Boys**	167	23.8	4.0	19.2	5.6	41.2	5.5	16.7	7.0	100.9	17.9
	Girls	174	24.7	3.8	21.4	4.8	42.2	6.0	16.7	6.6	105.0	16.6

Table 4.36 shows in bold letters that, but for the large increase (33%) in the 4-grade girls' mean scores on part 4 and the medium increase (24%) on part 2, both with medium effect sizes, none of these differences are significantly strong in any of the parts or the total score, as calculated by means of Cohen's *d* and the percentage change in the scores. Although it is a fact that girls obtain higher mean total scores than boys consistently across grades, the increase in the means is small or negligible in all grades and none of the increases ever reaches a large effect size.

Table 4.36. Differences in the means on the part and total scores on the MLAT-E norming study across grades according to sex

MLAT-E PART 1 HIDDEN WORDS	N	Cohen's *d*	Effect size	Percentage change	Direction
GRADE 3 BOYS - GIRLS	227	0	negligible	0	negligible change
GRADE 4 BOYS - GIRLS	201	0.30	small	8	small increase
GRADE 5 BOYS - GIRLS	221	0.05	negligible	-1	negligible change
GRADE 6 BOYS - GIRLS	341	0.23	small	4	negligible change
MLAT-E PART 2 MATCHING WORDS	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 BOYS - GIRLS	227	0.04	negligible	2	negligible change
GRADE 4 BOYS - GIRLS	201	0.58	medium	**24**	**medium increase**
GRADE 5 BOYS - GIRLS	221	0.2	small	9	small increase
GRADE 6 BOYS - GIRLS	341	0.42	medium	11	small increase
MLAT-E PART 3 FINDING RHYMES	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 BOYS - GIRLS	227	0.08	negligible	3	negligible change
GRADE 4 BOYS - GIRLS	201	0.37	small	9	small increase
GRADE 5 BOYS - GIRLS	221	0.21	small	3	negligible change
GRADE 6 BOYS - GIRLS	341	0.17	small	2	negligible change
MLAT-E PART 4 NUMBER LEARNING	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 BOYS - GIRLS	227	0.11	negligible	10	small increase
GRADE 4 BOYS - GIRLS	201	0.48	medium	**33**	**large increase**
GRADE 5 BOYS - GIRLS	221	0.21	small	3	negligible change
GRADE 6 BOYS - GIRLS	341	0	negligible	0	negligible change
MLAT-E RAW TOTAL SCORE	**N**	**Cohen's *d***	**Effect size**	**Percentage change**	**Direction**
GRADE 3 BOYS - GIRLS	227	0.08	negligible	3	small increase
GRADE 4 BOYS - GIRLS	201	0.54	medium	14	small increase
GRADE 5 BOYS - GIRLS	221	0.13	negligible	3	negligible change
GRADE 6 BOYS - GIRLS	341	0.24	small	4	negligible change

Therefore, except for the better performance by 4-grade girls in parts 2 and 4, the sex variable did not seem to be crucial in the MLAT-E norming study. Maybe for this reason the division according to the sex variable disappears in the *MLAT-ES Manual* although no explanation is given for this omission in the *Manual*.

Harper and Kieser (1977) used the MLAT-ES to prove its concurrent validity with 133 7-grade girls and 141 7-grade boys, and with 137 8-grade girls and 145 8-grade boys. Girls were superior to boys in the total scores as well as in all part scores except for the girls in grade 7, who scored 0.23 less than boys in the same grade.

Table 4.37. Means and standard deviations of the MLAT-E scores for grades 7 and 8 (adapted from Harper & Kieser, 1977:25)

Parts			Part 1 (30 items)		Part 2 (30 items)		Part 3 (45 items)		Part 4 (25 items)		Total (130 items)	
Grade	Gender	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
7	Boys	141	23.66	4.85	21.79	4.30	42.78	4.19	22.50	4.02	110.35	12.72
	Girls	133	24.58	4.55	22.04	3.99	42.55	5.50	23.12	4.90	112.27	14.10
8	Boys	145	23.25	4.32	20.00	5.19	42.08	5.43	21.14	4.96	106.55	15.08
	Girls	137	24.66	4.12	22.36	3.88	42.28	4.04	22.35	4.40	112.76	10.60

If we calculate the differences in the means obtained by the boys and girls in that study using the data provided in the study, it can be seen that the differences between the mean scores are either negligible or small and that only two of them (on the MLAT-E Part 2 and the raw total score in grade 8) have medium effect sizes.

Table 4.38. Differences in the means on the part and total scores on the MLAT-E from the data in Harper & Kieser (1977) study comparing sex

	Grade Boys - Girls	N	Cohen's *d*	Effect size	Percentage change	Direction
MLAT-E PART 1 HIDDEN WORDS	7	274	0.20	small	4	negligible change
	8	282	0.33	small	6	small increase
MLAT-E PART 2 MATCHING WORDS	7	274	0.06	negligible	1	negligible change
	8	282	0.51	medium	12	small increase
MLAT-E PART 3 FINDING RHYMES	7	274	0.05	negligible	-1	negligible change
	8	282	0.04	negligible	0	negligible change
MLAT-E PART 4 NUMBER LEARNING	7	274	0.14	negligible	3	negligible change
	8	282	0.26	small	6	small increase
MLAT-E RAW TOTAL SCORE	7	274	0.14	negligible	2	negligible change
	8	282	0.48	medium	6	small increase

In the data of this dissertation, as opposed to the data in the MLAT-E norming study, girls do not consistently obtain higher scores than boys.[17] On the contrary, only the girls in grade 3 (M=65.85, SD=23.08) outperform the boys in the same grade (M=63.92, SD=173.95) on the MLAT-ES. Also, on the MLAT-EC boys obtain higher mean total scores consistently across grades and it is only in grade 7 where girls score higher (M=103.73, SD=17.07) than boys (M=95.94, SD=19.11). Nevertheless, girls outperform boys in some part scores, as shown in tables from 4.13 to 4.22 in section 4.2. While, as measured by the MLAT-E, girls seemed to have higher aptitude than boys (although they were not significantly superior), boys are the ones who seem to

[17] Descriptive results according to the variable sex are not explained in depth here because they are very similar to the descriptive results in sections 3.5.2.3 and 3.5.3.3. See tables of descriptive results on the MLAT-ES and the MLAT-EC according to sex in appendix R.

have higher aptitude as measured by the MLAT-ES and the MLAT-EC, except for the cases mentioned. Hence the second research question of this dissertation.

The following tables present the results of the Mann-Whitney U tests used to compare two independent samples (boys and girls) in the same grade. Seeing the inconclusive results as regards the gender division obtained Kiss and Nikolov (2005; Kiss 2009), no assumption was made as to who — whether boys or girls — would obtain higher scores, only the 2-tailed significance values are provided. Two-tailed significant values, large effect sizes and medium changes (both increases and decreases) are in bold letters. In contrast with the tables in section 4.2, medium changes — and not large ones — appear in bold since no large change is found in any of the parts. The direction compares the girls' performance with the boys'. That is, "small decrease" means that girls obtain lower mean scores than boys.

In most grades, negligible changes and small decreases are found when comparing the performances of girls and boys in part 1 of the MLAT-ES and the MLAT-EC. The only decrease noticeable is that of girls in grade 4 on the MLAT-ES (*Mdn*=16.00*, U*=464.00, *p*<.001, *r*=-.29) as compared to boys (*Mdn*=22.00).

Table 4.39. Differences in the mean scores by boys and girls on part 1 of the MLAT-ES and MLAT-EC across grades

Gr.	MLAT	N	M-W. U	Z	Asym. Sig. (2-t.)	*r* (eff. s.)	C.'s *d*	Effect size (Cohen)	% chg	Direction
3	**ES**	66	514.500	-.072	.942	-.01	0.04	negligible	2	negligible change
	EC	57	399.000	-.096	.924	-.01	0.03	negligible	-2	negligible change
4	**ES**	75	464.000	-2.484	**.013**	-.29	0.66	medium	**-20**	**medium decrease**
	EC	62	400.500	-.964	.335	-.12	0.31	small	-9	small decrease
5	**ES**	57	295.500	-1.687	.092	-.22	0.45	medium	-10	small decrease
	EC	61	462.000	.000	1.00	0	0	negligible	0	negligible change
6	**ES**	60	354.000	-1.399	.162	-.18	0.45	medium	-12	small decrease
	EC	60	360.000	-1.308	.191	-.17	0.29	small	4	negligible change
7	**ES**	67	534.000	-.305	.760	-.04	0.20	small	-2	negligible change
	EC	64	417.000	-1.277	.202	-.16	0.22	small	4	negligible change

Gr.: grade; M-W. U: Mann-Whitney U; Asym. Sig (2-t.): asymptotic significance, 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d:* Cohen's *d*; % chg.: percentage change

As regards part 2, the general tendency is, as in part 1, for girls to obtain lower scores than boys, one of which (5-grade girls on the MLAT-ES) of medium percentage (17%) and medium effect size as measured by Cohen's *d.* In this part, though, some increases are also found, two of which of medium change and medium effect size as measured by Cohen's *d*. These are 3-grade girls (*Mdn*=9.50*, U*=357.00, *p*<.001, *r*=-.10), who outperform boys (*Mdn*=9.00) on the MLAT-EC *Part 3* with a 17% increase,

and grade-7 girls (*Mdn*=26.00*, U*=351.00, *p*<.001, *r*=-.27), who get a score 19% higher than that of boys (*Mdn*=21.00) with a significant two-tailed value (.030).

Table 4.40. Differences in the mean scores by boys and girls on part 2 of the MLAT-ES and MLAT-EC across grades

Gr.	MLAT	N	M-W. U	Z	Asym. Sig. (2-t.)	*r* (eff. s.)	C.'s *d*	Effect size (Cohen)	% chg	Direction
3	**ES**	66	476.000	-.578	.563	-.07	0.16	small	8	small increase
	EC	57	357.000	-.770	.441	-.10	0.27	small	**17**	**medium increase**
4	**ES**	75	664.000	-.352	.085	-.04	0.10	negligible	-4	negligible change
	EC	62	424.000	-.629	.529	-.08	0.20	small	-8	small decrease
5	**ES**	57	293.000	-1.723	.085	-.23	0.47	medium	**-17**	**medium decrease**
	EC	61	453.000	-.123	.902	-.02	0.03	negligible	-1	negligible change
6	**ES**	60	379.500	-1.019	.308	-.13	0.32	small	9	small increase
	EC	60	447.500	-.007	.994	0	0.02	negligible	-1	negligible change
7	**ES**	67	543.500	-.184	.854	-.02	0.07	negligible	-2	negligible change
	EC	64	351.000	-2.164	**.030**	-.27	0.55	medium	**19**	**medium increase**

Gr.: grade; M-W. U: Mann-Whitney U; Asym. Sig (2-t.): asymptotic significance*,* 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d:* Cohen's *d*; % chg.: percentage change

Again in part 3 the tendency is for girls to obtain either very similar scores to those obtained by boys or to score lower than them except for one case on the MLAT-ES, in which 5-grade girls score 24% higher than boys with a significant two-tailed value (*Mdn*=31.00*, U*=267.50, *p*<.001, *r*=-.28).

Table 4.41. Differences in the mean scores by boys and girls on part 3 of the MLAT-ES and MLAT-EC across grades

Gr.	MLAT	N	M-W. U	Z	Asym. Sig. (2-t.)	*r* (eff. s.)	C.'s *d*	Effect size (Cohen)	% chg	Direction
3	**ES**	66	502.000	-.236	.813	-.03	0.08	negligible	3	negligible change
	EC	57	379.500	-.408	.683	-.05	0.03	negligible	-2	negligible change
4	**ES**	75	529.000	-1.790	.073	-.21	0.42	medium	-14	small decrease
	EC	62	432.500	-.508	.612	-.06	0.19	small	-6	small decrease
5	**ES**	57	267.500	-2.137	**.033**	-.28	0.61	medium	**24**	**medium increase**
	EC	61	391.000	-1.030	.303	-.13	0.30	small	-7	small decrease
6	**ES**	60	442.500	-.082	.935	-.01	0.03	negligible	-1	negligible change
	EC	60	427.000	-.313	.755	-.04	0.06	negligible	-1	negligible change
7	**ES**	67	544.000	-.178	**.001**	-.02	0.19	small	3	negligible change
	EC	64	408.000	-1.396	.163	-.17	0.32	small	8	small increase

Gr.: grade; M-W. U: Mann-Whitney U; Asym. Sig (2-t.): asymptotic significance*,* 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d:* Cohen's *d*; % chg.: percentage change

In part 4 girls consistently obtain lower scores than boys or there is virtually no change except for 7-graders on the MLAT-EC. One of the decreases on the MLAT-ES,

that of girls in grade 5, is medium (-20%) and significant as measured by both the two-tailed significance value (.003) and Cohen's *d* (0.92), (*Mdn*=21.00*, U*=218.00, *p*<.001, *r*=-.39).

Table 4.42. Differences in the mean scores by boys and girls on part 4 of the MLAT-ES and MLAT-EC across grades

Gr.	MLAT	N	M-W. U	Z	Asym. Sig. (2-t.)	*r* (eff. s.)	C.'s *d*	Effect size (Cohen)	% chg	Direction
3	**ES**	66	480.500	-.521	.603	-.06	0	negligible	0	negligible change
	EC	57	341.500	-1.018	.309	-.14	0.32	small	-14	small decrease
4	**ES**	75	686.000	-.118	.906	-.01	0.05	negligible	-1	negligible change
	EC	62	430.500	-.541	.589	-.07	0.09	negligible	-2	negligible change
5	**ES**	57	218.000	-2.978	**.003**	-.39	**0.92**	**large**	**-20**	**medium decrease**
	EC	61	397.000	-.968	.333	-.12	0.29	small	-6	small decrease
6	**ES**	60	371.500	-1.182	.237	-.15	0.28	small	-4	negligible change
	EC	60	381.500	-1.007	.314	-.13	0.09	negligible	-2	negligible change
7	**ES**	67	548.000	-.135	.892	-.02	0.01	negligible	0	negligible change
	EC	64	465.500	-.696	.486	-.09	0.23	small	4	negligible change

Gr.: grade; M-W. U: Mann-Whitney U; Asym. Sig (2-t.): asymptotic significance*,* 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d:* Cohen's *d*; % chg.: percentage change

It is precisely the medium decreases by girls in grade 5 on *Parte 2* and *Parte 4* that make girls be the only ones showing a medium decrease (*Mdn*=88.00) as compared to boys (Mdn=102.00) on the total score, in spite of the increase they appear to have on *Parte 3.* This decrease (15%) is significant as measured by the two-tailed value and has a large effect size as measured by Cohen's *d* 0.79, *Mdn*=102.00*, U*=231.50, *p*<.001, *r*=-.36.

Table 4.43. Differences in the mean scores by boys and girls on the total score of the MLAT-ES and MLAT-EC across grades

Gr.	MLAT	N	M-W. U	Z	Asym. Sig. (2-t.)	*r* (eff. s.)	C.'s *d*	Effect size (Cohen)	% chg	Direction
3	**ES**	66	484.500	-.466	.641	-.06	0.09	negligible	3	negligible change
	EC	57	380.000	-.400	.689	-.05	0.06	negligible	-2	negligible change
4	**ES**	75	538.500	-1.687	.092	-.19	0.43	medium	-11	small decrease
	EC	62	409.000	-.842	.400	-.11	0.26	small	-6	small decrease
5	**ES**	57	231.500	-2.711	**.007**	-.36	**0.79**	**large**	**-15**	**medium decrease**
	EC	61	425.000	-.528	.597	-.07	0.21	small	-4	negligible change
6	**ES**	60	425.000	-.341	.733	-.04	0.11	negligible	-2	negligible change
	EC	60	429.500	-.274	.784	-.04	0.15	small	-3	negligible change
7	**ES**	67	497.500	-.761	.446	-.09	0.01	negligible	0	negligible change
	EC	64	372.000	-1.876	.061	-.23	0.44	medium	8	small increase

Gr.: grade; M-W. U: Mann-Whitney U; Asym. Sig (2-t.): asymptotic significance*,* 2-tailed; *r* (eff. s.): *r* (effect size); C.'s *d:* Cohen's *d*; % chg.: percentage change

To sum up, while in the MLAT-E norming study, girls consistently showed a higher, though not significant, aptitude than boys, this is not the case of Catalan/Spanish girls in this study who, overall, tend to score lower than boys, with the exceptions already pointed out, on both the MLAT-ES and the MLAT-EC, especially from grades 3 to 5. Except for the medium increase in the MLAT-EC by girls in grade 7, boys and girls in grades 6 and 7 do not show significant differences in their means on any of the parts.

4.4. Research question 3: The MLAT-ES and the MLAT-EC as concurrent predictors of FL proficiency

The third research question aims to check the construct validity of the aptitude measures used and it is stated as follows:

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?

The participants' proficiency was determined by three different types of measures: teachers' criterion measures, students' self-assessment measures and objective language proficiency measures.

The teachers of the participants in this study were asked to fill in a form for each student in which they had to mark them on a 10-point scale according to the following criteria: aptitude for vocabulary, for speaking, for grammar and for listening; general academic aptitude; aptitude for FL learning; general EFL knowledge; and the mark they believed the students would obtain at the end of the course. The first variables mentioned (aptitude for vocabulary, speaking, grammar and listening) are more language-learning specific whereas the others are of a more general scope. Not all the teachers filled in the questionnaires and not all the teachers who filled them in answered all the questions because the two last ones (the students' FL general knowledge and the mark the teachers forecast their students would obtain at the end of the course) were originally indicated as optional in the forms designed for the norming study, which were also the ones used for this study. Descriptive statistics for the criterion variables are reproduced below. The N for the optional questions (the

students' general knowledge of the FL and their probable mark in the FL subject) is specified below the table.

Table 4.44. Means and standard deviations of the criterion variables ratings provided by the teachers

Grade	Stats.	Voc.	Speak.	Gram.	Listen.	Gral. apt.	FL apt.	FL gral knw[a]	Prob. FL mark. [b]
3	**Mean**	7.02	6.96	6.78	7.16	7.33	7.20	6.67	6.94
N=51	**SD**	2.17	1.97	1.96	1.90	1.56	1.96	1.81	1.91
4	**Mean**	6.84	7.00	6.65	6.49	7.05	7.00	5.89	6.71
N=37	**SD**	2.41	2.35	2.23	2.28	1.58	2.01	1.94	1.64
5	**Mean**	6.00	5.06	5.71	5.49	6.29	5.80	5.66	5.86
N=35	**SD**	2.17	1.96	2.08	2.06	1.93	2.07	1.96	1.67
6	**Mean**	6.24	5.55	6.00	5.90	6.29	5.88	5.79	6.08
N=42	**SD**	1.82	1.97	2.00	2.05	1.89	1.97	2.03	1.65
7	**Mean**	6.07	5.25	5.80	5.93	5.52	5.61	5.86	5.95
N=44	**SD**	1.95	1.77	1.90	1.78	1.85	1.73	1.88	1.70
All	**Mean**	6.46	6.00	6.22	6.25	6.52	6.33	6.00	6.35
N=209	**SD**	2.13	2.15	2.06	2.07	1.87	2.04	1.94	1.78

[a] Gr. 3 – N= 39 ; Gr. 4 – N= 19; Gr. 5 – N= 35; Gr. 6 – N= 42; Gr. 7 – N= 21; All grades – N= 156
[b] Gr. 3 – N= 51 ; Gr. 4 – N= 19; Gr. 5 – N= 35; Gr. 6 – N= 42; Gr. 7 – N= 21; All grades – N= 168
Voc.: Vocabulary; Speak.: Speaking; Gram.: Grammar; Listen.: Listening; Gral. apt.: General aptitude; FL apt.: Foreign language aptitude; FL gral knw.: Foreign language general knowledge; Prob. FL mark: Probable mark in the foreign language.

Except in grade 7, general aptitude is the aspect evaluated with the highest marks while general knowledge of the foreign language receives the lowest marks, followed by speaking. Overall, students in grades 5 and 7 obtain the lowest marks, followed by those in grade 6, whereas students in the lower grades are the ones whose teachers consider them to be not only better FL language learners, but also better at academic skills. The language skill that receives the lowest grade is speaking, closely followed by grammar and listening. Although teachers say they follow a communicative teaching approach, speaking is, surprisingly, the skill that seems to need further practice or the skill most students are worst at. Vocabulary, which is the skill that most depends on memory and traditionally tends to be tested using more traditional paper-and-pencil tests, is the one rated with the highest grades.

In order to check the construct validity of the aptitude tests, correlations must be run between the scores on these aptitude measures (the MLAT-ES and the MLAT-EC) and the criterion measures. Before that, the descriptive statistics for these criterion measures are presented. As the number of items in the questionnaires to be filled in by the teachers is quite high, they appear here in several tables, distributed according to the approach of the battery of questions. The first analyses to appear are those dealing

with the language-specific aspects (vocabulary, speaking, grammar and listening). As far as group 2 (the students who took the MLAT-EC in the first place) is concerned, only one teacher filled in the criterion measures questionnaires. Consequently, the descriptive statistics and correlations between the scores on the MLAT-EC and the criterion measures supplied by this teacher are not reported due to the small amount of data available (13 subjects in grade 3 and only one in grade 4) and, therefore, the construct validity of the MLAT-EC using the criterion variables cannot be checked.

Table 4.45. Group 1 descriptive statistics of the criterion variables ratings for specific language skills and raw scores on the MLAT-ES

Grade	Stats.	Voc.	Speak.	Gram.	Listen.	MLAT-ES 1	MLAT-ES 2	MLAT-ES 3	MLAT-ES 4	MLAT-ES Total
3	**Mean**	6.89	7.00	6.82	6.96	11.84	13.18	24.42	18.87	68.32
N=38	**SD**	2.33	1.97	2.12	2.04	6.43	5.54	8.88	6.25	19.80
4	**Mean**	6.83	7.03	6.67	6.49	17.72	17.81	24.39	19.72	79.64
N=36	**SD**	2.44	2.37	2.26	2.28	7.71	7.51	10.93	6.02	26.02
5	**Mean**	6.00	5.00	5.71	5.49	23.14	19.71	28.80	20.71	92.37
N=35	**SD**	2.17	1.96	2.08	2.06	5.21	7.42	9.94	5.69	21.57
6	**Mean**	6.24	5.55	6.00	5.90	21.81	23.74	30.79	22.62	98.95
N=42	**SD**	1.82	1.97	2.00	2.05	7.22	6.45	8.51	3.84	20.41
7	**Mean**	6.24	5.25	5.80	5.93	27.52	24.61	33.57	23.45	109.16
N=44	**SD**	1.82	1.77	1.90	1.78	2.44	5.96	5.26	2.97	9.04
All	**Mean**	6.39	5.95	6.18	6.15	20.64	20.06	28.64	21.20	90.54
N=195	**SD**	2.15	2.16	2.09	2.08	8.04	7.76	9.33	2.78	24.45

Voc.: Vocabulary; Speak.: Speaking; Gram.: Grammar; Listen.: Listening

Overall, the mean scores the teachers gave their students in vocabulary, speaking, grammar and listening are very similar. This should not be interpreted as if teachers had a general idea of their students' abilities, yet they did not consider their students stand out in one skill or another (speaking, listening, grammar, vocabulary) over others. For instance, while speaking is the skill that gets the highest mark in grades 3 and 4, it becomes the skill consistently marked with the lowest marks from grade 5 on as compared to the other skills, followed by listening, grammar and vocabulary in this order. Wilcoxon signed-rank tests between these variables show that teachers marked their students significantly differently in each of these skills only from grade 5 on, as shown in Table 4.46. This table includes the test statistic (T), the z-value, the two-tailed significance value and the effect size of the difference between tests. These two last measures appear in bold if they are significant, as for the former, and if the effect size is moderate or large as for the latter.

Table 4.46. Wilcoxon signed-rank test between specific language skills criterion variables

Skills	Statistics	Grade 3 N= 38	Grade 4 N= 36	Grade 5 N= 35	Grade 6 N= 42	Grade 7 N= 44	All grades N= 195
speaking / vocabulary	**T**	82.50	88.00	.00	13.00	9.50	1685.00
	Z	-.528[a]	-.639[a]	-4.586[b]	-4.746[b]	-4.578[b]	-5.315[b]
	Asym. Sig. (2-t.)	.597	.523	**.000**	**.000**	**.000**	**.000**
	Effect size	-.06	-.08	**-.55**	**-.52**	**-.49**	**-.27**
speaking / grammar	**T**	54.00	47.00	10.00	22.00	42.00	1674.50
	Z	-1.528[a]	-1.739[a]	-4.104[b]	-3.800b	-3.356[b]	-3.801[b]
	Asym. Sig. (2-t.)	.127	.082	**.000**	**.000**	**.001**	**.000**
	Effect size	-.02	-.20	**-.48**	**-.41**	**-.36**	-.19
grammar / vocabulary	**T**	83.00	128.50	34.00	.00	61.50	1351.00
	Z	-.513[b]	-.294[b]	-2.236[b]	-2.887[a]	-1.730[b]	-2.657[b]
	Asym. Sig. (2-t.)	.608	.769	**.025**	**.004**	.084	**.008**
	Effect size	-.06	-.09	**-.26**	**-.31**	-.18	-.13
listening / vocabulary	**T**	121.00	81.50	27.00	17.00	82.50	1611.50
	Z	-.194[b]	-1.209[b]	-3.105[b]	-3.130[b]	-.876[b]	-3.335[b]
	Asym. Sig. (2-t.)	.846	.226	**.002**	**.002**	.381	**.001**
	Effect size	-.02	-.01	**-.37**	**-.34**	-.09	-.17
listening / speaking	**T**	85.50	20.00	.00	18.00	21.50	1289.50
	Z	-.426[b]	-.426b	-3.419[a]	-3.273[a]	-3.647[a]	-3.101[a]
	Asym. Sig. (2-t.)	.670	.127	**.001**	**.001**	**.000**	**.002**
	Effect size	-.05	-.05	**-.41**	**-.36**	**-.39**	-.16
listening / grammar	**T**	18.00	59.50	63.00	66.50	91.00	1793.50
	Z	-1.508[a]	-1.237[b]	-1.789[b]	-.943[b]	-.910[a]	-.765[b]
	Asym. Sig. (2-t.)	.132	.216	.074	.346	.363	.444
	Effect size	-.17	-.15	-.21	-.10	-.10	-.04

[a]Based on negative ranks
[b]Based on positive ranks

It is noticeable that no significant differences between the means are found in grades 3 and 4 in any of the pairs of measures. This could be due to the fact that, as the teachers themselves reported, in the lower grades most of the learning takes place by reading aloud, listening and speaking in class. Consequently, the students are hardly ever marked on their grammar or writing proficiency. That is to say, these skills are practiced together with the other skills and no special attention is given to them in class, as the teaching approach is mainly oral.

Two-tailed Spearman correlations were run between the criterion measures and the part and total scores on the MLAT-ES to check the test's concurrent validity. Spearman correlations were chosen over Pearson correlations because of the non-normal distribution of the data.

Table 4.47. Two-tailed Spearman correlations between MLAT-ES scores and criterion variables related to language skills

		MLAT-ES score				
Grade	**Ratings**	***Parte 1***	***Parte 2***	***Parte 3***	***Parte 4***	***Total***
3 N=38	**vocabulary**	-.151	.094	.060	-.058	.036
	speaking	-.062	.129	.063	.090	.107
	grammar	-.115	.136	.060	.008	.065
	listening	-.037	.165	.119	.032	.129
4 N=368	**vocabulary**	.351*	.294	.322	.319	.384*
	speaking	.477**	.396*	.681**	.585**	.639**
	grammar	.684**	.497**	.648**	.602**	.746**
	listening	.623**	.325	.662**	.510**	.667**
5 N=35	**vocabulary**	.300	.372*	.466**	.776**	.564**
	speaking	.216	.447**	.452**	.689**	.527**
	grammar	.255	.399*	.493**	.685**	.549**
	listening	.262	.432**	.489**	.692**	.545**
6 N=42	**vocabulary**	.402**	.537**	.528**	.335*	.590**
	speaking	.246	.468**	.445**	.289	.463**
	grammar	.349*	.570**	.512**	.319*	.565**
	listening	.357*	.505**	.506**	.336*	.556**
7 N=44	**vocabulary**	.117	.438**	.390**	.285	.496**
	speaking	.096	.432**	.357*	.270	.445**
	grammar	.141	.495**	.432**	.290	.568**
	listening	.333*	.445**	.410**	.196	.570**
All grades N=195	**vocabulary**	.051	.194**	.261**	.241**	.228**
	speaking	-.076	.107	.198**	.205**	.116
	grammar	.098	.261**	.331**	.289**	.298**
	listening	.118	.214**	.334**	.266**	.287**

* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

No significant positive relationship between these language-related criterion measures and the MLAT-ES is found in grade 3. Actually, correlations between *Parte 4 Aprendamos números* and vocabulary and *Parte 1 Palabras ocultas* and all the language skills measures are even negative. It is important to remember that grade 3 was the grade that consistently had large differences in the means of all parts when compared to those obtained by 4-graders, while the differences in the scores obtained by the rest of grades were smaller and mostly with smaller effect sizes. In the other grades significant positive relations are found between the language skills measures and the partial and/or total scores on the MLAT-ES, but with some exceptions.

Considering each part on the MLAT-ES separately and without taking into account any of the correlations in grade 3, *Parte 1 Palabras ocultas*, which is aimed at measuring knowledge of Spanish vocabulary and sound-symbol association, presents

a significant positive relationship at a moderate and high level in grade 4 with all the language skills measures. Nevertheless, even if this part is aimed at measuring knowledge of vocabulary, this is the skill with which it has the lowest correlation in this grade (r_s=.351, p<.05). As in grade 3, *Parte 1* does not bear any significant positive relationship with any of the language skills measures in grade 5 and considering all the grades together. It only has a significant positive relation with listening in grade 7 (r_s=.333, p<.05). In grade 6, in contrast, it does significantly correlate with most language skills measures at a moderate level except with speaking.

Parte 2 Palabras que se corresponden is aimed at predicting sensitivity to grammatical structure, so it should correlate positively at least with the grammar measure. Indeed, except for grade 5, the skill with which this part has the highest correlations is grammar, ranging from r_s=.399, p<.05 to r_s=.570, p<.001. This part is also related to the other language skills measures across grades except for speaking when considering the whole group, as no significant positive relationship is found between these two measures (r_s=.107), and neither is found in grade 4 between *Parte 2* and listening (r_s=.294) and vocabulary (r_s=.325).

Parte 3 Palabras que riman is intended to measure the ability to hear speech sounds. Therefore, it would be expected to have a relationship with the listening measure at least. The correlations with this language skill range from r_s=.334, p<.001 to r_s=.662, p<.001. In addition, this part obtains from moderate to high significant positive correlations at p<.001 level with all the language skills measures except for speaking in grade 7, with which it is significantly correlated at a moderate level r_s=.357, p<.001.

Aimed at measuring some kind of rote memory component, *Parte 4 Aprendamos números* correlates highly and significantly with all the language skills measures in grades 4 and 5 at p<.001 and it has from a low to moderate significant relationship with all the measures when considering the whole cohort. However, it does not present any kind of relationship with any of the language skills measures in grade 7. Also, in grade 6 it is significantly correlated at a moderate level with all language skills measures but speaking. Since this part is intended to measure memory, it should in principle be related to the vocabulary measure because of the role memory plays in this skill as regards lexical storage and effective retrieval, but it could also be related with speaking because of the need of memory to enhance fluency. However, it is not related so strongly to speaking as it would be expected.

Leaving grade 3 apart as well as the speaking variable in the total group, the total scores on the MLAT-ES present from low to moderate-high correlations with all

the language skills measures across grades, ranging from r_s=.228, p<.001 (with vocabulary in all grades together) to r_s=.746, p<.001 (with grammar in grade 4). The lowest correlations, yet still significant at p<.001, are found considering the entire group. It is at this point when the MLAT-ES has no relationship any longer with the speaking measure (r_s=.116), but it does appear to have a significant positive relationship with this language skill measure in the other grades at a moderate level, ranging from r_s=.445, p<.001 to r_s=.639, p<.001.

The teachers were also asked about their students' aptitudes and marks at a more general level. These variables are general aptitude, foreign language aptitude, general FL knowledge and the probable mark the students would obtain at the end of the course (see Table 4.48). Significant differences were found in the mean scores provided by the teachers regarding their students' linguistic skills although means seem very similar, as are the means on those scores with a more general kind of scope (general aptitude, FL aptitude, general knowledge of the FL and the probable FL mark at the end of the course). The data available for general FL knowledge in grades 3 and 4 and of general FL knowledge and probable FL mark in grade 7 are smaller than in the rest of groups, as they were optional items in the questionnaire answered by the teachers. In those cases in which groups do not reach the minimum N desirable (N=30) the results should, consequently, not be taken into much consideration.

Table 4.48. Descriptive statistics and correlations between MLAT-ES scores and general aptitude and general language criterion variables

		Rating			MLAT-ES score				
Grade	**N**		**Mean**	**SD**	***Parte 1***	***Parte 2***	***Parte 3***	***Parte 4***	***Total***
3	38	**Gral apt.**	7.29	1.69	-.009	.194	.259	.236	.274
		FL apt.	7.26	2.08	-.054	.172	.160	.075	.170
	26	**FL gral knw.**	6.54	1.92	-.029	.141	.258	.135	.278
	38	**Prob. FL m.**	6.92	2.00	.178	.141	.218	.164	.279
4	36	**Gral apt.**	7.03	1.59	.569**	.390*	.663**	.557**	.674**
		FL apt.	7.00	2.04	.585**	.455**	.669**	.601**	.706**
	18	**FL gral knw.**	5.83	1.98	.345	.158	.500*	.584*	.521*
		Prob. FL m.	6.64	1.65	.176	.234	.334	.536*	.418
5	35	**Gral apt.**	6.29	1.93	.310	.475**	.605**	.646**	.643**
		FL apt.	5.80	2.07	.212	.462**	.491**	.682**	.555**
		FL gral knw.	5.66	1.96	.226	.439**	.486**	.684**	.558**
		Prob. FL m.	5.86	1.67	.335*	.493**	.532**	.711**	.645**
6	42	**Gral apt.**	6.29	1.89	.387*	.504**	.518**	.422**	.572**
		FL apt.	5.88	1.97	.280	.489**	.425**	.282	.479**
		FL gral knw.	5.79	2.03	.300	.498**	.451**	.304	.496**
		Prob. FL m.	6.08	1.64	.368*	.532**	.501**	.353*	.574**
7	44	**Gral apt.**	5.52	1.85	.098	.435**	.422**	.268	.556**
		FL apt.	5.61	1.73	.170	.432**	.385**	.270	.505**
	21	**FL gral knw.**	5.86	1.88	.275	.602**	.310	.113	.564**
		Prob. FL m.	5.95	1.70	.154	.621**	.306	.150	.610**
All	195	**Gral apt.**	6.45	1.89	-.005	.168*	.298**	.263**	.218**
		FL apt.	6.28	2.07	-.017	.153*	.252**	.233**	.183*
	142	**FL gral knw.**	5.91	1.96	.134	.296**	.350**	.353**	.355**
	154	**Prob. FL m.**	6.29	1.78	.048	.222**	.279**	.296**	.260**

* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)
Gral. apt.: General aptitude; FL apt.: Foreign language aptitude; FL gral knw.: Foreign language general knowledge; Prob. FL m.: Probable mark in the foreign language.

The criterion measures that correlated significantly with the MLAT-ES are of a general kind, so it would be expected that they correlate at least with the total score on the MLAT-ES. As can be seen in Table 4.47, except in grade 4, where the total score shows no significant correlation with the probable FL mark at the end of the course probably because of the low N, all the criterion measures show from moderate to high significant correlations with the total score from grade 4 to grade 7. When considering the group as a whole, these correlations are, as compared to each grade separately, rather low (especially with FL aptitude, r_s=.183, p<.05), yet still significant. Exactly as it happened with the language skills variables, though, neither partial scores on the MLAT-ES nor the total score correlate with any of the criterion measures in grade 3.

Regarding each part separately, there are almost no significant positive relationships between *Parte 1 Palabras ocultas* and the criterion measures. Exceptions are the significant positive relationship with general aptitude in grade 4 (r_S=.569, p<.001) and grade 6 (r_S=.387, p<.001), with FL aptitude in grade 4 (r_S=.585, p<.001) and with the probable FL mark at the end of the course in grade 5 (r_S=.335, p<.05) and grade 6 (r_S=.368, p<.05). *Parte 2,* which taps a rather concrete aptitude aspect, grammar sensitivity, is the one that, after *Parte 1,* obtains the lowest correlations with the general criterion measures for the whole group, especially with general aptitude (r_S=.168, p<.05) and with FL aptitude (r_S=.153, p<.05). However, all the other correlations between this part and the general criterion measures are from low to moderate and significant at p<.001. *Parte 3 Palabras que riman* also correlates significantly and positively with most measures except for the probable FL grade in grade 4 (r_s=.334) and this same measure plus general FL knowledge in grade 7. *Parte 4 Aprendamos números* loses its relationship power with the criterion measures the higher the grade. While all the correlations between this part and the general criterion measures are from moderate to high in grades 4 and 5, in grade 6 it only correlates with general aptitude (r_S=.422, p<.001) and with the probable FL mark (r_S=.353, p<.05) and it does not correlate significantly with any of the criterion measures in grade 7.

The Wilcoxon signed-rank tests between these language skills variables (see Table 4.49) show that, overall, the teachers considered the probable mark in the FL subject at the end of the course related to general FL aptitude. Nevertheless, it should be reminded that, school FL grades can depend on many variables other than just FL aptitude, such as the students' attitudes, effort, accomplishment with the school assignments, and so on. Instead, these factors are perhaps most reflected in the comparison between probable FL grade at the end of the course and the general FL knowledge. The teachers tend to rate their students lower in the general knowledge of the FL than in the probable mark they will obtain at the end of the course. The differences between both means are significant in most grades except in grade 5 and grade 7. In the same line, mean scores on general FL knowledge do not seem to be related to general FL aptitude in grades 3 and 4, with large effect sizes (r=.40 in grade 3 and r=.45 in grade 4) and in the group taken as a whole. General aptitude is the variable which is consistently marked most differently from the others, especially from general FL knowledge, reaching from medium to large effect sizes in all grades but 7, and from general FL aptitude, whose mean scores are significantly different in grades 5 and 6 and considering the group as a whole. Mean scores on general aptitude are,

however, very similar to those on general FL knowledge in all grades but 5 and in the whole cohort.

Table 4.49. Wilcoxon signed-rank test between general aptitude, general FL aptitude, general knowledge of the FL and probable FL mark at the end of the year.

Skills	Statistics	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	All grades
general FL aptitude / general aptitude [1]	**T**	72.50	104.00	5.50	48.50	178.00	1784.00
	Z	-.206[b]	-.039[b]	-3.082[b]	-2.712[b]	-.069[b]	-2587[b]
	Asym. Sig. (2-tailed)	.837	.969	**.002**	**.007**	.945	**.010**
	Effect size	-.02	-.01	**-.36**	-.29	-.01	-.15
general FL knowledge / general aptitude [2]	**T**	6.00	6.50	6.50	31.50	30.00	396.00
	Z	-3.095[b]	-3.231[b]	-3.458[b]	-3.479[b]	-.277[b]	-6.188[b]
	Asym. Sig. (2-tailed)	**.002**	**.001**	**.001**	**.001**	.782	**.000**
	Effect size	**-.43**	**-.54**	**-.41**	**-.38**	-.04	-.35
probable FL mark / general aptitude [2]	**T**	120.00	53.50	70.50	163.50	79.00	2215.00
	Z	-1.450[b]	-.774[b]	-2.511[b]	-1.439[b]	-.592[b]	-2.931[b]
	Asym. Sig. (2-tailed)	.147	.439	**.012**	.150	.554	**.003**
	Effect size	-.20	-.13	**-.30**	-.16	-.09	-.17
general FL knowledge / general FL aptitude [2]	**T**	6.50	.00	4.00	3.50	6.00	85.00
	Z	-2.887[b]	-2.724[b]	-1.890[b]	-1.633[b]	-.378[a]	-4.511[b]
	Asym. Sig. (2-tailed)	**.004**	**.006**	.059	.102	.705	**.000**
	Effect size	**-.40**	**-.45**	-.23	-.18	-.06	-.26
probable FL mark / general FL aptitude [3]	**T**	117.00	63.50	132.50	171.00	66.50	2804.50
	Z	-.965[b]	-.238[a]	-.507[a]	-1.532[a]	-.379[b]	-.752[a]
	Asym. Sig. (2-tailed)	.334	.812	.612	.125	.705	.452
	Effect size	-.11	-.04	-.06	-.17	-.06	-.04
probable FL grade / general FL knowledge[2]	**T**	46.00	.00	96.00	120.50	74.50	1245.50
	Z	-2.025[a]	-3.346[a]	-1.571[a]	-2.346[a]	-.876[b]	-4.704[a]
	Asym. Sig. (2-tailed)	**.043**	**.001**	.116	**.019**	.381	**.000**
	Effect size	-.28	**-.56**	-.19	-.03	-.13	-.27

[a]Based on negative ranks
[b]Based on positive ranks
[1]Gr. 3 – N= 38 ; Gr. 4 – N= 36; Gr. 5 – N= 35; Gr. 6 – N= 42; Gr. 7 – N= 44; All grades – N= 195
[2]Gr. 3 – N= 26 ; Gr. 4 – N= 18; Gr. 5 – N= 35; Gr. 6 – N= 42; Gr. 7 – N= 21; All grades – N= 142
[3]Gr. 3 – N= 38 ; Gr. 4 – N= 18; Gr. 5 – N= 35; Gr. 6 – N= 42; Gr. 7 – N= 21; All grades – N= 154

The students were also asked to write the mark they had obtained the previous year in the EFL subject and the mark they thought they would obtain at the end of the course. Many students in group 1 (N= 239) and fewer in group 2 (N=149) answered the question regarding the mark they thought they would obtain in English. Fewer students in both groups (N= 188 in group 1 and N= 102 in group 2) answered the question regarding the mark they had obtained in the EFL subject the previous year. This is probably because they could not recall it. Table 4.50 includes the means provided by the students regarding the mark they had obtained the previous year and the mark they thought they would obtain the same year they took the aptitude tests.

Table 4.50. Group 1 and 2 students' criterion variables ratings

Criterion variables	Stats	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	All grades
Last year's mark according to the students	**N**	45	54	41	38	92	290
	Mean	8.72	8.02	7.32	6.85	6.76	7.40
	SD	1.36	1.72	1.76	1.84	2.04	1.94
This year's mark according to the students	**N**	83	68	51	61	125	388
	Mean	8.57	8.28	7.56	7.15	7.40	7.62
	SD	1.33	1.28	1.52	1.48	1.94	1.61

Most students, except for some in grade 3, gave themselves a slightly higher mark than the one they had obtained the previous year. These means, though, are to be regarded with caution, as students answered the question about their last year's mark from memory, and the way the question related to the mark they expected to obtain at the end of the course was worded ambiguously (see sections 2.3.4 and 3.6.3).

Restricting the sample to only the participants who both marked themselves in relation to the mark they thought they would obtain in the present course and those whose teacher decided to give a probable mark for them in FL, it can be seen that the tendency is for teachers to give their students lower marks, as shown in the table below. The teachers' marks (Probable FL mark) and the students' self-assessment (this year's mark) differ almost up to two points in grade 3 and almost 1.5 in grade 4. The differences between these means are a little smaller between the other grades although the teachers' marks are consistently lower.

The correlations between these criterion variables (see Table 4.51) change a little from the ones presented above regarding the criterion measures of a general scope. In addition to grade 3, in grade 4 there are no significant positive correlations between any of the part or total scores on the MLAT-ES and these criterion measures. Besides, the significant positive correlations between these measures in grade 7 are also scarce. Notice, though, that the number of grade-4 and grade-7 teachers who provided their students' probable FL mark is very small (16 and 21 respectively) and, therefore, the correlations in these groups are not to be taken much into account. What is worth highlighting from these correlations is the fact that the significance and impact of the correlations differs very much between the teachers' and the students' marks considering the whole sample, but it does not change so much considering each grade separately. Correlations between the MLAT-ES score and the mark provided by the students are always lower but in the same direction, impact and significance (if any) except with *Parte 1*. It is also remarkable that this part only presents a significant

positive correlation with the probable FL mark in grade 6 and that there is a significant negative correlation with the marks provided by all the students in the group (r_s=-.348, p<.001). Moreover, there are some discrepancies between the scores which correlate significantly and those which do not, as *Parte* 4 in grade 6, which does correlate significantly with the mark provided by the teachers, but not with the mark provided by the students. Therefore, it can be concluded that self-reported measures should be used with caution, as results do not appear to be consistent when comparing them.

Table 4.51. Descriptive statistics and correlation between MLAT-ES scores and the marks provided by the students themselves and their teachers (pairwise)

Rating				MLAT-ES score				
Grade		**Mean**	**SD**	***Parte 1***	***Parte 2***	***Parte 3***	***Parte 4***	***Total***
3	**This year's mark**	8.84	1.09	-.012	-.115	.283	.183	.165
N=37	**Probable FL mark**	6.97	2.00	.141	.113	.181	.145	.250
4	**This year's mark**	8.22	1.37	.131	.049	-.171	.125	.015
N=16	**Probable FL mark**	6.78	1.54	.229	.157	.281	.475	.383
5	**This year's mark**	6.88	1.29	.095	.421*	.424*	.604**	.482**
N=33	**Probable FL mark**	5.83	1.69	.300	.507**	.539**	.709**	.635**
6	**This year's mark**	6.82	1.54	.179	.426**	.405**	.257	.414**
N=41	**Probable FL mark**	6.14	1.61	.387*	.512**	.481**	.316*	.567**
7	**This year's mark**	5.45	1.53	.320	.527*	.143	.034	.383
N=21	**Probable FL mark**	5.95	1.70	.154	.621**	.306	.150	.610**
All	**This year's mark**	7.30	1.76	-.348**	-.129	-.033	.031	-.149
N=148	**Probable FL mark**	6.32	1.78	.028	.219**	.260**	.274**	.243**

* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

In fact, the differences in the correlations between the marks provided by the students and those provided by the teachers could be due to the fact that the differences between these means were significant in all grades but 7, as proved the Wilcoxon signed-rank test between these marks (see Table 4.52). However, the MLAT-ES total score in grade 7 has a significant positive correlation with the grades provided by the teachers (r_s=.610, p<.001), but not with those provided by the students themselves (r_s=.383). In Table 4.52, some values appear in bold. These are the two-tailed significance value when it is significant and the effect size of the difference between tests if it is moderate or large.

Table 4.52. Wilcoxon signed-rank test between the FL marks provided by the students and those provided by the teachers

Group 1	Statistics	Grade 3 N= 37	Grade 4 N= 16	Grade 5 N= 33	Grade 6 N= 41	Grade 7 N= 21	All grades N= 148
This year's mark (student) / Probable FL mark (teacher)	**T**	31.50	11.00	.00	38.00	25.00	903.50
	Z	-4.253[a]	-2.612[a]	-4.227[a]	-3.880[a]	-1.450[b]	-6.116[a]
	Asym. Sig. (2- t.)	**.000**	**.009**	**.000**	**.001**	.147	**.000**
	Effect size	**-.49**	**-.46**	**-.52**	**-.43**	-.22	**-.36**

[a]Based on negative ranks
[b]Based on positive ranks

The number of questionnaires returned by the teachers of participants in group 2 was very small (only 14 forms were filled in). Consequently, Spearman rho correlations were not run with the teachers' criterion measures in this group. On the same note, as for the criterion measures reported by the students themselves, it was not possible to conclude anything from the correlations between the MLAT-EC scores and these marks, as most groups do not have 30 subjects, the number needed for a correlation to be reliable.

Besides the general and the language-related criterion measures, several language proficiency measures were administered in order to prove the concurrent validity of the MLAT-ES and the MLAT-EC. These measures were different depending on the grade to which they were distributed. To start with, subjects in grade 3 and 4 took the same cloze passage and listening tests whose histograms appear below on a 100-point scale.

Figure 4.11. Histogram of the cloze passage – grade 3

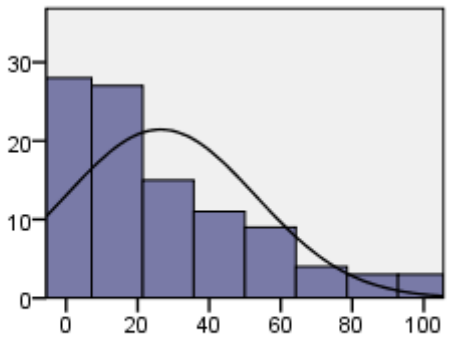
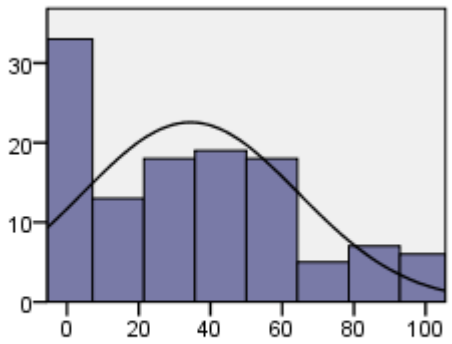


Figure 4.12. Histogram of the cloze passage – grade 4



The mean *p*-values of the cloze passage were .26 for grade 3 and .34 for grade 4. This means this test was difficult for these grades, especially for 3-graders. Although

the children were not used to this type of activity, in both grades some children got the highest score possible. Hence the large standard deviations, as the mean in grade 3 is lower than the standard deviation (N=100, M=26.43, SD=26.58) and results are positively skewed (*S*=1.059) and distributed in a pointy shape (*K*=.441). In contrast, the standard deviation becomes smaller in grade 4 (N=119, M=48.25, SD=22.18), and although results are still positively skewed (*S*=.522), their distribution tends to be platykurtic (*K*=-.657).

Correlations between the cloze passage and the scores obtained on the MLAT-ES by group 1 appear in Table 4.53 below along with the percentage descriptive statistics of the cloze passage. The cloze passage is a proficiency measure considered to be integrative, that is to say, representative of general proficiency in the FL language. There is a significant positive correlation between the partial and total scores on the MLAT-ES and the cloze passage in all the parts but *Parte 1* in grade 3 and *Parte 4* in grade 4. *Parte 1* is meant to measure vocabulary, which is a relevant aspect to solve the cloze passage, and sound-symbol association, which was a linguistic skill not required in order to complete the cloze passage successfully. Nevertheless, in grade 4 and when considering grades 3 and 4 together, *Parte 1* presents some kind of relationship with the scores on the cloze passage. Sensitivity to grammatical structure, the ability measured by *Parte 2*, is needed to fill in the blanks of the cloze passage. This part presents significant moderate correlations in grade 3 and considering both grades 3 and 4 together, and two significant low correlations in grade 4 (r_s=.276 and r_s=.277, p<.05). *Parte 3* also presents significant positive correlations with scores on the cloze passage even though this part measures the ability to hear speech sounds, which is not needed to answer the cloze test. *Parte 4*, which measures rote memory, correlates moderately with the cloze passage scores in grade 3 and considering both grades 3 and 4 together, but it does not in grade 4. Along with *Parte 3*, the total score correlates significantly with the cloze passage in grades 3 and 4 at p<.001 with moderate values ranging from r_s=.359 to r_s=.518.

Table 4.53. Descriptive statistics and correlations between MLAT-ES scores and the cloze passage for grades 3 and 4

Grade	N	Mean	SD	Parte 1	Parte 2	Parte 3	Parte 4	Total
3	43	20.93	24.62	.051	.499**	.426**	.401**	.518**
4	57	39.85	28.79	.277*	.276*	.495**	.118	.359**
3 & 4	100	31.71	28.54	.387**	.432**	.542**	.293**	.516**

* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

The distribution of the Spearman rho correlations obtained between the cloze passage scores and the MLAT-EC scores is very similar except for two salient differences. First, while *Parte 1* did not significantly correlate with the cloze passage in grade 3, there is a significant moderate correlation between its analogous *Part 1* and the cloze passage (r_S=.517, p<.001). Second, there is no significant correlation between *Part 2* and the cloze passage (r_S=.155) although both measures correlate significantly at a low-moderate level in grade 4 and considering both grades together. As for the rest of correlations between the scores on the MLAT-EC and the cloze passage, they are consistently lower in group 2, but most of them are moderate, except for the following correlations: between *Part 2* and the cloze passage considering grades 3 and 4 together (r_S=.217, p<.05) and between *Part 4* and the cloze passage considering again both grades together (r_S=.246, p<.001). As happened in group 1, there is not significant correlation (r_S=.160) between *Part 4* and the cloze passage in grade 4. The total score presents once again significant positive correlations across grades and considering both grades 3 and 4 as one.

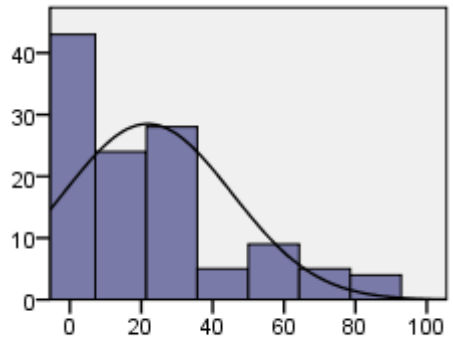
Table 4.54. Descriptive statistics and correlations between MLAT-EC scores and the cloze passage for grades 3 and 4

Grade	N	Mean	SD	Part 1	Part 2	Part 3	Part 4	Total
3	57	30.58	27.46	.517**	.155	.319*	.383**	.509**
4	62	29.49	30.59	.387**	.346**	.332**	.160	.423**
3 & 4	119	30.01	29.02	.353**	.217*	.301**	.246**	.362**

* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

The cloze passage administered to children in grades 5, 6 and 7 varied in the number of items. The 5-graders' version had 7 items and the one for 6- and 7-graders had 24 items. This test was rather difficult for all groups, as the mean *p*-value for grade 5 was .22, and .20 and .29 for grades 6 and 7 respectively. Mean scores are low in all grades because this test, along with the listening and the dictation tests, is meant to be used by adolescents and adults learning English as a FL. The 118 5-graders' percentage mean was 21.79, with a standard deviation larger than the mean (SD=23.60) due to the extreme non-normal distribution of the data. Scores are positively skewed and with a mid-leptokurtic distribution (S=1.094, K=.457).

Figure 4.13. Histogram of the cloze passage – grade 5



Grades 6 and 7 took the same version of the cloze passage. 6-graders' percentage mean score is lower than 7-graders' (grade 6, N=120, M=20.00, SD=18.41; grade 7, N=131, M=29.13, SD=21.20). The scores are positively skewed in both grades (grade 6 *S*=1.371, grade 7 *S*=.615), but they are more spread in grade 7 (*K*=-.336) than in grade 6 (*K*=1.920).

Figure 4.14. Histogram of the cloze passage – grade 6

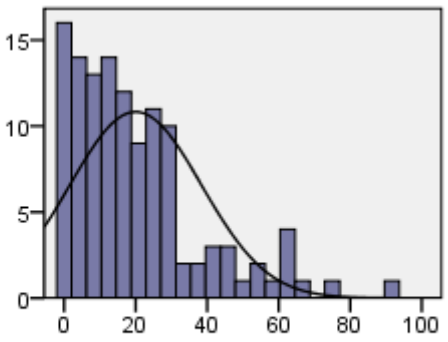
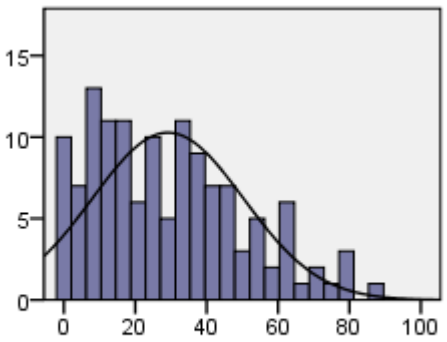


Figure 4.15. Histogram of the cloze passage – grade 7



Grades 5, 6 and 7 took the same listening test, which, after the discrimination and reliability analysis, was rendered with 12 items. If this test was mid-difficult for 5-graders (N=118, mean *p*-value=.47, M=46.69, SD=18.78), it was easier for 6- and 7-graders (mean *p*-value .58 and .63 respectively). The difference between the means of these two grades is smaller than the one between grade 5 and grade 6 (grade 6, N=120, M=58.19, SD=18.27; grade 7, N=131, M=63.30, SD=17.21). In all the grades the asymmetry in the means scores is negatively skewed (grade 5 *S*=-.092, grade 6 *S*=-.129, grade 7 *S*=-.480) and kurtosis values are also negative for the three grades (grade 5 *K*=-.285, grade 6 *K*=-.076, grade 7 *K*=-.400).

Figure 4.16. Histogram of the listening test – grade 5

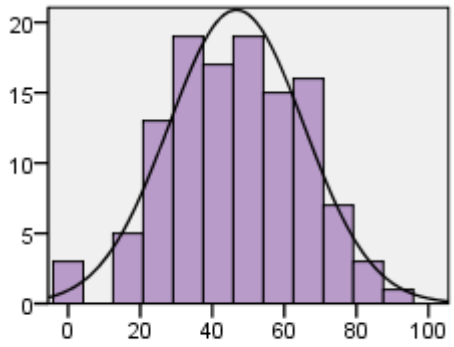


Figure 4.17. Histogram of the listening test – grade 6

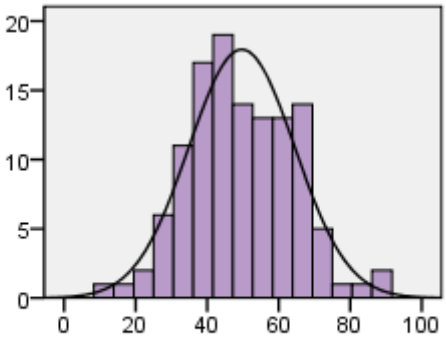
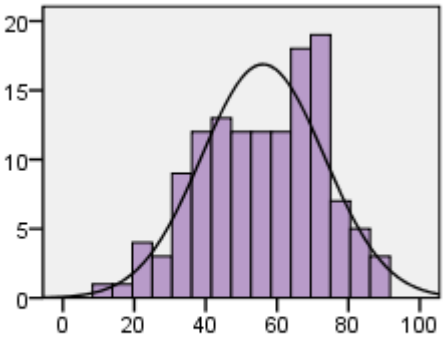


Figure 4.18. Histogram of the listening test – grade 7



The dictation was also easier the higher the grade. It was a difficult test for 5-graders (mean *p*-value .39) and mid-difficult for 6- (mean *p*-value .48) and 7- graders (mean *p*-value .57). The means increase 10 points average across grades (grade 5 N=82, M=37.96, SD=18.44; grade 6 N=96, M=47.95, SD=20.03; grade 7 N=94, M=57.51, SD=20.28). The tendency is for scores to be positively skewed in grade 5 (*S*=0.29) but negatively skewed for 6- and 7-graders' (*S*=-.126 and *S*=-.242 respectively). In the three graders, the distribution tends to be platykurtic as kurtosis values are negative in grade 5 (*K*=-.892), grade 6 (*K*=-.689) and grade 7 (*K*=-.671).

Figure 4.19. Histogram of the dictation – grade 5

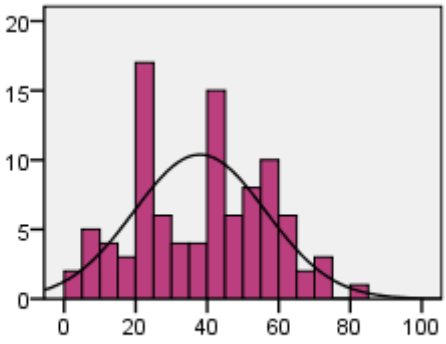


Figure 4.20. Histogram of the dictation – grade 6

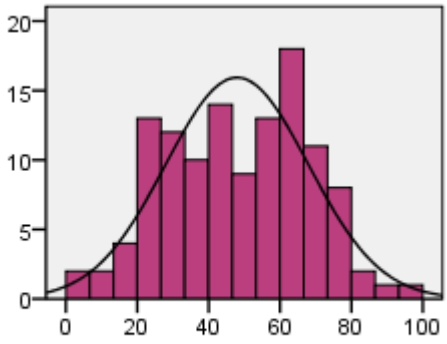
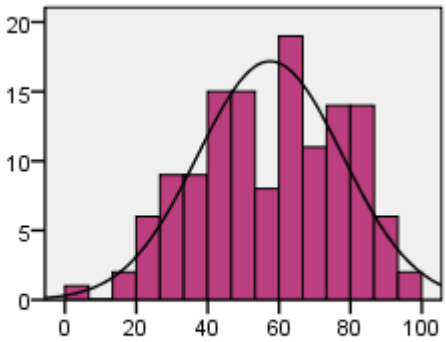


Figure 4.21. Histogram of the dictation – grade 7



Descriptive statistics and two-tailed Spearman correlations between the aptitude tests and the proficiency measures for grades 5, 6 and 7 appear in Table 4.55. Notice that only 35 children did the dictation in grade 5, while the number of children in the other groups is larger.

Table 4.55. Descriptive statistics and correlations between MLAT-ES scores and the proficiency measures for grades 5, 6 and 7

Proficiency measures	Grade	N	Mean	SD	*Parte 1*	*Parte 2*	*Parte 3*	*Parte 4*	*Total*
Cloze passage	**5**	57	20.05	24.28	.490**	.351**	.511**	.517**	.581**
	6	60	18.75	18.40	.214	.376*	.163	.192	.244
	7	67	21.33	18.04	.344**	.371**	.330**	.486**	.466**
	6&7	127	20.11	18.19	.299**	.379**	.245**	.350**	.366**
Listening	**5**	57	48.10	20.35	.338*	.324*	.274*	.386**	.409**
	6	60	61.94	18.50	-.003	.358**	.231	.145	.157
	7	67	59.20	19.36	.189	.278*	.257*	.113	.315**
	5, 6 & 7	184	56.66	20.17	.361**	.331**	.339**	.359**	.420**
Dictation	**5**	35	41.49	19.66	.402*	.359*	.487**	.628**	.589**
	6	60	52.23	18.45	.327*	.458**	.462**	.279*	.446**
	7	67	53.37	17.28	.513**	.360**	.337**	.229	.434**
	5, 6 & 7	162	50.38	18.74	.414**	.426**	.433**	.364**	.493**

* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

Table 4.55 shows that there are significant moderate correlations between the cloze passage and all the part and total scores on the MLAT-ES, except in grade 6, in which there is a significant positive moderate correlation between this proficiency measure and only *Parte 2* (r_s=.376, p<.05). However, the cloze passage does not correlate with the rest of the parts, not even the total score. In principle, the cloze passage should be prone to correlate with *Parte* 2 due to its grammatical nature, with *Parte 1* because it measures vocabulary, and with the total score because of it being of an integrative sort. *Parte 4* is also considered to measure some kind of analytic skills in order to be able to encode the numbers and to retrieve them easily later. Along with the total score, *Parte 4* is the part that overall correlates the highest with the cloze passage, except for grade 6.

The listening test is the proficiency measure that obtains the lowest correlations with the partial and total scores on the MLAT-ES, especially in grades 6 and 7, leaving the non-significant correlations with the cloze passage in grade 6 aside. More specifically, correlations between this proficiency measure and the total score in grade 6 are not significant (r_s=.157), they are moderate-low in grade 7 (r_s=.315, p<.001). The rest of correlations with this proficiency measure are, with some exceptions, mostly from low to moderate, ranging from r_s=.257, p<.05 with *Parte 3* in grade 7 to r_s=.429, p<.001 with the total score considering grades 5, 6, and 7 together. *Parte 2* is the only part with which this test correlates significantly across grades. It is remarkable that there are no significant correlations between the listening measure and *Parte 4* in grades 6 (r_s=.145) and 7 (r_s=.113), even though this part is supposed to tap for auditory

alertness. The correlations between the listening and *Parte 1* are also particularly low, reaching a negative value in grade 6 (r_s=-.003). No significant correlation is found either between the listening and *Parte 1* in grade 7 (r_s=.189) nor with *Parte 3* in grade 6 (r_s=.231).

The dictation is the proficiency measure that, with some exceptions, has the highest correlation with both the partial and the total scores on the MLAT-ES, the coefficients ranging from r_s=.279, p<.05 with *Parte 4* in grade 6 to r_s=.589, p<.001 with the total score in grade 5. Although the format of *Parte 4* is a dictation, it does not correlate significantly with this measure in grade 7 (r_s=.229, p<.001) and it obtains the lowest correlation in grade 6 (r_s=.279, p<.05). This could be due to the fact that this measure became extremely easy for the higher grades and, therefore, it may not have enough discriminatory power at this level. *Parte 1*, which aims at measuring sound-symbol association ability, and *Parte 3*, which aims at measuring the ability to hear speech sounds, consistently obtain significant moderate correlations with this proficiency measure, and so does *Parte 2* across grades. No clear pattern can be observed as to whether the correlations are higher for one part in particular. None of the grades seems to stand out for obtaining higher correlation coefficients over the rest either. Taking into account the cases of no significant correlations already mentioned, *Parte* 4 seems to be the one that correlates the least in grades 6 and 7 but not in the other grades, either separately or grouped.

As can be seen in Table 4.56, most correlations between the MLAT-EC and the proficiency measures are moderate and significant, with some exceptions.

Table 4.56. Descriptive statistics and correlations between MLAT-EC scores and the proficiency measures for grades 5, 6 and 7

	Grade	N	Mean	SD	Part 1	Part 2	Part 3	Part 4	Total
Cloze passage	**5**	61	23.42	23.04	.227	.480**	.295*	.344**	.445**
	6	60	21.25	18.50	.491**	.349**	.474**	.135	.497**
	7	64	37.30	21.30	.610**	.648**	.572**	.293**	.680**
	6&7	124	29.53	21.48	.595**	.521**	.464**	.329**	.620**
Listening	**5**	61	45.36	17.25	.075	.226	.262*	.193	.237
	6	60	54.03	16.98	.416**	.306*	.403**	.136	.423**
	7	64	68.36	17.26	.474**	.388**	.387**	.341**	.485**
	5, 6 & 7	185	56.13	19.58	.392**	.395**	.416**	.368**	.468**
Dictation	**5**	61	35.93	17.55	.413**	.356**	.454**	.597**	.544**
	6	60	43.67	20.77	.596**	.557**	.544**	.133	.656**
	7	64	61.84	22.33	.581**	.644**	.585**	.378**	.704**
	5, 6 & 7	185	47.41	23.03	.559**	.521**	.484**	.416**	.631**

** Correlation is significant at 0.01 (two-tailed)

Starting with the cloze passage, and as it happened in group 1, it does not correlate significantly with *Part 4* in grade 6 and it does not correlate either with *Part 1* in grade 5. The rest of the correlations with this proficiency measure are significant and mostly moderate, the highest one being r_s=.680 at p<.001 with the total score.

Once again, the listening measure is the one that correlates overall the lowest with the part and total scores on the aptitude measure. There are no significant correlations between this test and any of the aptitude test parts in grade 5 except in *Part 3* (r_s=.265 at p<.05). Perhaps the listening test was too difficult in relation to the knowledge this grade had of the FL at the moment of taking the test. The correlation between the listening and *Part 4* in grade 6 is also low (r_s=.136), but the rest are all moderate, including those with *Part 3*, which has to do with the abilities needed in listening activities, and the total score across grades except in grade 5, as already pointed out.

The correlations between part and total scores on the MLAT-EC and the dictation are all moderate and significant at p<.001 except for one case: *Part 4* in grade 6. This lack of significance in this part could be explained in this case due to the fact that grade 6 showed an unexpected behaviour in this part, scoring less than grade 5 (grade 5 M=19.44; grade 6 M=19.28). Actually, of all the parts in the MLAT-EC, *Part 4* is the one that correlates the lowest with the proficiency measures in this study.

Comparing the correlations between the aptitude measures and the criterion measures on the one hand, and the proficiency measures on the other hand, correlations between the aptitude measures (both the MLAT-ES and the MLAT-EC) and the criterion measures were rather homogeneous except for those with the FL probable mark at the end of the course, which is a measure that implies factors other than ability in the FL. Most of the correlations were moderate and significant across grades. Nevertheless, it must be pointed out that the teachers' scores are not entirely reliable, as they may take into account more academic-specific aspects. In addition, the teachers of grades 3 and 4 failed to see significant differences in their students' language-specific abilities. Therefore, their conception of their students' abilities may not be discriminative enough regarding the different skills involved in FL learning. Actually, while the MLAT-ES (except *Parte 1* and *Parte 4* in some cases) tends to correlate significantly with the language-specific criterion measures, this correlation is lost when using the variables of a more general scope, provided both by the teachers and the students, not only in grade 3 but also in grades 4 and 7. Moreover, while no

significant correlations between the language-specific criterion measures and the MLAT-ES (both part and total scores) are found, the MLAT-ES seems to be related to the grade-3 participants' FL proficiency as measured by the cloze passage, which is considered to be an integrative FL proficiency measure.

The construct validity of the MLAT-ES and the MLAT-EC tested by using the language proficiency measures appears to be proved in all grades. Partial and total scores correlate significantly and moderately across grades except some specific cases (mainly grade 6 in group 1 with the cloze passage and the listening test, and grade 5 in group 2 with the listening test). The listening test does not present correlations as high as the other proficiency measures, especially with part 4. This could be due to the characteristics of the proficiency measure (a rather difficult one) and/or due to part 4 itself, as it proved to be the easiest of the parts in the validity study.

4.5. Summary of the findings

The first research question asked whether the MLAT-ES and the MLAT-EC are equally suitable for grades 3 to 7, as not only in this study, but also in the *Manuals* of the MLAT-ES and the MLAT-E, it had been observed that mean scores did not develop in the same way across all grades. That is, between grades 3 and 4 the increase in the means was always higher than the increase between grades 4 to 6. In addition, the results seemed to reach a plateau between grades 6 and 7. After running Mann-Whitney U tests, it was confirmed that the tendency is for the differences in the mean scores on the MLAT-ES between grades 3 and 4 to be significant most times except in *Parte 4.* Across the rest of grades, if the difference between the means is significant, it appears most times when looking at the 1-tailed value, not in the 2-tailed. The differences in the means obtained by 6- and 7-graders are the smallest and they were only significant in *Parte 1* (2- and 1-tailed) and in the total score (1-tailed). When calculating the differences in percentages and the effect sizes, the largest values are found between grades 3 and 4, especially in *Parte 1* and *Parte 2.* The percentage changes between grades 3 and 4 in *Parte 3* and *Parte 4* are small, yet larger than in the rest of grades, and in the total score the difference expressed in percentages is medium with a large effect size.

Regarding the MLAT-EC, similar patterns to those of the MLAT-ES are found: the most significant differences between the means when comparing contiguous grades are found between grades 3 and 4, although the grades that obtain the least significant differences — actually none — are grades 5 and 6, not 6 and 7, as opposed to what happened on the MLAT-ES. When calculating the differences between the means in percentages, this time the means in grade 4 are consistently larger than those obtained by 3-graders in all parts and the effect sizes are from medium to very large. The differences between grades 6 and 7 are small while those between grades 5 and 6 are negligible in all parts but in *Part 3.*

Part 2 in both the MLAT-ES and the MLAT-EC is the only part that presents a medium increase across other grades. The means increase at a medium level between grades 5 and 6 on the MLAT-ES and between grades 4 and 5 on the MLAT-EC.

The results obtained on the MLAT-ES and the MLAT-EC are consistent with those in the data provided in the *Manuals*. Comparisons between groups were possible by calculating the percentage change and the effect size using Cohen's *d.* Several issues deserve to be discussed as to why such large differences are found between grades 3 and 4 and why they gradually disappear across grades, except in part 2, in which mean scores intermittently experience medium increases across certain grades. Also, the Mann-Whitney U tests run comparing the mean total scores of grades 5 and 7 show that there are significant differences between these two grades. Therefore, if we had to determine which point is the onset of aptitude stability as measured by the MLAT-ES and the MLAT-EC, this point would be grade 6.

The second research question was whether there were significant differences between the mean scores obtained by boys and girls. The question aroused when comparing the *Manuals* of the MLAT-E and the MLAT-ES, as the former makes this distinction while the latter does not. Results show that, as opposed to the results of the MLAT-E norming study, the mean scores obtained by boys tend to be higher than those of girls'. When running a Mann-Whitney U test, only one significant difference is found at a time in each part and in different grades and tests, and this difference sometimes favours boys and sometimes favours girls. When calculating the percentage changes in the means of the MLAT-E norming study, only one medium increase and one large increase is found as well, always in grade 4, the rest being either negligible or small increases. Consequently, sex does not seem to be a relevant variable when measuring aptitude with the elementary version of the MLAT in any language.

Checking the concurrent construct validity of the MLAT-ES and the MLAT-EC is the aim of the third research question. As for the MLAT-ES, this was done by running

correlations between the teachers' criterion variables, the students' self-ratings and the proficiency measures collected for this study. After describing the criterion variables provided by the participants' teachers and those provided by the participants themselves, it has been seen that the criterion the teachers use is different from that of their students and, consequently, the correlations between these two different kinds of measures with the MLAT-ES do not fully coincide. Nevertheless, there is one constant found in both correlations: the correlations for grade 3 are the lowest and not significant with any of the aptitude test part scores, and correlations tend to be lower when considering the whole group together, not per grades. *Parte 1* is also the part with the lowest, if any, significant correlations, while *Parte 2*, *Parte 3* and the total score are those which obtain the highest correlations with the MLAT-ES across grades.

The proficiency measures collected (a listening test and cloze passage in all grades, in addition to a dictation in grades 5 to 7) make it possible to check the construct validity not only of the MLAT-ES but also of the MLAT-EC. In both cases, the listening test is the one that seems to correlate the lowest with the aptitude measures while the dictation is the one that seems to correlate the highest. As opposed to the criterion measures and the students' self-ratings, the proficiency measures used in grade 3 do obtain significant moderate correlations with the aptitude measures. Both in the MLAT-ES and the MLAT-EC, part 1 and part 4 are those that obtain the lowest correlations with the proficiency measures used. The total score on the aptitude measures, in contrast, is the score that correlates significantly and the highest with the proficiency measures administered except with the listening test.