UNIVERSITAT DE BARCELONA

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

CHAPTER 3: METHODOLOGY

3.1. Introduction

This chapter deals with the methodology of the current dissertation. First, the characteristics of the research context and of the participants in the study are described. This is followed by a presentation of the aptitude measures used, how they were administered and a thorough validation study of these aptitude measures taking into account the context in which they were administered. Following the information about the aptitude measures, the instruments used to measure the participants' proficiency in English are described. These involve English proficiency measures administered to the participants themselves and of questionnaires that their teachers answered rating their students. This chapter finishes with the procedure followed to administer all the measures described and how these were corrected taking into account the age of the participants.

3.2. Research context

The population sample for this study is a convenient sample of the bilingual context in which the study was carried out, Catalonia. The majority language in Catalonia is Spanish, known by practically the totality of the population. Catalan is the minority language. The data were collected in seven different schools, five of which located in the metropolitan area of Barcelona (in the cities of Barcelona and Badalona) and two of which in the province of Lleida (in the towns of Bellpuig and Castellnou) so as to have some dialectal variability in the sample. The latter have been considered as one school because most students of the primary school go to the same secondary school from grade 7 on. Although in both areas both Catalan and Spanish are spoken, the percentage of use of each language as a means of communication and instruction varies depending on the teacher, the school and environmental and demographical factors. Thus, the subjects from the area of Barcelona may be more influenced by Spanish, because of the influence of the media and of the high rate of Spanish-speaking immigration, whereas the majority of subjects from Lleida are Catalan-dominant, including the immigrant population. On the other hand, the participants from Lleida speak a different variety of Catalan from the participants from Barcelona. While those in Barcelona speak what is known as standard Catalan, the one that is also

spoken in most media, people from Lleida speak the North-western variety. Among other specific features, what distinguishes both varieties is the pronunciation of unstressed vowels. While in the central variety unstressed <e, a> are pronounced /ə/ and unstressed <o> is pronounced /u/, in the North-western variety they tend to be pronounced as they are written, i.e. /e, a/ and /o, u/, except in some specific cases.

In Catalonia, both Catalan and Spanish are taught formally at primary and secondary school and the language used to teach the rest of subjects is Catalan. Compulsory infant education begins at the age of 3, and it constitutes an immersion period for children from Spanish-speaking families. English is the first foreign language of the schools in this study. It is taught formally 3 hours per week from grade 1 although, in some cases, the onset age was earlier or later. The total number of hours varies from school to school according to the policy they have adopted regarding EFL. Some schools have decided to offer some other curricular subjects in English (such as Maths, Arts and Crafts, Natural Sciences, etc.) besides the English language class adopting a Content and Language Integrated Learning (CLIL) approach. Table 3.1 exposes the number of EFL hours that the participants in this study received at school per week.

Table 3.1. Number of curricular hours of EFL per week according to school

Schools	Infant school			Primary school						Secondary school
	P - 3	P – 4	P – 5	1	2	3	4	5	6	7
1	3h	3h	3h	3h	3h	3h*	3h*	3h*	3h *	3h *
2	2h	2h	2h	3h	3h	3h	3h	4h	4h	4h
3				3h	3h	3h	3h	3h	3h	3h
4				3h	3h	3h	3h	3h	3h	3h
5			45'	1h 30'	2h	2h 30'	2h 30'	3h	2h30'	-
6	1h	1h 30'	1h 30'	3h	3h	3h	3h	4h	4h	3h

* Plus 1h of CLIL exposure to the FL

As not all the participants in the study had always attended the same school, they were also asked at what age they had started learning English. 156 participants reported they had started when they were in *Pàrvuls – 3* (P-3), which is the equivalent to Foundation Stage 1 in the United Kingdom education system or Pre-school in the United States. In this grade children are between 3 and 4 years old. 231 reported they had started when they were in Grade 1 while, 74 said they had started at Grade 3 and the rest of participants affirmed they had started at different ages (see Table 3.2).

Table 3.2. Participants' age of onset of FL learning at school

Grade (at data collection)	Starting age of FL learning at school							
	P - 3	P - 4	P - 5	Gr. 1	Gr. 2	Gr. 3	Gr. 4	Gr. 5
3	45	5	10	40	18	5	0	0
4	51	9	7	38	16	15	1	0
5	25	5	15	45	12	15	1	0
6	18	7	19	48	3	22	2	1
7	17	3	27	60	7	17	0	0
Total N	156	29	78	231	56	74	4	1

Besides exposure to the FL in the school context, 253 out of 629 subjects in the study were also attending or had attended extra-curricular classes of English, most of them during 2 or 3 hours per week (see subjects who were receiving extracurricular exposure in Table 3.3).

Table 3.3. Subjects receiving extracurricular exposure to EFL across grades

Extracurricular exposure	Grade					
	3	4	5	6	7	All
yes	35	68	45	47	58	253
no	88	69	73	73	73	376

In all grades there is a great variety in the participants regarding the type and amount of exposure to the FL they have had. In addition, P-3, grade 1 and grade 3 are the grades in which most students started to learn English as a foreign language, although the rest of participants started in courses in-between. This makes it difficult to distribute the population in homogeneous groups, as no general trends are observed in both these variables at once. As this variety in these factors related to FL learning and teaching is present in all grades, the groups of subjects will be kept following the conventional variable "grade". Besides, the studies published using the MLAT-E divide their cohorts according to this same criterion.

3.3. Participants

Although the battery of tests was administered to approximately 750 subjects, only those subjects who did the whole test series were included for the reliability and validation analysis of tests. The subjects whose L1s were not Catalan and Spanish were not included in the sample and neither were those who deliberately answered the

tests wrongly, that is, those who were caught in the act of crossing any option without reading the test or they left almost everything blank. The participants are considered as two groups (group 1 and group 2) as far as the aptitude measures are concerned. Group 1 is formed by those students who took the Spanish first and then the Catalan version of the MLAT-E, and group 2 is formed by those who took the tests in reverse order. This division is necessary so that the norming study of both versions of the test can be done.

The age range of these subjects is 8.3 – 14.9. They are distributed as follows according to grades:

Table 3.4. Mean age and SD across grades according to the order of administration of aptitude measures

Grade	All subjects			Group 1			Group 2		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
3	123	8.8	.52	66	8.8	.33	57	8.8	.66
4	137	9.8	.43	75	9.9	.48	62	9.7	.36
5	118	10.8	.33	57	10.9	.33	61	10.8	.34
6	120	11.8	.33	60	11.7	.34	60	11.9	.33
7	131	12.9	.45	67	12.9	.45	64	12.8	.44

Although all the participants are bilingual, 237 participants declared themselves to be Catalan-dominant, 221 said they were Spanish-dominant and 172 declared themselves to be balanced bilinguals. Table 3.5 shows the distribution of participants in group 1 and 2 and all together depending on their language preference.

Table 3.5. Subjects' language preference across grades

Grade	All subjects			Group 1			Group 2		
	Catalan	Spanish	Both	Catalan	Spanish	Both	Catalan	Spanish	Both
3	46	47	30	25	25	16	21	22	14
4	51	46	40	21	30	24	30	16	16
5	45	40	33	16	24	17	29	16	16
6	45	37	38	19	21	20	26	16	18
7	48	52	31	7	35	25	41	17	6
All	235	222	172	88	135	102	147	87	70

The questions that determined their language dominance were the language they usually speak with their mother, father, brothers, sisters and rest of the family, and with their friends. They were also asked which language they feel more comfortable

with. To confirm their language preference, they were asked the same questions twice, in the first and in the last data collection sessions.

In Table 3.6 is the distribution of subjects according to their language preference and school. School 4, the school in Lleida which forms part of group 2, is the school with most students preferring Catalan over Spanish (88%). Also, the cohort of school 3, in Barcelona, is a more Spanish-dominant one (59.8%). The students of this school are distributed in group 1 and group 2. In contrast, the distribution according to language preference taking the group as a whole is more balanced.

Table 3.6. Subjects' language preference per schools

Language preference	Schools						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
Catalan	41	19	22	73	51	29	235
Spanish	31	57	119	2	7	6	222
Both	33	28	58	8	19	26	172
Total	105	104	199	83	77	61	629

Due to the balanced distribution of the population regarding their language preference, the participants will be considered as one group regarding this aspect, although it is true that the participants' language preference may play some kind of role in their performance on the aptitude tests, as they are language-based.

3.4. Instruments of aptitude: the MLAT-E in Spanish and Catalan

As the subjects of this study are Catalan and Spanish bilinguals, two versions of the MLAT-E were used: one in Spanish and one in Catalan. The MLAT-E was chosen because, besides being grounded on psychological constructs known in the literature of the field, the translation into Spanish was at hand thanks to the work by the SLTI staff. The author of this dissertation offered herself to administer some tests to help the validation process of this translated version. This new version in Spanish made it easier to translate it into Catalan, as Catalan and Spanish are close Romance languages.

The Catalan version of the MLAT-E (from now on MLAT-EC) was adapted and translated from the Spanish one (not from the English original version), not only

because of the linguistic proximity, but also because the cultural references in the test were also proximal. In addition, since the MLAT-ES was distributed in the first place during the first year of data collection, the adaptor of the test in Catalan could already take advantage of the observations she made when administering the test as well as of some of the students' reactions when taking the MLAT-ES. The test adaptor, and author of this dissertation, worked on her own but distributed the test to two other experienced researchers and to her thesis advisor to receive some feedback on the test development.

The parts of the MLAT-EC are the same as the ones in the MLAT-E and the MLAT-ES. What follows is an overview of the decisions taken when adapting and translating the test into Catalan so as to avoid the minor mishaps that occurred in the administration process of the MLAT-ES.

Even though some problems had been detected in relation to the format of the test (font size and type, instructions at the bottom of the page and so on), the format of the Catalan version was kept as it was in the original version of the MLAT-ES. This is because both tests had to be comparable for further analysis and no major changes could be made in order to be able to consider both tests as something as close as possible to a form A and a form B of the same test. The only two changes that were made regarding format were adding "*El meu nom i cognoms són*" (My first name and family names are) — where test takers have to write their name and surnames — and adding the missing line for item 25 in part 4. In this part, letter <ñ> was skipped, as it does not exist in Catalan. The directions of the test were adapted from the MLAT-ES and recorded on a CD in the Phonetics Laboratory of the University of Barcelona. The speaker was a student of Romance Philology at the University of Barcelona (UB) who was an experienced radio presenter. In it, the pause time in between parts and directions appears in red[13]. The speaker read the instructions with no stops and then the pauses were added using the software available at the Phonetics Laboratory in the *Facultat de Filologia* of the UB (see appendix C for the transcription of the text on the compact disc and some sample items from the MLAT-EC).

Part 1 is entitled *Paraules ocultes* and it contains essentially the same stems and distractors as the MLAT-ES. Translating them implied, in some cases, substantial changes in the word form because, although Catalan and Spanish are Romance languages, not all the words in the MLAT-ES are cognates in Catalan. Consequently, in order to keep the same cognitive challenges in the items while keeping the original distractors as much as possible, some stem words were changed. For instance, item 7

on the MLAT-ES ddo – *está en la mano*, which refers to "*dedo*" (finger) was changed for ungl, which refers to "*ungla*" (fingernail), because the correspondent word in Catalan "*dit*" did not offer many possibilities to change its spelling. Then, the distractor "*para jugar*" did not make much sense, so it was changed for "*part d'un triangle*", taking into account that *ungl* could be wrongly interpreted as "*angle*" (angle), which is a word that belongs to the semantic field of geometry.

7. ungl ☐ despertador ☐ està a la mà

☐ part d'un triangle ☐ dur

So as to keep phonetics and spelling closer, the stem *esepzional* in item 19 on the MLAT-ES was changed into *eccpcional*.

19. eccpcional ☐ tipus de triangle ☐ únic

☐ emocionant ☐ mur alt

Item 18 *yielo* ("*hielo*" - ice)*,* which corresponds to "*congelado*" was not kept as Catalan "*gel*", but was changed for the adjective of the same family ("*gelat*" hidden as "*glat*") because, if kept as "*gel*", the only alternative to create the stem was to remove the vowel (gl). Then, these two letters could have been confusing because of the confusing graphic presentation of <l> and because they could have corresponded to several words (*gel, gol, gla*), not just one as it happens with "*yielo*".

18. glat ☐ molt fred ☐ gros ☐ sucre ☐ perfumat

One of the distractors in item 29 was also changed. Stem 29 *cirrena* had as a distractor "*cerrada*" (closed), which shares the <c> and the <rr> with the stem. The translation into Catalan "*tancada*" did not bear any phonetic similarity with the stem. One word that was phonetically similar to "*sirena*" (siren) was "*serena*" (calm), which could have shown an unexpected behaviour as a distractor because it is too close phonetically speaking to the stem. Consequently, the distractor was changed for "*fruita*" (fruit), thinking that test takers could think that *cirrena* referred to "*cirera*" (cherry).

---

[13] The directions are in appendix C and the audio file is on the CD enclosed.

29. cirrena   ☐ camina de pressa   ☐ estri

             ☐ fruita   ☐ viu al mar

The confusion detected between "*àguila*" (eagle) and "*aglà*" (acorn) was avoided by changing the target of item 11 for "*corb*" (crow), and the right response alternatives became "*ocell negre*" (black bird) instead of "*pájaro grande*" (big bird).

11. korb   ☐ corre ràpid   ☐ ocell negre

            ☐ aliment   ☐ pregunta

The stem of item 12 in Spanish *eccitoso* was also changed, as the adaptor considered that the Catalan equivalent "*reeixit*" was not so common a word, so it would add extra difficulty to the Catalan version that the MLAT-ES did not have. The final stem was *fmos*, which refers to the word "*famós*" (famous), and all the response alternatives were kept except for the correct answer, which became "*conegut*" (well-known).

12. fmós   ☐ oliós   ☐ salat   ☐ poc comú   ☐ conegut

The same rationale was followed when changing the stem of item 17 *rmozo* for "bnic", which refers to "*bonic*" (nice), as "*formós*" is not so common a word, and "*bell*", which is a synonym, is homophone with *vell* (old), and that would have involved changing the whole item.

17. bnic   ☐ jove   ☐ maco   ☐ ample   ☐ fresc

One of the distractors of the hidden work *bakka* ("*vaca*" – cow) was "*con manchas*" (spotted), which caused some sort of confusion among the test takers, as some cows are spotted. So as to avoid the confusion that "*con manchas*" arose, this distractor was changed for "*amb arrugues*" (wrinkled).

15. bakka   ☐ és dolça   ☐ part de la cara

            ☐ fa llet   ☐ amb arrugues

One distractor that was interchanged with the stem was the one in item 30. The Spanish stem is dzpacio ("despacio" – slow / slowly), which becomes an adverbial locution in Catalan *("a poc a poc")*. That is why the stem was changed for lntamnt ("*lentament*" – slowly), which corresponded to the response alternative "*a poc a poc*".

30. lntamnt ☐ ample ☐ blau ☐ a poc a poc ☐ rodó

Part 2 Paraules que es corresponen was translated literally from the Spanish version, assuming that the test takers would not remember the sentences in the test 3 or 4 months after the first administration. In the Catalan version, the words that had caused some kind of confusion in the administration of the Spanish version were modified, so "Leila" becomes "Laura" and "Perla" becomes "Paula". The use of the subject pronoun in item 24, although it sounds as redundant as in Spanish, was kept so as to keep the same number and position of the distractors of this item.

Part 2 posed some problems when adapting it from Spanish because, if translated literally from the Spanish version, the number of words increased when translating verbs, possessive adjectives and proper nouns into Catalan. The Spanish past tense is formed with only one word ("*ellos compraron*"), whereas the correspondent Catalan past tense used by default consists of two words: the auxiliary verb "*anar*" (go) plus the infinitive of the verb ("*ells van comprar*"). There exists, though, a past tense in Catalan with just one word ("*ells compraren*"*)*, which is used mainly in formal or literary texts. It was noticed that some subjects thought that the use of "*compraren*" had some kind of spelling mistake and some of them even modified its spelling because they mistook it for a future tense misspelled ("*ells compraran*" – they will buy). If the compound form of the verb was used but it was not the target answer, the checkbox signalling the different options for that item was placed under the content word of the verb, i.e., in "*va agafar*" the checkbox was placed under "*agafar*".

Another modification which implied the use of more words was the translation of possessive adjectives. In Spanish they consist of just one word, while in Catalan they are composed by the determinate article and the possessive adjective. In this case, the checkbox was placed under the possessive adjective. Finally, proper nouns in Spanish are not preceded by a determinate article, whereas they are in Catalan. In this case, the articles were added but no checkbox was put beneath them so as to keep the same number of distractors present in the MLAT-ES.

The target functions in the MLAT-EC are the same targeted in the MLAT-ES (subject, verb, adjective and direct object). In the published Spanish version, item 22, which aimed at recognising the subject function in a confusing sentence, was eliminated because it was miskeyed in the MLAT-ES correction spreadsheet and was finally removed. This item was kept in the MLAT-EC during the administration process as it appeared in the MLAT-ES so as to keep both tests as similar as possible. Thus, in the piloting version of the MLAT-EC, there were 31 items too. However, the number of items targeting the functions is different from the MLAT-ES because, by mistake, in item 31 the word that appeared capitalised in the stem was not the verb, as was in the MLAT-ES, but the subject. Consequently, out of the 31 items of this part, 7 aimed at the verb, while 8 items aimed at each of the remaining functions (subject, object and adjective).

Like *Part 1*, *Part 3 Paraules que rimen* contains many stems that coincide with those in the same part of the MLAT-ES. The distractors are mostly different, though, because what was important for this part in order for it to work like the Spanish version was to keep the same rhyme or false rhyme scheme and not so much the word. Besides, Catalan phonology varies from the Spanish one. To start with, depending on the regional variety, Catalan has up to 9 vowel phonemes. That is, <e> can be pronounced as /e/, /ə/ or /ɛ/, and <o> can be pronounced as /o/ or /ɔ/ without any spelling hints about their pronunciation unless they happen to carry a graphical stress. <o> can also be pronounced /u/ and <a> as /ə/ in the central and northern varieties of Catalan when these letters appear in unstressed syllables. Of course, there are many other regional pronunciation varieties, but they are not relevant for the purpose of this dissertation. Summarising, Catalan pronunciation is not so intuitive as the Spanish one because there is not always one-to-one correspondence between the vowel grapheme and the phoneme and, in addition, one grapheme can be pronounced in more than one way, as happens in English.

The different pronunciations of <e> were somehow the aim of item 24, as "*PROMESA*" (promise). The stem of this item is pronounced with an open /ɛ/ while the distractor "*peça*" (piece) is pronounced with the closed variance /e/.

24. PROMESA.. ☐ mossa …….. ☐ permesa …… ☐ camisa …..… ☐ peça

The different pronunciations of <o> were the aim of item 29 *ESTORA* (rug), which rhymes with "*rentadora*" (washing machine), both pronounced with a closed /o/, but it does not rhyme with "*perola*" (saucepan), containing an open /ɔ/ or "*cotorra*"

(parrot), which does have a closed /o/ but also a double <rr>, which makes it a wrong option.

29. ESTORA...... ☐ perola ……... ☐ cotorra ……. ☐ tara……........ ☐ rentadora

Supposing that some test takers would do the test paying attention only to the graphical representation of words, item 18 FORMÓS (nice) was designed using twice the <-mos> ending in the options, one stressed so that it sounded /'mos/ (*"famós"* - famous) and the other one without graphical stress in the plural word "*termos*" (vacuum flasks), which is pronounced /'termus/.

18. FORMÓS ... ☐ famós ……... ☐ carros ……... ☐ termos …....... ☐ formes

Item 45 does not contain any graphical stress, but it also works with the combinations of <-ol>. The stem "*FLABIOL*" (piccolo) rhymes with Oriol, with an open /ɔ/, but it does not rhyme with "*pèsol*" (pea), in which <o> is pronounced /u/ in the standard variety.

45. FLABIOL… ☐ pèsol ……… ☐ Oriol ……. ☐ flascó ……... ☐ avió

The *seseo* and *ceceo* phenomena do not exist as such in Catalan, although some Spanish-dominant bilinguals do have some difficulty when pronouncing /z/, which is the way the voiced <s> sounds, and so pronounce it unvoiced. Due to their complexity, the different pronunciation and spelling correspondences of the alveolar fricatives are taught formally at schools. Generally speaking, how a word with these sounds is pronounced can be easily recognised from the spelling except for some particular cases, which were avoided in the design process of this part of the test because it was considered that the test takers may not have been taught these exceptions.

In Catalan, the voiced fricative alveolar consonant /z/, summarising the explanations of the electronic version of the *Gramàtica Catalana de l'Institut d'Estudis Catalans*, is spelled <-s-> between vowels except for some borrowings and words of Latin and Greek root, in which case it is written <z> (e.g. "*protozou*" – "protozoa"). The pronunciation of <z> is always voiced, no matter its position or contour phonemes in the word it appears.

In contrast with <z>, <s> is pronounced differently depending on its position in the word and/or the phonemes it has next to it. At the beginning of words, <s> is

always pronounced /s/ (unvoiced alveolar fricative consonant). When preceded by consonants, <s> is pronounced as /s/ except in some words that are exceptions, and when it appears closing a syllable followed by a voiced consonant. The spelling rules of the unvoiced /s/ are rather complex, but this phoneme poses almost no pronunciation problems to Spanish dominant speakers, except to those who overcorrect their pronunciation. The graphical representation between vowels is generally <-ss->, although there are some exceptions too. /s/ is also represented by <- ç- + a, o, u> in the middle of words as in "*caçar"* (to hunt) as well as by <-c- + e, i>, as in "*cacera"* (the hunt), no matter whether they are preceded by consonant or vowel. When closing a syllable, <s> is the grapheme used. At the end of words only <ç> and <s> can represent graphically the unvoiced alveolar fricative except for some borrowings or some verb endings in the Balearic dialect. For further explanations of the graphical representations of fricative alveolar phonemes in Catalan, see appendix D.

One item playing with the different pronunciation of <s> is item 16 *CAMISA* (shirt), which is meant to be matched with option "*llisa"* (plain), both of them containing a voiced alveolar fricative /z/. This item had as distractors other words containing <-c-> ("*Patrícia"* – Patricia) and <-ss-> ("*cloïssa"* – clam), both of which are pronounced with the unvoiced /s/.

16. CAMISA…. ☐ llisa ……….. ☐ Patrícia …… ☐ camèlia …… ☐ cloïssa

Another item is item 33 *BRAÇ* (arm), which is very similar to the distractor "*braça"* (breaststroke) and finishes with the same consonant as "*arboç"* (a small tree), but it rhymes with "*cabàs"* (basket).

33. BRAÇ ……. ☐ arboç ……… ☐ cabàs………. ☐ braça ……… ☐ carabassa

A similar example is found in item 44 *PALLASSOS* (clowns), which rhymes with "*eficaços"* (efficient), but shares the same consonant as "*carbassons"* (courgettes) and "*pastissos"* (cakes).

44. PALLASSOS.. ☐ eficaços…..... ☐ carbassons… ☐ pastissos ….. ☐ pressupost

One example of voiced alveolar fricative is item 24 *PROMESA* (promise), which rhymes with the option "*permesa*" (allowed), but it could be matched as well with "*peça*" (piece).

24. PROMESA.. ☐ mossa …........ ☐ permesa …… ☐ camisa ……. ☐ peça

Item 5 (FLETXA – *fitxa – metxa – metge – pedra*) was designed to be easy but misleading if the test taker only looked at the word endings without paying attention to pronunciation. However, if the test taker pronounced the word "*metge*" (doctor), he or she could doubt whether this or "*metxa*" was the correct answer. This is so because some Catalan speakers tend to devoice <-tge> /dʒ/ and pronounce it as /tʃ/ instead. Also, like in Spanish, <b> and <v> sound no longer different. That is why the consonant was used in item 46 *VALL* (valley) to be matched with "*ball*" (dance).

46. VALL…….. ☐ bell …..…….. ☐ vella …....…… ☐ bala …....…… ☐ ball

Distinguishing between consonant and assonant rhyme was also the target of some items, although, as already pointed out when describing the Spanish version, the instructions were perhaps not very self-explanatory in this point. For instance, item stem 9 *TACA* rhymes with both "*tanca*" (fence) and "*tacte*" (touch) in an assonant way, but the target option was "*petaca*" (hipflask), which rhymes in a consonant way.

9. TACA ……... ☐ toca ………... ☐ petaca ……… ☐ tanca ……… ☐ tacte

Another example is item 22 *AUTOBÚS* (bus), which is to be matched with "*abús*" (overuse), but it also rhymes with "*obtús*" (obtuse)*,* "*barnús*" (bathrobe) and "*bust*" (bust).

22. AUTOBÚS.. ☐ abús ……..... ☐ obtús ……… ☐ barnús ……. ☐ bust

As in the Spanish version, some distractors contain similar consonant-vowel combinations to make them look like the beginning of the stem word. This is the case of, for example, item 23 *GAVINA* (seagull), which has as a distractor "*ganivet*" (knife) or item 12 *DAURAT* (golden), which has as a distractor its feminine form "*daurada*". One more item is 26 *LLENTIA* (lentil), two distractors of which start with the same syllable ("*llençol*" – blanket and "*llengua*" – tongue). The third distractor ("*lenta*" – slow) is very

similar to the stem too, but the correct answer only rhymes with the word without being so similar to it ("*eucaristia*" – Eucharist).

12. DAURAT … ☐ pesat ……… ☐ daurada …… ☐ benaurat…..... ☐ urani

23. GAVINA .... ☐ ganivet …..... ☐ petita …....... ☐ fauna .... ☐ petxina

26. LLENTIA … ☐ llençol .... ☐ eucaristia….. ☐ lenta ....…. ☐ llengua

      Part 4 *Aprenguem números* was kept exactly as it was in the Spanish version, since it was believed that the test takers would not remember the name of the numbers after 3 months, when they would take the test in the other language. However, a test with brand new items could have been developed for the MLAT-EC following the rationale behind the names of the numbers (i.e. the suffix –ca for the tens and the similarity between "*vein*", which refers to number "two" and "*veinte*", which means "twenty" in Spanish). Not changing the test in the Catalan version would threaten the validity of scores if the test was to be taken in both languages in a very short period of time. However, this is not the most common situation, so no changes were made to this part. Moreover, for the norming study, only the data of the test taken in the first place would be used.

      While the numbers are the same in both tests, the similarities between them and the real numbers in Catalan are different from those that appear between the MLAT-ES and the Spanish language. For instance, on the MLAT-ES "vein" is similar to "twenty" in Spanish. On the same note, "*vinca*" also shares the same syllable with the name of number twenty in Catalan "*vint*". Therefore, since the participants in this study are bilingual Spanish-Catalan, it was thought it would be interesting to see how test takers reacted to the similarity between "*vint*" and "*vinca*".

3.5. Piloting the MLAT-E in Spanish and Catalan

      Adapting a test does not only consist in translating it item per item, but there are also some cultural, context and linguistic differences that may alter the expected results if we take as a model the results of previous administrations of other versions. It is these differences that will help explain the administration design, the standardisation and norms obtained as well as the statistical information obtained of the MLAT-E in Spanish and Catalan. With this aim, following is a description of the administration

design for this study as well as an in-depth analysis of the items of the MLAT-ES and MLAT-EC in order to be able to check their validity and reliability.

3.5.1. Administration design

One possible approach to piloting two tests in different languages at the same time is using bilinguals, assuming they are fully proficient in both languages. Then both tests are administered to the same population so as to see whether these two tests are really equivalent. The tests should be administered in two different sittings leaving some time in-between. Nevertheless, assuming that two different language versions of one test are 100% equivalent would be a fallacy, as it is crucial to get to discern the nature of the effects that a double sitting can have. These could be not only group effects, but also test effects having to do with the adaptation, translation and administration of the test, which can cause several types of bias. Ideally, if well adapted and properly translated, no bias should appear in the validation process of the MLAT-E in Spanish and Catalan unless cross-linguistic and cross-cultural issues are involved but, as will be shown below, these issues do appear at some point.

Since the MLAT-ES and the MLAT-EC are in Romance languages but the original test was conceived in English, which is a Germanic language, the very language used in the test adaptation can threaten the construct validity and, consequently, the criterion-related validity of the test as well. Actually, using real language in FL aptitude testing has been long criticised, as one's proficiency in the language of the test could be a factor having some kind of impact on the test score. The subjects in this study are childhood bilinguals, which could be interpreted as "fully" proficient in both languages, but not as "equally" proficient in both languages, as language dominance on the one hand, and cross-linguistic influence on the other are factors present in any bilingual context. These are factors that threaten the assumption that there are no differences among bilinguals just because both languages are their L1. Therefore, item differential functioning is an issue to be examined when administering two versions in different languages of the same test to the same population.

The administration procedure was counterbalanced, that is, about half of the test takers (N=325) took first the MLAT-E in Spanish and then in Catalan and the other half (N=304) took it in the reverse order. The fact that the same test in different

languages was taken by the same subjects can be a drawback, since in the second sitting the test takers already know what the test is like and may be either extra-motivated or, on the contrary, feel unwilling to do the same task twice.

3.5.2. Statistical information of the MLAT-ES

There are manifold statistical procedures to carry out the validation of a norm-referenced test. The one adopted here mostly follows the Classical Test Theory (CTT) model, which helps us explain the difficulty and discrimination power of items and the reliability of scores, as well as the power of measurement errors and the test consistency. In addition, the CTT model proposes a kind of item analysis that is reliable when used with small sample sizes, that is relatively easy to comprehend and that can be computed using common statistics programmes such as SPSS.

The MLAT-ES had already been normed by the SLTI, Inc. However, it was deemed necessary to validate it again because the behaviour of some items could be sensitive to sample-dependence. That is, the influence of Catalan or any other variables that make the population of this study unique or special could be the cause of some variation in the test performance or the reason for some points worth making in relation to the validity of the MLAT-ES.

In this major section, the content validity for each part of the Spanish and Catalan versions of the test is provided. It is widely accepted that a test has content validity if it measures knowledge of the content domain of which it was designed to measure knowledge. That is to say, for a test to have content validity, the items it consists of should be an adequate and representative sample of the area to be measured. This information here is provided through several analyses: index of facility (IF), discrimination power of items (D_i), point-biserial correlation coefficient (r_{pb}) and a similar index, the corrected item-total correlation, which is provided when running reliability analyses on the SPSS.

The index of the facility (or difficulty) of an item informs us about the proportion of test takers who answered an item correctly. It is not to be mistaken with item discrimination, which is "the extent to which the item discriminates between different groups of test takers" (Bachman, 2004:122). It aims at discriminating those items that work properly according to the final score obtained. That is to say, if one item is answered correctly mostly by those who get the lowest scores and not so much by

those getting the highest scores among the rest of test takers, the discrimination power of this item is low. The same would happen if an item was failed by those who get the highest scores and not by those test takers whose test scores are among the lowest. The point-biserial correlation coefficient (r_{pb}) and the corrected item-total correlation are measures also used to measure the discrimination power of an item. They assume that the variable measured (in our case, every single item) is dichotomous and for it to be discriminating enough, it should correlate highly with the total score.

Construct validity (the quality of a test as a whole, not item-per-item) to predict someone's score on an external criterion measure) is an issue dealt with in section 4.4 and so is the test construct validity ("the meaningfulness and appropriateness of the *interpretations* that we make on the basis of test scores" – Bachman & Palmer, 1996:21) in section 3.5.2.2 for the MLAT-ES and in section 3.5.3.2 for the MLAT-EC. This validity is addressed by analysing the results obtained in the intercorrelation of parts. Low-moderate correlations between parts in a test prove that each part taps a different construct (see section 2.3.1.1). If correlations are moderate or high, this could be a sign of some type of overlap in the constructs each part is supposed to measure. This type of validity is also assessed when correlating the scores on each part with the teachers' criterion variables broken down into specific language aspects and with proficiency measures aiming at one particular skill, and not so much when using measures of general language proficiency.

3.5.2.1. Content validity

A close examination into items was gauged in order to identify and reduce the sources of error in measurement in the raw scores and thus be able to compare if the subjects in this study performed in the same way as did all the subjects in the major piloting study. The explanation of the test takers' performance of this study is backed up not only by the scores on each item, but also by a qualitative interpretation of the deviant items. It is considered, as Banerjee and Luoma (1997:275) put it, that "the character and extent of the interpretations and uses of scores depend on a *thorough understanding* of the test: the way in which it is constructed; how test takers give their responses; how these responses are evaluated; and, how the scores are used in making decisions about the test takers." With such aim, in this section a qualitative interpretation of some particular items as well as of their distractors is given for each

subtest. The choice of items commented is based on the deviant results, after having detected some irregularities from the analysis of their facility index and discrimination power as well as from the response patterns and the unidimensional internal structure of each part, computed by means of an item-total correlation analysis. Consequently, the qualitative analyses stem from the quantitative approach of item analysis.

The first three subtests are of selected-response type. The test takers were instructed to select only one answer to each item, as these subtests are dichotomously scored. That is to say, they are zero-one items in terms of correction. However, a few test takers selected more than one option for some items. In this case, the item was considered not passed.

As most test takers from grade 6 and 7 get to finish all subtests, it could be assumed that the elementary version of the MLAT was not meant to be speeded for these grades, in contrast with the MLAT, which is considered heavily speeded. Nevertheless, the instructions warn the test takers to answer the test accurately but quickly. The results for grade 3 and 4 do seem to have been influenced by the speed variable, as many of the test takers do not reach the end of the first three parts, especially part 1 and 3, but this tendency seems to diminish progressively though not linearly the upper the grade. This tendency may be due to the fact that, as Osterlind (1989) remarks, there is the consensus among test developers that children in the early stages of primary school need at least one minute to reflect on one multiple-choice item, whereas children from fourth grade on need one minute maximum. *Parte 2* is the one that most test takers finished. This could be due to the fact that the time allotted (18' 45'') was long enough for everybody to finish this part. Table 3.7 contains the percentage of the test takers who, for some reason, do not reach the end of the first three parts on the MLAT-ES. The rest of test takers not included in these percentages get to finish the test although they may have left some items unanswered on the way.

Table 3.7. Percentage of unfinished tests per parts (MLAT-ES)

Test	Grade	3	4	5	6	7	All grades
	N	66	75	57	60	67	325
Parte 1	Unfinished	54	27	10	15	2	108
	Percentage	81.8%	36%	17.5%	25%	2.9%	33.2%
Parte 2	Unfinished	11	15	1	0	2	29
	Percentage	16.7%	20%	1.8%	0%	3%	8.9%
Parte 3	Unfinished	36	32	10	7	5	90
	Percentage	54.5%	42.7%	17.5%	11.7%	7.5%	27.7%

In the analysis of IF, when counting unreached items as incorrect, the difficulty of the test is overestimated, as items which have not been attempted are considered to be difficult by themselves. However, they get a zero score because of their position in the test, not because of their inner difficulty. In contrast, when using only the data of those who attempted the items, the difficulty is underestimated since only the data from more able students are being considered. Ours is not an extreme case where so few subjects reach the last items on the test that there are not enough data to analyse the items that appear in final positions. Consequently, both ways to count unreached items provide us with a useful estimate. The potential error and the direction of this error of counting the missing items like this should, nevertheless, not be forgotten. Actually, while items left blank that appear at the end of the test are most probably so because of lack of time, items left blank at the beginning or in the middle of a test are a bit problematic, as they may not have been answered for several reasons, not only because of time constraints. One reason could be, for example, leaving what seems at first sight a difficult item to be answered at the end of the test if there is time left. Thus, the test taker tries not to waste time reflecting on each item and takes most advantage of the time allotted to answer as many items as possible. Other reasons for leaving an item blank could be because of distraction, loss of interest, or failing to see a particular page of the test booklet because of having turned the pages erroneously. As far as the IF analyses that follow are concerned, unreached items have been counted as missing data because we were interested in knowing the behaviour of each item across grades. In contrast, regarding the facility of the whole subtest, unreached items have been counted as failed, as speed does play a role in the performance of each part depending, mainly, on the test takers' grade or age, for this matter.

Parte 4 Aprendamos números is not a multiple-choice test but a constructed-response type of test. The item analysis for this type of test is slightly different regarding the calculation of IF and the response patterns, but overall, the same criteria can be adopted to interpret the results of the item analysis. This part could be considered not speeded, as it must be answered keeping the pace set on the CD recording, which is moderate enough to stand. In this case, it has been considered that if a test taker leaves an item blank, this is because of distraction or because they were simply unable to answer it. Therefore, missing values on *Parte 4* are considered incorrect values in all cases.

3.5.2.1.1. Parte 1 Palabras ocultas

The ideal *p*-values (or index of IF) of a test are between .40 and .80 and appear in a normal distribution. Following is the table showing the distribution of items of *Parte 1 Palabras ocultas* according to their facility across grades.

Table 3.8. IF of items on MLAT-ES *Parte 1 Palabras ocultas* across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=66	**45.1%** Items: 1, 3, 5, 10, 13, 15, 16, 18, 20, 23, 25, 27, 28, 31	**29%** Items: 2, 4, 6, 8, 9, 21, 22, 24, 26	**6.5%** Items: 14, 29	**12.9%** Items: 7, 11, 17, 30	**6.5%** Items: 12, 19
Grade 4 N=75	**64.5%** Items: 2, 3, 4, 5, 8, 9, 10,13, 15, 16, 18, 20, 21, 22, 23, 25, 26, 27, 28, 31	**19.4%** Items: 1, 6, 14, 24, 29, 30	**3.2%** Items: 17	**12.9%** Items: 7, 11, 12, 19	**0%** Items: -
Grade 5 N=57	**83.8%** Items: 1, 2, 3, 4, 5, 6, 8, 9, 10, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31	**6.5%** Items: 11, 17	**6.5%** Items: 7, 19	**3.2%** Items: 12	**0%** Items: -
Grade 6 N=60	**74.2%** Items: 1, 3, 4, 5, 9, 10, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31	**16.1%** Items: 2, 6, 8, 11, 19	**6.5%** Items: 7, 12	**3.2%** Items: 17	**0%** Items: -
Grade 7 N=67	**90.3%** Items: 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31	**9.7%** Items: 7, 12, 17	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grades 3- 7 N=325	**77.4%** Items: 1, 2, 3, 4, 5, 8, 9, 10, 13, 15, 16, 18, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31	**9.7%** Items: 6, 11, 14	**9.7%** Items: 7, 17, 19	**3.2%** Items: 12	**0 %** Items: -

As can be seen in the table, most items in *Parte 1* are very easy across grades, especially from grade 5 to 7. *p*-values are also considered to be very revealing regarding the behaviour of distractors and are good indicators of the discrimination power of an item. There is a change in the degree of difficulty of some specific items across grades (i.e. items 7, 11, 12, 17 and 19) except for item 30, which results in a particularly difficult item for those in grade 3 who reached the end of the test though not for the participants in other grades.

Taking a look at how many times some distractors are chosen across grades may also increase our understanding of some items (for a complete account of *Parte 1* answers and distractor facility indexes see appendix E). For instance, in item 30 dzpacio – *lento* (slow) (see Table 3.9), the distractor "*amplio*" (spacious) draws the attention of most of those who fail this item, probably because they relate dzpacio with "*espacio*", and the sentence "*hay espacio*" means "there is enough room / it is spacious" in English. Notice that all those distractors that are appealing in some way or another have an IF that is positive or very close to 0.

Table 3.9. Item 30 on the MLAT-ES *Parte 1 Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B	C*	D	Missing	Attempts	IF – A	IF – B	IF – C*	IF – D
3	66	3	2	4	0	57	9	0.111	-0.037	0.259	-0.333
4	75	8	0	35	3	29	46	-0.104	-0.333	0.681	-0.246
5	57	2	3	38	1	13	44	-0.273	-0.242	0.818	-0.303
6	60	3	0	39	0	18	42	-0.238	-0.333	0.905	-0.333
7	67	3	0	62	2	2	65	-0.282	-0.344	0.928	-0.303
all	325	19	5	178	4	119	206	-0.210	-0.301	0.819	-0.307

30. dzpacio A – amplio; B – azul; C – lento; D – redondo
* correct answer

Those items that are very easy (item 3, 5, 10, 13, 16, 18, 20, 25, 26, 27, 28 and 31) have consistently ineffective distractors, while those items that are easy or mid-difficult either have distractors that are more or less popular in the same way across grades (item 6, 9, 21, 22), or present some distractor that attracts the test taker for some reason. Item 6 stands out by the amount of subjects (22%) that leave it blank even though it is among the first items. This could be because of not being able to decode the very stem (qlvra), which contains a lower-case <l> that could have been mistaken for a capital <i> (I), which looks like this <l> in Times New Roman font.

Some subjects, especially in the lower grades, seem not to have understood the task properly and that is the reason why whenever they fail items in which one distractor rhymes with or sounds similar to the stem, they go for it. That is the case of item 1 (lfant – *elegante*), item 2 (kstiyo – *castigo*), item 8 (tlbzión – *división*), item 14 (nmigo – *hormiga*), item 23 (hinzendio – *indio*), (skushar – *estudiar*) and item 29 (cirrena – *herramienta*).

Some items present some response patterns that are especially relevant in the lower grades. This is the case of item 4 and 7. Item 4 refers to a geometrical concept, which could still be unknown or not completely clear for students in grades 3 and 4.

Though appearing at the beginning of the test, it is left blank by 17% of learners in grade 3. 20% of them go for option A, which refers to a geometrical concept, too.

Table 3.10. Item 4 on the MLAT-ES *Parte 1 Palabras ocultas: p*-**values and distractor behaviour analysis**

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	13	38	2	2	11	55	-0.018	0.588	-0.285	-0.285
4	75	8	60	2	2	3	72	-0.185	0.778	-0.296	-0.296
5	57	2	52	0	1	2	55	-0.285	0.927	-0.333	-0.309
6	60	4	54	0	0	2	58	-0.241	0.908	-0.333	-0.333
7	67	0	65	1	0	1	66	-0.333	0.980	-0.313	-0.333
all	325	27	269	5	5	19	306	-0.216	0.839	-0.312	-0.312

4. sírqulo A – ángulo; B – circunferencia; C – fruta; D – peligroso
* correct answer

Item 7 (ddo), which hides the word "*dedo*" (finger), is interpreted by some as "*dado*" (dice), which varies in only one letter from the actual word hidden "*dedo*". The meaning of "*dado*" happens to be referred to in distractor C ("*para jugar*" – to play). This distractor is chosen by 17.9% of the test takers and it remains unanswered by 16 learners (24.2%) in grade 3 even though it appears at the beginning of the test.

Table 3.11. Item 7 on the MLAT-ES *Parte 1 Palabras ocultas: p*-**values and distractor behaviour analysis**

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	9	22	13	6	16	50	-0.093	0.253	0.013	-0.173
4	75	8	38	14	6	9	66	-0.172	0.434	-0.051	-0.212
5	57	5	33	10	5	4	53	-0.208	0.494	-0.082	-0.208
6	60	2	37	11	8	2	58	-0.287	0.517	-0.080	-0.149
7	67	0	52	10	4	1	66	-0.333	0.717	-0.131	-0.253
all	325	24	182	58	29	32	293	-0.224	0.495	-0.069	-0.201

7. ddo A – despertador; B – está en la mano; C – para jugar; D – duro
* correct answer

In turn, items 12 and 19 involve abstract concepts and present distractors that are conflictive not only in the low grades but across all of them. In item 12 (see Table 3.12), the most popular distractor is the one that rhymes with the stem. In item 19 (see Table 3.13), "*único*" (unique) is considered to be the correct answer to *esepzional* ("*excepcional*" – exceptional), but the test takers are attracted by "*emocionante*" (exciting, thrilling), which is considered an overly plausible correct answer by twice as many subjects as those who choose the right option in grade 3 and by 5 subjects less than those who pass this item in grade 4.

Table 3.12. Item 12 on the MLAT-ES *Parte 1 Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B	C	D*	Missing	Attempts	IF – A	IF – B	IF – C	IF – D*
3	66	22	0	13	12	24	42	0.302	-0.397	0.111	-0.016
4	75	28	1	3	29	14	61	0.279	-0.311	-0.268	0.301
5	57	23	1	3	26	4	53	0.245	-0.308	-0.258	0.321
6	60	19	0	0	35	5	55	0.133	-0.327	-0.327	0.521
7	67	16	0	1	47	3	64	0	-0.333	-0.313	0.646
all	325	108	3	11	153	50	275	0.190	-0.319	-0.280	0.408

12. eccitoso A – aceitoso; B – salado; C – poco común; D – con éxito
* correct answer

Table 3.13. Item 19 on the MLAT-ES *Parte 1 Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	2	8	16	4	36	30	-0.244	0.022	0.378	-0.156
4	75	4	29	24	2	4	59	-0.243	0.322	0.209	-0.288
5	57	1	31	19	1	14	53	-0.302	0.453	0.151	-0.302
6	60	3	40	9	0	8	52	-0.256	0.692	-0.103	-0.333
7	67	1	52	10	0	4	63	-0.312	0.767	-0.122	-0.333
all	325	11	160	78	7	68	257	-0.275	0.498	0.073	-0.296

19. esepzional A – tipo de triángulo; B – único; C – emocionante; D – muro alto
* correct answer

A minor problem found when administering the test was distractor B in item 5 (Ikóptero – "*insecto*"), the correct answer of which is "*avión"* (plane). In this case, the test taker's additional un-knowledge impaired them as they were probably thinking of "*coleóptero*" (coleopterous) instead of "*helicóptero*" (helicopter). When they asked about which option was the correct one, because they stated they were hesitating between option "*insecto*" and option "*helicóptero*", the test takers were told out loud that there was only one possible answer and that they had to choose the one they thought was the most common one. That is why not many test takers chose "*insect*" as the correct answer, although they did ask about this ambiguity when taking the test in class*.*

When answering item 15 bakka (see Table 3.14), which refers to "*vaca*" (cow), the answer to which is – "*da leche*" (it produces milk), tests takers were probably affected by proactive inhibition, i.e. the case in which test takers make use of knowledge beyond what the item writer had initially in mind in order to answer an item (Lefrancois, 1988, in Osterlind, 1989). Option D "*con manchas"* (spotted) was chosen over *"da leche"* by 9% of the test takers because some students thought that some cows are spotted and so they chose D as the correct answer. Again, when the test takers asked, they were warned that only one answer was possible and that they

should choose the most plausible one, but there were still some test takers who chose option D.

Table 3.14. Item 15 on the MLAT-ES *Parte 1 Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B	C*	D	Missing	Attempts	IF – A	IF – B	IF – C*	IF – D
3	66	1	0	45	4	16	50	-0.307	-0.333	0.867	-0.227
4	75	1	1	59	7	7	68	-0.314	-0.314	0.824	-0.196
5	57	2	0	45	5	4	53	-0.277	-0.327	0.805	-0.201
6	60	0	0	50	7	3	57	-0.333	-0.333	0.836	-0.170
7	67	1	0	61	4	1	66	-0.313	-0.333	0.899	-0.252
all	325	2	5	260	27	31	294	-0.324	-0.311	0.846	-0.211

15. bakka A – es dulce; B – parte de la cara; C – da leche; D – tiene manchas
* correct answer

A controversial distractor that affects the response pattern in item 17 is rmozo – *bonito*. Instead of choosing "*bonito*" (nice) as an equivalent of "*hermoso*", the hidden word, 75 test takers (23%) chose "*joven*" (young) and 56 (17.2%) left it blank even though item 17 is not at the end of the test (see Table 3.15). The reason for this response pattern is probably because they interpreted rmozo as the sum of the determiner "*el*" (the) and the noun "mozo" (the young boy).

Table 3.15. Item 17 on the MLAT-ES *Parte 1 Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	12	23	4	3	24	42	0.048	0.397	-0.206	-0.238
4	75	19	36	0	4	16	59	0.096	0.480	-0.333	-0.243
5	57	7	38	2	5	5	52	-0.154	0.641	-0.282	-0.205
6	60	21	29	1	2	7	53	0.195	0.396	-0.308	-0.283
7	67	16	43	0	4	4	63	0.005	0.577	-0.333	-0.249
all	325	75	169	7	18	56	269	0.038	0.504	-0.299	-0.244

17. rmozo A – joven; B – bonito; C – amplio; D – fresco
* correct answer

To finish with the analysis of distractors of this part, it is important to mention the case of item 11. This item presents a differential item functioning, as linguistic interference affects the pattern of the answers to this item. The stem "ágla" refers to "*águila*" (eagle), which corresponds to option B "*pájaro grande*". This was a difficult item, especially for the lower grades, because, in this case, spelling demands two written vowels <ui> in order to form the phonemic chain /gi/. 27% of those who answered this item chose distractor C "*alimento*" (food), as "*aglà*" in Catalan is "acorn", which is a kind of food. Choosing this distractor was constant across grades; therefore,

it could be considered that item 3's is a uniform kind of bias irrelevant to the construct the test is intended to measure. Consequently, this bias should make the test adaptors ponder changing the wording of option C if the test is to be administered in Catalonia.

Table 3.16. Item 11 on the MLAT-ES Parte 1 *Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	5	24	24	0	13	53	-0.208	0.270	0.270	-0.333
4	75	1	33	24	3	14	61	-0.311	0.388	0.191	-0.268
5	57	2	39	11	2	3	54	-0.284	0.630	-0.062	-0.284
6	60	2	42	12	2	2	58	-0.287	0.632	-0.057	-0.287
7	67	0	52	4	0	10	57	-0.386	0.830	-0.292	-0.152
all	325	10	190	75	7	42	283	-0.285	0.563	0.021	-0.299

11. ágla A – corre rápido; B – pájaro grande; C – alimento; D – pregunta
* correct answer

The index of item discrimination (D_i) is useful to compare *p*-values for high- and low-achieving subpopulations. The cut-off point chosen in this study is 33% of the population, although the item-analysis literature acknowledges that the range could go from 27% to 33%. Choosing one cut-off point or another is a matter of convenience depending on the number of subjects, although some consider it more accurate to restrict the group tails to 27%. In contrast with IF, whose indexes should be between .40 and .80 and normally distributed, most items' D_i should ideally be above .40. If their D_i is below .40, their removal or modification should be considered for future test sittings. It should not be forgotten that, since Parts 1, 2 and 3 are speeded, all the items at the end of the test will be good discriminators because of the effect of time constraints in the lower grades. Should they appear at the beginning of the test, their D_i would probably be different. Table 3.17 shows the items on *Parte 1* distributed according to their discrimination quality depending on their D_i (see exact D_i and number of correct answers in the upper and lower 33% tails in Table F.1 in appendix F).

Table 3.17. Percentage of items on the MLAT-ES *Parte 1 Palabras ocultas* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=325	78.3% Items: 2, 4, 6, 7, 8, 9, 10, 11, 12, 14, 17 - 31	9.7% Items: 1, 15, 16	6.5% Items: 5, 13	3.2% Items: 3

Item 3 was eliminated in the published version of the test. The author of this dissertation hypothesised that this item could have been eliminated due to distractor A "*con mucha gente*", as it means the same as "*va llena*", as two words, contained as one word in distractor A. However, this distractor does not seem to appeal to many test takers and neither do the rest of options, as shown in Table 3.18.

Table 3.18. Item 3 on the MLAT-ES *Parte 1 Palabras ocultas: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	2	60	3	0	1	65	-0.292	0.897	-0.272	-0.333
4	75	3	68	2	0	2	73	-0.279	0.909	-0.297	-0.333
5	57	0	54	1	1	1	56	-0.333	0.952	-0.310	-0.310
6	60	0	59	0	0	1	59	-0.333	1	-0.333	-0.333
7	67	1	66	0	0	0	67	-0.313	0.980	-0.333	-0.333
all	325	6	307	6	1	5	320	-0.308	0.946	-0.308	-0.329

3. vallena A – con mucha gente; B – animal marino enorme; C – barbilla; D – peluca
* correct answer

Adding up the results of the D_i and IF index, it is confirmed that item 3 is a deficient item and, therefore, should be removed. As its IF is very high, it does not provide much information by itself, given the fact that most items in *Parte 1* are also very easy. Figure 3.1 is the graphical representation in a non-standard quintile plot of the behaviour of item 3. The x-axis of non-standard quintile plots shows item options (0 in this case stands for blank answer) while the y-axis shows the percentage of the test takers who chose an option. The test takers have been divided into five groups or quintiles according to their total score on *Parte 1*. Option B is the correct answer to item 3.

Figure 3.1. Quintile plot of item 3 on the MLAT-ES *Parte 1 Palabras ocultas*
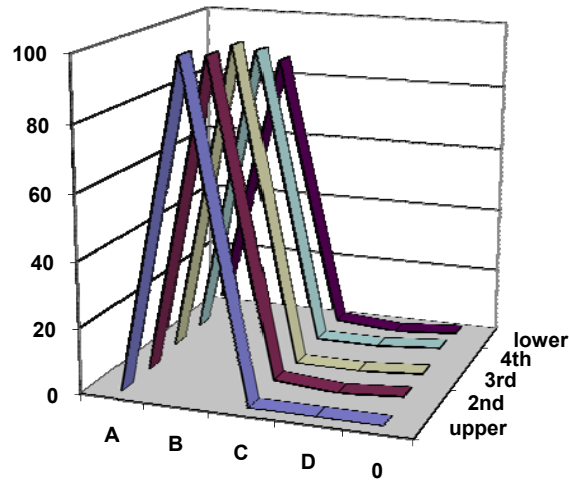
	A	B	C	D	0
□upper	0	100	0	0	0
■2nd	0	96.9	1.5	0	1.5
□3rd	1.5	96.9	0	0	1.5
□4th	1.5	92.3	1.5	1.5	3.1
■lower	6.2	86.2	1.5	0	1.5

Other items that should be improved on the basis of their D_i are item 5 and 13, whose response alternatives did not cause problems but were not effective either. Both of them are also very easy items, as is item 3. Therefore, if we represented these items in quintile plots, they would result in very similar plots to that of item 3.

One more index that informs us about the discrimination power of an item is the point-biserial correlation coefficient (r_{pb}), which estimates the degree of association between the total test score and a single item, which has to be a dichotomous variable (either 0 or 1) (Bachman, 2004). This coefficient should ideally be higher than .300. As can be seen in Table 3.19, all items but 3 fulfil this requirement. The correlations increase positively for the items that appear towards the middle and end of this part. This is so because this part is speeded and only those who reach the items that appear towards the end of the test are those who obtain higher scores while there is more variance on the scores of those who attempt the first items on the test. Consequently, those who reach the end of the test and score on the end-of-test items make the correlation be higher. It is important to say that this index is also used to evaluate a test's internal consistency, as if a test has construct validity, scores on the individual test items should correlate highly with the total test score, which proves that all items are measuring the same construct.

Table 3.19. Item-total correlation of MLAT-ES *Parte 1 Palabras ocultas*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
1	.310**	6	.409**	11	.399**	16	.496**	21	.750**	26	.782**
2	.398**	7	.442**	12	.527**	17	.358**	22	.654**	27	.790**
3	.221**	8	.426**	13	.386**	18	.600**	23	.749**	28	.715**
4	.478**	9	.451**	14	.553**	19	.531**	24	.732**	29	.681**
5	.332**	10	.464**	15	.388**	20	.729**	25	.778**	30	.692**
										31	.713**

** p<.001

Further confirmation that item 3 should be removed is the corrected item-total correlation obtained from the Cronbach's alpha analysis. Although the index of reliability increases only slightly (from .926 to .927) when excising this item, its low corrected item-total correlation (.192) reveals that it is a deficient item and, therefore, should be removed (see Table F.2 in appendix F).

3.5.2.1.2. Parte 2 Palabras que se corresponden

In *Parte 2 Palabras que se corresponden* the test takers had to recognise four different functions in the stem sentence and match the target function with the word performing the same function in the sentence below the stem. The items and functions were distributed as shown in Tables F.3, F.4, F.5 and F.6 in appendix F. In these tables, the word in the stem that appears in capital letters is the word whose function the test takers had to identify so as to match it with the word of the target sentence performing the same function. This word appears underlined. Seven items were aimed at matching the subject[14], eight items had the main verb as a target word, eight more items were aimed at recognising the word performing the function of object and the eight items left had an adjective as a target word.

Before interpreting any item analysis, it should be pointed out that, although the aim of this part was tapping grammatical sensitivity by matching words with the same function regardless of their position in the sentence, the choice of the right or wrong option could have been influenced by the arrangement of the words in the sentence. That is, while the sentences in the MLAT-E follow the SVO order most times, this is not the case of the sentences in the MLAT-ES because, in Spanish, the word order within

---

[14] Notice that item 22, which was miskeyed, appears in Table F3 classified in the group aiming at recognizing the function of subject, although the test developers in the SLTI Inc. first thought it performed the function of direct object.

a sentence is much more flexible and so it is very common to find the VSO pattern. Consequently, the test takers could present a variety of reactions when facing an item because recognising the target function could depend, up to a certain extent, on the position of the words in either the item stem or the sentence below it. Besides, the position of the target word in a sentence can make its function change. For instance, although in item 2 and 4 the test takers have to recognise which word is acting as an adjective, the adjectives to be recognised are qualitatively different in the pairs of sentences.

<div align="center">

2. El perro PEQUEÑO rompió el florero de cristal.
La casa roja tiene las ventanas abiertas.

☐ ☐ ☐ ☐ ☐   ☐   ☐

4. La GRAN mansión del presidente es blanca.
En la clase de matemáticas hay pocos alumnos.

☐☐ ☐ ☐　☐　　☐ ☐　☐

</div>

In item 2 the adjectives "*PEQUEÑO*" (little) and "*roja*" (red) appear after the noun they modify both in the stem and in the target sentence. In contrast, in item 4 the adjectives "*GRAN*" (big) and "*pocos*" (few) appear in front of the noun they modify but their function is different. Though both words are adjectives, they should not be considered to belong in the same group: in item 4 the adjective "*gran*" describes a quality of the mansion while "*pocos*" is a quantifying adjective. In the tables below (and all the tables referring to items on *Parte 2*) one can see the behaviour of the target item and the distractors that were more popular. These are most often content words as, regardless of the grade, those distractors which were function words were hardly ever chosen.

Table 3.20. Item 2 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	B	C*	F	G	Missing	Attempts	IF – B	IF – C*	IF – F	IF – G
3	66	8	26	17	13	2	64	0.026	0.307	0.167	0.104
4	75	2	51	7	11	2	73	-0.023	0.648	0.046	0.100
5	57	1	45	6	3	1	56	-0.015	0.771	0.074	0.021
6	60	1	53	3	2	0	60	-0.003	0.864	0.031	0.014
7	67	2	61	2	1	0	67	0.015	0.896	0.014	0
all	325	14	236	35	30	5	320	0	0.694	0.066	0.050

2. B – casa; C – roja; F – ventanas; G – abiertas
* correct answer

Table 3.21. Item 4 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	C	E	G*	H	Missing	Attempts	IF – C	IF – E	IF – G*	IF – H
3	66	15	12	21	17	1	65	0.134	0.088	0.226	0.165
4	75	19	11	32	8	2	73	0.180	0.070	0.358	0.029
5	57	22	3	27	3	0	57	0.311	-0.023	0.398	-0.023
6	60	12	8	33	4	1	59	0.140	0.073	0.496	0.005
7	67	11	13	38	1	2	65	0.110	0.141	0.525	-0.044
all	325	79	47	151	33	6	319	0.172	0.072	0.398	0.028

4. C – clase; E – matemáticas; F – pocos; G – alumnos
* correct answer

From Tables 3.20 and 3.21 above, it can be inferred that the order in which words appear either in the stem sentence or in the target sentence somehow determines the response pattern. While in item 2 the adjective in the stem sentence appears after the noun it complements and so does the adjective in the target sentence — chosen by 74% of those who answer this item —, this does not happen in item 4. In this item, the qualitative adjective "*GRAN*" should theoretically be matched to the word "*pocos*" (a few), which appears in front of "alumnos" (students). Yet 24.8% of the test takers who attempt at the item choose "*clase*" (class) probably because of the order in which this word appears in the target sentence. It is also important to say that 47 out of 319 test takers who attempt this item seem to notice the fact that "*pocos*" is a quantifying adjective (it does not inform us of a quality of "*alumnos*" as does "GRAN" in the stem sentence) and, therefore, choose "*matemáticas*" (mathematics), which does inform us of a quality of "*clase*" despite being a noun, not an adjective.

Actually, taking a look at Table 3.22, it can be seen that item 4 never becomes an easy item while item 2, among other items aiming at recognising adjectives, decreases in difficulty across grades. As happened in *Parte 1 Palabras ocultas*, the difficulty of the items does not appear to be normally distributed. Instead, in grade 3 most items are difficult or very difficult while the distribution of items according to their IF changes radically from grade 4 on (for a complete account of all the distractors' IF see appendix G).

Table 3.22. Percentage of items on the MLAT-ES *Parte 2 Palabras que se corresponden* according to their facility across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=66	**0%** Items: -	**12.9%** Items: 1, 12, 18, 26	**19.4 %** Items: 8, 13, 14, 19, 28, 30	**38.7%** Items: 2, 3, 5, 6, 7, 11, 15, 17, 20, 23, 27, 29	**29%** Items: 4, 9, 10, 16, 21, 22, 24, 25, 31
Grade 4 N=75	**6.5%** Items: 12, 26	**51.5%** Items: 1, 2, 3, 5, 6, 8, 13, 14, 15, 18, 19, 20, 25, 28, 29, 30	**22.6 %** Items: 7, 9, 10, 11, 17, 23, 31	**9.7%** Items: 4, 16, 27	**9.7%** Items: 21, 22, 24
Grade 5 N=57	**12.9%** Items: 2, 5, 12, 18	**48.4%** Items: 1, 3, 7, 8, 9, 13, 14, 17, 19, 20, 23, 25, 26, 28, 30	**9.7%** Items: 6, 15, 29	**25.8%** Items: 4, 10, 16, 21, 22, 24, 27, 31	**3.2%** Items: 11
Grade 6 N=60	**54.8%** Items: 1, 2, 3, 5, 6, 7, 8, 9, 12, 13, 14, 18, 19, 20, 26, 28, 29	**38.7%** Items: 10, 11, 15, 16, 17, 21, 23, 24, 25, 27, 30, 31	**6.5%** Items: 4, 22	**0%** Items: -	**0%** Items: -
Grade 7 N=67	**58.1%** Items: 1, 2, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18, 20, 26, 28, 29	**35.5%** Items: 3, 10, 16, 19, 21, 23, 24, 25, 27, 30, 31	**3.2%** Items: 4	**3.2%** Items: 22	**0%** Items: -
Grades 3 - 7 N=325	**12.9%** Items: 1, 12, 18, 26	**54.8%** Items: 2, 3, 5, 6, 7, 8, 9, 13, 14, 15, 17, 19, 20, 23, 25, 28, 29, 30	**19.4%** Items: 10, 11, 16, 21, 27, 31	**9.7%** Items: 4, 22, 24	**0%** Items: -

From Table 3.22, it can be inferred that this test is much more difficult for students in grade 3 than for the rest of grades. There are, nevertheless, some items that are also difficult for some other grades. For example, item 25 is difficult only for 3-graders, but it is not for the rest of grades. This could be either because of the fact that it appears towards the end of the test and most test takers who did not get to finish the test belonged to grades 3 and 4 (see Table F.7 in appendix F for a complete account of the IF distribution across grades according to the target function).

In item 22, which was miskeyed in the correction worksheet, test takers have to identify the subject. However, since the type of subject that appears in this item is different from the standard subject because of the thematic roles of the verb "*parecer*" and of its position within the sentence, as explained in section 2.3.2.1, this item cannot be considered in the same way as the other items aiming at words performing the function of subject. As happens to item 4, item 22 is consistently difficult across grades, which makes it a red-flagged item liable to be eliminated from the test.

22. ¿Qué te pareció el nuevo COMPAÑERO?

Mi tía salió y no apagó el televisor.

☐☐ ☐☐☐ ☐ ☐   ☐

Almost the same number of test takers as those who chose the correct answer "*tía*" (aunt) chose the distractor "*televisor*" (television). One of the reasons for this choice could certainly be its final position in the target sentence (as both "*compañero*" and "*televisor*" appear at the end of the sentence) as well as the fact that "*compañero*" and "*tía*" refer both to human beings ("mate" and "aunt"). It would be, therefore, interesting to see if the response pattern would be the same if either "*tía*" or "*compañero*" were changed for a word referring to an animal or, rather, an object. Notice as well the IF values of "*tía*", "*televisor*" and "*apagó*" in Table 3.23. Both distractors have a positive IF value, which is further proof of the inadequate functioning of this item.

Table 3.23. Item 22 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	B*	C	F	H	Missing	Attempts	IF – B*	IF – C	IF – F	IF – H
3	66	18	9	14	17	6	60	0.200	0.050	0.133	0.154
4	75	22	8	10	26	7	68	0.227	0.021	0.050	0.243
5	57	24	3	5	22	0	57	0.338	-0.030	0.005	0.281
6	60	32	2	4	21	0	60	0.467	-0.033	0	0.365
7	67	28	2	7	26	1	66	0.342	-0.052	0.024	0.253
all	325	124	24	40	112	14	311	0.313	-0.009	0.043	0.261

22. B – tía; C – salió; F – apagó; H – televisor
* correct answer

Although not explicitly stated in the test directions, the marking system considers that the target items should only appear in main clauses. This causes a problem when the target sentence has two words performing the same function, one in the main clause and the other in the subordinate clause. If the test taker chooses the word in the subordinate clause that performs the same function as the word in capital letters in the stem sentence, this item is considered failed. We could group the items in three ways, bearing in mind the order of appearance of the words both in the stem and the target sentences, and taking into account whether there is a subordinate clause preceding the main clause that also has a word performing the target function. The first group could be formed by those items keeping the canonical word order; the second group, those which have some order of sentence elements changed in the stem sentence, and the third group would be those which have the order of the elements disrupted in the target sentence (see Table F.8 in appendix F). Word order in Spanish

is much more flexible than in English. Consequently, "canonical" here does not mean following the SVOCA rule, but should be interpreted as "most common" order, that is, the order that sounds more common or natural to the Spanish native speaker's ear.

All the red-flagged items commented on so far appear under the label of changed order, be it in the stem or in the target sentence. Consequently, it seems it would be advisable to take a look at the behaviour of the rest of items under these categories.

Starting with those sentences targeting the function of subject, in item 15 the subject appears, as in item 22, at the end of the stem sentence while the subject in the target sentence is the second word from the beginning. Contrarily to what happens in item 22, most test takers choose the correct answer. Nevertheless, it is worth mentioning that 20% of the 210 test takers who attempted at the item chose "*comida*" (food), the last word in the target sentence, coinciding with the position of the word in capitals in the stem. The IF of this item decreases the upper the grade, so it functions as most other items in this test.

15. Dime una cosa, ¿todavía está enferma tu HERMANA?
Las hormigas trabajan duro acarreando comida.

☐   ☐   ☐   ☐   ☐   ☐

Item 23 is one example of sentence in which two words have the same function as the one targeted, yet only one of them is considered to be correct because it appears in the main clause.

23. SUSANA le quitó el sombrero a Juan.
Cuando se acerca el invierno, los pájaros empiezan a volar hacia el sur.

☐   ☐ ☐   ☐   ☐   ☐   ☐   ☐   ☐

While it is true that 189 test takers got their answer right, 42 test takers chose "*invierno*" (winter), the other subject that appears in the target sentence. Strictly speaking, this answer should be considered right. However, the only subject that was considered to be correct in the target sentence is "*pájaros*" (birds), which is the subject that appears in the main clause.

The label "order of sentence elements changed" may not be the most accurate for item 24, also aiming at the subject function, or at least, may not be the only label under which this item should appear. Certainly, making the subject explicit in the target sentence of this item is either redundant or somehow emphatic in Spanish, as it

appears in the form of a strong personal pronoun, not a noun phrase containing a noun (common or proper).

24. ENRIQUE puso una campanilla en la puerta de entrada de su casa.

¿A qué horas crees tú que llegarás a cenar?

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

When facing this item, many test takers were either disconcerted or just chose the matching word at random. The word order in the target sentence was not so prominent, probably because the second sentence starts with a preposition, a function word that is not so meaningful as the proper noun "*Enrique*". Consequently, 163 out of 299 test takers who answered this item incorrectly did not choose the first word in the target sentence, but other content words before "*tú*" (you).

Table 3.24. Item 24 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	C	D	E*	F	H	Missing	Attempts	IF – C	IF – D	IF – E*	IF – F	IF – H
3	66	8	5	10	21	10	9	57	0.037	-0.015	0.072	0.265	0.072
4	75	10	5	18	14	11	12	63	0.069	-0.010	0.196	0.133	0.085
5	57	9	1	24	16	7	0	57	0.086	-0.055	0.349	0.208	0.050
6	60	4	2	37	6	5	4	56	0.029	-0.007	0.618	0.065	0.047
7	67	2	4	47	7	5	1	66	-0.006	0.025	0.676	0.070	0.040
all	325	33	17	136	64	38	26	299	0.042	-0.011	0.387	0.146	0.059

24. C – horas; D – crees; E – tú; F – llegarás; H – cenar
* correct answer

As so many test takers chose distractors "*llegarás*" and "*cenar*", the IF index of these items is positive across grades and the IF index of "*llegarás*" is even higher than the IF of the correct target word "*tú*".

Four sentences aiming at verbs also presented some kind of disruption regarding word order in the sentence. These are items 3, 7, 9 and 17. In item 3, the indirect object appears reduplicated and fronted and the subject is omitted while the stem sentence follows the normal order.

3. Olga CAMINÓ trescientos metros ayer.

A mi primo Francisco le regalaron un automóvil nuevo.

☐☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

57.5% of those who failed the item (N=113) chose the penultimate word, the noun "*automóvil*" (automobile) in the target sentence before any other option. However, this item is not difficult for any grade but for grade 3 (see Table 3.25).

Table 3.25. Item 3 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	C	D	F*	H	I	Missing	Attempts	IF – C	IF – D	IF – F*	IF – H	IF – I
3	66	2	4	26	23	9	0	66	-0.045	-0.015	0.318	0.273	0.061
4	75	5	2	46	16	4	1	74	0.020	-0.020	0.574	0.169	0.007
5	57	4	4	38	9	1	0	57	0.029	0.029	0.625	0.126	-0.024
6	60	1	1	47	8	2	1	59	-0.008	-0.008	0.771	0.110	0.008
7	67	4	1	51	9	1	1	66	0.032	-0.013	0.744	0.108	-0.013
all	325	16	12	208	65	17	3	322	0.005	-0.007	0.602	0.158	0.009

3. C – primo; D – Francisco; F – regalaron; H – automóvil; I – nuevo
* correct answer

In item 9, it is in the stem sentence where the indirect object appears before the verb ("*A Juan le COMPRARÁN un regalo el lunes*" - John will be bought a present on Monday), while in the target sentence, the Subject + Verb order is kept although they appear preceded by a time adverbial. This item, like item 3, is especially difficult for 3-graders, and it is still difficult for those in grade 4. Again, most who failed it (34 test takers) chose the last word in the target sentence over any other option.

9. A Juan le comprarán un regalo el lunes.
El fin de semana pasado, María y José jugaron fútbol.
☐　　☐　　☐　　☐ ☐☐　　☐　　☐

In item 7, in which the subject in the target sentence appears at the end, the distribution between those who choose a wrong word is even (except, of course, the function words, which were hardly ever chosen). That is to say, content-word distractors "*verano*" (summer), "*vientos*" (winds) and "*calientes*" (hot) are popular in approximately the same degree, as they are chosen by 38, 37 and 22 test takers respectively (see Table 3.26). This item is not easy for grades 3 and 4, grades in which most test takers go for the first content word in the target sentence "*verano*" (summer). Therefore, the postposition of the subject in the target sentence does not seem to affect the answers in any grade. Instead, what seems to determine the response pattern is the coinciding order of appearance of the elements in both sentences.

7. Pedro PONE el despertador todas las noches.
En el verano soplan vientos calientes.
☐☐ ☐　　☐　　☐　　☐

Table 3.26. Item 7 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	C	D*	E	F	Missing	Attempts	IF – C	IF – D*	IF – E	IF – F
3	66	10	29	11	11	3	63	0.051	0.352	0.067	0.067
4	75	10	45	12	7	0	75	0.053	0.520	0.080	0.013
5	57	13	36	4	1	1	56	0.161	0.571	0	-0.054
6	60	2	54	2	2	0	60	0.013	0.880	0.013	0.013
7	67	3	52	8	1	2	65	0.006	0.760	0.083	-0.025
all	325	38	216	37	22	6	319	0.055	0.613	0.051	0.004

7. C – verano; D – soplan; E – vientos; F – calientes
* correct answer

Contrary to what happens in item 7, in item 17 one distractor ("*canción*" - song) is much more popular than the others, probably because in the printed version of the test it appears right beneath the word to be matched with. Also, as opposed to what happened in items 3 and 9, the test takers were probably aware of the fact that "*bebé*" (baby), the last word in the target sentence, could not be matched with "*llevó*" (carry), whose category they may not have been able to recognise but, for sure, they knew the right answer was not a person and, therefore, "*bebé*" was not likely to be the correct answer. That could be the reason why they did not choose the last word and went for "*canción*" instead.

17. Cuando llueve, yo siempre LLEVO botas de goma.
La mamá le cantó una canción de cuna a su bebé.
☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Although it does not present any change in the expected order of words in the sentence, the response pattern in item 31 should be analysed as well, as it is an item that is still difficult for the test takers in grade 5.

31. Marcos BAILA merengue con Marcela.
Los niños descuidados dejan los libros en el suelo.
☐ ☐ ☐ ☐ ☐ ☐ ☐☐ ☐

21% of those test takers who attempted the item chose "*descuidados*" (careless) over "*dejan"* (leave). One possible reason for this choice could be that "*descuidados*" is a deverbal adjective and, therefore, is close to "*baila*" ((he) dances), grammatically speaking.

Only one item aiming at the function of direct object and one item aiming at the function of adjective present some kind of disruption as regards word order. These are

items 4 (see Table 3.21) and 11. In item 11, which appears to be still difficult for 5-graders, the verb "*recogió*" (picked up) was chosen by as many as 85 test takers, which is 45.7% of those test takers who attempted the item. In this case, the position of the word "*recogió*" is the second content word that appears in the target sentence and so is the word in capitals in the stem.

<p style="text-align:center">11. Por tirar BASURA, castigaron a Juan.
Susana recogió su muñeca del suelo.
☐   ☐   ☐   ☐   ☐ ☐</p>

Table 3.27. Item 11 on the MLAT-ES *Parte 2 Palabras que se corresponden*: *p*-values and distractor behaviour analysis

Grade	N	A	B	D*	F	Missing	Attempts	IF – A	IF – B	IF – D*	IF – F
3	66	6	19	28	8	3	63	-0.016	0.190	0.333	0.016
4	75	3	19	43	8	0	75	-0.045	0.168	0.488	0.021
5	57	6	23	20	7	0	57	-0.025	0.274	0.221	-0.007
6	60	0	14	42	4	0	60	-0.060	0.173	0.640	0.007
7	67	1	10	53	3	0	67	-0.027	0.107	0.749	0.003
all	325	16	85	186	30	3	322	-0.035	0.178	0.493	0.009

11. A – Susana; B – recogió; D – muñeca; F – suelo
* correct answer

As for the D_i, all items reach an acceptable index, even item 22, which was removed in the final version of the MLAT-ES. The only item that could be improved according to this index — though it is still good — is item 1 (see Table F.9 in appendix F for a complete account of D_i and number of correct answers in the upper and lower 33% tails).

Table 3.28. Percentage of items on the MLAT-ES *Parte 2 Palabras que se corresponden* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=325	**96.8%** Items: 2-31	**3.2%** Items: 1	**0%** Items: -	**0%** Items: -

Both item 1 and item 22 are closer to the limit than the rest of items for an item to be acceptable (r_{pb} .355** and r_{pb} .381** respectively) in terms of its correlation with the total score, as shown in Table 3.29.

Table 3.29. Item-total correlation of MLAT-ES *Parte 2 Palabras que se corresponden*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
1	.355**	6	.507**	11	.472**	16	.545**	21	.553**	26	.609**
2	.571**	7	.601**	12	.580**	17	.527**	22	.381**	27	.640**
3	.610**	8	.558**	13	.570**	18	.489**	23	.571**	28	.575**
4	.416**	9	.667**	14	.442**	19	.484**	24	.619**	29	.693**
5	.683**	10	.472**	15	.645**	20	.528**	25	.563**	30	.596**
										31	.630**

*** *p*<.001

The Cronbach's alpha coefficient without deleting item 2 was excellent, as it was .922, and it actually stays the same if deleted. The internal consistency of the test was fairly good as well (see Table F.10 in appendix F).

3.5.2.1.3. Parte 3 Palabras que riman

As far as the difficulty of *Parte 3* is concerned, Table 3.30 shows that, as happened in parts 1 and 2, the test gets easier the higher the grade. Nevertheless, while some items prove to be, as expected, increasingly easier across grades (item 6, 14, 28, 29, 38, 39) or keep the same level of facility, some others (1, 11, 16, 20, 22, 46) present oscillating patterns, which leads to questioning their correct functioning within the test. Actually four of these six items disappear in the final version of the test.

Table 3.30. Percentage of items on the MLAT-ES *Parte 3 Palabras que riman* according to their facility across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=66	**28.2%** Items: 2, 3, 5, 10, 15, 19, 20, 26, 27, 31, 32, 33, 36	**50%** Items: 1, 4, 8, 9, 11, 12, 16, 17, 18, 21, 22, 23, 24, 25, 30, 34, 35, 37, 40, 41, 42, 44, 45	**10.9%** Items: 6, 14, 28, 29, 43	**8.7%** Items: 7, 13, 38, 39	**2.2%** Items: 46
Grade 4 N=75	**52.2%** Items: 1, 2, 3, 5, 10, 15, 17, 18, 19, 21, 23, 25, 26, 27, 30, 31, 33, 36, 40, 41, 42, 43, 44, 45	**37%** Items: 4, 7, 8, 9, 12, 13, 14, 20, 24, 28, 29, 32, 34, 35, 37, 38, 46	**6.5 %** Items: 6, 16, 39	**4.3%** Items: 11, 22	**0%** Items: -
Grade 5 N=57	**54.3%** Items: 2, 3, 5, 8, 15, 18, 19, 20, 21, 25, 26, 27, 28, 30, 31, 32, 33, 35, 36, 37, 40, 41, 42, 43, 45	**41.3%** Items: 1, 4, 7, 9, 10, 11, 12, 13, 14, 16, 17, 22, 23, 24, 29, 34, 38, 39, 44	**2.2%** Items: 6	**2.2%** Items: 46	**0%** Items: -
Grade 6 N=60	**76.1%** Items: 2, 3, 4, 5, 8, 9, 10, 12, 13, 14, 15, 18, 19, 20, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 40, 41, 43, 44, 45	**21.7%** Items: 1, 6, 7, 16, 17, 24, 38, 39, 42, 46	**2.2%** Items: 11	**0%** Items: -	**0%** Items: -
Grade 7 N=67	**91.3%** Items: 1, 2, 3, 4, 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45	**8.7%** Items: 6, 7, 11, 22, 46	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grades 3-7 N=325	**67.4%** Items: 2, 3, 4, 5, 8, 9, 10, 15, 17, 18, 19, 20, 21, 23, 25, 26, 27, 29, 30, 31, 32, 33, 35, 36, 37, 40, 41, 42, 43, 44, 45	**30.4%** Items: 1, 6, 7, 11, 12, 13, 14, 16, 22, 24, 28, 34, 38, 39	**2.2%** Items: 46	**0%** Items: -	**0%** Items: -

A total of eight items (item 7, 11, 13, 15, 16, 17, 20 and 22) were eliminated in the final version of the test.

7. BUENO…..... ☐ tengo………. ☐ muerdo……. ☐ heno............. ☐ pino

11.YESO........... ☐ paso.............. ☐ piso …......… ☐ tropiezo........ ☐ ocio

13.AMIGO …… ☐ obligó.........… ☐ ombligo…... ☐ refugio...…… ☐ hormiga

15.TESORO....... ☐ oro ………... ☐ tijera ……... ☐ taro.............. ☐ sordo

16.REPISA........ ☐ tiza …..……... ☐ repasa…....… ☐ piso ....… ☐ reposa

17.BARCO........ ☐ charco.........… ☐ barato…....... ☐ burrito...…… ☐ terco

20.RIZO............. ☐ quiso...……... ☐ pozo.......…… ☐ precio..........… ☐ pizza

22.BUZO.......... ☐ dispuso.........… ☐ hizo…….......... ☐ aburre...…… ☐ liso

In items 7, 15 and 17 test takers are expected to discern between assonant and consonant rhyme. That is, they face two or more distractors that rhyme with the stem, but they are supposed to choose the distractor that rhymes consonantically. Since choosing the consonant rhyme over the assonant was perhaps not clearly explained in the directions, most test takers who failed these items chose the distractors which rhyme with the stem in an assonant way. For instance, in item 7 distractors A "*tengo"* and B "*muerdo*" rhyme in assonance with the stem ("*bueno*"), but what is considered the only true answer is option C "*heno*", which rhymes consonantically. Only 14 test takers choose distractor D, which shares the last two letters with the stem but does not rhyme with it (see appendix H for the complete item behaviour analysis of Parte 3).

Table 3.31. Item 7 on the MLAT-ES *Parte 3 Palabras que riman*: *p*-values and distractor behaviour analysis

Grade	N	A	B	C*	D	Missing	Attempts	IF – A	IF – B	IF – C*	IF – D
3	66	4	20	34	6	2	64	-0.250	0.083	0.375	-0.208
4	75	10	11	52	2	0	75	-0.156	-0.138	0.591	-0.298
5	57	8	10	38	1	0	57	-0.147	-0.099	0.556	-0.310
6	60	6	6	44	2	2	58	-0.195	-0.195	0.678	-0.287
7	67	4	9	51	3	0	67	-0.254	-0.154	0.682	-0.274
all	325	32	56	219	14	4	321	-0.200	-0.100	0.576	-0.275

7. BUENO A – tengo; B – muerdo; C – heno; D – pino
* correct answer

Item 13, which has the word "*amigo*" (friend) as stem, appears to be confusing for the test takers. 16.8% of those who attempt this item seem not to have paid attention to rhyme but to the spelling of the word-ending and choose distractor A "*obligó"* instead of option B "*ombligo"*. It is worth mentioning that 70.4% of those who choose A are in grades 3 and 4 while in most other items, the choice of a wrong distractor is not so salient in the lower grades.

Table 3.32. Item 13 on the MLAT-ES *Parte 3 Palabras que riman*: *p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	22	34	4	2	2	64	0.135	0.385	-0.240	-0.281
4	75	16	52	2	5	0	75	-0.049	0.591	-0.298	-0.244
5	57	6	45	2	3	1	56	-0.190	0.738	-0.286	-0.262
6	60	8	47	2	3	0	60	-0.156	0.711	-0.289	-0.267
7	67	0	59	0	6	2	65	-0.333	0.878	-0.333	-0.210
all	325	54	237	10	19	3	322	-0.108	0.650	-0.290	-0.253

13. AMIGO A – obligó; B – ombligo; C – refugio; D – hormiga
* correct answer

Four of the items removed (11, 16, 20 and 22) aim at recognising assonant rhymes disguised with the use of sibilant fricative consonants represented graphically like this: <s, c, z>. In these items, the test takers who failed them forget about the stressed syllable and about finding the rhyming partner and focus instead of the last letters of the distractors. These items coincide with those which present different degrees of facility across grades, but, at the same time, have good discrimination power, as shown in Table 3.33 (see Table F.11 in appendix F for a complete account of the D_i and correct answers in the upper and lower 33%).

Table 3.33. Percentage of items on the MLAT-ES *Parte 3 Palabras que riman* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=325	63% Items: 7, 11, 12, 13, 14, 16, 20, 22, 24, 28-46	26.1% Items: 1, 4, 6, 8, 9, 17, 18, 19, 21, 25, 26, 27	10.9% Items: 2, 3, 5, 10, 15	0% Items: -

In contrast with what happens in *Parte 1*, the items that have average discrimination power according to the data of this study are all at the beginning of the test and, therefore, were answered by most test takers. Consequently, since most test takers attempted these items, their lack of discrimination power is a rather acute matter that should definitely be solved by finding more effective distractors. The eight items removed in the published version appear all in the first half of the test as well, although only item 15 coincides with the items that only have average discrimination power according to these data.

The distribution of the item-total correlation in this part (see Table 3.34) is very similar to that of *Parte 1 Palabras ocultas.* This is so because of the speeded nature of this part, which is a feature common to the first three parts. Despite the fact that the correlations are lower for the beginning-of-test items, they are all over .300 and are significant, which supports the discrimination power of this test.

Table 3.34. Item-total correlation of MLAT-ES *Parte 3 Palabras que riman*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
1	.322**	9	.393**	17	.475**	25	.495**	33	.679**	41	.713**
2	.423**	10	.361**	18	.569**	26	.512**	34	.677**	42	.687**
3	.339**	11	.440**	19	.556**	27	.549**	35	.735**	43	.700**
4	.350**	12	.513**	20	.505**	28	.610**	36	.676**	44	.694**
5	.362**	13	.460**	21	.446**	29	.618**	37	.691**	45	.643**
6	.392**	14	.520**	22	.486**	30	.618**	38	.501**	46	.616**
7	.392**	15	.392**	23	.553**	31	.650**	39	.623**		
8	.408**	16	.507**	24	.512**	32	.680**	40	.666**		

** *p*<.001

Item 1 presents the lowest correlation and is, indeed, the item that in the corrected item-total correlation of the reliability analysis does not reach the desirable .300 value, although it is very close to it. The reliability of the test without removing only item 1 is very high (.946) anyway (see Table F.12 in appendix F).

The items removed in the final version are items 7, 11, 13, 15, 16, 17, 20 and 22. According to the analyses run with the data of 325 test takers, which is not as big a population as the one used for the norming study, the items that present some kind of flaw coincide partially with those removed. All the items removed had a good D_i. However, item 1, which has been kept in the final version, presents too low an item-total correlation as compared to the rest of items. This could be due to the fact that, as mentioned before, the stem is meant to be matched to a word that rhymes consonantly, yet two more distractors present a vowel rhyme with the stem. Strictly speaking, three options rhyme with the stem, as it is never explicitly stated that the words that are to be matched are only those that share the last letters.

Table 3.35. Item 1 on the MLAT-ES *Parte 3 Palabras que riman*: *p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	66	12	44	5	5	0	66	-0.091	0.556	-0.232	-0.232
4	75	9	58	6	2	0	75	-0.173	0.698	-0.227	-0.298
5	57	4	40	4	9	0	57	-0.240	0.602	-0.240	-0.123
6	60	5	45	6	4	0	60	-0.222	0.667	-0.200	-0.244
7	67	5	52	8	1	1	66	-0.232	0.717	-0.172	-0.313
all	325	35	239	29	21	1	324	-0.189	0.650	-0.214	-0.247

1. RESPUESTA A – pueda; B – fiesta; C – nueva; D – pista
* correct answer

3.5.2.1.4. Parte 4 Aprendamos números

Unlike the other parts, which are multiple-choice-like, this part requires the test taker to produce the answer rather than choose it from a range of options. Precisely because of the relative freedom of the item format on this part, there appear number combinations and figures that were not in the instructions, although these are the fewer. In principle, the test taker is supposed to write figures that are made up of numbers 1, 2, 3, 10, 20 and 30 or the combination thereof (11, 12, 13, 21, 22, 23, 31, 32, 33). In addition to the numbers in Table 3.36, other numbers appeared as answers as well, even though they were not mentioned in the instructions. These are numbers made up of ciphers 4, 5, 6, 7 and 8 and were numbers 14, 15, 16, 17, 24, 34, 38, 4, 40, 42, 44, 5, 50, 52, 53, 54, 60.

The items in the test are distributed as follows. Notice that the item letters refer to the order of appearance in the test. Supposedly, numbers were avoided to mark the order of the items so as to avoid confusion with the numbers to be written.

Table 3.36. Numbers aimed at, number names and order of appearance on the MLAT-ES *Parte 4 Aprendamos números*

Number	Number in letters	Item (letter)	Number	Number in letters	Item (letter)
1	co	j	20	vinca	l, w
2	vein	b, x	21	vinca-co	e, p
3	ras	f	22	vinca-vein	i, r
10	silca	n	23	vinca-ras	a, u
11	silca-co	c, v	30	rasca	h
12	silca-vein	q	31	rasca-co	k
13	silca-ras	g, ñ, s	32	rasca-vein	m, t
			33	rasca-ras	d, o

The figures presented in the instructions appeared combined in numbers of two figures maximum (i.e. tens and units) combining four numbers: 0, 1, 2, and 3. Nevertheless, test takers wrote other combinations with up to three figures. These are numbers 110, 120, 130, 210, 220, 222, 230, 310, 320, 330. As can be seen, all these three-figure numbers except for 222 are built from the combination of the first figure plus 10, 20 or 30 and, precisely because of that, in most cases the number aimed at was the one in the first two figures of these 3-figure combinations. For instance, when being asked to write number 13 (*silca-ras*), test takers interpreted that they had to write 130.

Most mistakes were not made due to writing three-figure combinations, but due to three other reasons. The first mistake pattern, which is the least recurrent, was made

by writing numbers belonging to the same ten. For instance, some test takers (4.6% and 3.7%) wrote 12 or 13 respectively for number 11 the first time this number is dictated on the test.

The second type of mistake was writing a two-figure number that has the same unit as the intended number, so the mistake is in the tens (see Table F.13 in appendix F, in which one can see the numbers that were written most often instead of the ones aimed at and the percentage of test takers that wrote them wrong, both the first and the second time they appear in the test). Apparently, the mistakes were complementary, that is, if number 11 was mistaken by number 21, 21 was also mistaken by number 11.

It is worth mentioning that no combination with number 30 plus any unit appears in this table. This could be due to the fact that number 30 in Spanish, as already mentioned, has a meaning of its own in Spanish ("*rasca*" means "it scratches"), so this could have contributed to a more effective memorisation of the number and, hence, less chances of making mistakes with this number.

The third most salient reason for making a mistake in this test is the confusion between the unit and the corresponding ten. In this type of mistake, while number 1 (*co*) is not mistaken for 10 (*silca*) very often (their names are very different indeed), *rasca* and *ras* (30 and 3) are mistaken more often (almost 10% of the test takers wrote 3 instead of 30 and 30 instead of 3) and so are 20 (*vinca*) and 2 (*vein*), probably due to their similarity. The most common mistake by far was writing 20 (*vinca*) instead of 2 (*vein*). The first time *vinca* appears in the test as target number 25.2% of the test takers wrote 2 instead, and so did 26.8% of the test takers the second time number 20 (*vinca*) appears in the test (see all this type of mistakes and the corresponding percentages in Table F.14 in appendix F).

The name for number 2 (*vein*) in the artificial language used in this part is very similar to the real name for number 20 in Spanish (*veinte*), which could have led to this confusion. What is surprising, though, is that even if these two numbers appeared one after another at the end of the test (items w, x), 50 test takers wrote number 20 as the answer to both item w and x, while only 3 test takers wrote 2 (*vein*) as the answer to both item w and x (see all response patterns to items w and x in Table F.15 in appendix F).

Table 3.37 below shows precisely that number 2 (*vein*) is the item that appears to be consistently most difficult at grades 3 and 4, especially the second instance (IF=.47 at grade 3 and IF=.55 at grade 4) although, generally speaking, all the items on this test are very easy. For a complete account of the index of facility of all the items, see appendix I.

Table 3.37. Percentage of items on the MLAT-ES *Parte 4 Aprendamos números* according to their facility across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=66	**36 %** Items: 12 (q); 13 (ñ); 23 (a, u); 30 (h); 32 (m, t); 33 (d, o)	**60%** Items: 1 (j); 2 (b); 3 (f) ; 10 (n); 11 (c, v); 12 (g); 13 (s); 20 (l, w); 21 (e, p); 22 (i, r); 31 (k)	**4%** Items: 2 (x)	**0%** Items: -	**0%** Items: -
Grade 4 N=75	**76%** Items: 1 (j); 3 (f); 10 (n); 12 (g, q); 13 (s); 20 (l, w); 21 (e, p); 22 (i, r); 23 (u); 30 (h); 31 (k); 32 (m, t); 33 (d, o)	**24%** Items: 2 (b, x); 11 (c, v); 13 (ñ); 23 (a)	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grade 5 N=57	**82%** Items: 1 (j); 3 (f); 11 (c, v); 12 (q); 13 (ñ, s); 20 (l, w); 21 (e, p); 22 (i, r); 30 (h); 31 (k); 32 (m, t); 33 (d, o)	**8%** Items: 2 (b, x); 12 (g)	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grade 6 N=60	**96%** Items: 1 (j); 2 (b); 3 (f); 10 (n); 11 (c); 12 (g, q); 13 (ñ, s); 20 (l, w); 21 (e, p); 22 (i, r); 23 (a, u); 30 (h); 31 (k); 32 (m, t); 33 (d, o)	**4%** Items: 2 (x)	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grade 7 N=67	**100%** Items: 1 (j); 2 (b, x); 3 (f); 10 (n), 11 (c, v); 12 (g, q); 13 (ñ, s); 20 (l, w); 21 (e, p); 22 (i, r); 23 (a, u); 30 (h); 31 (k); 32 (m, t); 33 (d, o)	**0%** Items:-	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grades 3-7 N=325	**92%** Items: 1 (j); 3 (f); 10 (n); 11 (c, v); 12 (g, q); 13 (ñ, s); 20 (l, w); 21 (e, p); 22 (i, r); 23 (a, u); 30 (h); 31 (k); 32 (m, t); 33 (d, o)	**8%** Items: 2 (b, x)	**0%** Items: -	**0%** Items: -	**0%** Items: -

All the parts of the test prove to be very easy, even in the lower grades. This part is, though, the easiest of them all, as there are no difficult items at any grade. Moreover, the percentage of very easy items is extremely high from grades 5 to 7. It is, therefore, questionable whether the test works properly for the upper grades.

As for the item discrimination power, once again, those items that need some kind of improvement are those which consistently have to do with number 30 (see Table 3.38), which is the only number that does actually mean something by itself (see D_i and the number of correct answers in the upper and lower 33% tail in F-16 in appendix F).

Table 3.38. Percentage of items on the MLAT-ES *Parte 4 Aprendamos números* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=325	**68%** Items: 1 (j); 2 (b, x), 3 (f); 10 (n); 11 (c, v); 12 (g, q); 13 (ñ, s); 20 (l, w); 21 (e), 22 (i, r)	**20%** Items: 21 (p); 23 (a, u); 31 (k); 32 (t); 33 (d)	**12%** Items: 30 (h); 32 (m); 33 (o)	**0%** Items: -

The reliability of *Parte 4* is excellent (.925) and, according to the corrected item-total correlation, no item should be removed. Both Table 3.39 and Table F.17 in appendix F, which contain the item-total correlation of *Parte 4*, confirm that item h (the one that aims at number 30 - *rasca*) functions differently from the rest, as its correlation coefficient is the lowest of all. Nevertheless, it is still above the threshold of deficient items in relation to the rest of items in this part, as its correlation coefficient is above .300.

Table 3.39. Item-total correlation of the MLAT-ES *Parte 4 Aprendamos números*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
a. 23	.432**	**f. 3**	.624**	**k. 31**	.633**	**o. 33**	.609**	**t. 32**	.683**
b. 2	.501**	**g. 12**	.669**	**l. 20**	.534**	**p. 21**	.647**	**u. 23**	.635**
c. 11	.617**	**h. 30**	.371**	**m. 32**	.581**	**q. 12**	.720**	**v. 11**	.684**
d. 33	.630**	**i. 22**	.588**	**n. 10**	.602**	**r. 22**	.678**	**w. 20**	.479**
e. 21	.665**	**j. 1**	.636**	**ñ. 13**	.664**	**s. 13**	.681**	**x. 2**	.519**

** *p*<.001

3.5.2.2. Reliability and intercorrelation of parts

The reliability of each test part is calculated with Cronbach's alpha for the reasons explained in section 2.3.2.3, as only one form of each test was administered and therefore, it was not possible to calculate the reliability using the split-half method. Besides, as the tests are supposed to be highly-speeded, it is not possible to use split-half scores or item variances, as test takers are not expected to be able to attempt all the items, especially at lower grades.

From the item analysis run, it was decided to remove item 3 from *Parte 1 Palabras ocultas*. In *Parte 2 Palabras que se corresponden,* it seemed more than sensible to remove the miskeyed item. The item analysis of *Parte 3 Palabras que riman* posed some more difficulty as regards the removal of faulty items, since the items removed from the published version of the test do not totally coincide with those that appear to be faulty in the data of this study. The items that discriminate only with average quality do not fully coincide with the items removed and neither do those which present irregular IF patterns. However, taking into account that this study involves only 325 test takers, while in the official version 1,186 test takers were analysed, and that the items removed do not work perfectly well in these data either, it also seems sensible to remove the same items as the SLTI removed. After removing these items, the index of alpha reliability diminishes a little bit (.877 in grade 7), which is still a very good reliability index, though. *Parte 4 Aprendamos números* has very good reliability indexes in all grades, although it is an extremely easy test. However, no item seems to function badly, so all items will be kept in further analyses.

In order to complete the statistical information of this validation phase, it is necessary to include here a table with the reliability coefficients and standard errors of measurements of the total raw scores on the MLAT-ES. While the data in the *MLAT-E Manual* appears split according to the sex variable, this division has not been kept here due to the reduced sample of the population, which would result in too small a number when split according to sex. Besides, boys and girls are considered together in the same group in the statistical information of the *MLAT-ES Manual* too. The reliability indexes in Table 3.40 below are excellent across grades and so is the reliability index for all parts and grades, which does not appear in the table (Cronbach's alpha .969). The standard error of measurement decreases regularly from grades 3 to 7. In contrast, the mean increases across grades. Notice that between grades 3 and 4, the mean increases more than 16 points, while the increase is not so sharp across the other grades, ranging from 9.28 points between grades 4 and 5 to 6.76 between grades 6 and 7. Actually, the difference between means diminishes as the grade increases. These differences across grades in the means have to be taken into account when analysing the validity of the test across the grades the test is meant to be addressed to. Also, while standard deviations are rather stable from grades 3 to 6 (the maximum difference being 3.2 points), the difference in standard deviation between grades 6 and 7 is of 5.8 points, which almost doubles the difference between the other grades. This also shows that the variance of the scores in grade 7 is much

smaller than in other grades and, perhaps, the discrimination power of the test (or of one or more of its parts) is not optimal.

Table 3.40. Reliability coefficients and standard errors of measurements of the total raw scores on the MLAT-ES administered in Catalonia

Statistics	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7
N	66	75	57	60	67
Reliability	.948	.957	.958	.959	.930
SE$_M$	4.81	4.60	4.33	3.85	3.49
Mean	65.09	81.84	91.12	99.93	106.69
SD	21.1	22.0	21.1	19.0	13.2

Following is the table that summarises the intercorrelations, the reliability coefficients, the part-total correlations, the mean p-values and the mean item total of parts on the MLAT-ES run in this study after removing the items that disappear from the original version in the published test.

Table 3.41. Intercorrelations, reliability coefficients, part-total correlations, mean *p*-values, and mean item total of parts of the MLAT-ES administered in Catalonia

	GRADE 3 (N=66)				GRADE 4 (N=60)			
	1	2	3	4	1	2	3	4
1. *Palabras ocultas*	.896				.896			
2. *Palabras que se corresponden*	.229*	.825			.366**	.895		
3. *Palabras que riman*	.369**	.397**	.932		.605**	.418**	.938	
4. *Aprendamos números*	.330**	.356**	.392**	.925	.443**	.338**	.437**	.909
Part–total correlation	.629**	.637**	.832**	.686**	.792**	.695**	.844**	.658**
Mean *p*-value	.63	.36	.65	.70	.76	.55	.74	.78
Mean item total	.43	.42	.58	.70	.65	.59	.67	.78
	GRADE 5 (N=57)				GRADE 6 (N=60)			
	1	2	3	4	1	2	3	4
1. *Palabras ocultas*	.849				.912			
2. *Palabras que se corresponden*	.442**	.921			.456**	.908		
3. *Palabras que riman*	.458**	.500**	.938		.471**	.526**	.934	
4. *Aprendamos números*	.354**	.349**	.348**	.925	.458**	.383**	.496**	.863
Part–total correlation	.718**	.822**	.792**	.570**	.810**	.783**	.811**	.634**
Mean *p*-value	.80	.56	.75	.83	.83	.74	.81	.91
Mean item total	.76	.62	.76	.83	.76	.78	.82	.91
	GRADE 7 (N=67)				GRADES COMBINED (N=329)			
	1	2	3	4	1	2	3	4
1. *Palabras ocultas*	.766				.927			
2. *Palabras que se corresponden*	.328**	.905			.579**	.925		
3. *Palabras que riman*	.477**	.301*	.877		.622**	.572**	.942	
4. *Aprendamos números*	.180	.194	.286*	.923	.506**	.466**	.517**	.925
Part–total correlation	.626**	.703**	.748**	.496**	.839**	.821**	.848**	.690**
Mean *p*-value	.90	.77	.85	.91	.80	.60	.77	.83
Mean item total	.89	.80	.87	.91	.69	.64	.74	.83

Note. – Reliability coefficients are indicated in bold-face type.
* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

Despite there being more than 30 participants in each grade, the intercorrelations shown in Table 3.41 above are not Pearson but Spearman, as the distribution of scores was not normal. This is shown by the fact that, in the Kolmogorov-Smirnov test, which was also run, all significant values of all parts and for all grades were less than .05 $p<.05$, which indicates a deviation from normality. However, the Shapiro-Wilk test, which is generally more accurate than the Kolmogorov-Smirnov, presents a non-significant value for the total score in grade 3, D(66)=0.825, $p<.05$ and in grade 4 D(75)=0.119, $p<.05$ (see Table F.18 in appendix F).

The correlations between all parts and the total score are high and significant in all grades but for *Parte 4 Aprendamos números* in grade 7, which is moderate though still significant. In fact, of all parts, *Parte 4* is the one that has the lowest correlations with the total except in grade 3. In this grade, the correlations of parts 1 and 2 are lower

than in the rest of grades except for *Parte 1* in grade 7. Intercorrelations between parts are from low to moderate in all parts and grades (ranging from .180 *ns* between parts 1 and 4 in grade 7 to .605 at *p*<.001 between *Parte 1* and *Parte 3* in grade 4). *Parte 4 Aprendamos números* is the part that consistently has the lowest correlations with the rest of parts, probably because it is the only part meant to measure memory from aural input.

The results change substantially when combining all the grades, as all the correlations appear to be moderate, the highest one being between *Parte 1* and *Parte 3* (r_s=.622 *p*<.001), which is something expectable, as both are meant to tap phonetic coding ability. *Parte 1* and *Parte 2* are also correlated moderately, which is something unexpected, as they are meant to tap different abilities. That is, while *Parte 1* is supposed to measure both knowledge of English vocabulary and sound-symbol association ability, *Parte 2* taps sensitivity to grammatical structure. The lowest correlation (though still a moderate one) in this part is between *Parte 2* and *Parte 4*, although they share as a target the ability to recognise the grammatical function of words (especially *Parte 2*) and to discover the relationships between words (especially *Parte 4*). *Parte 4* is also supposed to measure some memory component, which is what theoretically makes this intercorrelation be lower than the others.

The mean item-total correlation data (which can be interpreted as an index of item discrimination) by test part and grade level go from moderate (in *Parte 1*, *Parte 2* and *Parte 3* in grade 3) to high (in the rest of parts and grades), which is a sign that all items are working similarly. The mean item-total correlation for the whole test and all the grades, which does not appear in the table, was .72. It can be concluded that all parts, as the mean item-total correlation indicates, are very good at discriminating across grades.

Mean *p*-values inform us about the difficulty of the tests for each grade, as they are the average percent correct score on a test. On the basis of its mean *p*-values, *Parte 2 Palabras que se corresponden* appears to be the most difficult for all grades, including when all grades are combined. The other mean *p*-values indicate that all the other parts of the test are easy, as they are above .55. Following *Parte 2*, in an ascending order regarding mean *p*-values, is *Parte 3 Palabras que riman* in all grades but in grade 3, although its value is very close to that of *Parte 1 Palabras ocultas*. *Parte 4 Aprendamos números*, besides always being the easiest one, reaches an extremely high mean *p*-value from grade 5 on and is, therefore, extremely easy for all grades. In fact, the mean *p*-value of this part is the same in grades 6 and 7, which indicates that

the difficulty of this part is similar for both these grades. The mean *p*-value for all the test and grades was .75.

3.5.2.3. Descriptive statistics of the MLAT-ES administered in Catalonia

Below are the descriptive statistics of the MLAT-ES. These have been made including all the participants in the study, even the so-called outliers. This is so because the data are not normally distributed and, therefore, non-parametric tests will be used for further analyses. These make few or no assumptions about the distributions, and do not rely on distribution parameters. Consequently, they are robust enough to avoid using outlier tests as they down-weigh the effect of the extreme values in the population. However, it must be said that all those outliers, for reasons other than low ability (i.e. for their aberrant behaviour as when not finishing the tests deliberately or answering at random on purpose), were automatically dropped from the initial sample and were not included in the previous item analyses either.

The descriptive statistics of *Parte 1 Palabras ocultas* (see Table 3.42) show that, while the difference in the means from grades 4 to 7 is on a more or less gradual increase, it is between grades 3 and 4 where the difference between means is of almost one standard deviation. Medians are fairly close to means across grades. Standard deviations in this part are round 6 points for all grades except in grade 7, in which it is only 3.22. Actually, the mean of this group is very close to the maximum possible score. Also, for this grade, the minimum score is markedly larger than for all other grades. In contrast, some test takers reach the maximum score already in grade 4.

As for the distribution of scores, skewness and kurtosis are presented along with their standard errors and their z-score values. There are two reasons for including the z-score of the skewness and kurtosis. First, the values obtained are not out of round numbers and not all parts have the same amount of items; consequently, it is difficult to interpret what the raw skewness and kurtosis values represent. Second, for small samples like these here, it is not enough to look at the shape of the distribution, although the raw value of skewness and of kurtosis are informative of the shape and direction of the deviance from normality. However, it is more significant to look at its significance rather than at the value of the skewness and kurtosis statistics alone. The significance is calculated by dividing the skewness and the kurtosis minus the mean, which is always 0 in z-scores, by their standard error of measurement. In small

samples, the z-skewness and z-kurtosis values are considered to be significant when they are above (-)2.58 at $p<.01$. These appear in bold letters in the tables. While not significant, the only group which presents positive skewness is grade 3. From grade 4 on, skewness becomes negative and it is significant from grade 5 on. As for kurtosis, the distribution of scores is platykurtic in grades 3, 4 and 6 and leptokurtic in the rest of grades, especially in grade 7.

Table 3.42. Descriptive raw scores for *Parte 1 Palabras ocultas* on the MLAT-ES (30 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SE S[a]	$z_{skewn.}$	Kurt.	SE K[b]	$z_{kurtosis}$
3	66	12.82	11.00	6.81	2	28	26	.532	.295	1.803	-.592	.582	-1.017
4	75	19.45	21.00	6.81	4	30	26	-.368	.277	-1.329	-.986	.548	-1.799
5	57	22.93	25.00	5.33	8	30	22	-.943	.316	**-2.984**	.301	.623	.483
6	60	22.67	25.00	6.65	8	30	22	-.853	.309	**-2.761**	-.529	.608	-.870
7	67	26.76	28.00	3.22	17	30	13	-1.342	.293	**-4.580**	1.267	.578	2.192

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

In *Parte 2 Palabras que se corresponden,* while not so acute as in *Parte 1*, there is a sharper increase in the means between grades 3 and 4 and grades 5 and 6 while the difference is smaller between grades 4 and 5 and minimal between grades 6 and 7. The minimum score and the standard deviation are similar across grades, as well as the range of scores. As in *Parte 1*, the only grade that presents scores positively skewed is grade 3, while in the other grades, skewness turns to the right tail of the distribution, though only significantly at $p <.01$ in grades 6 and 7, where scores pile up to the right tail with a leptokurtic distribution, which reaches a significant value in grade 7.

Table 3.43. Descriptive raw scores for *Parte 2 Palabras que se corresponden* on the MLAT-ES (30 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SES[a]	$z_{skewn.}$	Kurt.	SE K[b]	$z_{kurtosis}$
3	66	12.58	12.00	5.87	3	27	24	.484	.295	1.641	-.646	.582	-1.109
4	75	17.55	18.00	7.21	2	29	27	-.200	.277	-.722	-1.125	.548	-2.052
5	57	18.63	19.00	7.88	3	30	27	-.223	.316	-.705	-1.220	.623	-1.958
6	60	23.33	25.50	6.40	6	30	24	-1.498	.309	**-4.847**	1.390	.608	2.286
7	67	23.99	26.00	6.13	3	30	27	-1.633	.293	**-5.573**	2.304	.578	**3.986**

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

Parte 3 is the only part in which, at first sight, means increase gradually across grades and medians always appear slightly higher than means (see Table 3.44). Ranges are also similar across grades and all skewness values are negative (i.e. most

scores are piled up to the right tail). Skewness is significant at *p*<.01 from grade 5 on. Kurtosis changes shape across grades. While in grades 3 and 4 the distribution is flattish, it starts being leptokurtic in the other grades, reaching a significant value in grade 7, in which the distribution is very pointy.

Table 3.44. Descriptive raw scores for *Parte 3 Palabras que riman* on the MLAT-ES (38 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SES[a]	$z_{skewn.}$	Kurt.	SEK[b]	$z_{kurtosis}$
3	66	22.12	23.00	9.38	3	36	33	-.085	.295	-.288	-1.234	.582	-2.120
4	75	25.31	27.00	9.43	5	38	33	-.385	.277	-1.389	-.934	.548	-1.704
5	57	28.74	32.00	8.83	6	38	32	-1.060	.316	**-3.354**	.153	.623	.246
6	60	31.18	35.00	7.77	8	38	30	-1.207	.309	**-3.906**	.369	.608	.606
7	67	33.10	35.00	5.34	16	38	22	-1.748	.293	**-5.966**	2.751	.578	4.760

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

In *Parte 4 Aprendamos números,* means increase gradually from grades 3 to 6 (see Table 3.45) and seem to reach a plateau in the upper grades. In all grades the maximum score is reached in all grades and medians reach values close to the maximum score from grades 5 to 7. The minimum score in grade 6 is rather high (10 points) as compared to the rest of grades. This makes the range of scores in grade 6 be narrower than in the rest of grades. All distributions are negatively skewed (most scores accumulate on the right tail of the distribution) and their value is highly significant starting already from the lower grades. More significant is, though, the kurtosis value in grade 7, which is 8.747, *p* 15.133. The skewness value is highly significant in all grades, increasing gradually across grades. These values show that scores are mostly piled up at the right tail of the distribution. All kurtosis values are positive, though fairly low in grades 3 and 4. They increase significantly from grade 5 on, which shows that all distributions are pointy, reaching extremely high levels of significance in the upper grades.

Table 3.45. Descriptive raw scores for *Parte 4 Aprendamos números* on the MLAT-ES (25 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SE S[a]	$z_{skewn.}$	Kurt.	SE K[b]	$z_{kurtosis}$
3	66	17.58	19.00	6.79	1	25	24	-.921	.295	**-3.122**	.015	.582	.026
4	75	19.53	19.00	5.66	3	25	22	-1.146	.277	**-4.137**	.620	.548	1.131
5	57	20.82	23.00	5.52	2	25	23	-1.605	.316	**-5.079**	1.940	.623	**3.114**
6	60	22.75	24.00	3.42	10	25	15	-2.324	.309	**-7.521**	5.505	.608	**9.054**
7	67	22.84	25.00	4.14	3	25	22	-2.818	.293	**-9.618**	8.747	.578	**15.13**

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

Overall, mean scores increase from the lower grades to the upper ones. The increase is, however, not gradual. On the contrary, the difference between means decreases the upper the grade. Thus, while almost 17 points separate the mean score between grades 3 and 4, this difference is less than 10 points between grades 4 and 5, close to 9 points between grades 5 and 6, and only 6.76 points between grades 6 and 7, which is one third of the differences between the means of grades 3 and 4. Standard deviations are smaller in grade 7 than in the other grades. Minimum scores are very similar in grades 5, 6 and 7 (ranging from 41 to 49), while they are half these values in grades 3 and 4, the minimum scores being 21 and 24 respectively. Skewness values are negative in all grades but 3, which shows that from grades 4 to 7 distribution values are skewed towards the right, reaching significant values only in grades 6 and 7. Kurtosis values are negative from grades 3 to 5 and are leptokurtic in the other grades, reaching a high significant value in grade 7.

Table 3.46. Descriptive raw total scores on the MLAT-ES (123 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SES[a]	$z_{skewn.}$	Kurt.	SEK[b]	$z_{kurtosis}$
3	66	65.09	64.00	21.09	21	114	93	.078	.295	.264	-.553	.582	-.950
4	75	81.84	83.00	22.16	24	119	95	-.382	.277	-1.379	-.510	.548	-.931
5	57	91.12	97.00	21.12	41	120	79	-.622	.316	-1.968	-.604	.623	-.969
6	60	99.93	105.50	19.04	48	120	72	-1.045	.309	**-3.382**	.365	.608	.600
7	67	106.69	110.00	13.24	49	122	73	-1.711	.293	**-5.840**	4.366	.578	**7.553**

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

The MLAT-ES is meant to be an aptitude measure valid for children in grades 3 to 7. However, score patterns differ significantly across grades. A sharp increase between grades 3 and 4 is observed, especially in the first and second subparts of the test and the effect of this sharp increase shows again in the raw total score means. On the other hand, the rate of this increase diminishes the upper the grade, so it could be possible that at some later point (from grade 8 on) scores reached a plateau. The test takers' results on *Parte 3* increase gradually across grades, so there are no big differences between one particular grade and another. The difference between the means between the lowest grade (grade 3) and the highest (grade 7) is similar in the first three parts (about 11 points in *Parte 2* and *Parte 3* and 13.94 in *Parte 1*), although it is important to remember that both *Parte 1* and *Parte 2* have 30 items while *Parte 3* has 38. *Parte 4*, in contrast with the other parts of the test, shows to be fairly easy in most grades, as the difference between grades 3 and 7 is of only 5.26 points.

216

3.5.3. Statistical information of the MLAT-EC

The same descriptive scheme given about the parts of the MLAT-ES in section 3.5.2 applies here, as both tests contain the same parts and aim to measure the same constructs. Our purpose, though, is to see if the test takers of the MLAT-EC show more or less the same response patterns and results as their counterparts on each part.

3.5.3.1. Content validity of the MLAT-EC

To start with, speed was an issue in the first three parts of the MLAT-ES and so it is in the MLAT-EC as well. The results on the MLAT-EC displayed in Table 3.47 show that while there is an abrupt decrease between grades 3 and 4 of the percentage of test takers who do not manage to finish parts 1 and 3, this decrease is not so abrupt on part 2. As for the rest of parts, the tendency is for the test takers between grades 5 and 7 to manage to finish these parts, but the decrease is not completely linear in any of the parts. This could be so because of the test answering strategy each test taker adopted, which could not be necessarily dependent on grade or age.

Table 3.47. Percentage of unfinished tests per parts (MLAT-EC)

Test	Grade	3	4	5	6	7	All grades
	N	57	62	61	60	64	309
Part 1	Unfinished	46	21	16	18	11	112
	Percentage	80.7%	33.9%	26.2%	30%	17.2%	36.2%
Part 2	Unfinished	22	10	6	14	11	63
	Percentage	38.6%	16.1%	9.8%	23.3%	17.2%	8.9%
Part 3	Unfinished	39	13	16	15	12	95
	Percentage	59%	17.3%	28.1%	25%	17.9%	29.2%

3.5.3.1.1. Part 1 Paraules ocultes

Like in the item analysis of the MLAT-ES, below is the table with the items of *Part 1 Paraules ocultes* distributed according to their IF across grades.

Table 3.48. IF of items on MLAT-EC *Part 1 Paraules ocultes* across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=57	**61.3%** Items: 1, 3, 5, 7, 10, 11, 12, 13, 15, 16, 17, 18, 20, 21, 22, 25, 27, 28, 29, 31	**22.6%** Items: 6, 8, 9, 14, 23, 24, 26	**3.2%** Items: 4	**3.2%** Items: 2	**9.7%** Items: 12, 19, 30
Grade 4 N=62	**74.2%** Items:1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 22, 26, 27, 28, 29, 31	**16.1%** Items: 6, 9, 21, 23, 25	**6.5%** Items: 24, 30	**3.2%** Items: 19	**0%** Items: -
Grade 5 N=61	**71%** Items: 1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 25, 26, 27, 28, 31	**16.1%** Items: 2, 9, 23, 24, 29	**9.7%** Items: 6, 19, 30	**0%** Items: -	**0%** Items: -
Grade 6 N=60	**83.9%** Items: 1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 25, 26, 27, 28, 31	**12.9%** Items: 9, 19, 29, 30	**3.2%** Items: 24	**0%** Items: -	**0%** Items: -
Grade 7 N=64	**87.1%** Items: 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 23, 24, 26, 27, 28, 30, 31	**9.7%** Items: 6, 25, 29	**3.2%** Items: 19	**0%** Items: -	**0%** Items: -
Grades 3- 7 N=304	**74.2%** Items: 1, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 25, 26, 27, 28, 31	**22.6%** Items: 2, 6, 9, 23, 24, 29, 30	**3.2%** Items: 19	**0%** Items: -	**0 %** Items: -

While items 19 (eccpcional) and 30 (lntamnt) become increasingly easier the upper the grade, some items have different degrees of difficulty across grades (mainly items 2, 4, 6, 12 and 24). This is somehow disconcerting unless we take a close look at the distractors in them (see all distractor answer patterns and IF in appendix J). Item 19, which hides an abstract word whose meaning had also caused some confusion in the Spanish version, shows the same response pattern in both tests. That is, test takers are attracted by the distractor "*emocionant*" and do not go for what is considered to be the right answer ("*únic*"). Consequently, distractor C has a positive IF index in all grades but 6, as shown in Table 3.49.

Table 3.49. Item 19 on the MLAT-EC *Part 1 Paraules ocultes: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	57	0	8	10	2	37	20	-0.333	0.200	0.333	-0.200
4	62	6	25	15	2	14	48	-0.167	0.361	0.083	-0.278
5	61	2	29	18	0	12	49	-0.279	0.456	0.157	-0.333
6	60	1	37	11	1	10	50	-0.307	0.653	-0.040	-0.307
7	64	3	39	18	1	3	61	-0.268	0.519	0.060	-0.311
all	304	12	138	72	6	76	228	-0.263	0.474	0.088	-0.298

19. eccpcional A – tipus de triangle; B – únic; C – emocionant; D – mur alt
* correct answer

This is not the case of item 30 (lntament) (see Table 3.50), which has been left blank by many test takers, not only because it appears at the end of the test, but also because of its inner difficulty, especially for 3-graders, as indicated by its positive IF value.

Table 3.50. Item 30 on the MLAT-EC *Part 1 Paraules ocultes: p*-values and distractor behaviour analysis

Grade	N	A	B	C*	D	Missing	Attempts	IF – A	IF – B	IF – C*	IF – D
3	57	2	1	2	2	50	7	0.048	-0.143	0.047	0.048
4	62	6	2	15	1	38	24	0	-0.222	0.500	-0.278
5	61	7	1	20	3	30	31	-0.032	-0.290	0.527	-0.204
6	60	6	2	23	0	29	31	-0.075	-0.247	0.656	-0.333
7	64	1	0	41	3	19	45	-0.304	-0.333	0.882	-0.244
all	304	22	6	101	9	166	138	-0.121	-0.275	0.643	-0.246

30. lntament A – ample; B – blau; C – a poc a poc; D – rodó
* correct answer

Table 3.51 shows the amount of missing values of items 27, 28, 29, 30 and 31 so that it can be seen numerically that the lack of time to answer the test was not the cause of the difficulty of this item. Actually, in all grades but 3, which is the one most affected by the speed variable, the number of missing values in item 30 is almost twice the number of missing values of the preceding and forthcoming item.

Table 3.51. Missing values of items 27, 28, 29 and 31 on the MLAT-EC *Part 1 Paraules ocultes*

ITEMS	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7	ALL GRADES
27	43	21	13	13	9	99
28	41	17	14	12	7	91
29	41	19	15	13	9	97
30	**50**	**38**	**30**	**29**	**19**	**166**
31	46	21	16	18	11	112

The difficulty of this item could be due to the fact that the letter <l> appears at the beginning of the word, and this could have made test takers think that instead of a lowercase letter <l> it was a capital <i> (I), which made it very difficult to come up with the right answer to this item. Actually, this same letter appears at the beginning of the stem of items 1 (lfant) and 5 (lkòpter), items in which it also seems to have caused some confusion. Three test takers wrote in the test "*infant*" below the stem lfant of item 1 and 13 participants of grade 3 (23%) left item 5 blank even if it is at the beginning of the test, which leads us to think that the font type is disturbing the response pattern of these items.

Those items that are very easy (item 3, 5, 10, 11, 12, 13, 15, 16, 17, 18, 20, 22, 26, 27, 28 and 31) have consistently ineffective distractors, while those items that are easy or mid-difficult either have distractors that are more or less popular in the same degree across grades (items 14, 21 and 25), or present some distractor that attracts the test taker for some reason. As happened in the MLAT-ES, item 6 stands out by the amount of subjects (38%) that leave it blank even though it is among the first items. A vowel <o> was added between <l> and <v> in the MLAT-EC from the original qlvra so as to render the item easier, but this change has not proved to have this effect in all grades, as shows its fluctuating IF value.

Choosing the distractor that rhymes with the stem is also the strategy adopted by the subjects in the lower grades when they do not choose the right option. This response pattern can be found in item 2 (kstell – *clatell*), item 8 (tlbizió – *divisió*), item 23 (hincndi – indi) and item 24 (skltar – *estudiar*) which, incidentally, presents quite a high rate of missing values in the upper grades as compared to the items that follow it.

When translating the test, the similarity between the stem and one distractor that there was in item 14 (nmigo – *hormiga*) was lost, as *nmic* is no longer similar to the translation into Catalan "*formiga*"; that is why this item and its distractors have lost effectiveness. This could have been avoided if instead of *nmic*, the stem of this item was *nmiga*. This is, therefore, an improvement to take into account in future versions of the test. A similar phenomenon occurs in item 12, in which the similarity between "*aceitoso*" and "*exitoso*" (the target word) is lost and, as a result, the distractor is not appealing anymore and it also occurs in item 3 (vlena – *animal marí enorme*), as there is no ambiguity possible with "*va plena*" ("*va llena*" in Spanish). Item 11 (korb – *ocell* negre) and 17 (bnic – *maco*) have also become easier than in the MLAT-ES because the linguistic interference in item 11 and the ambiguity caused by rmozo ("*el mozo*") have disappeared. The change in the distractor "*con manchas*", which caused some confusion in item 15 "bakka" has also rendered the item slightly easier than it was in

the MLAT-ES, but at least the question about the appropriateness of "*tiene manchas*" did not arise among the test takers.

Table 3.52. Item 15 on the MLAT-EC *Part 1 Paraules ocultes: p*-values and distractor behaviour analysis

Grade	N	A	B	C*	D	Missing	Attempts	IF – A	IF – B	IF – C*	IF – D
3	57	1	1	33	1	21	36	-0.296	-0.296	0.889	-0.296
4	62	3	2	50	1	6	56	-0.262	-0.286	0.857	-0.310
5	61	0	3	54	1	3	58	-0.333	-0.264	0.908	-0.310
6	60	0	0	55	0	5	55	-0.333	-0.333	1	-0.333
7	64	0	1	60	1	2	62	-0.333	-0.312	0.957	-0.312
all	304	4	7	252	4	37	267	-0.313	-0.298	0.925	-0.313

15. bakka A – és dolça; B – part de la cara; C – fa llet; D – amb arrugues
* correct answer

Other distractors that seem to be appealing as well in the Catalan version, especially in the lower grades, are those in item 4, which was translated literally from the MLAT-ES, and those in item 7, in which *ddo* was changed for *ungl*. Item 4 is still difficult for students in grade 3 though not for 4-graders, as it happened in the MLAT-ES. Let us recall that this item refers to a geometrical concept, whose meaning may be obscure for students in grade 3. While on the MLAT-ES this item was left blank by 17% of learners in grade 3 and 20% of them chose option A, on the MLAT-EC only 5 test takers left it blank, but 35% of them marked A "*angle*" as the correct answer.

Table 3.53. Item 4 on the MLAT-EC *Part 1 Paraules ocultes: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	Double answer	IF – A	IF – B*	IF – C	IF – D
3	57	20	31	0	0	5	52	1	0.186	0.468	-0.327	-0.327
4	62	4	53	1	0	4	58	0	-0.241	0.885	-0.310	-0.333
5	61	7	54	0	0	0	61	0	-0.180	0.847	-0.333	-0.333
6	60	2	55	0	0	3	57	0	-0.287	0.953	-0.333	-0.333
7	64	0	64	0	0	0	64	0	-0.333	1	-0.333	-0.333
all	304	33	257	1	0	12	292	1	-0.182	0.841	-0.328	-0.332

4. crcle A – angle; B – circumferència; C – truita; D – perillós
* correct answer

Item 7 *ddo* in the MLAT-ES (see Table 3.11) was mid-difficult for grade 5 and 6 and difficult for grade 3 and 4. The Catalan alternative *ungl* was nowhere near as difficult as *ddo* although the expected behaviour regarding distractor C (*part d'un triangle* – part of a triangle) does take place, yet not with powerful attraction, as only 6% mark it. The percentage of missing values in grade 3 (22.9%) is very similar to that in the MLAT-ES (24.3%).

Table 3.54. Item 7 on the MLAT-EC *Part 1 Paraules ocultes: p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	57	0	41	2	1	13	44	-0.333	0.909	-0.272	-0.303
4	62	0	57	3	0	2	60	-0.333	0.933	-0.267	-0.333
5	61	0	54	5	1	1	60	-0.333	0.867	-0.222	-0.311
6	60	0	55	4	0	1	59	-0.333	0.910	-0.243	-0.333
7	64	0	61	2	0	1	63	-0.333	0.958	-0.291	-0.333
all	304	0	268	16	2	18	286	-0.333	0.916	-0.259	-0.324

7. ungl A – despertador; B – està a la mà; C – part d'un triangle; D – dur
* correct answer

Lastly, the modification of the distractor C of item 29 (*cirrena - fruita*), which was added so that test takers got confused with "*cirera*" (cherry) and they did not guess that the hidden word was "*sirena*" (siren), had some kind of effect, as this item became slightly more difficult than it was in the MLAT-ES in grade 5, 6, 7 and in all grades considered as a whole, though it remains a very easy item for grades 3 and 4. In this case 14% of those who answered chose distractor C, although it is uncertain whether their reason for choosing it was the one intended by the test adaptor.

Table 3.55. Item 29 on the MLAT-EC *Part 1 Paraules ocultes: p*-values and distractor behaviour analysis

Grade	N	A	B	C	D*	Missing	Attempts	IF – A	IF – B	IF – C	IF – D*
3	57	0	0	3	13	41	16	-0.333	-0.333	-0.083	0.750
4	62	2	2	3	36	19	43	-0.271	-0.271	-0.240	0.783
5	61	0	3	7	36	15	46	-0.333	-0.246	-0.130	0.710
6	60	1	0	8	38	13	47	-0.305	-0.333	-0.106	0.745
7	64	1	3	8	43	9	55	-0.309	-0.261	-0.139	0.709
all	304	4	8	29	166	97	207	-0.308	-0.282	-0.147	0.736

29. cirrena A – camina de pressa; B – estri; C – fruita; D – viu al mar
* correct answer

The D_i of the items on the MLAT-EC are on average good enough (see Table 3.56), although the discrimination power of some of them has diminished because their facility has increased in comparison with the Spanish version.

Table 3.56. Percentage of items on the MLAT-EC *Part 1 Paraules ocultes* according to their index of discrimination

D$_i$	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=304	61.3% Items: 2, 6, 9, 14, 17 - 31	22.6% Items: 4, 8, 10, 11, 12, 15, 16	12.9% Items: 1, 5, 7, 13	3.2% Items: 3

Once again, it is item 3 the one with the lowest D$_i$ and, therefore, the one that should be removed, as it is also a very easy item that does not provide the test with any extra meaningful information (see exact D$_i$ in Table K.1 in appendix K). The graphical representation in a non-standard quintile plot of the behaviour of item 3 is very similar to that of the same item in the Spanish version of the test (see Figure 3.1 in section 3.5.2.1.1).

Figure 3.2. Quintile plot of item 3 on the MLAT-EC *Part 1 Paraules ocultes*
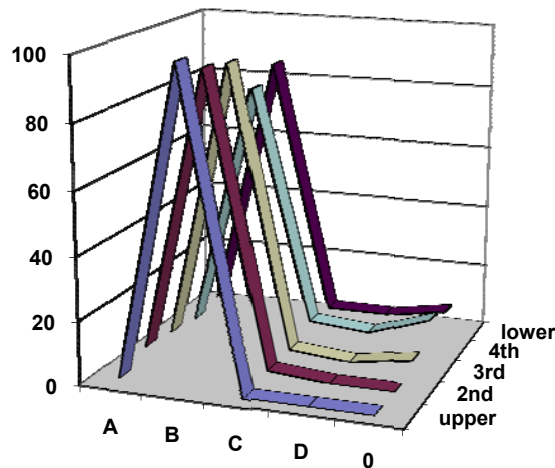
	A	B	C	D	0
upper	1.6	98.4	0	0	0
2nd	4.9	93.4	1.6	0	0
3rd	3.3	91.7	1.7	0	3.3
4th	1.6	80.3	4.9	3.3	10
lower	3.3	85	3.3	3.3	6.7

Some possible solutions to improve the items that do not have a powerful D$_i$ are looking for more appealing distractors that make the subjects actually have to discern

what the differences between the options are. If we always rely on the stems of the MLAT-ES, finding the right distractor is not always as easy so, perhaps, instead of sticking to the words on the MLAT-ES, brand new items should be created.

The point-biserial correlation coefficient (r_{pb}) of this part of the test (see Table 3.57) confirms that neither item 1 nor 3 reach the ideal .300 that is required for an item to be discriminant. However, as item 1 is close to .300 and only one item was removed in the MLAT-ES, it seems sensible to keep it so as to maintain the same number of items in both tests.

Table 3.57. Item-total correlation of MLAT-EC *Part 1 Paraules ocultes*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
1	.289**	6	.357**	11	.546**	16	.576**	21	.729**	26	.744**
2	.570**	7	.430**	12	.582**	17	.661**	22	.758**	27	.750**
3	.221**	8	.385**	13	.561**	18	.611**	23	.669**	28	.773**
4	.456**	9	.369**	14	.560**	19	.525**	24	.595**	29	.657**
5	.457**	10	.579**	15	.553**	20	.728**	25	.667**	30	.571**
										31	.716**

** $p<.001$

What confirms that item 3 is the one to be removed is the calculation of the reliability of this part. Initially, Cronbach's alpha for this part was .931, which is an excellent index. However, the correlation for item 3 is much lower than it would be desirable (.181) (see Table K.2 in appendix K).

3.5.3.1.2. Part 2 Paraules que es corresponen

In *Part 2 Paraules que es corresponen* test takers had to recognise the same four functions as in the MLAT-ES (subject, verb, object and adjective) and match them with the word performing the same function in the sentence below the stem. The number of items aiming at each function was supposed to be kept, but because of a miskey, it changed in the last item, which aimed at recognising the subject instead of the verb in the sentence. Item 22, miskeyed in the Spanish version, was translated literally and grouped in the sentences aiming at recognising the function of subject in order to check if the response pattern this item had in the MLAT-EC was the same as in the MLAT-ES. Thus, eight items aimed at the subject function (see Table K.3 in appendix K). Seven items aimed at the verb function, unlike the MLAT-ES, which had 8

items aiming at this function (see Table K.4 in appendix K). Eight were the sentences on the MLAT-EC *Part 2* in which test takers had to recognise the word performing the function of object in the target sentence (see Table K.5 in appendix K) and there were also eight sentences in which test takers had to match the adjectives in each pair of sentences (see Table K.6 in appendix K).

As in Spanish, Catalan is also a language in which word order is much more flexible than in English. Consequently, as happened in the Spanish test, the choice of the right or wrong option could have been influenced by the arrangement of the words in the sentence. Indeed, a tendency to choose the word right behind the word in capitals in the stem was also observed in this version of the test.

Adjective position and type was a controversial issue in items 2 and 4 in the MLAT-ES. In the Catalan version, while item 2 was translated literally, in item 4 the adjective in the target sentence was changed and, instead of a quantifier, a qualitative adjective was used.

2. El gos PETIT va trencar el gerro de vidre.

La casa vermella té les finestres obertes.

☐ ☐   ☐ ☐☐   ☐   ☐

4. La GRAN mansió del president és blanca.

A la classe de matemàtiques hi ha alumnes intel·ligents.

☐☐ ☐ ☐   ☐   ☐   ☐   ☐

Item 2 obtained a similar response pattern to the one it obtained in the Spanish version, regarding both the number of responses for each option and the missing attempts, as shown in Table 3.58.

Table 3.58. Item 2 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	B	C*	F	G	Missing	Attempts	IF – B	IF – C*	IF – F	IF – G
3	57	2	29	11	13	0	57	-0.047	0.427	0.111	0.146
4	62	2	42	8	10	0	62	-0.022	0.624	0.075	0.108
5	61	3	48	3	7	0	61	0.014	0.751	0.014	0.079
6	60	4	48	6	2	0	60	0.033	0.767	0.067	0
7	64	2	55	6	1	0	64	0.008	0.836	0.070	-0.008
all	304	13	222	34	33	0	304	-0.002	0.685	0.067	0.064

2. B – casa; C – vermella; F – finestres; G – obertes
* correct answer

The response patterns of item 4 in the Spanish version appeared to be influenced by the position of the adjectives, even though the type of adjective that was supposed to match the adjective in the stem sentence was different. In the MLAT-EC version of this item, the sentence was changed so that what prevailed was the type of adjective independent of its position. Therefore, "*GRAN*" was meant to be matched with "*intel·ligents*" (intelligent), whose order of appearance in the sentence (after the noun it modifies) differs from the one of "*GRAN*", which appears in front of the adjective it qualifies. "*intel·ligents*" is the most popular option (160 out of 292 test takers who attempted the item chose it), although "*classe*" (class) and "*matemàtiques*" (mathematics) were also popular, having been chosen by 46 and 54 test takers respectively.

Table 3.59. Item 4 on the MLAT-EC *Part 2 Paraules que es corresponen: p*-values and distractor behaviour analysis

Grade	N	C	E	G	H*	Missing	Attempts	IF – C	IF – E	IF – G	IF – H*
3	57	13	7	8	23	3	54	0.159	0.048	0.066	0.344
4	62	3	15	3	36	3	59	-0.005	0.199	-0.005	0.554
5	61	6	9	3	41	1	60	0.055	0.105	0.005	0.638
6	60	9	16	4	27	2	58	0.079	0.200	-0.007	0.389
7	64	15	7	5	33	3	61	0.180	0.049	0.016	0.475
all	304	46	54	23	160	12	292	0.093	0.120	0.014	0.484

4. C – classe; E – matemàtiques; G – alumnes; H – intel·ligents
* correct answer

From the results obtained in this item, it could be interpreted that the test takers who chose "*classe"* may have chosen it because of its position in the sentence, that those who chose "*matemàtiques*" may have chosen it because they noticed that it complements the noun "*classe*", but that most test takers chose "*intel·ligents*" because it coincides both in form and in category with the word in capitals "*GRAN*" in the stem sentence.

In Table 3.60, it is confirmed that item 2 in the MLAT-EC behaves in the same way as in the MLAT-ES. However, this is not the case of item 4. This item becomes easy in grades 4 and 5, but its IF value oscillates across grades, as in grade 6 it becomes a difficult item and in grade 7 it turns into an item with a mid-difficulty value. This inconsistent IF value makes item 4 a possible candidate item for removal.

Most items are more difficult in grades 3 and 4, and in grades 5 and 6 the percentage of very easy items is smaller in this part as compared to *Parte 2* in the MLAT-ES. These are easy items instead, which means that the difference in the IF values is not a large one (for a complete account of IF values, see appendix L). The

functions that appear most difficult to recognise by all grades are the adjective and the subject, followed by the verb and the object function (see Table K.8 in appendix K).

Table 3.60. Percentage of items on the MLAT-EC *Part 2 Paraules que es corresponen* **according to their facility across grades**

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=57	**0%** Items: -	**3.2%** Items: 1	**9.7 %** Items: 12, 25, 28	**61.3%** Items: 2, 3, 4, 5, 6, 7, 8, 9, 11, 13, 14, 15, 18, 19, 20, 23, 26, 29, 30	**25.8%** Items: 10, 16, 17, 21, 22, 24, 27, 31
Grade 4 N=62	**3.2%** Items: 1	**41.9%** Items: 2, 4, 5, 6, 8, 10, 11, 12, 14, 19, 20, 23, 26	**22.6 %** Items: 7, 9, 13, 18, 25, 27, 29	**22.6%** Items: 3, 15, 17, 22, 28, 30, 31	**9.7%** Items: 16, 21, 24
Grade 5 N=61	**22.6%** Items: 1, 2, 5, 12, 13, 14, 26	**54.8%** Items: 3, 4, 6, 7, 8, 9, 10, 15, 18, 19, 20, 21, 23, 25, 27, 28, 29, 30	**16.1%** Items: 11, 16, 17, 31	**6.5%** Items: 22, 24	**0%** Items: -
Grade 6 N=60	**25.8%** Items: 1, 2, 5, 9, 12, 14, 20, 26	**45.2%** Items: 3, 6, 7, 8, 10, 13, 15, 18, 19, 23, 25, 27, 28, 29	**16.1%** Items: 11, 17, 24, 30, 31	**9.7%** Items: 4, 16, 21	**3.2%** Items: 22
Grade 7 N=64	**51.6%** Items: 1, 2, 3, 5, 7, 8, 9, 12, 13, 17, 18, 19, 20, 26, 28, 29	**35.5%** Items: 6, 10, 11, 14, 15, 21, 23, 25, 27, 30, 31	**12.9%** Items: 4, 16, 22, 24	**0%** Items: -	**0%** Items: -
Grades 3 - 7 N=304	**3.2%** Items: 1	**58.1%** Items: 2, 3, 5, 6, 7, 8, 9, 12, 13, 14, 18, 19, 20, 23, 25, 26, 28, 29	**22.6%** Items: 4, 10, 11, 15, 17, 27, 30	**16.1%** Items: 16, 21, 22, 24, 31	**0%** Items: -

Other items that present an irregular pattern of difficulty in Table 3.60 are item 22, whose subject was confused by the object; item 11, in the stem sentence of which a subordinate clause preceded the main clause (see paragraph above), and item 21, in which there was a clear tendency for those who failed this item to choose the words that appear in the middle of the target sentence. 60 test takers chose "*dormir*" (sleep) and 54 chose "*coixí*" (pillow) instead of the matching word, which is the last word in the target sentence.

21. La mestra sempre té un somriure AMABLE per a tothom.

T'agrada dormir amb un coixí tou?

☐   ☐ ☐ ☐ ☐ ☐

Table 3.61. Item 21 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	A	B	E	F	Missing	Attempts	IF – A	IF – B	IF – E	IF – F*
3	57	3	16	15	15	7	50	-0.057	0.203	0.183	0.183
4	62	6	18	14	17	5	57	-0.012	0.199	0.129	0.181
5	61	5	7	10	36	2	59	0.020	0.054	0.105	0.545
6	60	4	10	11	28	6	54	-0.006	0.105	0.123	0.438
7	64	8	9	4	39	1	63	0.063	0.079	0	0.556
all	304	26	60	54	135	21	283	0.005	0.125	0.104	0.390

21. A – agrada; B – dormir; E – coixí; F – tou
* correct answer

Despite being miskeyed in the Spanish version, item 22 was kept in the MLAT-EC to keep both versions as similar as possible. For the same reasons as item 22 in *Parte 2* on the MLAT-ES, this item cannot be considered in the same way as the other items aiming at words performing the function of subject, as the thematic roles of the subjects in the stem sentence and in the target sentence are different. In the case of the MLAT-EC, item 22 is difficult in all grades but 7, in which it is a mid-difficult item. However, its IF value is not consistent across grades, as it is more difficult for 6-graders than for 5-graders (see Table 3.62). Therefore, besides the semantic reasons mentioned, the irregular IF pattern makes item 22 become a red-flagged item as it was in the MLAT-ES.

22. Què et va semblar el nou COMPANY?

La meva tia sortí i no apagà el televisor.

☐ ☐ ☐ ☐☐ ☐ ☐ ☐

Table 3.62. Item 22 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	C*	D	G	I	Missing	Attempts	IF – C*	IF – D	IF – G	IF – I
3	57	13	10	8	13	7	50	0.154	0.094	0.054	0.183
4	62	21	5	10	19	5	57	0.278	-0.003	0.085	0.286
5	61	28	3	6	21	2	59	0.400	-0.024	0.027	0.303
6	60	17	2	9	25	6	54	0.217	-0.061	0.068	0.283
7	64	32	5	3	20	2	62	0.447	0.012	-0.021	0.312
all	304	111	25	36	98	22	282	0.307	0.002	0.041	0.274

22. C – tia; E – sortí; G – apagà; H – televisió
* correct answer

Except in grade 7, in all grades "*tia*" (aunt) and "*televisor*" (television) have very similar IF values across grades. In addition, all of them are positive, which proves that the behaviour of this item is not the most appropriate one, as happened in the MLAT-ES, and that, therefore, it should be eliminated from the test.

Though not explicitly stated in the directions, only the word in the main clause with the same function as the word in capitals was considered to be the right option. Also, as word order in Catalan is not as strict as in English, the appearance of both the word in capitals in the stem and the target word is an aspect to take into account when analysing the items' behaviour. Table K.7 in appendix K shows which items follow the standard SVO word order (though it is not so common an order in Spanish and Catalan as it is in English) and those which present a particular ordering of items either in the stem or in the target sentence.

The main difference in word order as compared to the Spanish version is in item 9. While in Spanish the stem sentence had the indirect object fronted (*A Juan le comprarán un regalo el lunes*) and this made this indirect object sound emphatic, the indirect object in the Catalan version appeared in its most common position, i.e. after the verb, although an adverb phrase (*El cap de setmana vinent* – Next weekend*)* was placed before the verb (*El cap de setmana vinent, la Maria i en Josep jugaran a futbol*). However, this kind of complement before the verb is not emphatic. Hence, item 9 has been considered to maintain the "canonical" word order in Catalan.

In item 15, in which the subject appears at the end of the stem sentence while in the target sentence it is at the beginning, 50 out of 292 test takers who attempted the item chose the last word in the target sentence over other words that functioned as distractors. Nevertheless, a high percentage (59.9%) of those who answered this item scored on it.

15. Digue'm una cosa, encara està malalta la teva GERMANA?

Les formigues treballen molt transportant menjar.

☐   ☐   ☐   ☐   ☐   ☐

In item 23 in the MLAT-ES, targeting the function of subject, the target sentence had two subjects, although only one of them, the one in the main sentence, was considered to be correct. This inconsistency was avoided in the Catalan version by making the verb in the subordinate clause a noun in such a way that what was the subject in the Spanish version became the complement of this resulting noun. Thus, if in "*cuando se acerca el invierno*" (when winter is near), "*invierno"* is the subject of "*se acerca",* in Catalan it appears as the complement of the head of the prepositional phrase "*Al començament de l'hivern",* which is "At the beginning of winter" if translated literally. With this change in the syntactic categories of the words in the subordinate clause, the choice among the distractors was more widespread than it was in the

Spanish version. While in the MLAT-ES 42 test takers chose "*invierno*" (the subject "winter") as the matching word for "*Susana*", the subject in the stem, and the other distractors were chosen by a maximum of 24 test takers, in the Catalan version "*hivern*" (winter) was chosen by only 26 test takers.

Table 3.63. Item 23 on the MLAT-ES *Parte 2 Palabras que se corresponden*: ***p*-values and distractor behaviour analysis**

Grade	N	B	D	F	G	H	I	Misng.	Attmps.	IF – B	IF – D	IF – F	IF – G	IF – H	IF – I
3	66	7	9	22	5	7	6	9	57	0.046	0.081	0.309	0.011	0.046	0.029
4	75	1	10	39	4	9	3	8	67	-0.037	0.097	0.530	0.007	0.082	-0.007
5	57	2	6	39	1	3	0	0	57	-0.004	0.066	0.645	-0.022	0.013	-0.039
6	60	0	8	41	3	3	2	1	59	-0.038	0.097	0.657	0.013	0.013	-0.004
7	67	1	9	48	2	2	0	2	65	-0.017	0.106	0.706	-0.002	-	-0.033
all	325	11	42	189	15	24	11	20	305	-0.011	0.091	0.572	0.002	0.031	-0.011

23. B – acerca; D – invierno; F – pájaros; G empiezan; H – volar; I – sur
* correct answer

Table 3.64. Item 23 on the MLAT-EC *Part 2 Paraules que es corresponen*: ***p*-values and distractor behaviour analysis**

Grade	N	B	D	F	G	H	I	Misng.	Attmps.	IF – B	IF – D	IF – F	IF – G	IF – H	IF – I
3	57	4	5	21	1	3	6	7	50	0.008	0.028	0.054	0.183	-0.053	0.048
4	62	4	2	36	3	9	4	4	58	0.022	-0.013	0.085	0.286	0.004	0.022
5	61	5	4	37	3	6	4	1	60	0.035	0.019	0.027	0.303	0.002	0.019
6	60	3	4	35	4	4	1	8	52	0.017	0.036	0.069	0.283	0.036	-0.022
7	64	3	11	43	3	0	0	4	60	0.015	0.148	-0.021	0.312	0.015	-0.035
all	304	19	26	172	14	22	15	24	280	0.020	0.042	0.041	0.274	0.002	0.005

23. B – començament; D – hivern; F – ocells; G comencen; H – volar; I – sud
* correct answer

Similarly to what happened with this item in the MLAT-ES, item 24 was failed by a high number of those who attempted it. In this item, the subject "*tu*" is mentioned explicitly in the target sentence while, in a neutral statement, it would be omitted.

24. L'ENRIC posà una campaneta a la porta d'entrada de casa seva.

A quina hora creus tu que arribaràs a sopar?

☐ ☐ ☐ ☐ ☐ ☐   ☐   ☐ ☐

158 test takers out of 268 who attempted it chose either the word "*hora*" which appears at the beginning of the sentence, as does the word in capital letters in the stem sentence, or the verbs in the second part of the sentence. 60 test takers chose the finite form "*arribaràs*" ("will arrive") and 43 chose the non-finite infinitive form "*sopar*" ("have supper"). The high number of test takers who chose options other than

the right answer makes these distractors have a positive IF value, as shown in Table 3.65.

Table 3.65. Item 24 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	C	D	E	G	I	Missing	Attempts	IF – C	IF – D*	IF – E	IF – G	IF – I
3	57	8	5	5	18	11	10	47	-0.069	-0.005	-0.005	0.271	0.122
4	62	10	2	17	18	12	6	56	0.020	-0.051	0.217	0.234	0.127
5	61	9	5	27	7	11	4	57	0.022	0.022	0.408	0.057	0.127
6	60	4	0	28	9	6	11	49	0.048	-0.054	0.518	0.130	0.069
7	64	2	3	33	8	3	5	59	0.148	-0.004	0.504	0.081	-0.004
all	304	33	15	110	60	43	36	268	0.038	-0.018	0.337	0.150	0.087

24. C – hora; D – creus; E – tu; G – arribaràs; I – sopar
* correct answer

Three items aiming at the verb function presented a word order different from the conventional SVO. These are items 3, 7 and 17. In item 3 in the MLAT-ES, 65 out of 205 of the test takers who attempted the item chose the penultimate word in the target sentence before any other option in the target sentence. A similar pattern is found in the test takers' performance in the Catalan version of this item, as the most popular distractor is "*cotxe*" (car), chosen by 65 of those test takers who answered this item. This item was difficult for both grades 3 and 4.

3. L'Olga CAMINÀ tres-cents metres ahir.

Al meu cosí Francesc li regalaran un cotxe nou.

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Table 3.66. Item 3 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	C	D	F*	H	I	Missing	Attempts	IF – C	IF – D	IF – F*	IF – H	IF – I
3	57	1	6	22	23	9	1	56	-0.058	0.032	0.317	0.209	0.084
4	62	2	1	26	16	4	0	62	-0.040	-0.057	0.347	0.347	0.024
5	61	0	4	41	9	1	0	61	-0.041	0.025	0.631	0.189	-0.025
6	60	4	1	43	8	2	1	59	0.034	-0.017	0.695	0.102	0.017
7	64	2	3	52	9	1	0	64	0.008	0.023	0.789	0.055	-0.008
all	304	9	15	184	65	17	2	302	-0.019	0.001	0.560	0.180	0.017

3. C – cosí; D – Francesc; F – regalaran; H – cotxe; I – nou
* correct answer

Item 7, which is aimed at identifying the verb "*bufen*" (blow), presents the subject "*vents*" (winds) postponed. In the Spanish version of this item, content-word distractors, i.e. the noun "*verano*" (summer), the noun "*vientos*" (winds) and "*calientes*"

(hot), were popular in approximately the same degree, as they were chosen by 38, 37 and 22 test takers respectively, and the item was easy across grades. In the Spanish version, what seemed to exert more influence in the election of distractors was their position in the target sentence. In contrast, in the Catalan version, only the distractor "vents" (winds), which is the subject of the target sentence, is chosen by more than half of those who fail this item, while the rest of distractors are not so popular.

<div align="center">

7. En Pere POSA el despertador cada nit.

A l'estiu bufen vents calents.

□□□   □   □   □

</div>

Table 3.67. Item 7 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	C	D*	E	F	Missing	Attempts	IF – C	IF – D*	IF – E	IF – F
3	57	6	26	17	4	2	55	0.004	0.367	0.203	-0.033
4	62	3	34	14	7	4	58	-0.031	0.503	0.159	0.038
5	61	3	42	11	5	0	61	-0.013	0.626	0.118	0.020
6	60	2	41	13	4	0	60	-0.030	0.620	0.153	0.003
7	64	3	56	1	2	1	63	0.025	0.867	-0.006	0.010
all	304	17	199	56	22	7	297	-0.009	0.604	0.123	0.008

7. C – estiu; D – bufen; E – vents; F – calents
* correct answer

In item 17 the main clause of the stem sentence is preceded by a subordinate clause. The verb "*PORTO*" (carry), which is the word to be matched with the target word "cantà" (sang) appears in the middle of the sentence.

<div align="center">

17. Quan plou, sempre PORTO botes de goma.

La mare li cantà una cançó de bressol al seu bebè.

□ □□□ □ □   □   □ □

</div>

In this item, the distractor "*cançó*" (song) is chosen by more than half of those who failed this item, as happened in the MLAT-ES, and the IF value of this item becomes easy only from grade 6 on (see Table 3.68). Therefore, once more in this item what seems to be more attractive to the test takers who do not score on this item is the position of the distractor "*cançó*".

Table 3.68. Item 17 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	B	E*	F	G	I	Missing	Attempts	IF – B	IF – E*	IF – F	IF – G	IF – I
3	57	4	16	14	12	3	6	51	-0.007	0.228	0.189	0.149	-0.027
4	62	2	27	16	8	3	5	57	-0.031	0.408	0.215	0.075	-0.013
5	61	3	33	17	4	3	0	61	-0.008	0.484	0.221	0.008	-0.008
6	60	0	42	14	2	2	0	60	-0.038	0.663	0.196	-0.004	-0.004
7	64	2	49	6	3	0	3	61	0.008	0.779	0.074	0.025	-0.025
all	304	11	167	67	29	11	14	290	-0.015	0.523	0.178	0.047	-0.015

17. B – mare; E – cantà; F – cançó; G – bressol; I – bebè
* correct answer

Item 11 presents the subordinate clause preceding the main clause. "*BROSSA*" (rubbish), the word in capitals to be matched with its corresponding word in the target sentence, appears in the subordinate clause, towards the beginning of the sentence, while the matching word "*nina*" (doll) appears towards the end of the target sentence, which does not have any subordinate clause.

11. Per llençar BROSSA, van castigar en Joan.

La Susanna va agafar la seva nina de terra.

☐   ☐   ☐ ☐ ☐☐

In the MLAT-EC this item is only difficult for 3-graders and it oscillates between the mid-difficult and easy values in the rest of grades, while in the MLAT-ES it was difficult for test takers from grade 3 to 5.

Table 3.69. Item 11 on the MLAT-EC *Part 2 Paraules que es corresponen*: *p*-values and distractor behaviour analysis

Grade	N	A	B	D*	E	F	Missing	Attempts	IF – A	IF – B	IF – D*	IF – E	IF – F
3	57	6	14	22	4	8	2	55	-.0109	0.135	0.280	-0.048	0.025
4	62	4	11	38	2	4	2	60	-.007	0.110	0.560	-0.040	-0.007
5	61	4	11	37	0	8	1	60	-0.010	0.101	0.540	-0.077	0.057
6	60	1	12	36	3	5	0	60	-0.063	0.120	0.520	-0.030	0.003
7	64	11	6	40	2	4	0	64	0.097	0.019	0.550	-0.044	-0.013
all	304	26	54	173	11	29	5	299	0.002	0.096	0.494	-0.047	0.013

11. A – Susanna; B – agafar; D – nina; F – terra
* correct answer

In both the MLAT-ES and the MLAT-EC, a relationship between the order of appearance of the word in capitals and the distractors in the target sentence of this item can be observed, as the most popular distractor in the MLAT-EC is also the second word with content in the target sentence, which was chosen by 18% of those

who attempted the item. However, the distractor "*agafar*" (pick up) was not so popular as "*recogió*" in the MLAT-ES. This could be due to the fact that the meaning of "*recoger*" tends to be related to "picking up something which is not in its place", while the Catalan word "*agafar*" is more neutral, i.e. the thing that is "*agafada*" (picked up) does not necessarily have to be out of its place. Actually, the closest translation of "*recogió*" into Catalan is not "*va agafar*" but "*va recollir*", which does have the connotations that "*recogió*" has*.*

The D_i values of item 1 and 4 are rather low (0.22 and 0.19 respectively). Item 1 was also the one with the lowest D_i value in the MLAT-ES, but its D_i value is still acceptable and, though in need to be improved, it can be kept. Item 4, however, has too low a D_i value and, consequently, should be removed from the final version of the MLAT- EC (see in Table K.9 in appendix K all D_i and correct answers in the upper and lower 33% tails).

Table 3.70. Percentage of items on the MLAT-EC *Part 2 Paraules que es corresponen* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=304	87.1% Items: 2, 3, 5, 7-10, 12-31	6.5% Items: 6, 11	3.2% Items: 1	3.2% Items: 4

The point-biserial correlation between the items on the MLAT-EC and the total raw score confirms that item 4 should be removed from the test, as its correlation does not reach the minimum index .300. This value is also low for item 1, but, surprisingly, it is not for item 22, which is the only item that was removed from the MLAT-ES due to the miskey caused by the confusion of functions in the stem sentence.

Table 3.71. Item-total correlation of MLAT-EC *Part 2 Paraules que es corresponen*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
1	.263**	6	.314**	11	.307**	16	.497**	21	.530**	26	.564**
2	.505**	7	.602**	12	.644**	17	.577**	22	.505**	27	.664**
3	.668**	8	.656**	13	.574**	18	.564**	23	.562**	28	.713**
4	.189**	9	.592**	14	.464**	19	.467**	24	.653**	29	.747**
5	.608**	10	.439**	15	.584**	20	.608**	25	.443**	30	.697**
										31	.628**

*** $p<.001$*

All items but items 1, 4, 6 and 11 have a corrected item-total correlation value lower than the minimum desirable, which is .300 (see Table 3.71 and also Table K.10 in appendix K). While items 1 and 4 coincide with items that other indexes had pointed to as defective, neither item 6 nor 11 had appeared to be items in need of improvement or removal before. For the sake of comparability with the MLAT-ES, item 22 had to be removed from the MLAT-EC. Ideally, both tests should have the same number of items. However, since all indexes of validity redflag item 4 as an item that should be eliminated, it was decided to remove it from the test too. Items 6 and 11 were kept, though, as this one was the only index that did not reach the minimum value for them to be kept. Therefore, after removing item 4 and 22, the final version of the MLAT-EC has one fewer item than the MLAT-ES and its reliability index is excellent.

3.5.3.1.3. Part 3 Paraules que rimen

This part is a multiple-choice test in which test takers have to choose the word that rhymes with the stem out of 4 possible words. While the test had 46 items in its original version, only the items whose distractors caused some salient response patterns will be discussed below, although the response patterns of the whole test can be found in appendix M.

In the Table 3.72, the distribution of items according to their IF index shows that items 4, 9, 23, 26, 37, 38 and 41 were difficult for the lower grades, especially as they get easier from grade 5 on. In contrast, items 12, 22, and 28 are consistently difficult across grades and items 16, 29 and 38 present an oscillating IF pattern.

Table 3.72. Percentage of items on the MLAT-EC *Part 3 Paraules que rimen* according to their facility across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=57	**17.4%** Items: 2, 3, 7, 15, 21, 27, 33, 36	**45.7%** Items: 1, 5, 8, 10, 11, 13, 14, 17, 18, 19, 20, 24, 25, 31, 32, 35, 39, 40, 42, 43, 45	**6.5%** Items: 30, 44, 46	**15.2%** Items: 6, 16, 26, 34, 37, 38, 41	**15.2%** Items: 4, 9, 12, 22, 23, 28, 29
Grade 4 N=62	**51.6%** Items: 1, 2, 3, 7, 8, 10, 13, 14, 15, 17, 19, 20, 21, 24, 25, 27, 30, 31, 32, 35, 36, 39, 40, 42	**23.9%** Items: 5, 11, 18, 23, 26, 33, 34, 37, 41, 43, 46	**6.5 %** Items: 6, 38, 44	**15.2%** Items: 4, 9, 12, 16, 22, 29, 45	**2.2%** Items: 28
Grade 5 N=61	**67.3%** Items: 1, 2, 3, 5, 6, 7, 8, 10, 11, 13, 14, 15, 17, 19, 20, 21, 24, 25, 26, 27, 30, 31, 32, 34, 35, 36, 37, 39, 40, 42, 43	**17.4%** Items: 9, 18, 23, 33, 41, 44, 45, 46	**8.7%** Items:12, 16, 29, 38	**2.2%** Items: 4	**4.4%** Items: 22, 28
Grade 6 N=60	**67.3%** Items: 1, 2, 3, 5, 6, 7, 8, 10, 11, 14, 15, 17, 19, 20, 21, 24, 25, 26, 27, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 42, 43, 46	**21.7%** Items: 4, 9, 13, 16, 18, 23, 33, 38, 41, 45	**2.2%** Items: 44	**4.4%** Items: 12, 29	**4.4%** Items: 22, 28
Grade 7 N=64	**69.6%** Items: 1, 2, 3, 5, 7, 8, 10, 11, 13, 14, 15, 17, 19, 20, 21, 23, 24, 25, 26, 27, 30, 31, 32, 35, 36, 37, 39, 40, 41, 42, 43, 46	**13%** Items: 4, 6, 18, 33, 34, 38	**6.5%** Items: 9, 44, 45	**8.7%** Items: 12, 16, 22, 29	**2.2%** Items: 28
Grades 3-7 N=304	**60.8%** Items: 1, 2, 3, 5, 7, 8, 10, 11, 13, 14, 15, 17, 19, 20, 21, 24, 25, 27, 30, 31, 32, 35, 36, 37, 39, 40, 42, 43	**19.6%** Items: 6, 18, 23, 26, 33, 34, 38, 41, 45	**6.5%** Items: 4, 44, 46	**10.9%** Items: 9, 12, 16, 22, 29	**2.2%** Items: 28

A qualitative analysis of the items in this part makes it worthwhile to comment on how test takers react in front of the items containing the different phoneme-grapheme correspondence issues explained in the section describing this part (section 3.4.).

The items whose pronunciation one does not need to know in order to choose its right answer, as the sound-grapheme correspondence of the right answer is exact, were consistently easy across grades. For instance, item 24 *PROMESA – permesa* (promise – allowed), one of the items dealing with the different pronunciation of <e>, does not show to be conflictive. This is perhaps because the combination of consonants <ss, s, ç> in the distractors made the consonant pronunciation be more salient than the vowel to test takers, who may also have relied on the word spelling.

Neither are the items dealing with the sound-grapheme combination of <o> (items 18 and 45) conflictive, except for item 29 *ESTORA – rentadora,* which is never an easy item. In this item, as well as in items 4 and 9, which are very difficult for the lower grades and behave similarly, the test takers were expected to choose the distractor that rhymes consonantically with the stem despite there being distractors that also rhyme with it, but only assonantically. Consistently across grades, in item 29 almost 30% of those who reached this item chose the distractor "perola", that did not even rhyme in an assonant way with the stem, as the <o> in "*perola*" is pronounced /ɔ/ while the second <o> in "*cotorra*" is pronounced /o/.

Table 3.73. Item 29 on the MLAT-EC *Part 3 Paraules que rimen*: *p*-values and distractor behaviour analysis

Grade	N	A	B	C	D*	Missing	Attempts	IF – A	IF – B	IF – C	IF – D*
3	57	14	10	2	11	20	37	0.171	0.027	-0.261	0.063
4	62	18	13	0	26	5	57	0.088	-0.029	-0.333	0.275
5	61	18	9	0	30	4	57	0.088	-0.123	-0.333	0.368
6	60	20	9	2	26	3	57	0.135	-0.123	-0.287	0.275
7	64	10	13	2	35	4	60	-0.111	-0.044	-0.288	0.444
all	304	80	54	6	128	36	268	0.065	-0.065	-0.303	0.303

29. ESTORA A – perola; B – cotorra; C – tara; D – rentadora
* correct answer

Item 18 *FORMÓS* is to be matched with *famós* (famous)*.* In this item only one option besides the right answer seemed to appeal the test takers. This is the word "*termos*" (flasks), which has the same consonants as the stem word, but it also has a graphical stress (see Table 3.74)*.* This distractor appealed to only 16.6% of the test takers. The same response pattern can be found in item 45 *FLABIOL* (see Table 3.75)*,* to be matched with *Oriol,* although, if paying attention only to the consonant series, it could be matched to *pèsol,* which does not rhyme with the stem. This option was chosen by 25% of those who reached this item.

Table 3.74. Item 18 on the MLAT-EC *Part 3 Paraules que rimen*: *p*-values and distractor behaviour analysis

Grade	N	A*	B	C	D	Missing	Attempts	IF – A*	IF – B	IF – C	IF – D
3	57	43	5	2	2	5	52	0.769	-0.205	-0.282	-0.282
4	62	46	3	11	1	1	61	0.672	-0.268	-0.093	-0.311
5	61	45	2	13	0	1	60	0.667	-0.289	-0.044	-0.333
6	60	46	2	9	1	2	58	0.724	-0.287	-0.126	-0.310
7	64	48	2	14	0	0	64	0.667	-0.292	-0.042	-0.333
all	304	228	14	49	4	9	295	0.697	-0.270	-0.112	-0.315

18. FORMÓS A – famós; B – carros; C – termos; D – formes
* correct answer

Table 3.75. Item 45 on the MLAT-EC *Part 3 Paraules que rimen*: *p*-values and distractor behaviour analysis

Grade	N	A	B*	C	D	Missing	Attempts	IF – A	IF – B*	IF – C	IF – D
3	57	4	13	0	1	39	18	-0.037	0.630	-0.333	-0.259
4	62	14	27	5	1	15	47	0.064	0.433	-0.192	-0.305
5	61	14	30	1	0	16	45	0.081	0.556	-0.304	-0.333
6	60	5	35	3	2	15	45	-0.185	0.703	-0.244	-0.274
7	64	15	33	2	2	12	52	0.051	0.513	-0.282	-0.282
all	304	52	138	11	6	97	207	0.002	0.556	-0.263	-0.295

45. FLABIOL A – pèsol; B – Oriol; C – flascó; D – avió
* correct answer

The fricative alveolar phonemes also posed some kind of difficulty for the test takers. For instance, more than 41.7% of those who attempted item 16 *CAMISA – llisa – Patrícia – camèlia – cloïssa* failed it. The most powerful distractor was the last one, which shares the same ending as the stem as well as the number of syllables. Item 44 *PALLASSOS – eficaços – carbassons – pastissos – pressupost* also resulted in a mid-difficult item, as 47 test takers felt that "*pastissos*" was the right answer, probably because of the exact correspondence with the first syllable of the stem.

Actually, similar consonant-vowel combinations make the test takers in the lower grades, especially grade 3, choose words that sound similarly, but that do not necessarily rhyme with the stem. That is what happened in items 12, 23 and 26. These items have distractors that share entire syllables but whose endings are completely different from the stem. Also, item 5, the only item that deals with the distinction between /dʒ/ and /tʃ/, proved to be a good discriminator, but with an unexpected response pattern in this sense. It was expected that test takers would feel that "*metge*" rhymes with "*fletxa*" because of the tendency to devoice voiced fricatives. However, those who fail the item choose "*fitxa*" instead, which has the same consonants as the item stem *FLETXA.*

Item 7 *REBOST – carbó – esport - compost - vols* and item 22 were thought to follow the same rationale: using combinations of almost the same consonants and vowels in most distractors so that test takers had to choose the option that rhymed consonantically with the stem. While in item 7 the distractors were ineffective as a whole, in item 22 the similarity among them was such that the distractor that finished with a different letter was hardly chosen. That is the reason why all the distractors but D have an IF very close to 0.

Table 3.76. Item 22 on the MLAT-EC *Part 3 Paraules que rimen*: *p*-values and distractor behaviour analysis

Grade	N	A*	B	C	D	Missing	Attempts	IF – A*	IF – B	IF – C	IF – D
3	57	20	11	15	1	10	47	0.234	-0.021	0.092	-0.305
4	62	28	18	10	4	2	60	0.289	0.067	-0.111	-0.244
5	61	25	20	14	1	1	60	0.222	0.111	-0.022	-0.311
6	60	25	23	8	2	2	58	0.241	0.195	-0.149	-0.287
7	64	34	16	14	0	0	64	0.375	0	-0.042	-0.333
all	304	132	88	61	8	15	289	0.276	0.073	-0.051	-0.296

22. AUTOBÚS A – abús; B – obtús; C – barnús; D – bust
* correct answer

While item 22 also results in a very good discriminator (see Table 3.76), item 28 "*BOMBER*" (fireman) turned out to be a deficient item. In this item two distractors start with the same syllable as the stem ("*bomba*" – bomb, and "*bombeta*" – lightbulb) and the other two options are the ones that rhyme, one in an assonant way ("*cambrer*" - waiter), and the other in a consonant way ("*barber*" – barber). This item became deficient because the distractor "*cambrer*" was chosen over the right answer and, consequently, had to be removed with no hesitation.

Table 3.77. Item 28 on the MLAT-EC *Part 3 Paraules que rimen*: *p*-values and distractor behaviour analysis

Grade	N	A	B	C	D*	Missing	Attempts	IF – A	IF – B	IF – C	IF – D*
3	57	9	20	1	8	19	38	-0.018	0.368	-0.298	-0.053
4	62	3	34	7	15	3	59	-0.266	0.435	-0.175	0.006
5	61	1	40	0	18	2	59	-0.311	0.570	-0.333	0.073
6	60	3	40	1	14	2	58	-0.264	0.586	-0.310	-0.011
7	64	2	37	0	23	2	62	-0.290	0.462	-0.333	0.161
all	304	18	171	9	78	28	276	-0.246	0.493	-0.290	0.043

28. BOMBER A – bomba; B – cambrer; C – bombeta; D – bomber
* correct answer

Part 3 Paraules que rimen includes items with pairs of phonemes that do not exist in Spanish (/s/ and /z/, /e/ and /ɛ/, and /o/ and /ɔ/). Consequently, one possible different item functioning could be caused by the test takers' language preference, as those who prefer Spanish over Catalan could have some more difficulty in answering those items containing phonemes missing in Spanish. Items 16 and 24 dealt with the distinctions between sibilants /s/ and /z/. In Figure 3.3 it can be seen that the percentage[15] of those who answered the items properly are very similar regardless of

---

[15] The numbers on the columns stand for the number of test takers who answered the items right, wrong or left them in blank, while the coloured bits in the columns represent the percentages of each type of answer with respect to the number of subjects in each language preference group. Below each group of columns appears the stem of the item and the distractors that presented contrast with the phoneme.

their language preference. However, those who prefer Spanish over Catalan and got the item wrong (without taking into account those who did not attempt the item) double the number of those who prefer Catalan who did not score on this item either. In both items, those who declared no preference for one language over the other got very similar scores to those of the rest of groups.

**Figure 3.3. Performance on items with sibilant phonemes according to test takers' language preference**
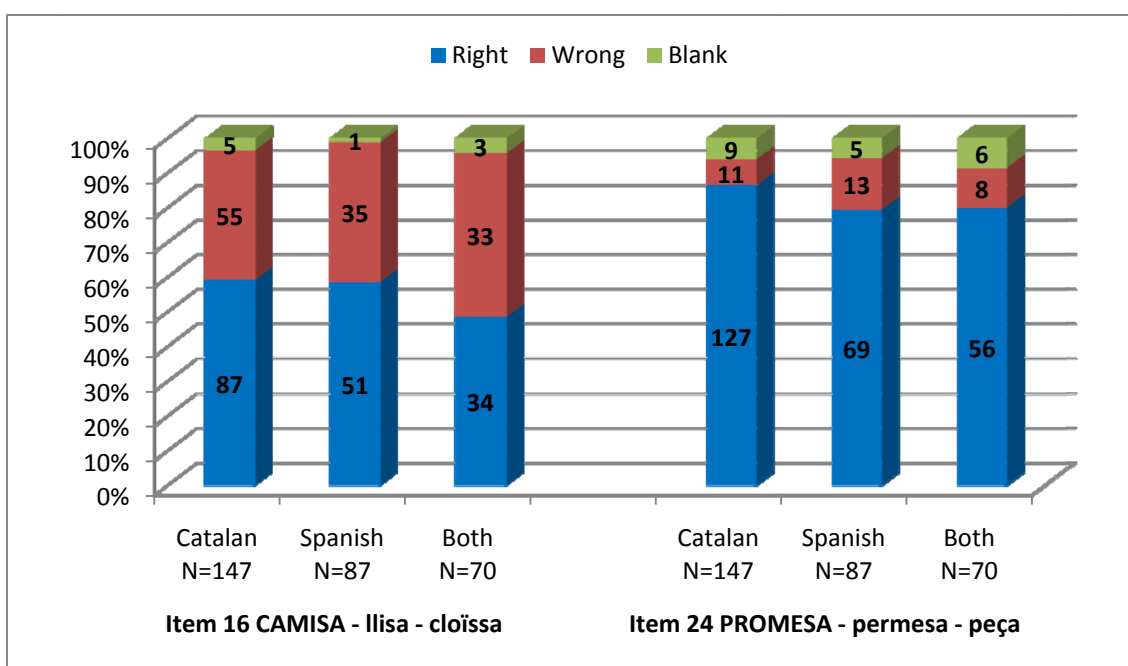
The pair of vocalic phonemes /e/ and /ɛ/ does not exist in Spanish. Therefore, this could also pose some kind of difficulty for those who prefer Spanish over Catalan. Figure 3.4 presents how the test takers answered the items containing these phonemes as main targets. This figure shows that language preference does not seem to affect much the answer patterns in the items where /e/ and /ɛ/ were contrasted, although those who do not show any preference for one language over the other and those who prefer Spanish over Catalan consistently get lower scores on these items, including item 24, which appears in the figure above.

Figure 3.4. Performance on items with phonemes /e/ and /ɛ/ contrasts according to test takers' language preference
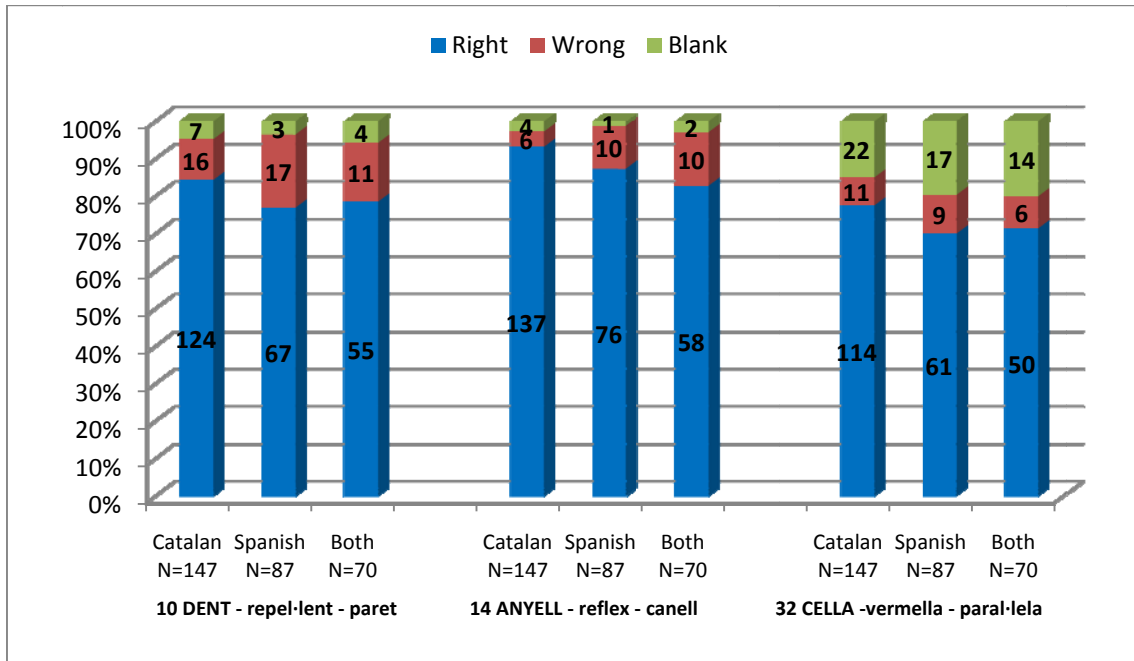
Phonemes /o/ - /ɔ/ are the last pair of phonemes contrasted in the test that exist in Catalan but not in Spanish. These do show a different item functioning depending on the language preference. While in item 25 the performance is similar across language preference groups, it is not in item 29, in which those who prefer Catalan performed much better than the other two groups, nor in item 45, in which those who declare to prefer one language over the other performed better than those who did not show any preference for any of the two languages.

Figure 3.5. Performance on items with phonemes /o/ and /ɔ/ contrasts according to test takers' language preference
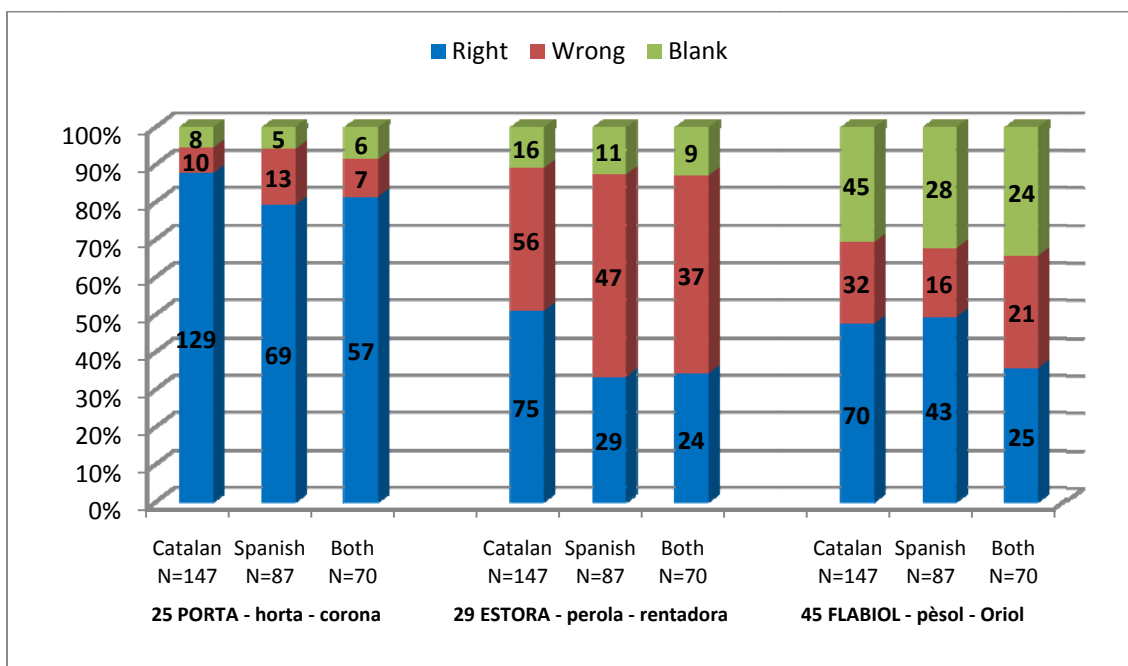
The amount of difficult and very difficult items makes this part more difficult than the Spanish version (see Table 3.30). Consequently, if both tests are to be rendered similar, difficult items should be targeted as items liable to be excised from this part. Therefore, although item 28 seems to have an average D_i which could be a reason to keep it, it is an item that should be removed, not only because of the comparability issue between tests, but also because it presents too different a behaviour from the rest of items in terms of facility.

The analysis of discrimination points at the items whose distractors presented unexpected response patterns or were not attractive enough to test takers (see Table K.11 in appendix K all D_i and the number of correct answers in the upper and lower 33% tails). These are the items which are deficient or average according to their D_i, as shown in Table 3.78.

Table 3.78. Percentage of items on the MLAT-EC *Part 3 Paraules que rimen* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=304	47.9% Items: 9, 22, 23, 26, 29-46	21.7% Items: 4, 5, 6, 10, 12, 20, 21, 24, 25, 27	17.4% Items: 1, 3, 8, 11, 14, 17, 19, 28	13% Items: 2, 7, 13, 15, 16, 18

As expected, the items in the first part of the test do not correlate extremely highly with the raw total score due to the speededness of the test. This statistical analysis helps us identify which items should be removed for not reaching the minimum expected in their correlation index. These are items 4, 13, 15, 16, 18, 22 and 28.

Table 3.79. Item-total correlation of MLAT-EC *Part 3 Paraules que rimen*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
1	.446**	9	.432**	17	.444**	25	.541**	33	.636**	41	.694**
2	.428**	10	.406**	18	.160**	26	.570**	34	.727**	42	.745**
3	.442**	11	.362**	19	.459**	27	.628**	35	.730**	43	.761**
4	.268**	12	.304**	20	.541**	28	.192**	36	.768**	44	.535**
5	.428**	13	.261**	21	.513**	29	.415**	37	.747**	45	.538**
6	.376**	14	.481**	22	.290**	30	.754**	38	.595**	46	.685**
7	.397**	15	.282**	23	.637**	31	.706**	39	.765**		
8	.436**	16	.074	24	.483**	32	.711**	40	.770**		

** $p<.001$

These items along with item 12 coincide with the eight items that stood out for their low corrected total-item correlation index, in spite of the fact that the initial Cronbach's Alpha index was very high (.938) (see Table K.12 in appendix K).

Considering the weaknesses of the items commented on as a whole, it was finally decided to remove items 2, 7, 13 and 15 because of their lack of discrimination power, as their distractors did not act as such; item 16 because of the test takers' choice when facing combinations of the same vowel and the fricative alveolar consonants; item 22 because, despite being a good discriminator, its functioning was very odd when compared to the rest of items in this part; item 18 because of its lack of discrimination power; and item 28 because of the confusion created by one of its distractors. Moreover, it was considered convenient to remove only eight items so that

this part could have 38 items, that is, the same number of items as the Spanish version of the test. After removing them, the Cronbach's alpha index was .945 for all grades, which is an excellent reliability index.

3.5.3.1.4. Part 4 Aprenguem números

The numbers targeted in this part are the same as the ones on the MLAT-ES *Parte 4 Aprendamos números.* However, the name of the items had to be different, since letter <ñ> is not included in the Catalan alphabet and so had to be skipped.

Table 3.80. Numbers aimed at, number names and order of appearance on the MLAT-EC *Part 4 Aprenguem números*

Number	Number in letters	Item (letter)	Number	Number in letters	Item (letter)
1	co	j	20	vinca	l, x
2	vein	b, y	21	vinca-co	e, q
3	ras	f	22	vinca-vein	i, s
10	silca	n	23	vinca-ras	a, v
11	silca-co	c, w	30	rasca	h
12	silca-vein	r	31	rasca-co	k
13	silca-ras	g, o, t	32	rasca-vein	m, u
			33	rasca-ras	d, p

In this test, numbers have to be built up from the combination of four numbers: 0, 1, 2, and 3. Other numbers besides the ones resulting from the combination of 0, 1, 2 and 3 were written as answers too. These are 14, 15, 18, 24, 25, 28, 34, 38, 4, 40, 42, 43, 5, 50, 51 and 64.

The figures presented in the instructions appeared combined in numbers of two figures maximum (i.e. tens and units). Some test takers wrote other combinations with three (101, 102, 120, 121, 130, 201, 202, 203, 212, 230, 301, 302, 303, 310, 320, 330) or four figures (1020, 1030, 2010, 2020). While in the MLAT-ES the two-figure numbers were built from the combination of the first figure plus 10, 20 and 30 (except number 222), in this test, other combinations are found. These are a ten, a twenty or a thirty in the first place and then number 1, 2 or 3, as in 101, 102, 201, 202, 203, 301, 302 and 303. Four-figure numbers consist of two tens together. For instance, if test takers had to write 12 (*silca-vein*), they wrote 1020 (*silca-vinca*), or if they had to write 13 (*silca-ras*), they wrote 1030 (*silca-rasca*) instead.

Some test takers also wrote the wrong number when they confused the unit. For instance, 4.6% wrote 22 (*vinca-vein*) when they should have written (23) and 3.6% wrote 22 (*vinca-vein*) when they were required to write 21 (*vinca-co*). Some others

wrote a two-figure number that has the same unit as the targeted number, so the mistake is in the tens place. This mistake was made more often in the MLAT-ES. While in the MLAT-ES the mistakes could be paired, this is not the case of the MLAT-EC (see Table K.13 in appendix K the most recurrent wrong number combinations and the mistake percentages). As happened in the MLAT-ES, number 30 did not cause much confusion among the test takers, probably due to the fact that the word "*rasca*", the name for number 30 in the test, has a meaning.

While the mistakes above were made by a small percentage of the test takers, the confusion between the unit and the corresponding ten happens more often regarding 2-20 and 3-30 combinations. Once again, most mistakes happen when test takers are supposed to write number 2 (*vein*), but they write 20 (*vinca*) instead as 20.7% of the test takers did so the first time number 2 (*vein*) appears in the test. So did 21.4% of the test takers the second time this number is mentioned (see all wrong answer percentages regarding units and corresponding tens in Table K.14 in appendix K).

The percentage of test takers who made a mistake writing number 20 instead of 2 is slightly lower in this test than in the MLAT-ES in the two instances this number appears. This could be due to the fact that while number 2 (*vein*) in the artificial language used in this part is very similar to the real name for number 20 in Spanish ("*veinte*"), it is not in Catalan, in which number 20 is "*vint*". Actually, number 20 in Catalan shares the same syllable with its name in the artificial language of the test. Quite a few test takers (44 out of 304 - 14.5%) wrote 20 (*vinca*) for both item x and item y even though the last two numbers in the test are different (item x is *vein* and item y is *vinca*) (see Table K.15 in appendix K).

Table 3.81 shows precisely that number 2 (*vein*) is the item that appears to be the most difficult in grade 3 in its two instances (IF=.46 the first time and IF=.37). While this item was also difficult for the 4-graders who took the MLAT-ES, it is still among the most difficult ones, but it never reaches the category of difficult. On the contrary, its IF stays among the range of easy items (IF=.60 in the first instance and IF=.69 the second time) along with item f, which corresponds to number 3 and whose IF is .71 (for a complete account of the index of facility of all the items and distractors, see appendix N).

Table 3.81. Percentage of items on the MLAT-EC *Part 4 Aprenguem números* according to their facility across grades

IF	Very easy > 0.74	Easy 0.55 - 0.74	Mid-difficult 0.45 - 0.54	Difficult 0.25 – 0.44	Very difficult <0.25
Grade 3 N=57	**8 %** Items: 1 (j); 32 (m)	**72%** Items: 3 (f) ; 10 (n); 11 (c, w); 12 (g, r); 20 (l, x); 21 (e, q); 22 (s); 23 (a, v); 30 (h); 31 (k); 32 (u); 33 (d, p)	**16%** Items: 2 (b); 13 (o, t); 22 (i)	**4%** Items: 2 (y)	**0%** Items: -
Grade 4 N=62	**88%** Items: 1 (j); 10 (n); 11 (c, w); 12 (g, r); 13 (o, t); 20 (l, x); 21 (e, q); 22 (i, s); 23 (a, v); 30 (h); 31 (k); 32 (m, u); 33 (d, p)	**12%** Items: 2 (b, y); 3 (f)	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grade 5 N=61	**100%** Items: 1 (j); 2 (b, y); 3 (f); 11 (c, w); 12 (g, r); 13 (o, t); 20 (l, x); 21 (e, q); 22 (i, s); 30 (h); 31 (k); 32 (m, u); 33 (d, p)	**0%** Items: -	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grade 6 N=60	**92%** Items: 1 (j); 3 (f); 10 (n); 11 (c); 12 (g, r); 13 (o, t); 20 (l, x); 21 (e, q); 22 (i, s); 23 (a, v); 30 (h); 31 (k); 32 (m, u); 33 (d, p)	**8%** Items: 2 (b, y)	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grade 7 N=64	**100%** Items: 1 (j); 2 (b, y); 3 (f) ; 10 (n), 11 (c, w); 12 (g, r); 13 (o, t); 20 (l, x); 21 (e, q); 22 (i, s); 23 (a, v); 30 (h); 31 (k); 32 (m, u); 33 (d, p)	**0%** Items:-	**0%** Items: -	**0%** Items: -	**0%** Items: -
Grades 3-7 N=304	**92%** Items: 1 (j); 3 (f); 10 (n); 11 (c, w); 12 (g, r); 13 (o, t); 20 (l, x); 21 (e, q); 22 (i, s); 23 (a, v); 30 (h); 31 (k); 32 (m, u); 33 (d, p)	**8%** Items: 2 (b, y)	**0%** Items: -	**0%** Items: -	**0%** Items: -

All the parts of the test prove to be very easy, except for grade 3, in which item y, aiming at number 2, is difficult and four more items are mid-difficult. These involve again number 2 in addition to number 22, which makes use of the name of number 2 to be formed, and number 13. Nevertheless, the percentage of very easy items is extremely high from grades 4 to 7. As happened with the same part on the MLAT-ES, it is important to see if the test works properly across grades as it seems to reach a ceiling effect from a very early stage.

As for the item discrimination power (see Table 3.82), in contrast with this part on the MLAT-ES, the item that should be improved is item j, which aims at number 1 and not the numbers formed using number 30. There are some items that could be

improved. They are those which make use of number 30, but also number 10, 20 and 21 (see Table K.16 in appendix K for all D_i and correct answers at the upper and lower 33% tails).

Table 3.82. Percentage of items on the MLAT-EC *Part 4 Aprenguem números* according to their index of discrimination

D_i	Very good > 0.40	Good 0.30 - 0.39 (though they could be improved)	Average 0.20 - 0.29 (they have to be improved)	Deficient <0.19 (they should be removed)
Grades 3-7 N=304	**60%** Items: 2 (b, y), 3 (f); 11 (c, w); 12 (g, r); 13 (o, t); 20 (x); 22 (i, s); 23 (a, v), 32 (u); 33 (d)	**36%** Items: 10 (n); 20 (l, x), 21 (e, q); 30 (h); 31 (k); 32 (m); 33 (p)	**4%** Items: 1 (j)	**0%** Items: -

The table containing the item-total correlation of this part 4 does not confirm the ill-functioning of item j. On the contrary, its item-total correlation value is above the minimum required, which is .300 and even some items show a lower coefficient than this item.

Table 3.83. Item-total correlation of MLAT-EC *Part 4 Aprenguem números*

Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}	Item	r_{pb}
a. 23	.635**	**f. 3**	.604**	**k. 31**	.685**	**p. 33**	.690**	**u. 32**	.715**
b. 2	.477**	**g. 12**	.633**	**l. 20**	.492**	**q. 21**	.613**	**v. 23**	.653**
c. 11	.598**	**h. 30**	.563**	**m. 32**	.710**	**r. 12**	.639**	**w. 11**	.662**
d. 33	.672**	**i. 22**	.653**	**n. 10**	.643**	**s. 22**	.687**	**x. 20**	.513**
e. 21	.640**	**j. 1**	.513**	**o. 13**	.736**	**t. 13**	.697**	**y. 2**	.554**

** $p<.001$

The reliability of *Parte 4* is excellent (.933) and, according to the corrected item-total correlation, no item should be removed either (see Table K.17 in appendix K).

3.5.3.2. Reliability and intercorrelation of parts

The reliability of each part is calculated using Cronbach's alpha, first, for the reasons explained in section 2.3.2.3, and second, for the sake of comparison with the MLAT-ES administered in Catalonia and with the data in the *MLAT-ES Manual.* Based

on the item analysis run, it was decided to remove item 3 from *Part 1 Paraules ocultes*. In *Part 2 Paraules que es corresponen,* apart from removing the item equivalent to item 22 in the original version, item 4 has also been removed. This item was intended to improve the performance in its equivalent in the MLAT-ES (*La GRAN mansión del presidente es blanca – En la clase de matemáticas hay pocos alumnos*) because the target word "*pocos"* is an adjective whose function does not correspond exactly to the adjective's in the stem sentence. Hence, in the Catalan version, the target sentence was changed into "*A la classe de matemàtiques hi ha alumnes intel·ligents*"*.* The test takers responded to the item as expected, choosing the qualitative adjective "*intel·ligents*" rather than "*matemàtiques*" or "*classe*", as happened on the MLAT-ES, but then this item turned out not to be discriminant at all in any grade. That is the reason why it was considered sensible to remove it. Unfortunately, this part of the MLAT-EC is rendered with one fewer item than *Parte 2* in the MLAT-ES. In *Part 3 Paraules que rimen*, 8 items were removed (2, 7 13 15, 16, 18, 22 and 28) after the results obtained in the IF and D_i analyses. Taking into account that this study involves only 309 test takers, while in all the other tests reviewed so far the cohorts had more than 1,000 test takers, the data regarding the norming study of the MLAT-EC should be taken as estimative but, certainly, not as definite. *Part 4 Aprenguem números* obtained very good reliability indexes in all grades despite being easy from grade 4 on, as its Spanish equivalent. Item j, which aims at remembering number 10 (*silca*), obtained an average D_i. This makes it highly recommended to improve it. However, its corrected item-total correlation is within the limits of acceptability, which goes in its favour. Moreover, removing it after the test has been taken and still using the data is out of the question due to the inherent characteristics of this test (a memory test in which the number of items goes hand in hand with the time test takers have to remember the items memorised for the task).

The reliability coefficients and standard errors of measurements of total raw scores on the MLAT-EC have been calculated without taking into account the sex variable, due to the reduced sample of the population resulting when split according to this variable. Actually, in the *MLAT-ES Manual* these statistics do not appear divided according to this variable either, although they do in the MLAT-E.

Table 3.84. Reliability coefficients and standard errors of measurements of total raw scores on the MLAT-EC

Statistics	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7
N	57	62	61	60	64
Reliability	.957	.960	.950	.944	.961
SE_M	4.73	4.36	4.11	4.09	3.63
Mean	55.12	83.77	93.13	94.20	99.95
SD	22.8	21.8	18.4	17.3	18.4

The reliability indexes of this test as a whole are excellent across grades. For the whole test and all grades it is .972. While this index is very similar across grades, the mean is not. Particularly, the mean between grades 3 and 4 increases more than one standard deviation, 28.65 points to be more precise, while the next biggest increase, between grades 4 and 5 is only 9.36 and even less between the other grades. Therefore, the test may not function in the same way between grades 3 and 4 as it does in the other grades, across which the increase is not nearly as sharp as in the lower grades.

The reliability of each part across grades, as shown in Table 3.85, is not always over .900 when checked per grade, the lowest being .856 in part 2 in grade 3 and .941 in part 4 in grade 7. Nevertheless, when all grades are combined, the indexes of reliability are all over .900 and it peaks when the test is treated as a whole and all the population is combined. This is understandable, as the number of subjects and items is much higher.

Table 3.85. Intercorrelations, reliability coefficients, part-total correlations, mean *p*-values, and mean item total of parts of the MLAT-EC

	GRADE 3 (N=57)				GRADE 4 (N=62)			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1. *Paraules ocultes*	**.934**				**.900**			
2. *Paraules que es corresponen*	.377**	**.856**			.525**	**.897**		
3. *Paraules que rimen*	.640**	.355**	**.934**		.700**	.536**	**.939**	
4. *Aprenguem números*	.331*	.151	.179	**.936**	.575**	.310*	.423**	**.909**
Part–total correlation	.844**	.573**	.779**	.594**	.869**	.755**	.854**	.654**
Mean *p*-value	.75	.32	.55	.62	.81	.50	.72	.82
Mean item total	.41	.36	.45	.62	.69	.54	.71	.83
	GRADE 5 (N=61)				GRADE 6 (N=60)			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1. *Paraules ocultes*	**.871**				**.878**			
2. *Paraules que es corresponen*	.513**	**.903**			.442**	**.910**		
3. *Paraules que rimen*	.509**	.557**	**.909**		.650**	.461**	**.919**	
4. *Aprenguem números*	.509**	.287*	.309*	**.922**	.038	.167	.210	**.893**
Part–total correlation	.798**	.818**	.824**	.541**	.736**	.786**	.822**	.406**
Mean *p*-value	.83	.65	.78	.88	.87	.67	.83	.85
Mean item total	.76	.67	.76	.88	.78	.67	.80	.85
	GRADE 7 (N=64)				GRADES COMBINED (N=304)			
	1	**2**	**3**	**4**	**1**	**2**	**3**	**4**
1. *Paraules ocultes*	**.873**				**.327**			
2. *Paraules que es corresponen*	.657**	**.915**			.635**	**.922**		
3. *Paraules que rimen*	.709**	.559*	**.913**		.737**	.607**	**.945**	
4. *Aprenguem números*	.223	.397**	.183	**.941**	.500**	.440**	.406**	**.935**
Part–total correlation	.860**	.852**	.851**	.408**	.877**	.840**	.867**	.636**
Mean *p*-value	.89	.74	.83	.91	.84	.58	.75	.82
Mean item total	.83	.74	.80	.93	.70	.60	.71	.82

Note. – Reliability coefficients are indicated in bold-face type.
* Correlation is significant at 0.05 (two-tailed)
** Correlation is significant at 0.01 (two-tailed)

As happened in the MLAT-ES, correlations between all parts and the total score are high and significant, although *Part 4 Aprenguem números* consistently gets lower correlations across grades. When combining each part separately, the correlations are from low to high in all parts and grades (ranging from .151 to .737**). The fact that all the intercorrelations are from moderate to high when the grades are combined is not favourable to the test, as this means that each part overlaps somehow with the construct that the other parts in the test are supposed to measure. Overall the intercorrelations of this test are higher than the ones of the MLAT-ES administered in Catalonia and the indexes provided in the norming *Manual* too. Significantly enough, the highest ones are found between *Part 1 Paraules ocultes* and *Part 3 Paraules que rimen*. This is something that can be explained by the fact that these parts depend very much on sound-spelling correspondences. Despite the fact that most correlations are

moderate, the ones that are low reach lower levels than in the other tests commented here. These are all the correlations that involve *Part 4 Aprenguem números*, which taps rote memory ability.

The mean item-total correlation data is much lower in grade 3 than in the other grades, especially in *Part 2 Paraules que es corresponen* and *Part 3 Paraules que rimen*. This could be related to the fact that the MLAT-EC proves to be much more difficult for 3-graders than for the rest of the cohort, regardless of the score the test takers in this grade obtain. In the other grades, however, the mean item-total correlations indicate that all the parts are very good at discriminating across grades. For the whole test and grades, the mean item-total value is .70.

Once more *Part 2 Paraules que es corresponen* appears to be the most difficult part for all grades as shown by the mean *p*-values, but it is especially so for test takers in grade 3, even when comparing this value to its equivalent value on the MLAT-ES. Also, as on the MLAT-ES, the other mean *p*-values show that all the other parts of the test are easy, as all of them are higher than .55 and *Part 3 Paraules que rimen* is the next part in an ascending order as far as difficulty is concerned across grades. If *Parte 4 Aprendamos números* reached a very high mean *p*-value from grade 5, the values are even higher in the MLAT-EC, reaching very high values starting from grade 4 on. Yet again, this part proves to be extremely easy. Curiously, and in contrast with the MLAT-ES, the mean *p*-value of this part decreases in grade 6 although it reaches exactly the same degree of difficulty in grade 7 as it did on the MLAT-ES. The mean *p*-value for the whole test and grades is .75, so it is overall an easy test.

3.5.3.3. Descriptive statistics of the MLAT-EC

As happened when exploring the data of the MLAT-ES, the distribution of the MLAT-EC is not normal (see Shapiro-Wilk normality test in Table K.18 in appendix K). Hence the use of non-parametric statistical techniques in all the analyses that will follow, as non-parametric tests are also very robust in the presence of outliers.

The descriptive statistics of *Part 1 Paraules ocultes* (Table 3.86) show that, while the difference in the means from grades 4 to 7 increases more or less gradually (maximum 2.10 points), the increase in the means from grades 3 to 4 is much sharper (slightly more than 8 points). Standard deviations are on the decrease across grades except between grades 5 and 6, where they are almost the same. The range of scores

is very similar across grades, as the maximum score is reached already by the test takers in grade 4. The minimum score is markedly larger in grade 7 than in the other grades. The skewness value shows a tendency for scores to pile up on the right tail and reach significant values from grade 5 on. These appear in bold letters in the table. As happened in *Parte 1 Palabras ocultas*, the only group that presents positive skewness is grade 3, although its value is not significant. Kurtosis is negative in grades 3, 4 and 5, but its values are relatively close to 0. From grade 5, kurtosis values would be represented in a leptokurtic shape and reach significant values in the two upper grades.

Table 3.86. Descriptive raw scores for *Part 1 Paraules ocultes* on the MLAT-EC (30 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SE S[a]	$z_{skewn.}$	Kurt.	SE K[b]	$z_{kurtosis}$
3	57	12.37	11.00	8.03	0	28	28	.338	.316	1.070	-.877	.623	-1.408
4	62	20.65	22.00	6.45	5	30	25	-.480	.304	-1.579	-.585	.599	-.977
5	61	22.75	24.00	5.46	8	30	22	-.832	.306	**-2.719**	-.905	.604	-1.498
6	60	23.47	24.00	5.47	4	30	26	-1.327	.309	**-4.294**	2.009	.608	**3.304**
7	64	25.16	26.50	4.82	10	30	20	-1.517	.299	**-5.074**	2.335	.590	**3.958**

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

In *Part 2 Paraules que es corresponen,* while not so acute as in *Parte 1*, the difference between the means in grades 3 and 4 is of 5.15 points, but it is not so large in the rest of grades. Actually, it diminishes a little between grades 5 and 6, but it continues on the increase between grades 6 and 7. Standard deviations are very similar across grades, as are minimum and maximum scores, the wider range of scores being in grade 3. Skewness has negative values in grades 3 and 4, but it is only significant in grade 3. Scores in the other grades are negatively skewed, but they reach a significant value in grade 7 only. Regarding kurtosis, values appear in a leptokurtic distribution in all grades but 3, although no kurtosis value is significant.

Table 3.87. Descriptive raw scores for *Part 2 Paraules que es corresponen* on the MLAT-EC (29 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SE S[a]	$z_{skewn.}$	Kurt.	SE K[b]	$z_{kurtosis}$
3	57	10.40	9.00	6.02	1	29	28	.982	.316	**3.108**	-.888	.623	-1.425
4	62	15.55	13.00	7.08	3	28	25	.416	.304	1.368	-1.074	.599	-1.793
5	61	19.44	20.00	6.96	6	28	22	-.454	.306	-1.484	-1.117	.604	-1.849
6	60	19.28	20.00	7.06	7	29	22	-.324	.309	-1.049	-1.184	.608	-1.947
7	64	21.31	23.50	6.81	6	29	23	-.951	.299	**-3.181**	-.353	.590	-.598

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

As happened in *Part 1,* means increase gradually across grades in *Part 3 Paraules que rimen,* reaching a plateau between grades 6 and 7 except between grades 3 and 4, where the difference between means is of 10.09 points. Grade 7 has the lowest range of scores, as the minimum score is higher than in the other grades. Maximum scores are reached already in grade 4 and continue to be reached in the rest of grades. 3-graders never reach the maximum score although their maximum score in this grade is very close to the limit (36 points). The smallest range of scores is found in grade 7 and the largest one in grade 4. The skewness values are mostly negative and all of them reach significant values at *p* <.01 except in grade 3, in which scores tend to be in the left tail though not with a significant value. Kurtosis presents itself differently and without a clear pattern when comparing grades. While it has negative values in grades 3 and 7, it is leptokurtic though with no significant values in the rest of grades.

Table 3.88. Descriptive raw scores for *Part 3 Paraules que rimen* on the MLAT-EC (38 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SE S[a]	$z_{skewn.}$	Kurt.	SE K[b]	$z_{kurtosis}$
3	57	16.89	16.00	9.41	1	36	35	.555	.316	1.756	-.617	.623	-.990
4	62	26.98	30.00	9.12	1	38	37	-.975	.304	**-3.207**	.376	.599	.628
5	61	28.97	31.00	7.37	9	38	29	-1.062	.306	**-3.471**	.444	.604	.735
6	60	30.32	32.50	7.19	8	38	30	-1.251	.309	**-4.049**	1.012	.608	1.664
7	64	30.28	33.00	7.04	13	38	25	-.956	.299	**-3.197**	-.352	.590	-.597

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

In *Part 4 Aprenguem números,* means are very high from grade 3 already and the maximum score, 25 points, is also reached by the test takers in this grade. Minimum scores are very low in all grades but 6, where the maximum score is 10. Medians are also very high from grade 3 on and reach values close to the maximum score from grade 5 to 7. The skewness value is significant in all grades, increasing gradually across grades. These values show that the scores are mostly piled up on the right tail of the distribution. As the kurtosis values are positive, this shows that all the distributions are pointy, reaching extremely high levels of significance in grades 5 and 7, in which both kurtosis and its significance soar dramatically.

Table 3.89. Descriptive raw scores for *Part 4 Aprenguem números* on the MLAT-EC (25 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SES[a]	$z_{skewn.}$	Kurt.	SEK[b]	$z_{kurtosis}$
3	57	15.46	16.00	7.56	0	25	25	-.4090	.316	-1.294	-1.140	.623	-1.830
4	62	20.60	22.50	5.25	4	25	21	-1.631	.304	**-5.365**	2.151	.599	3.591
5	61	21.97	24.00	4.66	1	25	24	-2.646	.306	**-8.647**	8.349	.604	**13.823**
6	60	21.13	23.50	4.76	6	25	19	-1.394	.309	**-4.511**	1.558	.608	2.563
7	64	23.20	25.00	4.03	3	25	22	**-3.589**	.299	**-12.003**	14.290	.590	**24.220**

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

Taking a look at the test as a whole, it can be seen that mean scores experience a sharp increase between grades 3 and 4 (28.65 points), that the increase is not so sharp between grades 4 and 5 (slightly less than 10 points), and that the increase is almost unnoticeable between grades 5 and 6, though it increases a little more again between grades 6 and 7. Medians follow exactly the same pattern as means, and standard deviations are fairly similar across grades. The minimum score is especially low in grade 3, but the maximum score reached is very similar in all grades. Actually, the maximum score in grade 4 is higher than in grades 5 and 6. Consequently, the range of scores in grade 3 is larger than in the other grades. The skewness values are negative in all grades but in grade 3, which shows that from grades 4 to 7 onwards, distribution values are skewed towards the right, reaching a significant value only in grade 7. The kurtosis values are all negative but in grade 7, where the distribution is pointy although not reaching a significant value at $p < .01$.

Table 3.90. Descriptive total raw scores on the MLAT-EC (122 items)

Gr.	N	Mean	Mdn	SD	Min	Max	Range	Skewn.	SES[a]	$z_{skewn.}$	Kurt.	SEK[b]	$z_{kurtosis}$
3	57	55.12	52.00	22.75	19	115	96	.439	.316	1.389	-.470	.623	-.754
4	62	83.77	85.00	21.77	38	120	82	-.257	.304	-.845	-.718	.599	-1.199
5	61	93.13	95.00	18.44	38	119	81	-.653	.306	-2.134	-.107	.604	-.177
6	60	94.20	98.00	17.32	49	119	70	-.566	.309	-1.832	-.529	.608	-.870
7	64	99.95	106.00	18.37	41	121	80	-1.204	.299	**-4.027**	1.000	.590	1.695

[a]Standard Error of Skewness
[b]Standard Error of Kurtosis

As in the MLAT-ES, in the raw total score there is a sharper increase between grades 3 and 4 than between the other grades. While in the MLAT-ES the increase was mainly in *Parte 1* and *Parte 2*, in the MLAT-EC this increase is found in *Part 1* and *Part 3* (for more information on the comparison between the means of both tests, see section 3.5.4). Particularly, in *Part 3* the difference in means between grades 3 and 4 is three times the difference in this same part of the Spanish version. On the other hand,
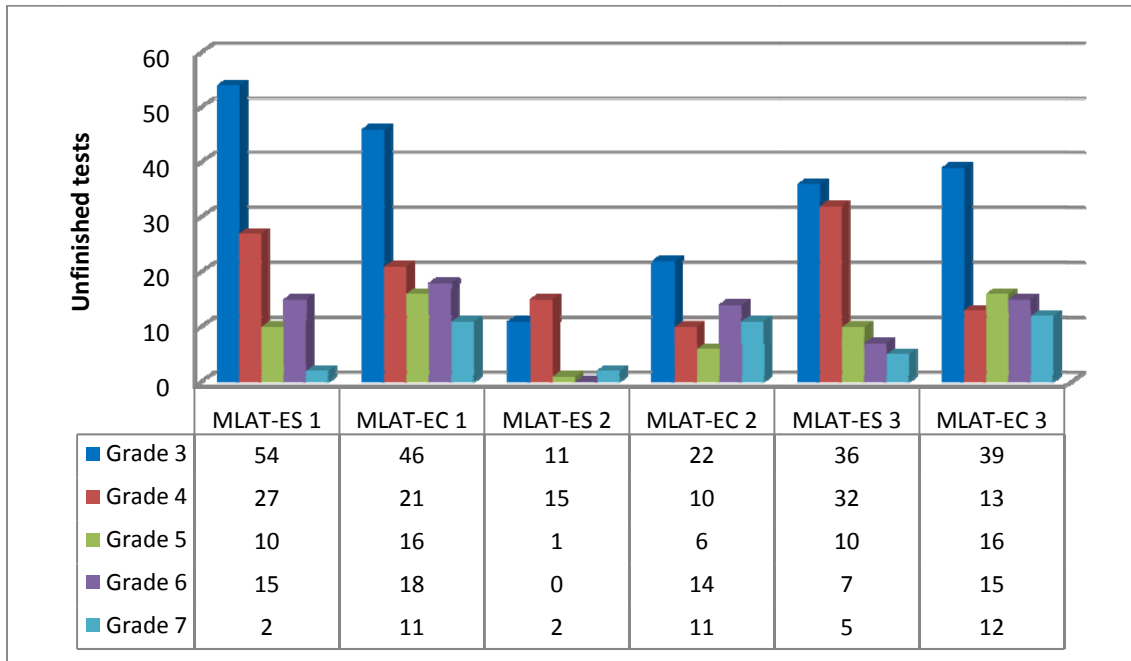
the increase rate in all parts diminishes the upper the grade, so it could be possible that at some later point (from grade 8 on), scores reached a plateau. While the test takers' results on *Part 1* and *Part 3* increase gradually across grades, except for between grades 3 and 4, the difference in the scores on *Part 2* between grades 4 and 7 is rather small. The range of scores between the means of grades 3 and 7 is similar in all parts but in *Part 4*, where the difference in the means between grades 3 and 7 is 7.74 points, which is slightly larger than in the same part of the MLAT-ES (5.26 points).

3.5.4. Comparing the MLAT-ES and the MLAT-EC

In order to check whether the MLAT-ES and the MLAT-EC are equivalent tests, all the aspects that have been examined separately in sections 3.5.2 and 3.5.3 should be comparable. These include test speed, difficulty, reliability and correlations between both tests.

To start with, the issue of speed was present in the first three parts of both tests. There was no clear pattern as for the decrease of the percentage of test takers unable to finish the test taking into account the grade variable. However, what can be observed is that both in Part 1 and Part 3 of both tests, subjects in grade 3 hold the highest number of unfinished tests. In contrast, the number of test takers who did not manage to finish Part 2 is significantly lower in both tests, especially in the MLAT-ES. In both tests, the number of unfinished tests decreases the upper the grade. It is important to mention, though, that the number of test takers in grades 3 and 4 who did not finish Part 2 is much smaller than the number of those who did not get to finish the other parts. Therefore, it can be concluded that, although the three parts are meant to be speeded, speed is indeed an issue in grades 3 and 4, with greater impact on Parts 1 and 3, but not so much on Part 2.

Figure 3.6. Number of unfinished tests across grades

	MLAT-ES 1	MLAT-EC 1	MLAT-ES 2	MLAT-EC 2	MLAT-ES 3	MLAT-EC 3
■ Grade 3	54	46	11	22	36	39
■ Grade 4	27	21	15	10	32	13
■ Grade 5	10	16	1	6	10	16
■ Grade 6	15	18	0	14	7	15
■ Grade 7	2	11	2	11	5	12

Once normed, both tests end up with a different number of items. In order to check the difficulty of each part of the tests, their mean *p*-values are to be compared. Below are these values represented together so that they can be compared across grades.

Figure 3.7. Mean *p*-values on part 1 on the MLAT-ES and the MLAT-EC across grades
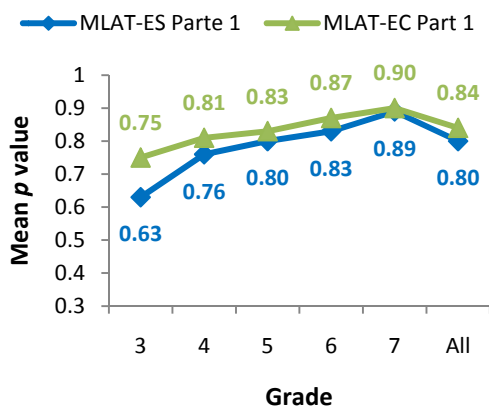
Figure 3.8. Mean *p*-values on part 2 on the MLAT-ES and the MLAT-EC across grades
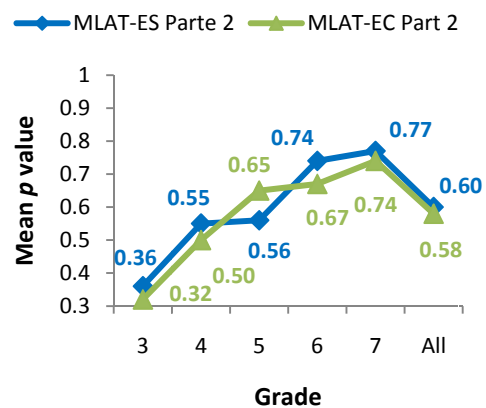
Figure 3.9. Mean p-values on part 3 on the MLAT-ES and the MLAT-EC across grades
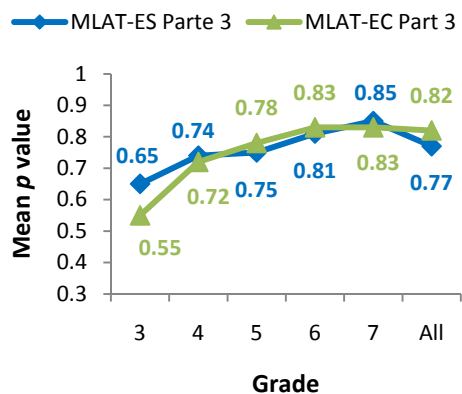
Figure 3.10. Mean *p*-values on part 4 on the MLAT-ES and the MLAT-EC across grades
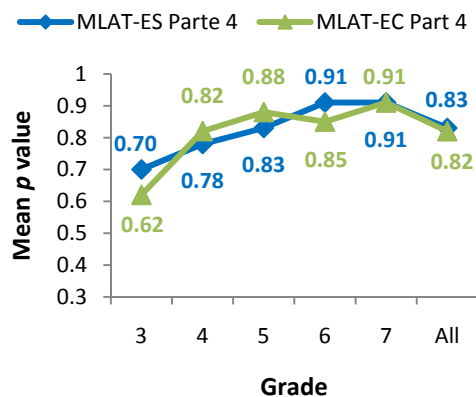
The mean *p*-value in *Parte 1* is consistently lower across grades but for grade 7 than it is in *Part 1 Paraules ocultes*, especially from grades 3 to 4. The reverse phenomenon is found between grades 3 and 4 in *Part 3 Paraules que rimen* and in *Part 4 Aprenguem números*. In these parts, the mean *p*-value of the MLAT-EC is lower than it is in the Spanish version. Also, in these parts, the gap between grades 3 and 4 is greater in the MLAT-EC than in the MLAT-ES.

From grades 5 to 7, the mean *p*-values of all parts do not increase so sharply as in between grades 3 and 4 and 4 and 5. This linearity cannot be found so clearly in Part 2, in which fairly sharp jumps are found between grades 3 and 4, 4 and 5, and 5 and 6, the latter being, nevertheless, smoother. All in all, it could be concluded that both the MLAT-ES and the MLAT-EC are fairly similar as far as their mean *p*-values are concerned from grades 4 to 7, while in grade 3 the mean *p*-value of the different tests differs depending on the language version of the test except for Part 2 Matching words. This could be due to the fact that this part is the least dependent of the language the test is in while the rationale of Part 1 and, especially, of Part 3 are based on the orthographic system of the language. While in Spanish the correspondence between sound and letter is almost exact, it is not so direct in Catalan, as some vowels correspond to more than one phoneme and, therefore, the Catalan spelling system is more complex than the Spanish one, especially for 3-graders. All in all, the sharp increase in the total raw score between grades 3 and 4 appears in both versions of the test, as can be seen in the figure below, but the one in the Catalan version is much sharper than the one in the MLAT-ES. Mind that the total raw score in Figure 3.11 is out of 123 in the MLAT-ES and out of 122 in the MLAT-EC.

Figure 3.11. Mean total raw scores on the MLAT-ES and the MLAT-EC across grades
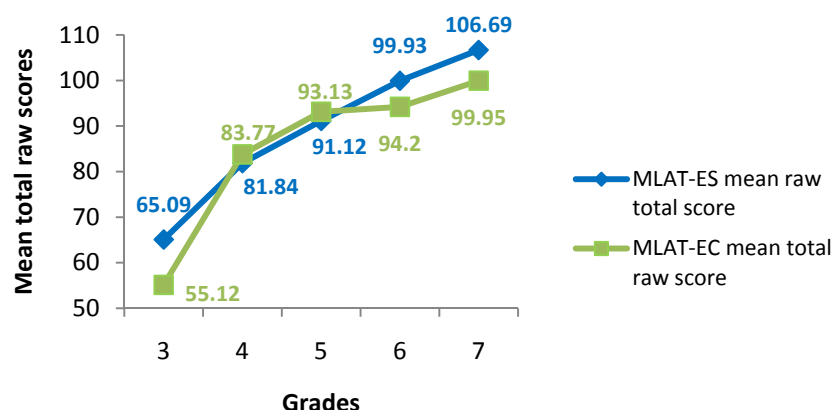
Despite the differences in specific parts and grades, both versions of the test had similar excellent reliability indexes in all parts and grades. There is slightly more variability in the indexes of part 1 in grade 3 and from grades 5 to 7, and in part 3 from grades 5 to 7, than in the other parts, but at no point is the reliability lower than acceptable in any of the parts.

Table 3.91. MLAT-ES and MLAT-EC Cronbach reliability indexes across grades

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 7
MLAT-ES *Parte 1*	.896	.896	.849	.912	.766	.927
MLAT-EC *Part 1*	.934	.900	.871	.878	.873	.932
MLAT-ES *Parte 2*	.825	.895	.921	.908	.905	.925
MLAT-EC *Part 2*	.856	.897	.903	.910	.915	.922
MLAT-ES *Parte 3*	.932	.938	.938	.934	.877	.942
MLAT-EC *Part 3*	.934	.939	.909	.919	.913	.945
MLAT-ES *Parte 4*	.925	.909	.925	.863	.923	.925
MLAT-EC *Part 4*	.936	.909	.922	.893	.941	.935
MLAT-ES *Total*	.948	.957	.958	.959	.930	.969
MLAT-EC *Total*	.957	.960	.950	.944	.960	.972

One of the statistical analyses to prove the reliability of a test is running a test-retest correlation. This technique is used when comparing the same test administered twice to the same population or when comparing two forms of the same test. This is not exactly the case of this study, although the tests are nearly parallel in content and difficulty, as proved so far in this section. Besides, both versions have similar estimates of reliability, means, standard deviations and standard errors of measurement across grades (see sections 3.5.2.2 and 3.5.3.2). Moreover, Catalan and Spanish, both Romance languages, are very similar types of languages. Therefore, it is very likely

that both the Catalan and the Spanish versions of the tests are so similar that they could be considered two forms of the same test, so it seems reasonable to run a Spearman correlation in order to see whether they correlate significantly and positively and thus confirm that they are equivalent tests. By doing this, though, we assume that, besides being parallel, the two sets of scores are independent, i.e. the second score is unaffected by the first (Bachman, 2004).

Table 3.92. Group 1 test-retest correlation across grades

Group 1 MLAT-ES / MLAT-EC	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	All grades
N	66	75	57	60	67	325
MLAT-ES 1 / MLAT-EC 1	.580**	.772**	.574**	.673**	.510**	.788**
MLAT-ES 2 / MLAT-EC 2	.647**	.705**	.790**	.798**	.573**	.783**
MLAT-ES 3 / MLAT-EC 3	.766**	.765**	.684**	.716**	.407**	.708**
MLAT-ES 4 / MLAT-EC 4	.637**	.654**	.746**	.704**	.476**	.634**
MLAT-ES / EC Total	.838**	.850**	.880**	.844**	.658**	.883**

While the entire group 1 took the MLAT-EC, not all the subjects in group 2 took the MLAT-ES, that is why the number of subjects for group 2 and for the whole population in these analyses is lower than in the descriptive statistics.

Table 3.93. Group 2 test-retest correlation across grades

Group 2 MLAT-EC / MLAT-ES	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	All grades
N	43	41	45	50	64	243
MLAT-EC 1 / MLAT-ES 1	.463**	.680**	.802**	.450**	.617**	.725**
MLAT-EC 2 / MLAT-ES 2	.505**	.545**	.776**	.780**	.745**	.754**
MLAT-EC 3 / MLAT-ES 3	.505**	.686**	.824**	.677**	.688**	.737**
MLAT-EC 4 / MLAT-ES 4	.429**	.621**	.474**	.457**	.397**	.555**
MLAT-EC / ES Total	.636**	.766**	.925**	.750**	.846**	.844**

From the correlations between tests, which are all positive and significant (see Table 3.92 and Table 3.93), it could be considered that the MLAT-ES and the MLAT-EC are equivalent only up to a certain extent, as most correlations are significant and from moderate to high, but not extremely close to 1. See that grade 3 obtains overall the lowest correlation coefficients in group 1 and in group 2 and so does grade 7 except when correlating total scores. Part 4 is the one that, in general, gets the lowest correlation coefficients across grades, especially in group 2, and when considering the whole cohort as one group. Besides, significant correlations appear only at a moderate

3. Methodology

level between *Parte 1 Palabras ocultas* and *Part 1 Paraules ocultes,* especially in grades 3, 6 and 7. The reason for this moderate correlation between tests could be due to the nature of these parts of the test, as they are very much language-dependent. That is to say, the way in which words were hidden in both tests varied according to the language, as explained in section 3.4. Part 4 also shows the lowest correlations of all the parts, especially in grade 7.

Both tests were administered to the same population leaving 3 or 4 months between test sittings so that test takers could not remember the exact contents of the test. This, though, did not eliminate the possibility that there could be task-training effect when taking the test for the second time. If this happens, no matter whether both versions are almost equivalent, only the score obtained in the first place can be used for further analyses. Obtaining high correlation between two scores, as shown in the tables above, only indicates that the scores move in the same direction, but there could be significant differences in the scores obtained because the test takers could have significantly obtained, in principle, higher marks in the second sitting because of being already familiar with the test format and contents. Below are the descriptive statistics in percentages obtained by the subjects on both the first time they took the test and the second time they took it in the other language. Since not all the students in the norming study of the MLAT-EC took the MLAT-ES, the number of participants in this table is smaller.

Table 3.94. Group 1 descriptive statistics of score percentages on the MLAT-ES and MLAT-EC

Grade	3		4		5		6		7	
N	66		75		57		60		67	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MLAT-ES *Parte 1*	42.73	22.69	64.84	22.70	76.43	17.76	75.56	22.16	89.20	10.71
MLAT-EC *Part 1*	50.40	22.30	77.73	19.23	83.80	14.69	89.33	13.21	92.74	10.66
MLAT-ES *Parte 2*	41.92	19.57	58.49	24.02	62.11	26.27	77.78	21.33	79.95	20.43
MLAT-EC *Part 2*	49.95	25.96	66.57	23.97	66.55	24.65	76.90	23.21	81.78	19.21
MLAT-ES *Parte 3*	58.21	24.68	66.60	24.81	75.62	23.23	82.06	20.44	87.12	14.04
MLAT-EC *Part 3*	61.44	25.78	76.28	17.76	74.98	19.95	80.92	18.09	86.68	12.79
MLAT-ES *Parte 4*	70.30	27.14	78.13	22.63	83.30	22.09	91.00	13.69	91.34	16.56
MLAT-EC *Part 4*	74.00	26.36	85.23	20.92	87.23	20.59	90.20	14.07	92.90	11.18
MLAT-ES *Total*	52.92	17.14	68.54	18.01	74.08	17.16	81.25	15.47	86.74	10.76
MLAT-EC *Total*	58.67	20.32	76.16	16.00	77.65	16.50	83.93	14.07	88.28	10.28

As expected, most scores on the MLAT-EC are higher since it was taken in the second place. Exceptions are grade 6 on *Part 2* and 4, grades 5 to 7 on *Part 3* and grade 6 on *Part 4,* in which the test takers obtained slightly lower scores. While the differences between means are consistently larger in Parts 1 and 3, this is not the case

of Parts 2 and 4. The fact that Parts 1 and 3 are heavily language-based, while Parts 2 and 4 measure constructs that are somehow more independent of language (grammatical sensitivity and memory), could be the reason for the smaller difference in means between these parts. Several factors could explain the lower scores in grades 5, 6 and 7 on *Part 3*. One of them could be possible language dominance issues, as group 1 is formed mainly by subjects whose language preference is Spanish (N=132) or both Catalan and Spanish, with no preference for one language over the other (N=102), while only 88 test takers declared themselves to be Catalan dominant. As this part is language-based, it is possible that it becomes easier if taken in the language one masters the most. Another factor for the scores on the MLAT-EC not being higher than those on the MLAT-ES could be the larger variety of phonemes in Catalan as compared to Spanish. This can increase the difficulty of this part even if taken in the second place.

Table 3.95. Group 2 descriptive statistics of score percentages on the MLAT-ES and MLAT-EC

Grade	3		4		5		6		7	
N	43		41		45		50		64	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MLAT-EC *Parte 1*	39.15	27.53	68.78	21.89	75.93	19.22	77.07	18.35	83.85	16.04
MLAT-ES *Part 1*	46.74	25.21	69.92	25.41	76.44	19.72	87.00	13.13	89.58	14.37
MLAT-EC *Parte 2*	39.62	21.11	60.08	24.68	66.74	24.25	65.10	24.67	73.49	23.47
MLAT-ES *Part 2*	44.03	24.98	70.92	25.53	66.52	24.49	73.80	23.24	77.50	21.17
MLAT-EC *Parte 3*	43.02	24.14	70.92	25.53	78.60	16.51	79.58	16.51	79.69	18.53
MLAT-ES *Part 3*	50.92	26.74	67.33	26.74	79.77	20.36	88.95	16.49	87.95	14.84
MLAT-EC *Parte 4*	64.56	29.82	77.95	53.73	86.49	20.94	84.48	19.87	92.81	16.11
MLAT-ES *Part 4*	78.42	24.72	86.63	16.50	90.04	22.11	93.28	8.56	85.94	8.47
MLAT-EC *Total*	45.67	18.63	68.91	19.51	76.74	14.89	76.52	14.28	81.93	15.05
MLAT-ES *Total*	53.81	19.60	70.12	20.57	77.81	17.64	85.66	12.01	87.42	11.41

The order of administration affected the means of group 2, who took the Catalan version first and then the Spanish version of the MLAT-E. Means are higher in the second sitting in all parts and grades except for grade 4 in *Parte 3*, for grade 5 in *Parte 2* (where the scores in the Spanish and the Catalan versions are almost identical) and for grade 7 in *Parte 4*.

In *Part 1* the increase in the means in the second sitting of group 2 (as compared to *Parte 1*) tends to be lower than in group 1 except in grade 7. This could be due to the fact that group 2 has more subjects whose language preference is Catalan (N=147) over Spanish (N=87) or both Catalan and Spanish (N=70) and, therefore, for them the Catalan version of this test is easier than it is for those who

declared themselves to be Spanish dominant or having equal preference for both languages. However, in *Part 3*, which, like *Part 1*, is very much language-dependent, the increase between the means obtained by group 2 is larger than that by group 1 except for grade 4, whose score on this part is lower in the second sitting. Given the fact that the Spanish orthographic system is more transparent than the Catalan one, a larger increase between sittings is somehow expectable on this part. 4-graders' performance on *Parte 3* is an exception, for which no plausible explanation could be found. In *Part 2* the results obtained in the two sittings increase for all grades but 5. On the whole, as compared to group 1, the increase experienced by group 2 is not so large as the one experienced by group 1, although the means of both groups are very similar. Consequently, although *Part 2* should not be language-dependent, the language preference of the test takers in group 2 could be having an influence on the results on this *Part.*

The increase in group 2 scores on *Parte 4* is much higher than on *Part 4* for 3- and 6-graders than it was in the scores obtained by 3- and 6-graders in group 1 taking the tests in reverse order. Surprisingly, grade 7 obtained even lower results on this subtest in the second sitting. Actually, the increase of means across grades on this Part is not regular in any of the groups.

On the whole, for both group 1 and group 2, the means on the test taken in the second place are higher than on the test taken first, in spite of the time span between sittings. Certainly, there could have been a task-training effect. In order to check the significance of this effect, we need to run the Wilcoxon Signed Rank test, the non-parametric alternative to the repeated measures t-test. Table 3.96 shows the test statistic (T), which corresponds to the smallest of the two sums of ranks, as well as the z-value, the two-tailed significance value and the effect size of the difference between tests, obtained from dividing the z-value obtained in the Wilcoxon Signed-Rank test by the square root of the number of observations, i.e. N subjects per two tests. The two-tailed significant values and large effect sizes appear in bold.

Table 3.96. Wilcoxon Signed Rank Test between MLAT-ES and MLAT-EC according to order of administration

Group 1 ES-EC	Statistics	Gr. 3 N=66	Gr. 4 N=75	Gr. 5 N=57	Gr. 6 N=60	Gr. 7 N=67	All grades N=325
MLAT-EC 1 / MLAT-ES 1	T	505.50	195.50	.239.00	56.50	294.50	6308.00
	Z	-3.580 [a]	-5.907 [a]	-4.111 [a]	-5.929 [a]	-3.485 [a]	-10.277 [a]
	Asym. Sig. (2-tailed)	**.000**	**.000**	**.000**	**.000**	**.000**	**.000**
	Effect size	-.31	**-.48**	-.38	**-.54**	-.30	-.40
MLAT-EC 2 / MLAT-ES 2	T	605.00	858.50	579.50	911.00	1067.00	19474.00
	Z	-3.197 [a]	-2.992 [a]	-1.963 [a]	-.029 [b]	-.246 [a]	-4.060 [a]
	Asym. Sig. (2-tailed)	**.001**	**.003**	**.050**	.977	.806	**.000**
	Effect size	-.28	-.24	-.18	-.02	-.02	-.16
MLAT-EC 3 / MLAT-ES 3	T	693.50	387.50	750.00	506.50	794.50	17703.50
	Z	-1.256 [a]	-4.137 [a]	-.557 [b]	-1.267 [b]	-.890 [b]	-1.643 [a]
	Asym. Sig. (2-tailed)	.209	**.000**	.578	.205	.904	.100
	Effect size	-.11	-.34	-.05	-.12	-.08	-.06
MLAT-EC 4 / MLAT-ES 4	T	590.00	458.00	210.00	274.00	541.50	10614.00
	Z	-1.318 [a]	-3.514 [a]	-2.153 [a]	-1.185 [b]	-.240 [a]	-3.228 [a]
	Asym. Sig. (2-tailed)	.188	**.000**	**.031**	.236	.810	**.001**
	Effect size	-.12	-.29	-.20	-.11	-.02	-.13
MLAT-EC total / MLAT-ES total	T	480.00	190.00	402.00	600.00	853.00	12078.00
	Z	-3.996 [a]	-6.522 [a]	-3.373 [a]	-2.319 [a]	-1.787 [a]	-8.500 [a]
	Asym. Sig. (2-tailed)	**.000**	**.000**	**.001**	**.020**	**.045**	**.000**
	Effect size	-.35	**-.53**	-.32	-.21	-.15	-.33
Group 2 EC-ES	Statistics	Gr. 3 N=43	Gr. 4 N=41	Gr. 5 N=45	Gr. 6 N=50	Gr. 7 N=64	All grades N=243
MLAT-ES 1 / MLAT-EC 1	T	313.00	217.00	262.50	164.50	230.00	6081.50
	Z	-1.734 [a]	-1.383 [a]	-.324	-4.239 [a]	-4.078 [a]	-5.444 [a]
	Asym. Sig. (2-tailed)	.083	**.020**	.167	**.000**	**.000**	**.000**
	Effect size	-.19	-.15	-.03	-.42	-.36	-.25
MLAT-ES 2 / MLAT-EC 2	T	273.50	301.50	324.00	100.00	335.50	6410.50
	Z	-2.233 [a]	-1.003 [a]	-9.26 [a]	-4.290 [a]	-3.788 [a]	-5.786 [a]
	Asym. Sig. (2-tailed)	**.026**	.316	.354	**.000**	**.000**	**.000**
	Effect size	-.24	-.11	-.04	-.43	-.34	-.26
MLAT-ES 3 / MLAT-EC 3	T	282.00	303.00	246.00	106.50	195.00	6651.00
	Z	-1.927 [a]	-.981 [b]	-1.373 [a]	-4.445 [a]	-4.729 [a]	-5.200 [a]
	Asym. Sig. (2-tailed)	.054	.326	.170	**.000**	**.000**	**.000**
	Effect size	-.21	-.11	-.15	-.45	-.42	-.24
MLAT-ES 4 / MLAT-EC 4	T	193.50	82.00	149.00	146.50	141.50	3424.00
	Z	-3.076 [a]	-3.414 [a]	-1.964 [a]	-3.104	-1.651 [a]	-6.002 [a]
	Asym. Sig. (2-tailed)	**.002**	**.001**	**.050**	**.002**	.099	**.000**
	Effect size	-.33	-.38	-.21	-.31	-.15	-.28
MLAT-ES total / MLAT-EC total	T	201.50	274.00	235.50	42.50	171.00	5033.50
	Z	-3.127 [a]	-1.620 [a]	-2.349 [a]	-5.597 [a]	-5.816 [a]	-8.350 [a]
	Asym. Sig. (2-tailed)	**.002**	.105	**.019**	**.000**	**.000**	**.000**
	Effect size	-.34	-.18	-.25	**-.60**	**-.52**	-.38

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

The results were overall significantly different although the distribution of these significant differences is not the same in all the parts and grades separately. The increase in the first subtest tends to be significant in all grades (it is not for grade 3 and 5 in group 2). The reason for this increase could certainly be the novelty of this activity, as students are not used to deciphering words and matching them with their synonyms.

What they are perhaps more used to is to just the first step required in this part, that is, decoding shortened words in, for example, text messages. The way significant values are distributed as for part 2 is disconcerting, as in part 2 significant values are found mainly in the lower grades of group 1 and in the upper grades of group 2. A similar phenomenon happens in part 3, where no significant values are found in group 1 except in grade 4, while in group 2 significant differences appear in the upper grades. Part-4 changes in the means in the second sitting are also inconsistent across grades comparing both groups, as in group 2 significant values are found in all grades but 7, while in group 1 the difference was only significant in grades 4 and 5.

When looking at the raw total score, more similarities between groups 1 and 2 are found when considering all the subjects as one cohort (except in part 3 of group 1) and also across grades. In group 1, the scores on the MLAT-EC were significantly higher than on the MLAT-ES across grades and the opposite was true for group 2 (scores on the MLAT-ES higher than on the MLAT-EC), although the difference was not significant for 4-graders (*Mdn* MLAT-ES=77.24, *Mdn* MLAT-EC=77.31), *T*=8, *p*<.05, *r*=-.18.

Since the Wilcoxon Signed-Rank Test shows there is a task-training effect between tests, from now on the only reference for aptitude taken into account will be the score on the aptitude measure taken in the first place.

3.5.5. Standardisation and norms of the MLAT-ES and the MLAT-EC administered in Catalonia as compared to those of the MLAT-E and MLAT-ES in the *Manuals*

The aim of piloting these aptitude tests is to get to establish standardised norms so that they can serve as a reference for anybody using the tests to interpret what a score means in their own sample. Following are the tables displaying the percentile norms for raw Total Scores of the MLAT-ES and MLAT-EC by grade along with the means and standard deviations. The percentile norms of part scores are in appendix O.

Table 3.97. Norms for students in grades 3, 4, 5, 6 and 7 on the MLAT-ES, Total Score. Raw total scores corresponding to designated percentiles

PERCENTILE	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7
99	108-123	118-123	120-123	120-123	122-123
97	102-107	116-117	119		
95	98-101	114-115			121
93	93-97	110-113	117-118	119	
90	91-92	109	116		120
87	89-90		115		118-119
84	88	105-108	113-114	118	
81	87	102-104	111-112		117
78	82-87	99-101	108-110		
75	79-81	98	107	117	116
72	77-78	96-97	105-106	116	
69	75-76	95	104	115	115
66	74	94	103	113-114	
63	70-73	92-93	102	112	114
60	68-69	90-91	100-101	110-111	
57	66-67	88-89	99	109	113
54	65	85-87	98	107-108	111-112
51		83-84	97-98	105-106	
48	64	80-82	92-96	103-104	110
45	61-63	78-79	91	101-102	
42	60	77	90	100	108-109
39	59	75-76	89	98-99	107-108
36	56-58	74	85-88	96-97	106
33	55	70-73	83-84	94-95	105
30	52-54	69	82	91-93	103-104
27	50-51	67-68	79-81	87-90	100-102
24	48-49	66	68-78	86	97-99
21	45-47	62-65	63-67	83-85	96
18	42-44	59-61	62	79-82	94-95
15	38-41	51-58	61	74-78	89-93
12	36-37	49-50		71-73	87-88
9	33-35	43-48	54-60	61-70	86
6	27-33	40-42	46-53	51-60	76-85
3	22-26	24-39	42-45	49-50	50-75
1	0-21	0-24	0-41	0-48	0-49
N	66	75	57	60	67
Mean	65.09	81.84	91.12	99.93	106.69
SD	21.09	22.16	21.12	19.04	13.24

Table 3.98. Norms for students in grades 3, 4, 5, 6 and 7 on the MLAT-EC, Total Score. Raw total scores corresponding to designated percentiles

PERCENTILE	GRADE 3	GRADE 4	GRADE 5	GRADE 6	GRADE 7
99	103-122	120-122	119-122	119-122	120-122
97	95-102	117-119	117	118	119
95	94	116		117	
93	90-93		116	115-116	118
90	84-89	110-115	115		
87	80-83		113-114	113-114	117
84	78-79	108-109	111-112		
81	76-77	106-107		111-112	
78	75	103-105	110	110	116
75	69-74	99-102	109	108-109	
72	66-68	97-98	108	107	115
69	65	96	107	106	
66		93-95	106		112-114
63	62-64	91-92	104-105	105	110
60	60-61	90	102-103	102-104	109
57	55-59	89	100-101	100-101	107-108
54	54	86-88	98-99	99	
51	51-53	85	94-97		106
48	50	84	93	97-98	104-105
45	48-49	81-83	92	94-96	101-103
42	47	80	90-91	90-93	98-100
39	45-46	78-79	87-88	89	97
36	44	72-77	85-86	85-88	95-96
33	43	70-71	83-84	84	94
30	39-42		82		93
27	35-38	68-69	80-81	81-83	92
24	34	66-67	79	79-80	89-91
21	32-33	65	75-78	77-78	87-88
18		60-64	76-74	74-76	86
15	30-31	56-59	71-72	72-73	74-85
12	26-29	51-55	67-70	70-71	66-73
9	23-25	47-51	63-66	66-69	63-65
6	22	39-46	54-62	56-65	58-62
3	20-21		39-53	50-55	42-57
1	0-19	0-38	0-38	0-49	0-41
N	57	62	61	60	64
Mean	55.12	83.77	93.13	94.20	99.95
SD	22.75	21.77	18.44	17.32	18.37

It is difficult to compare the percentiles of the MLAT-E (see section 2.3.1.2) and the MLAT-ES due to the difference in the quantity of items in both tests (130 items in the English version and 123 in the Spanish one). It is not so difficult, however, to compare the percentiles of the MLAT-ES that appear in the *Manual* (see section 2.3.2.2) with those obtained with the data of this study. The latter are consistently much higher in all grades, especially in the lower interquartile range from grade 3 to 5.

As compared to the percentiles of the MLAT-ES, the percentiles of the MLAT-EC tend to be lower in all the interquartile ranges except for the lowest one in grade 4, which starts with a difference of 14 points, although this difference disappears towards the 25% range. The largest differences in the percentile ranges are found in grade 3, but the rest of percentile ranges are fairly similar.

The norms tables for the raw Total Scores on the MLAT-ES and the MLAT-EC by grade presented here are not generalisable to the whole bilingual Catalan and Spanish community due to the reduced sample. Yet these norms can help as an initial reference for score interpretation. Means and standard deviations for the normative sample by grade are also shown at the bottom of each table for the sake of comparison with the norms of the MLAT-E and the MLAT-ES provided in the *Manuals.*

If we compare the mean raw total scores on the norming study of the MLAT-ES with the norms table of the MLAT-E, at first sight one sees that the means are lower than the ones on the MLAT-E of the norming sample even if there were more items and so the range of scores was wider. Score evolution follows the same pattern in both cases: a sharper increase between grades 3 and 4 (slightly smoother on the MLAT-ES) and not so sharp in the other grades, although in the *Manual* it is said that "mean total scores increase in a uniform way from one grade level to the next, while the standard deviation and standard error of measurement decrease for each grade level" (Stansfield & Reed, 2005:14). Besides, the tendency is for the raw total score to increase throughout all grades, but not at such a high rate between grades 6 and 7.

Figure 3.12 presents the mean raw total scores of the norms in the *Manuals* and of the cohort of this study corresponding to designated percentiles. Participants are distributed as follows:

Table 3.99. Number of participants included in the raw total scores corresponding to designated percentiles

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
MLAT-E boys	493	505	495	670	-
MLAT-E girls	528	510	500	640	-
MLAT-ES *Manual*	207	206	289	306	178
MLAT-ES Group 1	66	75	57	60	67
MLAT-EC Group 2	57	62	61	60	64

Figure 3.12. Norms for students on the MLAT-E (Manual), the MLAT-ES (Manual) and the MLAT-ES and MLAT-EC administered in Catalonia. Raw total scores corresponding to designated percentiles
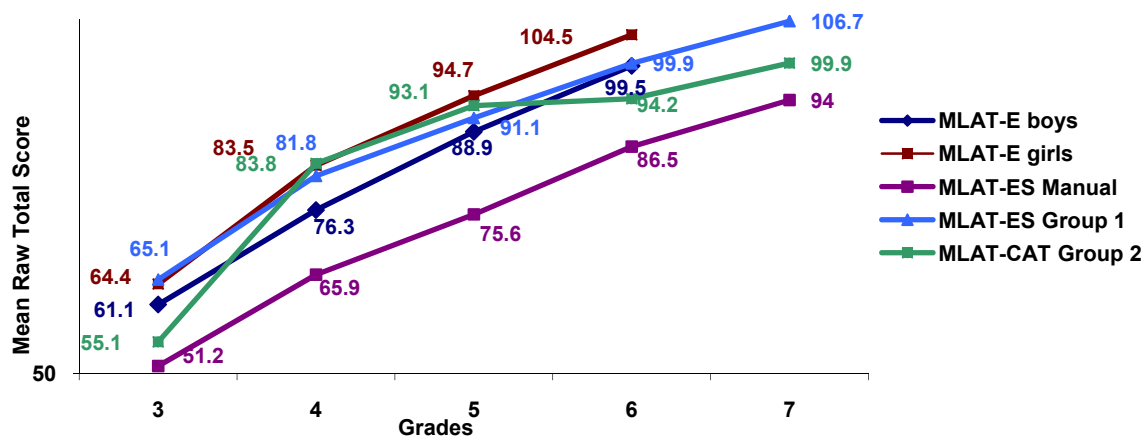
Regarding the scores on the MLAT-ES administered in Catalonia, they are markedly higher than those exposed in the norming study, but the pattern across grades is almost identical in both tests. Therefore, it could be said that the MLAT-ES administered in Catalonia presents a similar behaviour in both populations across grades though at a higher score level. The evolution of scores across grades on the MLAT-EC is not so regular as is on the MLAT-ES, although, on the whole, the tendency is similar: a large increase between grades 3 and 4 (the largest of all in this figure), and not so large an increase across the other grades. Actually, between grades 5 and 7 the range between the raw total scores in the grades is the lowest of all.

In conclusion, all the versions of the elementary MLAT present similar responses in the population regarding score evolution, which allows to claim that they are more or less equivalent tests. However, this is not completely true as far as the raw score itself obtained in each version is concerned. That is to say, differences are observed in the means depending on the context in which the test was administered and normed. As for the normed versions of the MLAT-E and the MLAT-ES, the language in which the test is in could play a role as well as the fact that almost 40 years have passed between the norming study of the two versions of the test. This variable is not relevant regarding the norming study of the MLAT-ES and the results obtained in Catalonia, as no more than five years had passed between the norming study and the administration of the test in Catalonia. Therefore, other variables such as the context or the bilingual condition of the Catalan subjects are factors to take into consideration. Finally, the MLAT-EC takes off with lower results in grade 3 than the MLAT-ES, then catches up with the MLAT-ES and, in grades 6 and 7, once again the results are a little lower than those obtained by the norming sample on the MLAT-ES.

As the context of administration was the same, these differences across grades between the MLAT-ES and the MLAT-EC should be explored in terms of the inner difficulty of each part of the MLAT-EC depending on the extent to which it can be considered language-based and depending on the test takers' linguistic preference.

3.6. English proficiency measures

In addition to the MLAT-E in Spanish and Catalan, the subjects were administered a battery of instruments in L2 English. These instruments were different in grades 3 and 4 from those used in grades 5 to 7. This is so because item types and tests had to be adapted to the test takers' age and FL knowledge. Shorter texts had to be used for the lower grades, as short immediate-memory span and attention are characteristics of young children. Although the subjects were used to dictations and compositions in their L1, they were not used to these in a FL. They were not used to filling in the gaps in cloze passages or to listening comprehension tasks in a multiple-choice form either in English or in their L1s.

3.6.1. Instruments used in Grades 3 and 4

The measures used in this study taken by children in grades 3 and 4 are two FL proficiency tests, a multiple-choice listening test and a cloze passage. Children this age are familiar with speaking and listening tasks in the FL classroom, but not so much with writing. They were nevertheless asked to fill in the cloze passage because it is considered one of the most integrative kinds of FL proficiency measures. The listening task in a multiple-choice format that participants in grade 3 and 4 took is the same used in the Barcelona Age Factor (BAF) project coordinated by Dr Carmen Muñoz, but it had never been used before for grades lower than grade 5. It originally consists of twenty-one items recorded on tape and the response alternatives are shown in pictorial form. Since items are arranged in an ascending order of difficulty, only the first eleven items were included in the form distributed to these grades. The reliability of this test with eleven items was very low as measured by the Cronbach's alpha index. After removing the two most deficient items, the reliability was still low (Cronbach's alpha .564). The

reason for this low reliability index could be the reduced number of items in the test. Consequently, it was decided to disregard this test in the main study[16]

3.6.1.1. Cloze passage in English

The text of the cloze passage was related to the *Little Red Riding Hood* tale. The subjects were asked to supply seven words represented by blanks in six sentences. The test included coloured pictures so that it was more appealing to the subjects and the text could be contextualised. This type of test was chosen because it involves not only receptive skills but also productive skills.

Although one common procedure to design these tests is deleting approximately one word every seven in a prose passage, this was not the procedure followed in this case. The words deleted were chosen assuming that they would be known by the subjects and they bore a relationship with the picture provided next to the text. Item deletion is actually a controversial topic in SLA studies, as it has been discussed whether the words deleted should be chosen on a rational basis (e.g. only function words, only nouns). In this way, we get three different types of cloze: syntactic, cohesive and strategic, depending on the kind of items deleted (Bachman, 1982). The words deleted in this text are mostly nouns, except for two verbs, both of which were full verbs, so it could be considered a test based on vocabulary knowledge.

As children this age are not used to writing much in English, all those items that were misspelled were considered correct. The reliability of this test was very good for both grades (Cronbach's alpha .732 and .757 for grade 3 and 4 respectively, and .746 for both groups) so it was not necessary to remove any item from the final count.

3.6.2. Instruments used in Grade 5, 6 and 7

Children in grades 5, 6 and 7 took three FL proficiency tests, a listening multiple-choice test, a dictation and a cloze passage. They all took the same tests, but the cloze passage was shorter for grade 5.

---

[16] The descriptive information and further statistical analyses that include the listening administered in grades 3 and 4 can be seen in appendix P.

3.6.2.1. Listening in English

The listening task used for grades 5, 6 and 7 is the complete version of the test used in the BAF project. This test originally consists of 21 items, but after obtaining the Cronbach's alpha reliability index and discrimination and IF tests, it was decided to leave it with nine items, thus rendering the test with an acceptable reliability index (Cronbach's alpha .726).

3.6.2.2. Dictation in English

The dictation used in the BAF project was also used in this study. It involves a taped descriptive test of 50 words distributed in short sentences. The text is played once without stops, then one more time bit by bit. Each bit is repeated twice with short time lapse in-between bits. Finally, the whole text is read again once more by the speaker.

The dictation had an excellent reliability for all grades (Cronbach's alpha .938). Apart from the fact that this test also proved to be a reliable measure in the BAF project, it was used because dictation texts are units with complete meaning and, as Burstall advises, "each item should be in the form of a complete and meaningful utterance (as) isolated words or incomplete phrases do not make satisfactory items and can only provide a partial indication of a child's ability to understand the target language" (1971:156). In this test misspelled items were considered incorrect, as accuracy was considered important for this part, as opposed to the cloze passage. In addition, "spelling ability includes at least implicit knowledge of conventional spelling rules and phoneme-grapheme correspondences" (Carroll, 1993:170).

3.6.2.3. Cloze passage in English

The cloze passage involved a text whose topic is the retelling of the *Little Red Riding Hood* tale. It consists of two parts in which there were 25 items initially. Validity and reliability tests, however, made it advisable to remove one item, thus rendering it with 24 items and a Cronbach's alpha reliability index of .861 and .862 for grades 6 and 7 respectively, and .867 for both groups together. Participants in grade 5 only had to complete the first part of the cloze passage, which consisted of 11 items. The

Cronbach's alpha reliability of this part was .694, which is very close to the value of a good reliability index.

3.6.3. Teachers' and students' criterion variables

The criterion variables of this study were obtained by the same means they had for the validation of the MLAT-E and the MLAT-ES. The teachers of the schools participating in the study were asked to fill out a form (see appendix Q) for each student in which they had to score their students' achievement in EFL. Only teachers from two schools filled out these forms. Hence the data available are for only 209 participants in this study. Moreover, the teachers did not answer all the aspects in the form, as the instructions on it explicitly indicate that they do not need to answer the last two questions, which are related to general proficiency and the mark the students will obtain at the end of the course. Consequently, the N for these measures varies depending on the criterion measure.

On the cover of the test booklet, the test takers were also asked about the mark they had obtained in EFL the previous year and the mark they thought they would obtain the same course when they took the tests in this dissertation. The marks of the previous year are to be taken with caution because they may not be reliable enough. This is because many students said they did not remember exactly what their mark was and some decided to leave this question blank. Moreover, some of them had received a mark that is very general and difficult to compute for our purposes because of its dichotomous nature: either "*puede mejorar*" (he/she can improve) or "*progresa adecuadamente*" (he/she makes proper progress). As a consequence, the self-reported grades of the previous years were not used in this study. Regarding the marks that appear as an answer to the question "*Si estudias inglés ahora, ¿qué notas esperas sacar?*", these were not very reliable due to the ambiguity of the verb "*esperar*". In addition, while some students answered in numbers the mark they expected to obtain at the end of the course, some filled the blank with "*puede mejorar*" or "*progresa adecuadamente*" and some others left this gap blank too. Consequently, the N for this criterion measures are also smaller than for other measures.

3.7. Procedure

3.7.1. Administration of test batteries

The test batteries were administered by the author of this dissertation and by other experienced research assistants of the University of Barcelona. Although the subjects' teachers were encouraged to stay in the classroom while the tests were being administered so as to reduce test-anxiety, not all of them did so. Only the teachers who foresaw some kind of disruptive behaviour during the test administration stayed in the classroom.

The series of tests was administered on three different days, during one hour or less each: the first two sessions took place on two consecutive days or with a span of time difference of three or four days, whereas the third session took place approximately three months later (February – May). For some grade-7 groups, the first two sessions were merged into one. The tests were usually administered in the morning, except for some grade 7 sessions, which took place in the afternoon.

The aptitude measures used in this dissertation are the elementary versions of the MLAT in Catalan and Spanish. As for the FL proficiency measures, the participants had to complete a cloze passage that changed depending on the grade, a dictation and a listening task. They also did some other tasks on the same day the tasks mentioned so far were administered. These were a composition in Catalan and in Spanish entitled "*El dia més feliç de la meva vida*" and "*El día más feliz de mi vida*" ("The happiest day of my life" respectively, a composition in English ("My life: past, present and future") and the oral tasks in English used in the BAF Project, that is, an interview, a story and a role-play (Muñoz, 2006a). A counterbalanced design with two equivalent groups of test takers was followed regarding the versions of the elementary MLAT in order to see if it was possible to estimate equivalence reliability. On the first day, the students had to fill out one of the versions of the elementary MLAT (either the Spanish or the Catalan one) and, if there was still time, they were also told to write the composition in Spanish or Catalan and to fill in the cloze passage in Catalan or Spanish accordingly. The following day or no later than a week after the first battery of tests, the English tests were administered. Three or four months later, when enough time had passed so as to hopefully avoid the task-training effect, the remaining version of the elementary MLAT was administered together with the remaining tests in the L1s. The oral tasks in English were performed on the first two sessions of data collection. A sub-set of each school participated in these tasks.

273

Regarding test organisation, the subjects were given the test booklets at the beginning of the test session and were informed that they were going to take a series of timed tests so that they did not worry too much about the amount of pages. Although the order of presentation of the subtests was most of the times maintained (see above), sometimes the sequence of parts was disrupted because of unexpected events, such as technical problems with some cd-players or because of time constraints (when testing had begun later than expected, and some subjects had to leave earlier, for instance).

The written tests were administered to intact groups in class-time under the supervision of the researcher and/or other trainees in the field and, sometimes, the teacher. For the oral tests, the students were called and placed in a separate room where they were informally interviewed. The participants were called at random if there was no time to interview them all and in alphabetical order if there was time enough to collect data from all of them.

The time allotted for each test was sufficient for the upper grades (some groups even finished before time had ticked out). However, the lower graders sometimes did not manage to finish some subtests of the aptitude tests. As for the compositions, the allotted time was from eight to nine minutes for each composition. However, taking into account that lower grade subjects are very young and are not used to writing for so long, if they were not willing to write for the time allotted, they were encouraged to write a minimum of 10 lines in the compositions in Catalan and Spanish and a minimum from 5 to 10 lines in the composition in English (depending on the subjects' grade).

3.7.2. Instructions

All the instructions for each test were given either in Catalan or Spanish, the subjects' L1s, so as to avoid misunderstandings, as the test administrators assumed the subjects' proficiency in English was not high enough to understand the procedure of taking the tests. If necessary, the blackboard was used to supplement the standard instructions. The directions to the oral tasks in English were given in English for students from grade 5 on, but in Catalan or Spanish for the lower grades if the researchers saw they were not being understood properly by the participants.

All the instructions as well as the specification of procedures and tasks were presented both in aural and visual channel. In most cases, both aptitude tests were presented with "canned" human input, that is, instructions were taped on a CD for both

tests. However, because of time constraints, on some occasions the instructions of some sub-tests in the elementary MLATs were presented in "live" human input. Then, the administrator of the tests controlled the timing with a stopwatch.

3.7.3. Explicitness of criteria for correctness

The subjects were asked to answer as many items as they could in all the tests. However, in the particular case of the elementary MLAT, if the subjects faced an item whose answer they did not know, they were explicitly encouraged to leave it blank rather than answer it at random or spend their time stuck in a particular item. The tests which had been answered clearly at random were eliminated from the sample. These were tests that either had systematically more than one answer per item or cases in which the test taker had finished the test in a few seconds, which is much less time than needed. Regarding the criteria of correction of the cloze passages, spelling mistakes were not taken into consideration if the answer was understandable. This criterion was adopted because children in grades 3 and 4 are not used to writing in English but to reading, speaking and listening. Also, several answers for the same item were considered to be correct as long as they were acceptable for any native speaker (Bachman, 1990). Another point to take into account when dealing with young populations is the fact that young children have great difficulty in complying with the norms of the tests. Consequently, if the cross of the answers on the elementary MLAT were not, for example, inside the checkbox but next to it consistently along the whole test, they were considered correct. They were not if the participant misunderstood the type of exercise and instead of ticking a checkbox they wrote the answers in letters.

It should be taken into account that, according to their teachers of English, none of the subjects were used to writing compositions in English, to completing cloze passages or to writing dictations in languages other than Catalan or Spanish. However, some students attending extra-curricular English classes may have been familiar with these kinds of tasks as well as with oral task performance.

3.8. Summary of Chapter 3

In this chapter, the research context and the participants of this dissertation have been described. Their most distinctive features are, on the one hand, the fact that the context of this study is bilingual, which supposes some questions that were perhaps not taken into account in the norming study of the MLAT-ES. On the other hand, the languages spoken in the context of this study are of a Romance origin, not Germanic, as is English, the language in which the original MLAT-E was written. Consequently, although this measure had already been validated, further exploration of the functioning of the items it consists of was considered worth doing.

The content validity of the MLAT-ES presents some divergences regarding the final version of the MLAT-ES that was published in 2005. That is to say, some of the items that, according to the statistical analysis (IF, discrimination and reliability), should be removed do not coincide with those removed in the published version. In addition, each part presents some features that may alter the item's functioning. For instance, some items in *Parte 1 Palabras ocultas* present extra difficulty for 3-graders, regarding both the font type used and the vocabulary of some items. *Parte 2 Palabras que se corresponden* presents some problems as well regarding, mainly, sentence word order and linguistic and cultural questions in the items' wording. In *Parte 3 Palabras que riman* the fact that Spanish has a transparent orthography may alter the construct this part was meant to measure when it was first designed in English. In addition, some items were designed taking into account phonemic phenomena that are present in the American varieties of Spanish but not in the Peninsular variety, such as *ceceo* and *seseo*. Finally, *Parte 4 Aprendamos números* is extremely easy from grade 5 on, which could be due to the strategies used when giving names to numbers, that is, using content words such as "*rasca*" and what could be analysed as the suffix *–ca.* The MLAT-EC has also presented some problems in the functioning of some of its items. While the troubles in *Part 1 Paraules ocultes, Part 2 Paraules que es corresponen* and *Part 4 Aprenguem números* are more or less the same as in the MLAT-ES, they are not in *Part 3*, which is the most language-specific part. In this case, the problems are due to the use of open and closed vowels that may not be as distinct as it was first assumed, even for native speakers.

The reliability, IF and intercorrelations of parts of both tests are very similar, which allows us, up to a certain extent, to say that they are quasi-equivalent tests, not exactly the same, as proved by the moderate - high correlations between them. Descriptive statistics also show that both the MLAT-ES and the MLAT-EC function in

the same way across grades, although the MLAT-EC is slightly more difficult than the MLAT-ES. When these two tests are compared to the norming study of the MLAT-ES and of the MLAT-E, it has been seen that they also behave in a similar way although the analysis of the content validity of the Spanish and Catalan versions should not be disregarded.

This chapter finishes with the description of the English proficiency measures used and how these and the aptitude tests were administered and corrected.