

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

Maria del Mar Suárez Vilagran

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

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

Tesi doctoral presentada per

**Maria del Mar Suárez Vilagran**

com a requeriment per a l'obtenció del títol de

**Doctora en Filologia Anglesa**

Programa de Doctorat *Lingüística Aplicada*

Bienni 2002-2004

Departament de Filologia Anglesa i Alemanya

Universitat de Barcelona

Barcelona, 2010

Directora: **Dra. Carmen Muñoz Lahoz**

# **CHAPTER 6: IMPLICATIONS, CONCLUSIONS AND FURTHER RESEARCH**

## **6.1. Introduction**

This last chapter will look, first of all, at the implications and limitations that the discussion of the results found in this study regarding, mainly, two aspects: the aptitude measures used and the characteristics of the population to whom they were administered. In the second part of this chapter some flaws of this study related to variables that could not be controlled will be pointed out and further research will be suggested that deals with aptitude in young learners, either using the aptitude tests used in this study or other instruments.

## **6.2. Implications regarding the language aptitude measures for Spanish/Catalan bilingual young learners**

The main issue of this dissertation was the validation of two versions of the same aptitude test, the elementary version of the MLAT, for young learners. These are the Spanish and Catalan versions of the MLAT-E. Research into aptitude in young learners is relatively recent. Consequently, there is a clear lack of valid measures that can be used to predict aptitude in this population. The translation and adaptation of the English MLAT-E to Spanish is a valid and very reliable version of the classical test, as presented in section 2.3. The standardisation and norms published of the Spanish version are supported by statistical results, which are, up to a certain extent, similar to those of the original English version. One must not forget, though, that Spanish was the only language considered when adapting the MLAT-E and that the target subjects of the standardisation process were speakers most of whom had Spanish as their main L1. In this study, however, the test was administered to Catalan/Spanish bilinguals, some of whom showed preference for Spanish, for Catalan or for both languages in the same degree. If item constructing and writing is a difficult task in itself, it is even more difficult to create appropriate, correct and exact items if the item writer, for whatever reason, does not take into account the fact that some test takers may have more than one L1, as is the case of the participants in this study. Consequently, some of the items

in the MLAT-ES do not work as expected (or as they do for monolingual Spanish speakers or speakers of a Spanish regional variety other than the Peninsular).

Besides the subjects' bilingualism, the regional variety they speak is also a factor to be taken into account, especially as for the functioning of those items which involve some aspect exclusive of a particular regional variety of the Spanish language. That is, there are "systematic differences in test performance that are the result of differences in individual characteristics, other than the ability being tested, of test takers" (Alderson, Clapham & Wall, 1995:271). Nevertheless, despite there being linguistic issues regarding the construction of items of the MLAT-ES version, they do not cause much trouble in terms of scores other than the higher results of the participants of this study, as seen in the comparison of results of the published version of the MLAT-ES and the one resulting from the validation of this study. One of the reasons could be that the test has enough items to neutralise the impact of the "problematic" items. Another reason could be the fact that when taking any test, many variables come into play, and crosslinguistic influence is only one of them which may yield to others. Actually, cultural issues also arose when looking at the items from a qualitative perspective, as most of the subjects who participated in the main validation study were Latin American, while the participants of this study are Catalan.

On this same note, in this study (and for that matter, in any other study comparing monolinguals and bilinguals), bilingualism has proved to be a relevant issue. From the results obtained, it can be concluded that, if the MLAT-ES is to be administered in bilingual communities, it should not mingle monolinguals and early bilinguals, unless the objective of the study is to identify anchor items to be used in tests to be taken only by monolinguals. In that case, the test could contain any item without taking into account languages other than the one the test is written in. If it is to be taken by bilinguals, consequently, the test should not contain any of those items with different functioning depending on the language spoken in the community. Anchoring items for the construction of the MLAT-EC to be taken by monolinguals would make no sense, though, as it is practically impossible to find pure Catalan monolinguals.

The study had three main research questions. The first one tackled the issue of equal validity of the aptitude tests used in children from grades 3 to 7. The second one asked whether performance by boys was significantly different from that of girls. The third and last research question aimed at checking the concurrent construct validity of the test. In what follows, the main conclusions regarding these three questions will be exposed.

As for the first research question, it was observed that the tests functioned, as a whole, in a different way for 3-graders than for the rest of grades. Several possible factors that could have had an influence on 3-graders' performance, most of them related to their age, were exposed. 3-graders are just at the beginning of their concrete operational stage, which limits their cognitive process. From an information-processing perspective, children at this stage still have to develop problem-solving strategies as well as strategies for encoding the information they receive and for memorising it. They still lack some of the memory categorisation and storage strategies that older children already master. Moreover, they have become literate relatively recently and show to have lower metalinguistic awareness than their older counterparts.

Below is a summary of how each MLAT-ES and MLAT-EC part seemed to pose some kind of extra difficulty for 3-graders. Other general validity issues are also commented.

Part 1 does not show consistently concurrent construct validity with the proficiency measures used in this study. Issues such as the font type or the vocabulary used in some of the distractors have already been mentioned and, perhaps, should be revised in future administrations of the tests, as they seem to pose some extra difficulty for the lower grades, especially for grade 3, due to the use of some vocabulary as well as of the font type in which the tests are.

As regards part 2 *Palabras que se corresponden / Paraules que es corresponen*, regardless of the grade, the words chosen the most are content words. This response pattern implies that the participants, including the youngest ones, have the grammatical sensitivity needed to distinguish content from function words. Therefore, it may not be necessary to place a checkbox below any of the function words. It was also noticed that many of those who fail to choose the right word matching the function of the word in capitals in the stem sentence choose the word that appears approximately below the word in capitals. These happened mainly in the lower grades. Choosing the words according to their position is a behaviour that should be further explored because, if the position of the matching words coincides, the test taker scores on this item, although this does not necessarily mean that he or she is aware of the function of the words. In addition, the introduction of formal grammar instruction at school may also have affected the performance on this part, so this is a factor that should be controlled.

Part 3 is totally original, i.e. it does not derive from the MLAT for adults. It has proved to be one of the parts which correlates strongest with the concurrent construct validity measures used in this study. However, the way some of the items have been

designed has been questioned repeatedly due to the fact that both Spanish and Catalan, as opposed to English, are languages with a transparent orthography. In fact, most of the items removed from the original version of the piloting test included the features suggested in section 3.5.2.2 as possible threats to the construct validity of this part. One possible way to measure the ability tapped by part 3, the ability to hear speech sounds, could be designing an oral test, although the time needed to administer it would certainly increase.

As for Part 4, Carroll (1990) himself admits that it is usually skewed negatively, that is, it is a part that tends to be very easy. In the Spanish and Catalan version, several reasons for this facility, already from grade 5, were suggested in the discussion of the first research question, including the fact that content words were used as names for numbers and then *-ca* could have been interpreted by some test takers as a suffix. This would have alleviated their memory load when encoding the numbers and retrieving them. A suggestion for further research could be a retrospective think aloud to see the strategies used by test takers in Part 4 Number Learning, as it proved to be a very easy part already from grade 5. Actually, perhaps what makes the performance by the lower grades different from those in the upper grades is the memory strategies that children used in order to encode and later on retrieve the numbers learned.

The scarce research into aptitude published which uses the MLAT-E or a derivate form considers the total aptitude score, without taking into account the tests administered part per part. It would be interesting, not only in the studies already published but also in this one, to see which factors tapped by the MLAT-E or any other aptitude measure for young learners relate to each skill individually (writing, listening, speaking and reading) with objective proficiency measures and not just FL proficiency in general. Focusing on the MLAT-E and its versions, what each subtest is meant to measure has not been clearly stated, especially as for part 3 Rhyming Words (and its equivalents). Secondly, depending on the language versions of the MLAT-E tests, each subtest has a different number of items and this is not corrected in any way by means of ponderation. Instead, the total score of the test is the arithmetic sum of the subtests. No explanation is included in the *Manuals* that explains this decision either. Another problem related to the interpretation of correlations between aptitude scores and FL proficiency is the fact that scores in some subtests, especially *Parte 4 Aprendamos números*, seem to reach a plateau from grade 5 on, while proficiency in English does not. FL input could be a variable that could help to explain this issue. Therefore, it would be desirable were we, by means of aptitude tests, able to not only predict success in language instruction but also to compare actual achievement with the

achievement expected from the results in aptitude tests. If we assume that aptitude tests are reliable measures of achievement, the factors that make the actual achievement lower or higher should be identified. Certainly, these conclusions have implications in teaching methodology for young learners, as depending on the diagnostic results, it should be easier to identify individual learning styles so that pedagogical materials can be created accordingly.

The sex variable has also been explored in the second research question. The conclusions drawn from the results obtained are that, in principle, aptitude does not seem to be affected by the participants' sex. However, it was suggested that perhaps a more cognitive-oriented approach into the way each of the parts of the elementary MLAT is answered would shed light on sex differences, if any. These should be explored along with different proficiency measures, those that, traditionally, have given an advantage to girls over boys (such as fluency) and those which have favoured boys over girls (e.g. receptive tasks).

As regards the third research question, related to the construct validity of the tests, it has been seen that both the MLAT-ES and the MLAT-EC seem to be valid measures of predictability of general concurrent validity although not in all skills, since the speaking skill, as estimated by the teachers' criterion, sometimes did not correlate with the aptitude measures. Also, correlations in grade 3 were consistently non-existent or lower than those in other grades. This phenomenon could be related to the issues discussed in relation to the participants' age and cognitive development and/or to the validity of the proficiency measures used. Consequently, the MLAT-ES and the MLAT-EC may not be the most appropriate aptitude test for this population or, at least, should be revised if they are to be administered to children this age.

Some researchers defend the idea that language aptitude cannot be trained while some others defend the opposite (see section 1.6.4) It is also believed that previous FL language learning experience may result in higher aptitude for several reasons and that is why it is highly recommended to take aptitude tests before any exposure to FL teaching. This is not the case of the participants in this study, as all of them had received EFL formal instruction before the administration of the aptitude tests. They had started learning EFL at different ages, from pre-school to grade 3. Besides, some participants were also receiving extracurricular English classes in addition to the compulsory classes at school. Not just that, some students also had some CLIL exposure to the FL, which may have some kind of influence on their FL acquisition process. Consequently, both the age of onset and the amount and type of extracurricular exposure to English could have exerted an influence not only on the

aptitude scores but also on the FL proficiency measures. This is, therefore, an issue that deserves further exploration.

Last but not least, the proficiency measures used to explore the construct validity of the MLAT-ES and the MLAT-EC proved to be rather difficult for all learners. Moreover, the type of tests used was relatively new for the students, who were not used to completing cloze passages or to doing dictations in English. Oral data were collected from most participants, so perhaps a deep analysis of these oral data will shed more light on the relationship between FL aptitude in young learners and their true FL proficiency in further research.

### **6.3. Further research and limitations of the present study**

Some formal improvements to the aptitude tests were already commented in section 2.3.4. However, there are still other improvements that could be suggested, not only in relation to the tests themselves but also to the design of this research. Some further research is also suggested in the lines that follow.

This study has presented the follow-up validation of the MLAT-ES in a bilingual context as well as the validation of the MLAT-EC. The main MLAT-ES validation study was carried out by the SLTI foundation using a large population. This study, in contrast, was conducted by an individual researcher, and so the population participating in this study is not so large.

The framework of the item analysis carried out in this dissertation is the Classical Theory of Testing, which allows research into validation, consistency and reliability of a test using small samples. The sample in this dissertation is not too small, but it is not large enough to establish norms with full confidence, which would have allowed the use of the Item Response Theory (IRT) paradigm in all its dimensions and would be more generalisable. This paradigm overcomes the limitations that the classical framework has, among which is, for instance, the fact that the estimations of the tests are sample-dependent. That is, any change in the population has some impact on the reliability of the test. This, however, does not happen when the sample is large and IRT is the paradigm adopted, as this paradigm is based on the probabilities of response patterns, so the population is not as relevant as in the classical paradigm. Nevertheless, response patterns have been tackled when considered pertinent and analysed qualitatively relying, at times, on the author's analysis criteria and some other



times, on the feedback the test takers supplied when taking the tests. This feedback was not computed in any way, but collected in the form of notes that the author of this dissertation took in all the test sittings.

The tables of norms of both the MLAT-ES and the MLAT-EC included in chapter 3 and in appendix O are to be taken only as an initial reference for score interpretation due to the reduced sample used, especially when split into grades. Consequently, in order to obtain a more accurate indication of the average score and the range of scores at each level, a larger sample should be tested. Moreover, the bilingual condition of the participants of this study requires a deeper study into the participants' degree and type of bilingualism so as to establish generalisable local norms.

The raw data of this study included only bilingual subjects and their data was contrasted with that of the population in the published standardisation of the MLAT-ES, in which both monolinguals and bilinguals (the fewer) participated. The raw data of these bilinguals were also included in the MLAT-ES norming study. Consequently, although the results obtained from the analysis do not threaten either the direction of the results or the conclusions reached from them, all the figures used when comparing the population of the official norming study with the population of this study are only indicative of a tendency, but they are not to be taken literally as they are. That is to say, ideally, the analysis comparing Stansfield's and Reed's norming study and the MLAT-ES and MLAT-EC administered in Catalonia should have been run with the data of the norming study containing only the subjects that did not participate in the main norming study. Nevertheless, all the other issues approached in this dissertation are not contaminated by any other data.

Degree of bilingualism and language dominance seem to be an issue that could lead to further investigation into the relationship between these variables and language aptitude. Besides, from the item analysis it was perceived that some items, especially in *Part 3 Paraules que rimen*, needed some improvement, so this is also a matter that should be solved in the future.

In this study all the subjects were Catalan-Spanish bilinguals and all of them have been treated as one single group. Despite being bilinguals from birth, the background questionnaire revealed that they had some language preference over Catalan, Spanish or both of them equally. This information could have been used to divide the population into three subgroups and see whether flagged items for crosslinguistic issues functioned in the same way depending on the participants' language preference. This and also language dominance (being truly balanced bilingual or not) are possible material for further research into aptitude and bilingualism.

In section 3.5.5, when comparing the MLAT-E, the MLAT-ES and the MLAT-EC, it was hinted that while the evolution of scores across grades followed similar patterns regardless of the test version, what made the norming studies different was that the mean raw total scores on the norming study of the MLAT-ES were lower than those of the MLAT-E and of the MLAT-ES and MLAT-EC administered in Catalonia. In appendix S it can also be seen that the mean total scores on the MLAT-E were not so different from those of the population of this study and that, in the case of girls, they scored even higher than the Catalan/Spanish cohort. In terms of bilingualism and formal schooling, the scarce amount of biodata in the *MLAT-E Manual* and the *MLAT-ES Manual* regarding the amount of languages that the population of the norming studies spoke or regarding their sociocultural and educative context makes it very difficult to compare the results of the MLAT-E norming study with those of this study. However, preliminary results (Suárez & Stansfield, 2007) showed that in spite of the inclusion in the norming study of 227 Catalan/Spanish bilingual participants, the raw mean total scores on the MLAT-ES of the bilingual Catalan/Spanish population were remarkably higher than the mean total scores of the MLAT-ES norming study across all grades. Given the advantages of bilinguals over monolinguals when it comes to FL learning (see section 1.6.5) and seeing that descriptive results are consistently higher for the Catalan/Spanish population in all parts of both the MLAT-ES and the MLAT-EC (see section 4.2.3), but that results on the MLAT-EC are similar to those on the MLAT-ES taken by the bilingual participants, it would be interesting to further examine if Catalan/Spanish bilingual speakers show higher aptitude than other communities due to their bilingualism or due to other variables such as their social, educative or cultural context, as the educational systems of the participants' main validation study could have been exerting some kind of influence on the test takers' performance.

L1 native skills, word recognition and literacy in general are believed to be directly related to FL aptitude, as the research by Sparks and colleagues has widely demonstrated. This is an aspect that, though reviewed in the theoretical framework in this volume (see sections 1.2.2.1 and 1.6.1), has not been dealt with in this dissertation. Therefore, this is a possible line of further research to complete the linguistic profile of the participants in this study. In relation to literacy, what has been highlighted several times all through this dissertation is the extent to which literacy seems to affect not only test performance but also L1 and FL acquisition. In addition, the author of this dissertation did not ask the schools participating in this study if any of the children had some problems of dyslexia or reading disabilities. This could have been of use to, for instance, see if the outliers, especially in Part 1, had some kind of

disability. It is also unknown if any of the participants was an early or a late talker, which could also have had some influence on the scores. The students' marks in their L1 language subjects were not asked either. One's ability in their L1 has also been found to be related to FL aptitude, so this is also a variable that could be included in further research.

The criterion variables used in research question 3 did not prove to be completely reliable, as contrasted with Wilcoxon signed-rank tests or contrasting the correlations between the MLAT-ES with the teachers' and the students' estimates of achievement. This was especially the case in grade 3 and 4. Perhaps teachers should be provided with clearer instructions in order to fill in the questionnaires related to their students' measures so as to avoid the apparent mismatches in these measures.

All the participants in this study had been studying English for some time. It is a challenging task for future research to establish some criteria to create groups of students with similar FL exposure, both in terms of quality and quantity. However, it has been defended (e.g. Winke, 2005b) that aptitude measures should be administered before exposure to the FL. Otherwise, it is very difficult to determine whether the results obtained in the aptitude measures are true scores of aptitude or whether they are due to either L1 or FL learning experience. However, as all the participants had received some FL instruction, their initial aptitude is supposed to have affected their FL performance, too.

In order to determine the construct validity of the aptitude measures, correlations were run, as they were in the norming studies. However, it would be interesting to run a stepwise multiple regression analysis to see how much each part separately contributes to concurrent validity. In the same way, further research could also start by classifying the subjects not only according to the grade they are in, but also according to the "type of learners" they are by performing a cluster analysis with the data and thus be able to trace learners' profiles.

Kasper and Rose (2002:277) also highlight the dearth of studies, not only in young learners but also in adults, that deal with language aptitude and interlanguage pragmatics, as "(T)he literature on individual differences in second language acquisition commonly distinguishes such categories as age, gender, language aptitude, personality variables, and cognitive, social, and affective factors. To the extent that interlanguage pragmatics research addresses individual differences at all, it has considered all of these variables, with the exception of aptitude." Consequently, a future line of research could also involve the study of IL pragmatics.

Finally, new FL teaching approaches are arising which entail issues such as the use of ICT tools inside and outside the classroom, the advanced introduction of FL learning, or teaching more than one FL at school, among others. The English version of the elementary MLAT was normed in the mid-sixties, when these socioeducative trends had not appeared yet and, therefore, the context of the norming study was necessarily different from the context in which the MLAT-ES was normed. Therefore, not only the test takers' age itself, but also the new socioeducative trends could affect the results obtained in the elementary version of MLAT if taken in the 21<sup>st</sup> century, more than 40 years after it was first normed.