

MacArthur-Bates Communicative Development Inventories. A Possible Means of Lexical Longitudinal Assessment of Very Young EFL learners

by Cristina Alexandrescu, Liceul Teoretic “Nicolae Balcescu”, Cluj-Napoca

Key words: progress reports, longitudinal tests, standardized tests, young EFL learners, lexical reception, lexical production, MacArthur-Bates Communicative Development Inventories

Abstract

Progress reports are one of the documents beginner Romanian teachers of English need to have included in their portfolios. Unfortunately, measuring progress is not as easy as it looks at first glance, as teachers need a testing tool that can be used longitudinally. Standard tests are easy to use and they have proved quite reliable with older students. Things become more complicated with young learners as the EFL standard tests targeting this age group are less complete and complex. Furthermore, they have hardly been scientifically tested. Under these circumstances, the aim of the present study was to make use of The MacArthur-Bates Communicative Development Inventories (MB-CDIs), a standardized and widely researched first language testing tool, in order to assess children's lexical progress in English as a FL. Accordingly, the receptive and productive vocabularies of a group of 51 young learners acquiring English as a FL in a formal immersion setting were measured across time by means of this tool. Afterwards and by means of the same test, the results were compared against those obtained by a control group made of 15 young learners of similar ages but who were learning English in a regular school setting with less hours of exposure per week. These outcomes were contrasted against those of other studies conducted with the same test, leading us to conclude that indeed the MB-CDIs could yield reliable longitudinal results with young EFL learners and hence could be used when writing the class or students' progress reports.

1. Introduction

Assessing students' progress in a foreign language such as English has become quite an issue in today's Romanian educational system, especially for those teachers who have to obtain their permanent teaching degree. Indeed, among others, the beginner teachers are required to build up a portfolio which needs to contain a students' progress report. At first sight, one might be tempted to provide a picture of student's linguistic development by presenting tables containing the qualifications that a particular student obtained in the different tests, measuring different items across the year. The only problem is that those qualifications are not reliable in measuring progress since they were not obtained by testing the same item across time. Researchers agree that progress can indeed be measured and reported if we use the same testing tool several times across a certain period of time. A possible way to do it is by using standard tests. One of their advantages is that they provide a yardstick against which academic performance could be measured. Moreover, if repeatedly used, they can serve as longitudinal testing instruments that can actually measure student's progress. For instance, with older EFL learners, a score in TOEFEL or IELTS proves one's competence of English first of all by measuring and comparing it against that of other testees, and secondly by quantifying progress (the same person's performance can be measured at different times, as the same test can be taken more than once).

Unfortunately, the standard ELF tests made for young learners are less used and hence, less convincing. Although tools such as the Early Language Listening and Oral Proficiency Assessment ELLOPA (WIDA Consortium, 2004) or the Test of Reception of Grammar TROG 2 (Bishop, 2003) have all been normed for young EFL learners, the number of studies dwelling on their relevance and validity is quite limited. Furthermore, these tests either address a specific population range (ELLOPA targets 5 year olds only), or they are incomplete (TROG 2 only tests comprehension). At this point and if we, indeed, acknowledge the desirability of standardized tests for measuring progress in a foreign language such as English, we either need to create, norm and validate such a tool or make use of standardized tests normed for much younger English monolinguals, which of course cannot provide comparison to the norms but at least might offer a more complete and valid longitudinal picture of our learners' performance in English.

2. MacArthur -Bates Communicative Development Inventories (MB-CDIs)

The MB-CDIs are such a standardized assessment tool that provides information of early language skills of children between 8 and 36 months old by assessing infants' and toddlers' vocabulary sizes in their mother tongue. The original version was first created in American English and then it was adapted to 60 other languages (<http://mb-cdi.stanford.edu/adaptations.html>) The American MB-CDIs have three forms. The first one, *MB-CDI-Words and Gestures*, is used for children between 8 and 18 months old, but can also be used in the case of children with language delays who are older than 18 months. The test measures the production and comprehension of early words as well as symbolic and communicative gestures. It consists of two parts. The first one evaluates the lexical comprehension and production of infants' early words in four different sections, that is, "first signs of understanding", "phrases", "starting to talk" and "vocabulary checklist". This last part consists of a list of 396 typical words in children's vocabulary divided according to different lexical categories. The evaluator is asked to mark whether the child is able to understand the word or if the child both understands and produces it. Separate scores for comprehension and production are calculated out of a total of 396 words. The second part "Actions and Gestures" evaluates infant's receptive and productive language (through gestures)

The second form, *MB-CDI-Words and Sentences*, is addressed to toddlers between 18 and 30 months of age. It evaluates language production and early grammatical constructions. The assessment is performed in two parts. The first part consists of two sections, that is, "words children use" and "how children use words". The second part of the *MB-CDI-Words and Sentences* refers to "Sentences and Grammar" and targets word endings, word forms, examples of sentences the child is able to produce and sentence complexity. The third form, *CDI III*, is quite new and it's used for children between 30 and 37 months. It focuses on lexical production, sentence complexity and language use but it has not been used much in research and therefore its validity could not be checked.

The validity of this instrument was tested through a significant amount of studies carried out with monolingual participants both healthy and language impaired (Galeote et al. 2006; Heilmann et al., 2005; Luyster et al., 2007; Nicholas & Geers, 2008; Mayor & Plunkett, 2010) as well as with simultaneous bilingual learners (David & Wei, 2005; Hoff & Elledge, 2005; Mancilla-Martinez, et al. 2010; Pearson, Fernandez & Oller, 1993; Pearson & Fernandez, 1994; Pearson et al 1997; Pearson and Fernandez , 2001; Poulin-Dubois et al., 2012, 2013; Serrat et al. 2010;).

Given the relevance of this tool with such diverse populations and acknowledging the lack of reliable standards for young EFL learners, the decision was taken to test the applicability of MB-CDIs with 3 to 6-year-old children studying English as a foreign language in formal settings.

3. The Study

3.1 Objectives of the study and research questions

This study aims at seeing if the MB-CDIs yield reliable results with older children who acquire English in a formal immersion setting where the amount of the hours of exposure is less than that in informal immersion settings. At the same time, we also aim to check whether the MB-CDI can provide reliable results when comparing two groups of similar ages but with different hours of exposure to English. If indeed, MB-CDIs prove reliable, then we have reasons to believe they could also serve for drawing students' progress reports. Consequently, our research was split between two major research questions as follows:

RQ1: Can MB-CDI capture lexical progress in the case of EFL learners acquiring English in immersion formal settings, where the amount of hours of exposure is less than in informal settings but much more than in regular school settings?

RQ2: Can MB-CDI compare groups of EFL learners of similar ages of onset but with different hours of exposure to English?

3.2 Methodology

3.2.1 Participants

The two groups of participants in this research were part of a three-year longitudinal study which aimed at analyzing early lexical development in English in terms of lexical comprehension and production. At the beginning of the study, all participants were 4 years old and had been acquiring English for a period of 9 months. They were all selected on the basis of the same criteria, meaning average students with no English native au-pairs and no extracurricular English lessons. The first group, the immersion

school group (IG), consisted of 51 children who had been learning English through formal immersion in an international private school in Barcelona. In this case, the number of hours of exposure to English amounted to 25 hours per week. The lexis acquired was both incidental and also targeted. This group was compared against a control group or a regular group (RG) made of 15 students of the same age as the children in the first group, but who follow regular school curricula where English is taught for two hours a week. These participants were exposed to English mostly through explicit instruction and had little exposure to incidental lexis as they were mainly addressed in Catalan by their teacher. Therefore, the vocabulary they acquired during the lessons was targeted being related strictly to the thematic units studied.

3.2.2 Instruments

In spite of the MBCDIs being used to assess lexical development of children between 0 and 36 months in their mother tongue, we wanted to see if it could also be used with older participants to evaluate lexical comprehension and production in English as a FL. Taking into account the age of onset of our participants, we selected what we believe to be the most relevant measures from the MB-CDI Words and Gestures and MB-CDI Words and Sentences. They are summarized in the following table.

Table 1: The MBCDI measures used for the study

Test	Section	Measures	Labels	Max. score
MB-CDI <i>Words and Gestures</i>	Phrases	Phrases understood	PhsU	28
	Vocabulary checklist	Words produced	VdsU	396
		Phrases produced	VdsP	396
MB-CDI <i>Words and Sentences</i>	Examples	Mean length of the three best sentences	M3L	Not obtained
	Complexity	Sentence complexity	SC	37

3.2.3 Data collection procedure

The data for the MB-CDIs was collected towards the end of each academic course during three consecutive years (May-June 2012, 2013, 2014). In the case of the IG

group, data was collected by the researcher (who was thoroughly familiar with the linguistic profile of the participants in English, which was the language of testing) with the assistance of the participants' class tutors who spent approximately 6 hours a day in the presence of the children. The researcher filled in each form. Next, the same form was handed in to the participants' respective tutors who were asked to double check the questionnaires and in case of disagreement to make pertinent remarks with a different colour pen. The lexical items which were valued differently by either rater (researcher/tutor) were subjected to the calculation of cumulative score which "count the best score for an item of the CDI that any one rater gives" (de Houwer et al., 2006; pp.336). For example, if a word was marked as understood/produced only by one rater, it was accepted as such by the other. If one rater marked a word as understood and the other rater marked it as produced, the word was counted as produced. For the regular school group (RG1), data was collected by the class English teacher and the researcher who became familiar with these children's linguistic profile in the L3 by observing some English language sessions in the school. A unique questionnaire was used, on which the teacher performed the evaluations and the researcher made possible observations in case of disagreement. Because the two raters coincided in their evaluations, no cumulative scores needed to be calculated. Afterwards, the scores obtained were introduced into the SPSS Statistics programme and processed accordingly.

3.3 Results

3.3.1 RQ1: The MB-CDI results for the IG

In order to answer RQ1 stated in the previous section as follows: "Can MB-CDI capture lexical progress in the case of EFL learners acquiring English in immersion formal settings, where the amount of hours of exposure is less than in informal settings but much more than in regular school settings", we conducted a One-way RM ANOVA statistic analysis with Bonferroni pairwise comparisons. The descriptive scores for the lexical receptive measures, summarized in Table 2, revealed that the means for *PhsU* and *WdsU* increased steadily from T1 (year 2012) to T2 (year 2013) and from T2 to T3 (year 2014). For the vocabulary production measures (*WdsP* and *SC*), we observed considerable growth from T1 to T2 and from T2 to T3. Regarding the M3L scores, we detected an important increase from T1 to T2 and little change from T2 to T3. The

statistic results (One -way RM ANOVA) showed an overall significant effect of *Time* for the lexical reception measures. As summarized in table 2, the Bonferroni pairwise comparisons revealed significant differences between T1/T2, T2/T3, and T1/T3 in all the measures of lexical reception. With respect to lexical production, the one-way RM ANOVA also showed an overall effect for *Time* in all the measures. As was the case with lexical reception, the pairwise comparisons showed significant differences between T1/T2, T2/T3 and T1/T3 for all measures. For the variable *M3L*, the Friedman test also ascertained the parametric results. Thus, an overall *Time* effect was observed: *M3L* [$\chi^2(2) = .87, p = .000$]. The further Wilcoxon Signed Rank test also confirmed the significant differences. The results were as follows: *M3L*: [T1/T2 ($Z = -6.181, p = .000$); T1/T3 ($Z = -6.255, p = .000$); T2/T3 ($Z = -4.284, p = .000$).

3.3.2 RQ2: The MB-CDI results for the IG as compared to the RG

To answer RQ2 (Can MB-CDI compare groups of EFL learners of similar ages of onset but with different hours of exposure to English?) a Mann-Whitney U Test was performed. As summarized in Table 3, the mean scores for the vocabulary receptive measures show the IG scored above the RG at all data collection times. Results for productive knowledge were in the same line. The statistical analysis of receptive vocabulary indicates an overall *Time* effect for all measures with the largest effect size obtained for *PhsU* ($\eta^2 p = .55$). A large school effect in favour of the IG was also observed for all measures, the largest effect size being registered for *WdsU* ($\eta^2 p = .60$). The results for the vocabulary production measures follow the same pattern. Thus, a large overall *Time* effect was found for all measures with the largest size effect for *WdsP* ($\eta^2 p = .87$). There is also a *School* effect in all measures of lexical production in favour of the IG (the largest effect size was indicated in the *SC*: $\eta^2 p = .63$). These differences between groups are summarized in Table 3. As for the *M3L*, which presented a non-normal distribution, the results of the Mann-Whitney U Test ascertained the previous outcomes. The mean ranks obtained by the IG were significantly superior to those obtained by the RG1. The statistical results for the three data collection times are summarized as follows: T1 ($Z = -4.200, p = .000$), T2 ($Z = -5.871, p = .000$), T3 ($Z = -4.867, p = .000$).

Table 2: Descriptive statistics and one-way RM ANOVA results for the MB-CDI total scores measures with Time (T) as the within-subjects factor for the participants in the IG (n= 51), with p≤ .01

Measures		Mean (<i>Sd.</i>)			One-way RM ANOVA						
		T1	T2	T3	F	<i>p</i>	η^2p	power	T1/T2	T2/T3	T1/T3
Receptive vocabulary	PhsU	10.65 (5.78)	11.92 (6.21)	12.61 (6.69)	157.8)=70.370	.000	.58	1	.000	.000	.000
	VdsU	(21.15)	6.41 (3.98)	3.71 (2.60)	10)=26.635	.000	.34	1	.000	.000	.000
Productive vocabulary	VdsP	11.94 (6.39)	11.61 (3.34)	12.35 (3.94)	161.80)=543.887	.000	.91	1	.000	.000	.000
	13L*	10.32 (3.02)	10.12 (2.73)	10.83 (2.06)	10)=177.951	.000	.78	1	.000	.000	.000
	SC	10.00 (3.00)	10.31 (2.42)	10.59 (2.74)	111.17)=452.214	.000	.90	1	.000	.000	.000

Note: *Normality assumption violated. Greenhouse-Geisser correction applied for variables in shaded cells

Table3: Descriptive results and the 3x2 RM ANOVA with Time (T) as the within-subjects factor and School (S) as the between-subjects factor, for the participants in the IG (n=51) and RG1 (n=15), with p≤ .01. Sphericity assumed.

Measures		Mean (Sd.)						3x2 RM ANOVA			
		T1		T2		T3		F	p	η^2p	power
Receptive vocabulary	PhsU	IG (5.78)	RG1 (5.51)	IG (5.21)	RG1 (5.88)	IG (3.69)	RG1 (2.92)	128)=80.538	.000	.55	1
								1)= 47.196	.000	.42	1
	VdsU	(21.15)	(13.26)	1 (31.98)	2.53 (1.24)	1 (17.60)	17 (5.02)	128) = 49.172	.000	.43	1
Productive vocabulary	VdsP	(16.39)	3 (3.88)	1 (18.34)	0 (3.39)	5 (33.94)	80 (.44)	129)=94.152	.000	.60	1
								128) =446.800	.000	.87	1

								)= 54.643	000	46	1
		.00						28) = 194.044	000	75	1
3L*	(1.02)	.00	1 (.73)	0 (.53)	! (1.06)	i7 (.35)		)= 57.507	000	48	1
		.00)						28) =306.712	000	82	1
SC	00	.00	.0.31	.87	9 (4.74)	.3 (.83)		)=111.039	000	63	1
	00)	.00)	2.42)	.91)							

Note: *Normality assumption violated

4. Discussion of results

In terms of vocabulary size, the results obtained in our study bring about further information on the young learner's receptive and productive vocabularies as measured by the MB-CDI. First of all, we noticed no ceiling effects, which means that indeed this tool might prove reliable when assessing EFL young learners.

In the absence of research testing the lexical comprehension and production in a FL with the MB-CDI, we will rely on other MB-CDI studies conducted with very young bilingual and trilingual children acquiring English as one of their first languages. In order to offer a reliable comparison of the results, we performed proportional calculations. Thus, with respect to the WdsU (see table 4) we have noticed that the scores obtained by the IG in English are lower than those obtained by other much younger bilingual and trilingual children. These results are predictable as we assumed that our participants would obtain poorer scores in English than bilingual and trilingual learners for whom English is one of their first languages. This might be an indicator of the fact that the WdsU measure of the MBCDI provides consistent results with FL kindergarten learners at different stages of lexical development.

Table4: Comparative results of research conducted on lexical comprehension by means of MB-CDI

Study	Participants	months (at testing time)	U in English
drescu (2018)	als learning English as	46	90.12
		58	106.41
		70	123.71
z Wrecker (2011)	als	17	156
	ilinguals	17	202
2001)	ilingual	16	52

With respect to the WdsP, when comparing the results in English obtained at the same test by the IG and other bilingual children we found that, on the one hand, the IG scores in English at the ages of 46 and 58 months are lower than those obtained by other much younger bilingual and trilingual children who acquire English as one of the first languages (see table5). This predictable outcome might indicate that the WdsP measure

is reliable and could be used to measure productive vocabulary of kindergarten children in a FL at different stages of lexical development. On the other hand, at the age of 70 months, the IG scores are comparable to those obtained by Byers and Wrecker (2011) (see shaded cells in table 5). All these results indicate that, with a few exceptions, our participants understand and produce fewer words in English as a FL than much younger bilinguals and trilinguals do.

Table 5: Comparative results of research conducted on lexical production by means of MB-CDI

Study	Participants	months (at testing time)	MP in English
Iordescu (2017)	Children learning English as FL	46	31.94
		58	75.61
		70	146.35
Byers & Wrecker (2011)	Monolinguals	17	37
	Bilinguals	17	75
Wolke (2001)	Trilingual	21	50

The outcomes of PhsU, M3L and SC could not be contrasted against those of other studies which instead of raw scores have used percentiles, a type of score we were not able to use as the MB-CDI has only been normed to test young learners in their first languages and not in a FL. Nevertheless, after carefully examining our results by comparing them against those of other studies conducted with monolinguals and bilinguals in their first languages, we can say that the MB-CDI yields relevant results when testing lexical acquisition in a FL.

The results in the MB-CDI **total scores** show the IG was superior to the RG in all measures of lexical reception and production. In other words, the IG understand and produce significantly more words than their regular school peers. Furthermore, they are also able to construct longer and more complex sentences. This predictable outcome might be due to the fact that the IG has received considerably more exposure to the L3 than the RG1. In the specialty literature these findings have also been confirmed by EFLA studies conducted in regular schools (Garcia Mayo, 2002) and immersion school settings (Bergst rm, 2016). More specifically, they showed that, when comparing students of similar ages at the time of testing and with similar ages of onset, the group

who benefited from more hours of instruction scored better than those who received less instruction.

After corroborating the results obtained against those of other studies we conclude that indeed MB-CDIs might measure lexical progress in EFL. The results of this study clearly apply to kindergarten children as participants in our research were between 4 and 6 years old at the time of testing. Yet, as no ceiling effects were registered in the scores (see tables 1 and 2) we tend to believe that the same test might serve to assess primary school Romanian learners, who during first school years are exposed to less English than the participants in our study. Nevertheless, a proper study on Romanian learners needs to be conducted in order for the results to be conclusive

5. Conclusions

The MB-CDIs is a standard tool which could be used to test lexical knowledge in EFL. As said in the previous sections there are reasons to believe it might also work with Romanian primary school learners, first of all because no ceiling effect was noticed with learners who had benefited of more hours of instruction than our students have. Second of all, the test covers a lexical range that Romanian learners acquire throughout their first years of school. This way, in order to use it, teachers don't need to teach to the test. Following the Romanian curriculum and using English as much as possible in the class, so that our students have as much exposure as possible and acquire as much incidental vocabulary as possible, is enough. This type of test can provide reliable results and can be used to assess students on several occasions (maybe two times in an academic year). It might be helpful especially for beginner teachers who are required to draw a student progress report which should be included in their end-of-the-year portfolio.

Nevertheless, using the MB-CDIs in class regularly, might have several drawbacks. One of them is that it is extremely time consuming for the teacher as on average it takes 15 minutes per student. This means that in order to assess a whole class of 30 students, one teacher would need approximately 7 hours and a half to complete the evaluation. Another drawback of the test is that it is completed upon recall and therefore, the teacher needs to know all the students very well in order to be able to fill in the test correctly. A possible solution to this problem might be adding an extra testing item which the student can

actually solve on a piece of paper such as writing a composition. This way the teacher would be able to provide a more thorough assessment.

To conclude the MB-CDI has yielded reliable results in research with L1 young learners. Our study has showed that it might be used with kindergarten EFL learners in immersion schools. Judging by our results we can say that it might be an option for testing Romanian primary school students if the drawbacks of the test are carefully considered.

References

- Bishop, D. V. M. (2003). *Test for reception of grammar TRGO2*, Pearson Assessment.
- Bergström, K., Klatte, M., & Steinbrink, C. (2016). Effects of language immersion in kindergarten. *Language Learning*, 66(2), 386-418.
- Byers, H.K. & Wreker, J. F. (2011). Bilingualism in infancy: first steps in perception and comprehension. *Trends in Cognitive Science*, 12(4), 144.151.
- Galeote, M., Soto, P., Serrano, A., Pulido, L., Rey, R., & Martinez-Roa, P. (2006). Un nuevo instrumento para evaluar el desarrollo comunicativo y lingüístico de niños con síndrome de Down. *Revista Síndrome de Down*, 23, 20-26.
- Heilmann, J., Ellis W. S., Evans, J., & Hollar, C. (2005). Utility of the MacArthur Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology*, 14(1), 40-51.
- Luyster, R., Qiu, S., Lopez, K., & Lord, C. (2007). Predicting outcomes of children referred for autism using the MacArthur-Bates communicative development inventory. *Journal of Speech, Language and Hearing Research*, 50(3), 667-681.
- Nicoladis, E., & Secco, G. (2000). The role of a child's productive vocabulary in the language choice of a bilingual family. *First Language*, 58, 3-28.
- Mancilla-Martinez, J., Alexander-Pan, B., & Banu, V. S. (2010). Assessing the productive vocabulary of Spanish – English bilingual toddlers from low income families. *Applied Psycholinguistics*, 32(2), 333-357.
- Mariscal Altares, S., Nieva, S., & Lopez-Ornat, S. (2010). Observar y medir el desarrollo gramatical temprano en español. *Psicothema*, 22(1), 51-56.

- Mayor, J., & Plunkett, K. (2010). A statistical estimate of infant total vocabulary size from CDI analysis. *Developmental Science*, 14(3), 1-17.
- Pallant, J. (2005). *SPSS Survival Manual. A step by step guide to data analyzing using SPSS for Windows (Version 12)*. Sydney: Allen& Unwin.
- Pearson, B. Z., & Fernandez, S. (1994). Patterns of interaction in the lexical growth in two languages of bilingual infants and toddlers. *Language Learning*, 44(4), 617-653.
- Pearson, B. Z., Fernandez, S., & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Language Learning*, 43, 93-120.
- Pearson, B. Z., Fernandez, S., Lewedeg, V., & Oller, D. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18(1), 41-58
- Poulin-Dubois, D., & Goodz, N. (2001). Language differentiation in bilingual infants: Evidence from babbling. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 95-106). Amsterdam: John Benjamins.
- Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2013). Lexical access and vocabulary development in very young bilinguals. *The International Journal of Bilingualism: Cross-Disciplinary, Cross-Linguistic Studies of Language Behaviour*, 17(1), 57-70.
- Quay, S. (2001). Managing linguistic boundaries in early trilingual development. In J. Cenoz & F. Genesee (Eds.), *Trends in bilingual acquisition* (pp. 149-200). Amsterdam: John Benjamins.
- Serrat, E., Sanz-Torrent, M., Badia, I., Aguilar, E., Olmo, R., Lara, F., Llorenc, A., & Serra, M. (2010). La relación entre el aprendizaje léxico y el desarrollo gramatical. *Infancia y Aprendizaje*, 33(4), 435-448.