DEVELOPING AN URDU CDI FOR EARLY LANGUAGE ACQUISITION

MARIAM DAR, HUMA ANWAAR, MARILYN VIHMAN, TAMAR KEREN-PORTNOY

University of York

Abstract

This paper reports on the development of a lexical checklist for parents to trace vocabulary advances in Pakistani children. Various cross-linguistic adaptations of the MacArthur-Bates Communicative Development Inventory (CDI) were consulted. Pilot testing was carried out using a sample of 17 children from middle-class homes, aged from 12 to 30 months. Cross-linguistic comparison reveals similarities between the vocabulary growth of the Pakistani children and children learning other languages. Plans for further pilot-testing and eventual validation are discussed.

1. Introduction

The MacArthur–Bates Communicative Development Inventory (CDI) has been used extensively for both clinical and research purposes. Various studies have testified to the effectiveness of this parent report in measuring lexical development in the early years (Dale, 1991; Dale, Bates, Reznick, & Morisset, 1989; Miller, Sedey, & Miolo, 1995; Thal, O'Hanlon, Clemmons, & Fralin, 1999). The instrument has been adapted into 38 languages (Bates, Dale & Thal, 1995), including some bilingual adaptations (Gatt, 2007; O'Toole & Fletcher, 2008; O'Toole & Fletcher, 2010; O'Toole, 2013). Cross-linguistic comparison shows similarities in the vocabulary growth of monolingual and bilingual children. The current study is a part of larger investigation exploring the development of the lexicon in children from multilingual homes in Pakistan, which has so far received little attention. The overall aim is to formulate an assessment tool for Urdu-speaking children, most of whom are learning it in a multilingual context; this can lead to advances in child language research as well as in clinical practice in Pakistan.

1.1. Significance of Urdu adaptation of CDI

Urdu is estimated to be spoken by 100 million people around the world (BBC, 2014). Apart from Pakistan, it is spoken and understood in parts of India, Bangladesh, Nepal, the Middle East, and many other countries around the world. For instance, the Urdu speaking community in the UK is comprised of about 400,000 speakers. Urdu is essentially the same language as Hindi, which is spoken by 425 million as a native language (The Constitution of India, 2007). Both Hindi and English are official languages in the major cities in India, along with 22 official regional languages throughout the country. Urdu and Hindi, despite having different names and orthography, have only minor differences in lexicon and share the same grammar. Due to these linguistic and cultural similarities, the Urdu adaptation of CDI could also be used (with modifications, notably in the spelling of Urdu/Hindi words) for multilingual children in India.

1.2. The linguistic situation in Pakistan

Pakistan is home to many different languages and cultures, which makes the development of an Urdu adaptation of the CDI for children's early words a challenging task. Pakistan has a total population of 186 million; roughly 60 languages are spoken over the region (Akram & Mahmood, 2007). Punjab is the largest province and Punjabi the regional language with the largest number of speakers (44% of Pakistanis speak Punjabi: Akram & Mahmood, 2007). Other widely spoken regional languages include Pashto (15%), Sindhi (14%), Saraiki (10%), Balochi (4%), and Hindko (1%; see Figure 1).
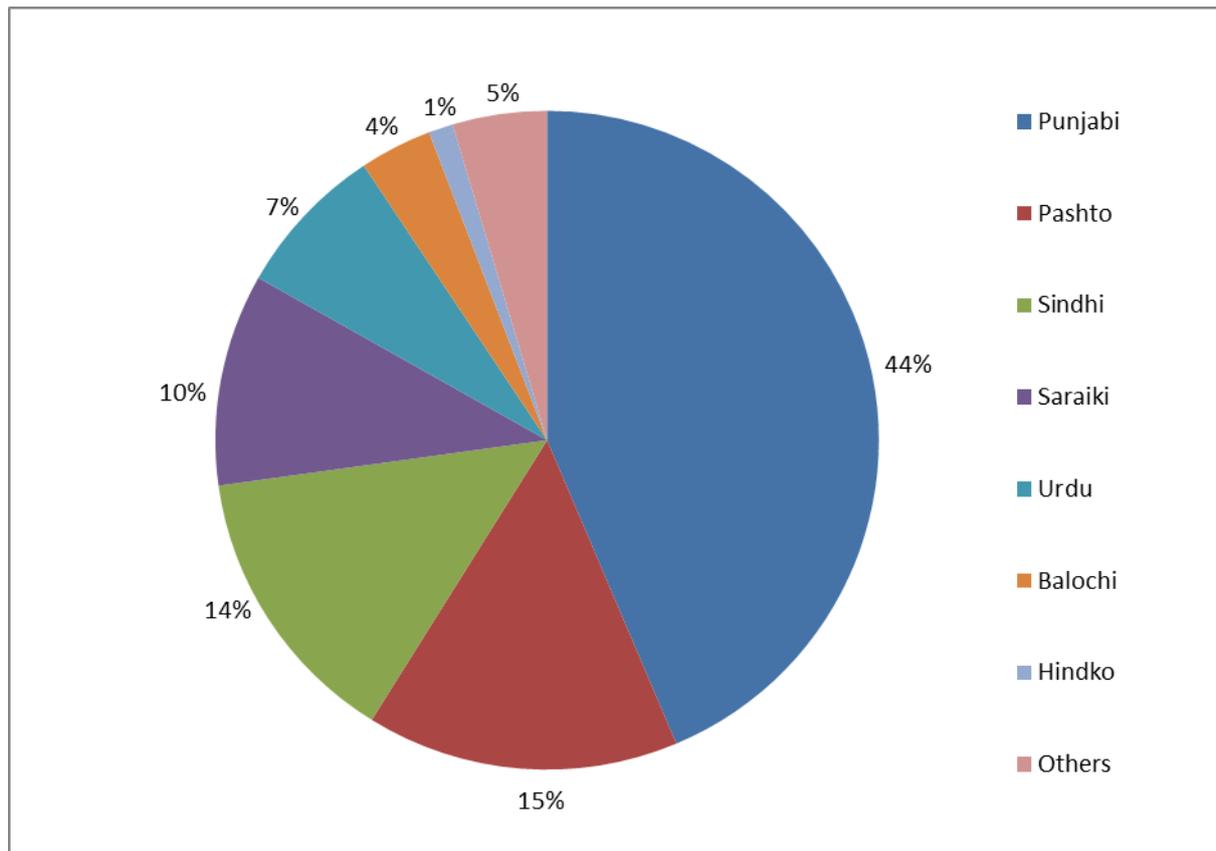

Figure 1: Distribution of native language use in Pakistan (Akram & Mahmood 2007: 2; adapted from table 1).

It is worth noting that Urdu, the national language of Pakistan, is native to only 7% of the population, predominantly those living in urban areas. However, it is used as a lingua franca among people from different ethnic backgrounds. There is no exact figure available on the number of L2 speakers of Urdu, although we estimate it to be roughly 60%. In the villages and small towns, children are exposed to Urdu as an L2 at school when they are aged three to five years.

Another language commonly spoken in the cities is English, also an official language. The majority of the population in urban areas speak either Urdu or English, which, as the languages of the educated and professional class, carry the most prestige.

English was introduced to the sub-continent through colonization by the British Empire in 1858. The British introduced a new educational system with English as the medium of instruction to create a link between the colonisers and the colonised (Evans, 2002). Later, English found its way into government administration, the law, the military, commerce and the mass media, where it continued to dominate, even after the end of colonization. It has surpassed

even the national language Urdu in popularity and has become a status symbol, particularly in urban areas. "It is the gateway to success, higher education and white collar jobs" (Ghani 2003: 105). Code-switching is very common in the cities in particular, where it is fashionable to insert English words into everyday speech.

The complex linguistic situation is reflected in the education system. Currently, four different school systems co-exist: Private Elite English medium, Private non-elite English medium, Government Urdu medium and Dini Madrasas (private religious schools focused on the teaching of the Quran, its interpretation, the sayings of the Prophet Muhammad and Islamic Law: Coleman & Capstick, 2012). Private elite English medium schools, which are very expensive and thus serve only the most prosperous families, are found in the larger cities; in these schools the medium of instruction is English only. The tutors are either native speakers of English or have received their training abroad. Children are extensively trained from the start to speak English 'like a native'. Private non-elite English-medium schools, which have local tutors, are tailored to the needs of the aspirational middle class, lower middle class and upper working class unable to afford the fees of the elite schools. These schools aim to develop reasonable proficiency in English. Children of the lower class in the urban and rural areas go to government-supported Urdu-medium schools, which are free; the Dini Madrasas, also Urdu-medium, provide religious education only. There are also some primary schools in Sindh and Khyber Pakhtunkhwa in which the medium of instruction is the regional language. Note that English is a compulsory subject in all but the Dini Madrasas.

1.3 The linguistic situation in homes

Because Urdu co-exists with English in a multilingual society there are abundant English loan words in Urdu as well as in the regional languages. A child raised in a middle-class home is typically exposed to Urdu, English and one or two regional languages. Parents encourage their children to learn and communicate in English since they regard it as the means to progress and prosperity. From a very early age infants are exposed to English through cartoons, books and rhymes. Parents and caretakers commonly insert English words into their everyday speech through code-switching. They usually switch to English when disciplining their children. Children are encouraged to learn and practice English politeness expressions such as *thank you, please, sorry* etc. Although children are exposed to English mostly through code-switching, we consider these children 'bilingual' because in most cases they produce more English than Urdu words (see Table 3). The children are fully exposed to English grammar only when they start school from the age of 3-5 years. Since they have already been exposed to English vocabulary at home, they begin to speak English with ease as soon as they start school.

However, it is important to note that exposure to a regional language is highly dependent on the region. People from rural areas or small cities or towns are proud of their regional affiliation and the regional languages are used frequently. However, regional languages do not hold much importance in big cities. Children are mostly exposed to Urdu and English and the regional language is introduced (if at all) after age two or three years because of the social stigma attached to it (speaking English/Urdu with a strong regional accent is associated with low social status). The Urdu CDI will initially be piloted and validated on a sample taken from the capital, Islamabad. For that reason, only English and Urdu will be included in the early versions of the CDI.

1.4 Research aims of the pilot study

The Urdu CDI is being developed with the aim of achieving a cultural and linguistic adaptation for Urdu-speaking children. Pilot testing was first carried out in order to test the appropriateness of the vocabulary items in the checklist in the context of a multilingual community. We wanted to make sure that we were choosing items that were relevant to the children's cultural and linguistic background. Another goal was to determine whether the vocabulary level would be comparable to that of other children of similar ages tested using other-language adaptations of the tool.

2. Development of an Urdu CDI

Due to the lack of data on Urdu language development, the preliminary checklist was based on the authors' own experience of living around children in the extended families typical of Pakistani homes. It was further validated by both face-to-face and Skype-interviews with mothers, all of whom come from middle-class Urdu backgrounds. In order to make the list more comprehensive, the American original (Fenson et al., 2007) and one UK adaptation of the CDI (Hamilton et al., 2010) were taken as models and all the vocabulary items included in those lists were screened for their relevance to lexical development in the Pakistani context. The Bengali adaptation (Hamadani et al., 2010) was also consulted for its cultural relevance to Pakistani languages. However, it proved less relevant, as it is very short (60 words) and was developed and tested with children from rural areas. The Maltese (Gatt, 2007) and Irish (O'Toole & Fletcher, 2008; O'Toole & Fletcher, 2010; O'Toole, 2013) adaptations were consulted for help in dealing with the multilingual aspect. A great number of lexical terms were added at this stage in order to make the list of words that children might know as complete as possible. A comprehensive checklist was developed based on these modifications and then further improved after feedback from colleagues.

2.1. Vocabulary Checklist

The initial checklist takes elements from both the "Words and Sentences" and "Words and Gestures" vocabulary inventories (CDI: WS), intended for use with infants aged 12 to 30 months. In order to reflect the linguistic and cultural features of Pakistani society the list includes 24 lexical categories (compared with 19 in the British and 22 in the American versions).

It was deemed essential to reflect the bilingualism characteristic of Pakistani homes in the Urdu CDI. Most concepts were represented as doublets, by including an English term in parallel with the Urdu one (e.g., *cat* and /bɪl:i/, *car* and /ga:ɽi/). Singlets were also added, including a) concepts specific to Pakistani culture, which have no equivalent in English (e.g. /mendi/ 'temporary tattoo', /ˈrɒʈi/ 'a kind of bread'), b) concepts present in Urdu but replaced by English words in everyday usage (e.g. /xa:ndan/ replaced by *family*, /geɪnd/ replaced by *ball)* and c) concepts adopted as inventions from the West with no equivalents in native speech (e.g. *computer, telephone),* which we consider loan words.

For expository purposes the lexical terms can be broadly divided into noun phrases, predicates and closed class words.

2.1.1. Noun phrases

The noun phrases in the early Urdu checklist correspond to the English-version categories of 'sound effects and animal sounds', 'animals', 'vehicles', 'toys', 'food and drink', 'clothing', 'body parts', 'small household items'.' furniture and rooms', 'outside things', ' 'places to go', and 'people'. Two new categories, 'festivals' and 'religion', were added to the existing ones to reflect the culture and life style of Pakistani people.

Religion plays an important role in the children's environment. For example, children see adults praying on the prayer mat (/ja: nəma:z/ 'praying mat'), raising their hands in a prayer (/du:ə/ 'prayer'). It is, for instance, a favourite routine of toddlers to bow down on the prayer mat in imitation of adults. Pakistani children also grow up witnessing religious and cultural festivals. For example, a wedding in Pakistan (/ʃadi/) takes place over three days, including /mendi/, the occasion of putting temporary tattoos on the bride's hands, /bəra:t/, the event in which the bride is married to the groom and is seen off by her family, and /vali:mə/, the official reception party offered by the groom's side. Other religious occasions include /i:d/, the most popular religious festival, celebrated after the holy month of Ramadan, and /tʃand ra:t/, the practice of moon sighting before /i:d/. These events and festivals are frequently celebrated in Pakistan, so all Pakistani children are familiar with them.

Additional culturally specific lexical items have been added to the existing categories. For example, words related to the concept of power shutoff, a common phenomenon in Pakistan that may be less often witnessed and spoken about in other societies, include such early-understood and used terms as *lantern, generator, UPS* and *load shedding*. These concepts are deeply integrated into the Pakistani way of life and every child is exposed to them. In fact, while doing the transcriptions in a recent project on 18-month-old-Pakistani children, we found that the children were already producing these words. One of the mothers in one of the video recordings actually switched off the light, to elicit *electricity gone* from the child.

The category 'people' was also extended by adding numerous kinship terms. Pakistani children are raised in a joint family system under the influence of grandparents, uncles, aunts and cousins and as a result Urdu has a complex system of kinship terms. Since respect and regard for elders is an important aspect of Pakistani society, special titles exist for almost all the maternal and paternal relationships (e.g. /na:na/ 'maternal grandfather', /da:da/ 'paternal grandfather'). Older and younger kin sometimes also have special terms (e.g., /ˈbʰəija/ 'older brother', /bʰa:i:/ 'younger brother', /tʃətʃə/ 'younger paternal uncle' and /tajə/ 'older paternal uncle').

2.1.2. Predicates

Predicates include verbs and adjectives. In Urdu both of these parts of speech carry number and gender markers. Since all the forms could not be listed, parents are instructed to tick the infinitive form of verbs (e.g. /ka:ʈnə/ 'to cut', /dəba:nə/ 'to press'). That is, explicit instructions are given to parents to mark the form even if the child does not use the infinitive but another form.

2.1.3. Closed class

These words include pronouns, question words, prepositions, quantifiers, auxiliaries. An additional category of 'tenses' was added, since Urdu tenses are quite complex. The tense is expressed by both an auxiliary and a main verb, each of which takes different forms according to tense and gender.

Within each category the Urdu words and their English equivalents were presented together, separated by a diagonal line (see Figure 2). Urdu words were also given in the Roman alphabet used for English; it is common practice in Pakistan to use the Roman alphabet for informal communication in Urdu over emails, text messages and Internet chat. Mothers were instructed to underline the form(s) used by the child. This was intended to minimise the length of the Urdu CDI.

6. BODY PARTS	U	U/S		U	U/S
Haath/Hands	☐	☐	Baal/Hair	☐	☐
Naaf/Belly Button	☐	☐	Angootha/Thumb	☐	☐
Pait/Taund/Tummy	☐	☐	Aiiree/Heel	☐	☐
Pair/Paoun/Feet	☐	☐	Tang/Leg	☐	☐
Ankh/Eye	☐	☐	Sir/Head	☐	☐
Moo/Face	☐	☐	Naak/Nose/Nosey	☐	☐
Ongli/Finger	☐	☐	Hont/Lips	☐	☐
Dant/Teeth	☐	☐	Moonchain	☐	☐
Zaban/Tongue	☐	☐	Daree/Beard	☐	☐
Thori/Chin	☐	☐	Nakhun/Nails	☐	☐
Bummies	☐	☐	Goadi/Goad/Lap	☐	☐

Figure 2: Sample of items listed under 'body parts' ('U' means understands only and 'U/S' means both understands and says).

3. Methodology

3.1. Participants

The data was collected in two phases. Eight Urdu-speaking mothers with children in the age range of 12 – 30 months participated in the first pilot. Gender was not controlled for. Only full-term children from middle-class homes with no hearing disabilities were included. All the children were reportedly spoken to primarily in Urdu within the home setting. The families

come from the two most important cities of Pakistan, Islamabad and Peshawar. Friends and acquaintances were contacted for the study. All had completed postgraduate education. Nine Urdu-speaking mothers participated in the second pilot. The same criteria were used for sample selection as in the first pilot. One child from the first pilot (18 months of age at that time) also took part in the second, at 30 months of age. All the families were settled in Islamabad, the capital city of Pakistan and Urdu was mainly spoken in all households. All parents had completed postgraduate education (see Table 1).

	Infant age in months	Gender	Siblings	Birth order	Mother's profession	Father's profession	% Urdu spoken
				First Pilot Test			
1	16	F	0	1	auditor	banker	45%
2	18	M	0	1	house-wife	banker	50%
3	24	F	1	2	linguist	engineer	65%
4	24	F	1	2	IT	IT	50%
5	24	M	0	1	lecturer	lecturer	55%
6	30	F	0	1	lecturer	lecturer	45%
7	30	M	1	2	house-wife	businessman	60%
8	30	M	1	2	teacher	armed forces	60%
				Second Pilot Test			
1	12	F	1	2	House-wife	banker	50%
2	13	F	0	1	house-wife	Engineer	69%
3	17	M	1	1	School teacher	Businessman	65%
4	18	M	0	1	House wife	IT	70%
5	22	M	0	1	lecturer	Sales	45%
6	24	M	0	1			
7	28	F	0	1	lecturer	lecturer	55%
8	28	M	1	2	HR	businessman	60%
9	30	M	0	1	teacher	armed forces	65%

Table 1: Background of participants

3.2. Data Collection

In both the first and the second pilot the Urdu CDI was submitted to the parents electronically and returned within two weeks of submission. Background information was collected by phone prior to submission, at which time mothers were also familiarised with the checklist. Detailed instructions were sent along with the CDI a few days later. Mothers were asked to make a note of the time it took them to fill in the Urdu CDI and also to rate the difficulty level. There was a comment box after each category to allow parents to add words not mentioned in any given category. The checklist data, including additional lexemes contributed by the parents, were coded manually.

4. Analysis

4.1. Challenges in assessing lexicon size for multilingual children

Assessing lexicon size in multilingual children is challenging due to the likely overlap of lexical knowledge in two (or more) languages. We wanted to compare the lexical level of Pakistani children with the lexical levels of monolingual and bilingual children learning other languages at similar ages. That would be helpful in determining how the language development reported through the tool relates to general learning trends in children all over the world. In previous adaptations of the MacArthur Communicative Development Inventories (CDI), the Total Vocabulary (TV) score – the total number of words a child knows – was computed to assess lexical growth in monolinguals; in bilinguals the total would logically be the words known in two languages. However, measuring lexical development in multilinguals is more complex than in monolinguals. In monolinguals the number of concepts can be taken to be roughly equal to the number of words, whereas in bilinguals or multilinguals one concept can be mapped to two or more words. The approach generally taken is to calculate the total number of concepts mapped onto two (or more) languages; Total Conceptual Vocabulary (TCV) is arrived at by counting only once the different labels for a single concept in two (or more) languages (Pearson et al., 1993). However, this measure may also be misleading, in that the translation equivalents might not be semantically equivalent for the child. Volterra and Taeschner (1978) showed that a bilingual German-Italian child was using doublets to talk about separate concepts (*barco* for sailboats and *boat* for other boats). In order to analyse the extent to which lexical knowledge in one language overlaps with lexical knowledge in the other(s) a separate comparison between Total Vocabulary (TV) and Total Conceptual Vocabulary (TCV) is advisable. This means that the vocabulary size for each of the languages has to be determined separately. This not only demands more time but also raises the issue of determining the source language for phonetically similar e.g., onomatopoeic forms like 'crow call', *caw* (English) and *kaẽ-kaẽ* (Urdu).

4.2. Analysis for the current study

The analysis could not be carried out as planned for the first pilot. In the checklist, Urdu words and their English equivalents were presented on the same line, separated by a diagonal line. Parents were instructed to underline the form(s) used by the child, which placed an additional demand on them: They had to tick the box as well underline the words. In the event, only three mothers followed the instructions properly, which made it impossible to judge the children's Total Vocabulary (TV). Only the Total Conceptual Vocabulary (TCV) score, in which each column was counted only once even though a child could potentially produce/comprehend two different forms for that concept, was calculated.

In the second pilot, the format of the checklist was modified to resolve the issue identified in the first one. The Urdu words and their English equivalents were given on separate lines and Urdu words were written in Urdu orthography (see Figure 3). This allowed the calculation of Total Vocabulary (TV) along with Total Conceptual Vocabulary (TCV). Also, in the open class categories only words reported by at least one parent from the first pilot were retained in the second. Additional lexemes contributed by parents were also incorporated. All closed-class items were retained, irrespective of whether or not they were scored by any parents.

11. PEOPLE (لوگ)

	U	U/S	سمجھنا/بولنا	سمجھنا			U	U/S	سمجھنا/بولنا	سمجھنا	
sister	☐	☐	☐	☐	بہن	uncle	☐	☐	☐	☐	چاچو
brother	☐	☐	☐	☐	بھائی	boy	☐	☐	☐	☐	لڑکا
baby	☐	☐	☐	☐	بچہ	girl	☐	☐	☐	☐	لڑکی
mama	☐	☐	☐	☐	امی	man	☐	☐	☐	☐	بندہ
baba	☐	☐	☐	☐	ابو	maid	☐	☐	☐	☐	ماسی
police	☐	☐	☐	☐	فوج				☐	☐	مأموں
			☐	☐	باجی				☐	☐	بہنوپو
			☐	☐	ناٹا				☐	☐	تایا
			☐	☐	ناٹی				☐	☐	چاچی
			☐	☐	مولوی صاحب				☐	☐	دادا
			☐	☐	قاری صاحب				☐	☐	دادی
			☐	☐	کام والی				☐	☐	خالہ
doctor	☐	☐							☐	☐	جن
aunty	☐	☐				teacher	☐	☐			
						sir	☐	☐			

Any other words:

Figure 3: Sample of revised Urdu checklist items listed under 'people'. The figure shows how two kinds of vocabulary items, doublets and singlets, are represented in the CDI.

Table 2 summarises the descriptive statistics for the seventeen children whose lexical development was recorded in the pilot study. Depending on the age of the child and the level of expressive vocabulary, the Urdu CDI took from 30 to 45 minutes to complete. The difficulty level was rated as medium by the parents. As the table shows, the sample size is very small in each group, ranging from 1 – 3 children.

It is also important to mention the Total Vocabulary (TV) scores in English and Urdu separately. As discussed in the earlier section, Urdu and English are the most often used languages in the major cities in Pakistan. Although children are too young to be formally exposed to English at school, the English learning that occurs through exposure to TV, books and code-switching from adults plays an important role in the children's development. Table 3 reports the vocabulary scores of children from the second pilot.

First Pilot Test			
Child ID	Age in Months	Total Conceptual Vocabulary (TCV)	Receptive-only Vocabulary
1	16	125	292
2	18	544	32
3	24	162	124
4	24	430	119
5	24	251	132
	Mean	281	125
6	30	350	540
7	30	407	148
8	30	398	371
	Mean	385	353
Second Pilot Test				
Child ID	Age (Months)	Total Vocabulary (TV)	Total Conceptual Vocabulary (TCV)	Receptive-only Vocabulary
1	12	68	67	92
2	13	12	12	144
3	17	27	23	87
4	18	109	102	207
5	22	324	270	93
6	24	285	227	62
7	28	468	385	93
8	28	705	574	180
	Mean	586	479	136
9	30	818	627	129

Table 2: Vocabulary level reported for sampled children

Child ID	**Age (Months)**	**Total Vocabulary (TV) in English**	**Total Vocabulary (TV) in Urdu**
1	12	38	30
2	13	7	5
3	17	15	12
4	18	43	56
5	22	169	155
6	24	140	145
7	28	269	199
8	28	400	305
9	30	574	244

Table 3: Vocabulary level in English and Urdu reported for sampled children

It is significant to mention that one of the aims was to reduce the size of the vocabulary of Urdu CDI. As was pointed above, only those words in the open class categories were retained, after both the pilots, which were reported by at least one of the parents. The words not reported by any of the parents were taken out. This was done to make the word list compact and comparable to word lists in other languages. To start with, the CDI had a word count of 2356 which has been reduced to 1290 words after the two pilots. Table 4 provides the word count of some other monolingual and bilingual CDIs for comparison.

Monolingual	Vocabulary count	Bilingual	Vocabulary count
English	680	Irish	826
Italian	670	Maltese	1681
Danish	725	Urdu	1290

Table 4: Comparison of the word count of monolingual and bilingual versions of the CDI.

A comparison of the Pakistani children's vocabulary level with that of American, Italian, Danish, Maltese and Irish children is presented in Figure 4 (Total Vocabulary scores (TCV) of Irish, Maltese and Urdu children were compared). The results can be viewed in terms of the vocabulary growth typical of monolingual as compared with bilingual children. First, the vocabulary development of the Urdu children is in line with that of monolingual children in both pilots. For instance, at 24 months Urdu-speaking children knew an average of 281 (first pilot) and 227 (second pilot) words; this compares with 307 for American (Dale & Fenson, 1996), 292 for Italian (Caselli et al, 2001) and 262 for Danish children (Anderson et al., 2006). The vocabulary sizes were also in line with those reported in the studies involving bilingual children: At 24 and 30 months the Urdu-speaking children knew an average of 281 and 385 words (in the first pilot), which is comparable to Irish (240 and 440 words at similar ages) and Maltese (356 words at 30 months). Similarly, in the second pilot, Urdu speaking children knew an average of 68 and 227 words at 16-18 and 24 months, which is comparable to Irish (81 and 240 at similar ages) and Maltese (58 words at 16-18 months). At 16-18 months (first pilot) and 30 months (second pilot) the TCV scores from the Urdu CDI considerably surpass the mean scores from either monolingual or bilingual children in the other studies. This is because one child with an unusually high TCV score (512) at 16-18 months in the first pilot also took part in the second pilot at 30 months of age.
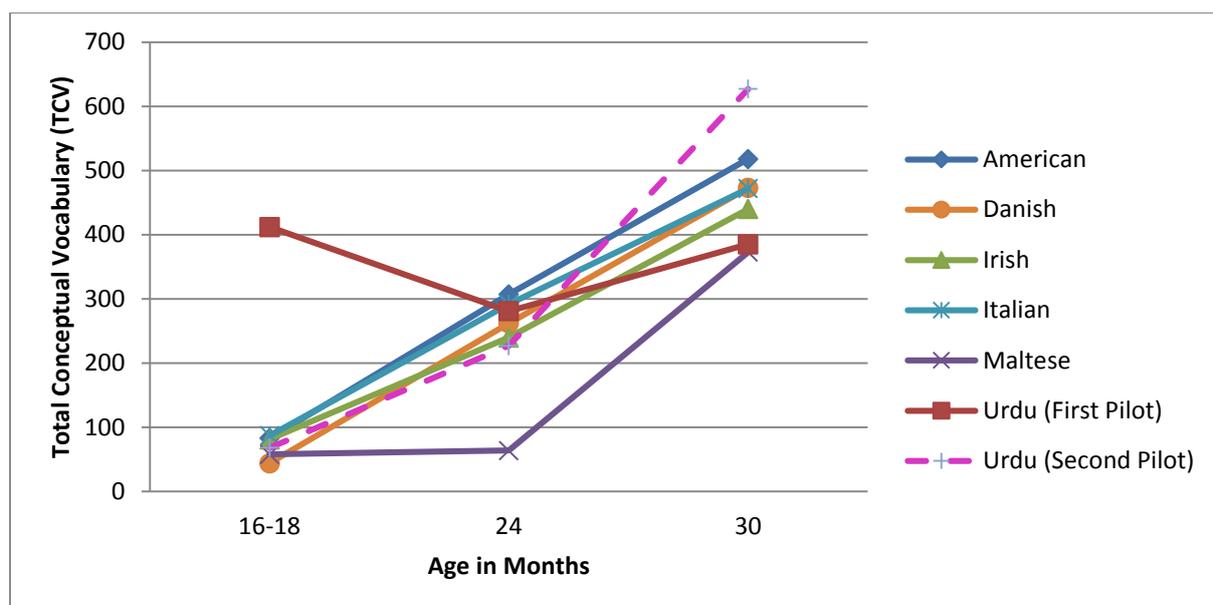
Figure 4: Comparison of mean expressive vocabulary scores by age from CDI adaptations. Sources: American (Dale & Fenson, 1996), Irish (O'Toole & Fletcher, 2010), Maltese (Gatt, 2007), Italian (Caselli et al., 2001), Danish (Anderson et al., 2006).

5. Discussion and conclusions

The basic aim of the study was to measure the lexical development of Pakistani children using an Urdu CDI, with a view to designing a larger study on the basis of the outcome. Both monolingual and bilingual adaptations of the MacArthur Communicative Inventory (CDI) were consulted. The preliminary Urdu checklist was pilot tested on eight children from middle class homes and the Total Conceptual Vocabulary (TCV) score was established. In the second pilot testing, data was collected from nine children from middle-class homes. Both Total Vocabulary (TV) and Total Conceptual Vocabulary (TCV) scores were calculated. The sample size is still too small to permit any conclusions about general trends in lexical development in Pakistan. However, these two stages of pilot testing have helped to refine the CDI and increase its appropriateness for use with the bilingual children of major cities of Pakistan. The results from the CDI are also comparable to those obtained from monolingual and bilingual children who are learning other languages. The research has now moved to the next stage, validation, which started in January 2015. The instructions have been prepared in both languages, as recommended by O'Toole (2013). A sample of 20-25 Urdu-learning children are being selected for the initial validation process. Care will be taken to include at least five age groups, with no fewer than two children in each group. Parental reports will be supplemented by video-recordings of a conversational sample (of approximately 30 minutes) involving each parent and child and using toys and picture books.

As in the pilot studies reported in this paper, only words reported by at least one parent will be retained in the open class categories in subsequent studies. Additional lexemes contributed by parents will also be incorporated. All closed-class items will be retained, irrespective of whether or not they have been checked by any parents.

References

AKRAM, M. & MAHMOOD, A. (2007). Status and teaching of English in Pakistan. *Language in India*, 7(12), 1–7.

ANDERSEN, C., VACH, W., WEHBERG, S. & BLESES, D. (2006). Conduct of the Danish CDI studies. *Working Papers,* 2. [Online], Available at: http://www.sdu.dk/Om_SDU/Institutter_centre/C_Boernesprog/e-prints.aspx [Accessed 20 July 2014].

BATES, E., DALE, P. & THAL, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. Macwhinney, (Eds). *The handbook of child language*. London: Blackwell, pp. 96–151.

BBC (2014). *A guide to Urdu – ten facts about Urdu language*. [Online] BBC, UK. Available at: http://www.bbc.co.uk/languages/other/urdu/guide/facts.shtml [Accessed 11 August, 2014].

CASELLI, M.C., CASADIO, P. & BATES, E. (2001). Lexical development in English and Italian. In M. Tomasello, (Ed). *Language development: essential readings in developmental psychology*. London: Blackwell, pp. 78–110.

COLEMAN, H. & CAPSTICK, T. (2012). *Language in education in Pakistan: recommendations for policy and practice*. Islamabad, Pakistan: British Council.

DALE, P. S. (1991). The validity of a parent report measure on vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research, 34,* 565–571.

DALE, P. S., BATES, E., REZNICK, J. S. & MORISSET, C. (1989).The validity of a parent report instrument of child language at twenty months. *Journal of Child Language, 16,* 239–249.

DALE, P. S. & FENSON, L. (1996). Lexical development norms for young children. *Behavioral Research Methods, Instruments, & Computers,* 28, 125–127.

EVANS, S. (2002). Macaulay's minute revisited: colonial language policy in nineteenth-century India. *Journal of Multilingual and Multicultural Development*, 23(4), 260–281.

FENSON, L., DALE, P., REZNICK, J.S., THAL, D., BATES, E., HARTUNG, J., PETHICK, S. & REILLY, J. (1993). *The MacArthur Communicative Development Inventories: user's guide and technical manual*. San Diego: Singular Press.

GATT, D. (2007). Establishing the concurrent validity of a vocabulary checklist for young Maltese children. *Folia Phoniatr Logop*, 59, 297–305.

GHANI, M. (2003). The status and position of English language in Pakistan. *Journal of Social Sciences and Humanities*, 1(1), 103–115.

HAMADANI, J.D., BAKER-HENNINGHAM, H.,TOFAIL, F., MEHRIN, F., HUDA, S.N. & GRANTHAM-MCGREGOR, S.M. (2010). The validity and reliability of mothers' report of language development in one year old children in a large scale survey in Bangladesh. *Food Nutr Bull* , 31(2), 198–206.

HAMILTON, A., PLUNKETT, K. & SCHAFER, G.(2000). Infant vocabulary development assessed with a British Communicative Development Inventory: lower scores in the UK than the USA. *Journal of Child Language*, 27, 689–705

MILLER, J. F., SEDEY, A. L. & MIOLO, G. (1995). Validity of parent report measures of vocabulary development for children with Down syndrome. *Journal of Speech and Hearing Research, 38,* 1037–1044.

O'TOOLE, C. & FLETCHER, P. (2008). Developing assessment tools for bilingual and minority language acquisition. *Journal of Child Speech and Language Studies*,16, 12-27.

O'TOOLE, C. & FLETCHER, P. (2010). Validity of a parent report instrument for Irish-speaking toddlers. *First Language*, 30(2), 199–217

O'Toole, C. (2013).Using parent report to assess bilingual vocabulary acquisition: A model from Irish. In V.C.M. Gathercole, (Ed). *Solutions for the assessment of bilinguals*. Bristol: Multilingual Matters, pp. 81–102.

Pearson, B.Z., Fernandez, S.C. & Oller, D.K. (1993). Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Language Learning*, 43(1), 93-120.

Thal, D. J., O'Hanlon, L., Clemmons, M. & Fralin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research, 42,* 482–496.

The Constitution of India (2007). *Ministry of Law and Justice: Government of India*. [Online]. Available at: http://lawmin.nic.in/coi/coiason29july08.pdf [Accessed 11 August, 2014].

Volterra,V. & Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, 5, 311-326.

Mariam Dar, Huma Anwaar, Marilyn Vihman, Tamar Keren-Portnoy
Department of Language and Linguistic Science
University of York
Heslington
York
email: md738@york.ac.uk