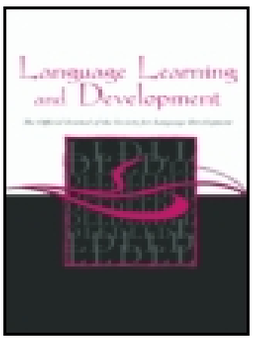
2021-04-25

An Adaptation of the MacArthur-Bates CDI in 17 Arabic Dialects for Children Aged 8 to 30 Months

Abdelwahab, AGS

10.1080/15475441.2021.1916502
Language Learning and Development
Taylor & Francis (Routledge)



Language Learning and Development

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hlld20

An Adaptation of the MacArthur-Bates CDI in 17 Arabic Dialects for Children Aged 8 to 30 Months

Alshaimaa Gaber Salah Abdelwahab, Samuel Forbes, Allegra Cattani, Jeremy Goslin & Caroline Floccia

To cite this article: Alshaimaa Gaber Salah Abdelwahab, Samuel Forbes, Allegra Cattani, Jeremy Goslin & Caroline Floccia (2021): An Adaptation of the MacArthur-Bates CDI in 17 Arabic Dialects for Children Aged 8 to 30 Months, Language Learning and Development, DOI: 10.1080/15475441.2021.1916502

To link to this article: https://doi.org/10.1080/15475441.2021.1916502

LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2021.1916502

Routledge
Taylor & Francis Group

An Adaptation of the MacArthur-Bates CDI in 17 Arabic Dialects for Children Aged 8 to 30 Months

Alshaimaa Gaber Salah Abdelwahab[a,b], Samuel Forbes[c], Allegra Cattani[a], Jeremy Goslin[a], and Caroline Floccia[a]

[a]University of Plymouth, Plymouth, UK; [b]Damietta University, Damietta, Egypt; [c]University of East Anglia, Norwich, UK

ABSTRACT

Assessing a child's language in the early years is critical to plan for an early intervention and maximize their learning potential. In a unique pan-Arabic approach to language development, we developed a new Arabic assessment tool, usable by parents and Early Years professionals to screen vocabulary in children between 8 months and 30 months across 17 Arab countries. Departing from the two relevant original Communicative Development Inventory forms (CDI: Words and Gestures and CDI: Words and Sentences, Fenson, Marchman, Thal, Reznick, & Bates, 2007), our Arabic CDI focuses on Words Only (Short Form), and assesses comprehension and production of a list of 100 words in 17 main dialects or Arabic, through a parental report. Data were collected from 436 Egyptian children and 168 children from the remaining 16 countries. Quasi-binomial model fits on Egyptian and Other Dialects comprehension and production data showed that Egyptian vocabulary norms could be reasonably extrapolated to the Other Dialects sample, as a first indication that the tool might be usable across the different countries.

Introduction

Language disorder is the most common developmental problem in pre-schoolers (in the US; Rossetti, 2001; in China: Lam, 2006). Screening for language delay in early childhood is an essential step for deciding on any subsequent intervention: the earlier the enrollment in intervention, the more positive the prospects of a healthy language development (Moeller, 2000; Vohr et al., 2008; Yoshinaga-Itano et al., 1998). In the Middle East, a range of country-specific tools have been developed to assess early language development in Arabic-learning children, relying mostly on face-to-face interactions with a practitioner (e.g., Egypt: Arabic Language Test: Rifaie, 1994; CALT: Abo Ras et al., 2010; Jordan: Arabic Token Test for Children, Alkhamra & Al–Jazi, 2016; Saudi Arabia: Language Comprehension Test: Al-Akeel, 1998). In this context where resources are often scarce, it is desirable to complement these approaches with more cost-effective solutions such as parental questionnaires, to pave the way for large-scale screenings. Here, we report an adaptation of the MacArthur-Bates Communicative Development Inventory (Fenson et al., 1994, 2007) for the Arabic language in 17 dialects, with the aim of providing the research and clinical communities with a simple tool (Short Form of the CDI) for assessing language development in Arabic across a variety of dialects.

CDIs have long been established for their reliability and validity in research (Mancilla-Martinez et al., 2016; Marchman et al., 2008) and, to some extent, for clinical assessment (Charman et al., 2003; Heilmann et al., 2005; Kim et al., 2014). Crucially, they have been adapted in up to 60 languages, with the aim of mirroring the structure and culture of the reference language as much as possible

CONTACT Alshaimaa Gaber Salah AbdelWahab alshaimaaabdelwahab@gmail.com PSQ A226, School of Psychology, University of Plymouth, Drake Circus, Devon PL4 8AA

Supplemental data for this article can be accessed on the publisher's website.

(Wordbank: Frank et al., 2017). CDIs are cost-effective and quick-to-use parental questionnaires, which do not require a professional for administration and can be used with children whose assessment would be problematic in a face-to-face interaction. In addition, such reports can reflect skills across a wide range of contexts (Crais, 2011; Fenson et al., 2007). The original American CDI (Fenson et al., 1994, 2007) was developed in three separate forms for different age ranges; the *CDI: Words and Gestures* form is aimed at typically developing children between 8 and 16 months. It assesses comprehension, production, and the use of communicative and symbolic gestures. The second form, the *CDI: Words and Sentences*, is aimed at children between 16 and 30 months and assesses productive vocabulary and grammatical knowledge. The final form, the *CDI-III*, assesses vocabulary (with 100 words) and sentence production as well as language use in toddlers aged 30 to 37 months. Subsequent developments of the CDI, and in particular, the Oxford CDI for British English (Hamilton et al., 2000), often assess only vocabulary knowledge in comprehension and production. Importantly, shortened versions of the CDIs have been developed over the years (e.g., American English: Fenson et al., 2000; British English: Floccia et al., 2018; German: Mayor & Mani, 2019), allowing for a more rapid assessment of children's language skills when time, resources and maybe parental literacy levels are to be considered. Following this trend, here we ultimately developed the Short Form Words Only version (the existing format of the CDI), for infants aged 8 to 30 months, spanning an age range originally covered by two separate versions (CDI:WG and CDI:WS).

Assessing language development in Arabic-learning children is a complex task for a number of reasons. First, the Arabic language poses a unique challenge as it comprises a large range of dialectal variations, to the extent that some are mutually non-understandable (e.g., Morocco versus Gulf region). Phonological, morphological, and lexical differences between Arabic dialects can make it challenging for a citizen from an Arab country to understand the dialect of another. For example, the three fricatives/θ/,/ð/, and/đ/from Modern Standard Arabic (MSA) would be used phonologically in Gulf Arabic (the Gulf Coast from Kuwait to Oman), Iraqi, and Yemeni Arabic, in addition to Jordanian and Tunisian Arabic (Khamis-Dakwar et al., 2012). However, Egyptian, Levantine, and Moroccan Arabic would instead use the dental stops/t/,/d/, and/ḍ/respectively. Lexical variations can be maximal, as with the English word "very" which be realized as/kulliš/in Iraqi Arabic,/killiš/in Gulf Arabic,/ktīr/in Levantine Arabic,/ʾawi/in Egyptian Arabic, and/ʿad/in Moroccan Arabic (e.g., Grigore & Bițună, 2016).

Another potential challenge stems from the fact that children in the Arab world are typically exposed to two languages: MSA, which is not a native language, and their local Arabic dialect. The former is the official language in school books (though the dialect of each country is used for teaching), newspapers, media, and official communications. Unlike MSA which is the primary language of writing, dialects are rarely used for literacy as there are no established rules for writing them (however the introduction of texting has changed this: see Al-Walaie & Khan, 2017). MSA and dialects can differ on morphological, grammatical, and lexical levels. For example, MSA has a dual form in addition to the singular and the plural forms, which some dialects lack. Usually children's television programs use the country's own dialect and not MSA, yet the children are still exposed to MSA when watching videos, using tablets and mobile phones. The impact of this passive exposure to MSA is unknown, but past research on children's learning of foreign-language speech sounds through television suggests that it might be minimal. In a seminal study, Kuhl et al. (2003) examined the impact of social interaction on non-native phonetic learning at the age of 9 months. Only a group exposed to the non-native language (Mandarin Chinese) through face-to-face interactions showed evidence of perceptual learning, in contrast to a group exposed to audio or audio-visual recordings of these interactions. Therefore, it is unlikely that the exposure to MSA through tablets will modify children's early representations of dialectal Arabic gained through interactions with their family. For this reason, it appears critical to assess the development of the children's language based upon the dialect of their original country, or the dialect they are mostly familiar with.

Lastly, very few culturally adapted assessment tools exist for evaluating the development of Arabic-learning children, with the existing ones tailored to a specific Arabic dialect such as Egyptian Arabic

(e.g., Rifaie, 1994) or Jordanian Arabic (Alkhamra & Al–Jazi, 2016), with no knowledge of their exportability, presumably because of the obstacles mentioned before.

In this paper, we adopted a pragmatic approach for the development of Arabic assessment tools, by developing a Short Form of the parental questionnaire MacArthur-Bates CDI (Fenson et al., 2007) to 17 different Arabic dialects, in a novel attempt to examine if a single list of words with their dialectal variations could capture lexical development across the Middle East. Given that all Arabic-speaking countries share to a large extent a common Arabic culture, we anticipated that parents would be using the same words with their children across the different dialects. At the same time, and as previously mentioned, the only common language between all Arabic-speaking countries is MSA, which is highly unlikely to be used or even understood by a child within the age range of 8 to 30 months. Therefore, we decided to create a dialect-specific short version of the common word list for each country. Our starting point was the Egyptian dialect, chosen because of Egypt's unique cultural influence across Arabic-speaking countries, and because of the availability of other Egyptian-specific language assessment tools that were needed to establish the Arabic CDI's validity. From there, we developed the tool in 16 other dialects of Arabic.

As in any CDI adaptation, we measured the impact of the main factors known to affect language development in the early years: SES and gender. It is well established that children from low SES background typically score lower in all measures of language skills as compared to children from higher SES background (Campbell et al., 2003; Hart & Risley, 1995; Hoff et al., 2002). Maternal education seems to be the main factor explaining why children from low SES have poorer language skills than their high SES counterparts (Rack et al., 1992) as it tends to affect the quality and quantity of the maternal language input (Hoff, 2003). Highly educated mothers tend to produce more word tokens, more word types and higher Mean Length of Utterances, which seem to boost children's productive vocabulary (Hoff, 2003); in addition they use more diverse vocabulary, in terms of nouns, tenses, and verbs than parents from low SES (Feldman et al., 2000; Hart & Risley, 1995). Here we predict a similar effect of SES on the vocabulary knowledge of Arabic-speaking children.

The effect of child gender on language skills is reported rather consistently in the literature, with girls outperforming boys in verbal abilities at least from the age of 1 to 7 years (e.g., Bornstein et al., 2004), which may explain a similar girl advantage in reading and writing skills (see e.g., the large sample analyzed by Reilly et al., 2019). Regarding very early language skills, Bornstein and Cote (2005) reported that girls at the age of 20-months-old in three countries (USA, Italy, and Argentina) acquire more expressive vocabulary than boys in all word categories at the same age. This was further supported by Eriksson et al. (2012) who, using adapted versions of the MacArthur-Bates CDI across a range of European languages, observed that while boys and girls scored similarly in relation to their comprehension performance, girls outperformed boys in gesture production, word production, and word combinations, which pointed to a biologically determined origin to this gender effect. However because all these countries are quite similar in terms of cultural profiles (i.e., mostly Western Christian), it is an open question as to whether gender impacts similarly on the development of vocabulary knowledge in Arabic-speaking children in the Middle East. Possible moderators of an advantage of girls over boys in terms of language development could relate to differences in parenting styles and attitude toward education. For example, male Arab adolescents across the Middle East report higher level of authoritarian parenting than females (Dwairy et al., 2006; Smetana & Ahmad, 2018), which contrasts with Western or Asian societies where the effect of child gender on parenting style is small to negligible (see the meta-analysis by Endendijk et al., 2016). In addition, in some rural areas such as Upper Egypt, girls are less likely to enroll in primary schools than boys, suggesting that girls' education is less valued than boys' (Iqbal & Riad, 2004). These parenting-related differences may contribute to minimize the language development advantage seen in girls in Western societies.

In summary, we will develop, through translation and cultural adaptation, a language assessment tool, the Short Form of the Arabic CDI Words Only, from English (Alcock et al., 2017; Fenson et al., 2007; Hamilton et al., 2000) to MSA, and then to the main 17 Arabic dialects of 17 Arabic countries, to

be used widely and freely with Arabic-speaking children aged 8 to 30 months. Impact of core predictors (age, SES and gender) will be examined. Most importantly, we will attempt to determine if the rate of vocabulary development is equivalent across the 17 dialects, by comparing a large data set from the Egyptian dialect to a smaller dataset obtained across the 16 other countries.

It must be noted that there are currently five (full form) dialect-specific Arabic CDIs: Saudi (Dashash & Safi, 2008), Kuwait (Abdalla & Shaalan, unpublished), Tunisian (Bourgou, unpublished), Israeli Palestinian (Essa & Armon-Lotem, unpublished) and Lebanese trilingual (Messarra & Kouba El Hreich, unpublished), which were developed independently. We were not aware of the last two when we started this research. Researchers from Saudi, Kuwait and Tunisia were contacted at the onset of the project to discuss a collaboration, which was positively received by the Tunisian team only (we agreed to pursue two Tunisian versions in parallel). As a result, we have not been able to include a Kuwaiti version, but we have developed a Tunisian and a Saudi version, with the agreement of the MacArthur-Bates CDI advisory board.

The current study

The study contains three separate parts, presented in more detail below.

In an initial Preparatory Study, we compiled a Full Form of the Arabic CDI Words Only (404 words) which was translated in 17 dialects (and in MSA). This was initially tested with a small sample of children from 10 of these countries to examine comparability, and concurrent validation was run by comparing data from a group of 23 Egyptian toddlers aged 24 months to those obtained through the Egyptian Arabic Language Test (Rifaie, 1994) used as gold standard.

Given the high attrition rate in terms of non-completion of the questionnaire in the preparatory study, in the next step we decided to develop a Short Form of the Arabic CDI Words Only (100 words) using data collected on the initial Full Form. In this second step, the Egyptian Arabic Study, we describe how we developed the Short Form, and assessed its test–retest reliability and effect of mode of completion (paper versus online). This was followed by the main data collection and the production of norms in this dialect.

Finally, we initiated a normative study using online data collection with 436 Egyptian infants, described in the third step, the Mixed Arabic Study. Data collected in the remaining 16 dialects (168 children) and word learning trajectories in the Mixed Arabic sample were compared to the Egyptian sample.

Preparatory study

Word selection for a full form Arabic CDI word only and translation in 17 dialects

The starting point of the selection of words was to create a list common to the MacArthur-Bates CDI: Words and Sentences (Fenson et al., 2007), the Oxford CDI (Hamilton et al., 2000), and the UK-CDI Words and Gestures version (Alcock et al., 2017), to include most of the words that might be familiar to young children. Words in common (419 words) between the three of them were included to form the initial Full Form of the Arabic CDI Words Only. These words were initially translated into MSA by the first author and checked by an experienced Arabic teacher to ensure that the correct standard Arabic words were used. Using feedback from Arabic-speaking parents in the UK, some words were deleted (15 words) from the three English CDIs because of cultural inappropriateness (pig, penguin, owl, pony, puppy, kitty, sweater, jumper, coat, butter, toast, cot, snow, penny, and pat-a-cake). The decision not to add additional words was taken not to inflate the total word count, especially given that the removal of these items did not deplete existing categories. An exception was made for the cultural-specific word "mosque" which was added. This Arabic list was then translated into the dialects of 17 Arabic countries by sending the MSA list to acquaintances living in these countries speaking or familiar with the most common dialect in that country. It was then given to a second speaker of the

same country to back translate the list from the country's dialect to MSA, independently from the first translator, to ensure consistency of word use. When a mismatch occurred (rarely), all suggested variants, whether in MSA or in the dialect itself, were included in the list because they gave the same meaning (with the agreement of both translators). By the end of this process, we obtained a word list that consisted of 404 words translated into 17 dialects: Algerian, Bahraini, Egyptian, Emirati, Iraqi, Jordanian, Lebanese, Libyan, Moroccan, Omani, Palestinian, Qatari, Saudi, Sudanese, Syrian, Tunisian, and Yemeni. This word list was divided into 19 categories: 12 animal sounds, 33 animal names, 11 vehicle names, 8 names of toys, 34 names of food and drinks, 19 clothing names, 22 names of body parts, 26 names of furniture and rooms, 34 names of small household items, 24 names of outside things and places, 22 people names, 20 names of games and routines, 65 verbs, 8 words related to time, 35 adjectives, 10 pronouns, 6 question words, 8 prepositions, and 7 quantifiers.

Comparability of the full form Arabic CDI words only across 10 Arabic dialects

To examine whether the Full Form Arabic CDI Words Only provided comparable estimates of vocabulary knowledge across countries, and whether it would be valid for use in children within the age range of 8 to 30 months, we recruited a group of parents of 24-month-old children from a range of geographically spread countries. We collected data from 33 Arabic-speaking children: 19 girls and 14 boys (from 33 families) living in their original countries of birth, which included between 3 and 4 children from each Arab country (10 countries selected on an opportunity sample based on the availability of acquaintances willing to support participant recruitment: Algeria, Egypt, Emirates, Iraq, Jordan, Libya, Morocco, Palestine, Saudi, and Syria). The age of 24 months was chosen as the midpoint (in terms of vocabulary size) of the 8–30 months target range. The 404-word list was sent to parents via Facebook, friends, relatives, mailings and phone contacts with the specific dialect for the selected country. Parents' task was to tick the words their children understood only, and the words they understood and said. Comprehension and production scores were calculated by simply adding up the number of times the parent, respectively, ticked "Understand only" or "Understands and says." Data are presented in Figure 1, and visual inspection suggested no obvious outliers amongst the different groups of participants.

Initial data collection with the full form Arabic CDI words only in all dialects

Following the recent standardization of the UK-CDI (Alkhamra & Al–Jazi, 2016) with an online platform, as well as the CDI-based UKBT bilingual norms at age 2 (Floccia et al., 2018), we opted for an online data collection to establish norms for the Arabic CDI Words Only (this assumption was further verified in step 2–2). This was further justified by the geographical spread of the target countries. A bespoke website was designed by University technicians, where parents could sign up using a unique and secure link and provide information about their children's vocabulary knowledge. The website was available in English and MSA. For signing up, personal information about only one parent (the one filling in the questionnaire), at least, was mandatory (e.g., contact details, parent's nationality, parent's dialect, country of current residence, education, occupation, and optional annual income). The decision not to request mandatory information about the second caregiver/parent (if applicable) was made to ensure maximum participation from the parent filling in the CDI. Some demographic information about the child was also collected (date of birth, gender, nationality, dialect of use, country of birth, country of current residence) together with background information. This included questions related to the primary caregiver, the number and age of siblings (if any), whether the child had any hearing problems, developmental delay, or was born more than 6 weeks premature (these last three questions are typically found in all CDIs). We also asked whether the child was exposed to an additional language inside home or outside home (in each case, which one, spoken by whom, and for how many hours per week). Modern Standard Arabic was used for all information provided to and by the parent preceding the selection and completion of the dialect-specific CDI. After

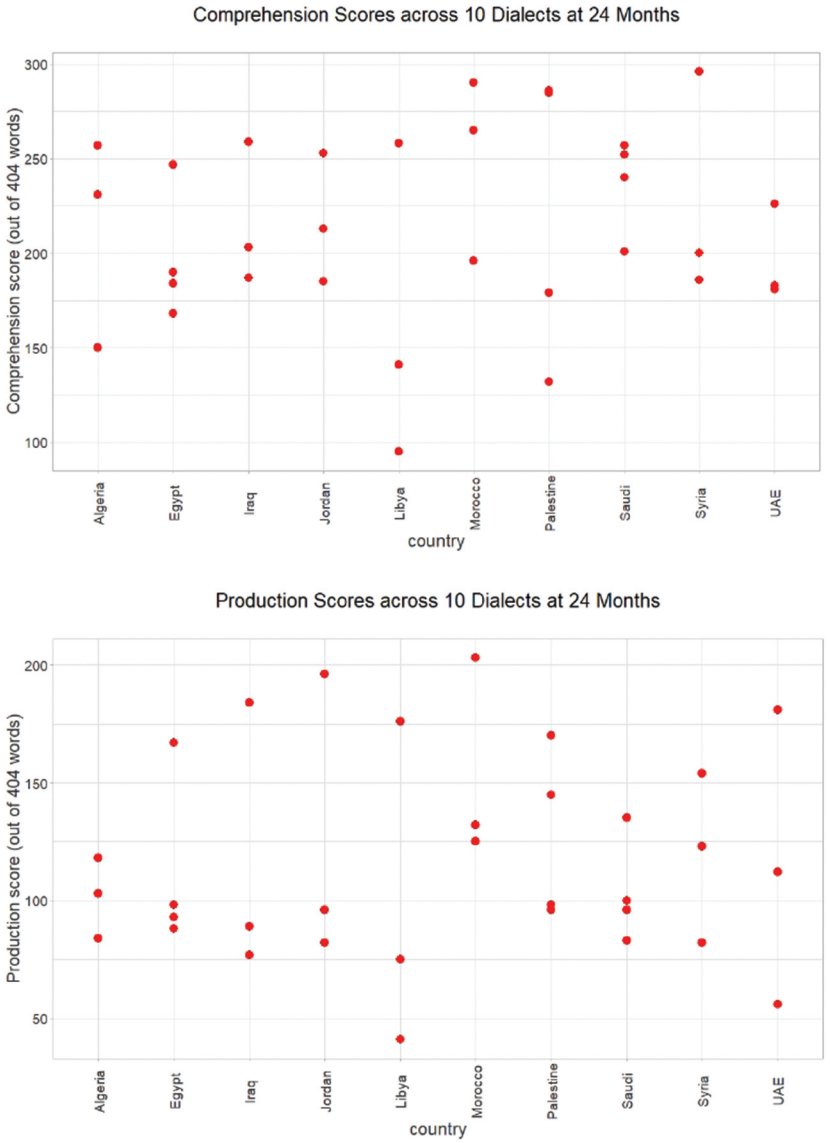


Figure 1. Mean scores for comprehension (top) and production (bottom) across 10 Arab countries with a sample of 24-month-olds (N = 33). Each dot represents a child.

providing this information and selecting the child's dialect, parents were presented with the list of 404 words divided into 19 categories (and presented in an alphabetical order within each category), translated into the child's dialect (based on the dialect used by the child). Parents were asked to tick the words their children only understood and the words they understood and said. As a thank you, after completing the word list, a word cloud was sent by e-mail to the parent to provide them with a memory of the words the child understood and/or said at that age. All data entered by the parent was automatically stored in a secure database that only authors could access. All information provided was further processed anonymously and confidentially.

A total of 343 participants from the 17 selected countries signed up on the initial Full Form Arabic CDI Words Only platform, of whom 205 participants completed the word list. The

other 138 were excluded either because the parent did not complete the list or because the child was aged younger than 8 months or older than 30 months at the time of conducting the study. Out of these 205 participants, 50 participants provided useful data relating to the inclusion criteria. Examples of participants who did not provide useful data are those who were exposed to another language than Arabic for more than 10 hours per week (134 children), following a very similar criterion used by Fenson et al. (2007, pp. 55, more than 12 hours a week). To be included, children also needed to be living in one of the target Arab countries and speak an Arabic dialect consistent with their country of origin (64 children were excluded for living in non-Arabic country and seven children excluded for speaking another foreign language). Expectedly, a positive correlation between age and CDI scores was found for comprehension (r (50) = 0.63, p < .001) and production (r (50) = 0.52, p < .001). Also, expectedly, a strong positive correlation was found between children's comprehension and production (r (50) = 0.78, p < .001) as children with higher performance in comprehension had also high scores in production and vice versa.

Concurrent validity of the full form Arabic CDI words only

Concurrent validity was used to check the validity of the Arabic CDI against the Arabic Language Test (Rifaie, 1994) used as a gold standard. This was done in Egyptian Arabic only, as this was the language of development of the Arabic Language Test.

The Arabic Language Test was developed to provide a broad picture of the child's expressive and receptive language in addition to assessing the child's semantics, pragmatics, and prosody. The test, administered in a face-to-face interaction with the child, is suitable for the age of 2 to 8 years and takes between 45 and 60 minutes. Twenty-three typically developing Egyptian healthy children (13 male, 10 female) with a mean age of 24 months took part in this study. The children were recruited from different Northern Egypt districts using word of mouth from the first author's personal contacts. They were raised in comparable middle- to higher-class backgrounds. The CDI was completed before the Arabic Language Test, or concurrently while the researcher administered the Arabic Language Test in an adjacent room. Pearson's correlations showed that both comprehension and production scores as measured by the Full Form Arabic CDI Words Only strongly and positively correlated with the raw comprehension and production scores measured by the Arabic Language Test as seen in Table 1.

The strong positive correlation between the two tests suggests that they measure similar abilities. Given that we established the Full Form Arabic CDI Words Only validity in the Egyptian dialect at age 2, we can reasonably assume that the translated versions in the other 16 dialects can be validly used to assess the vocabulary development of children from the other Arab countries. We can also assume that this validity extends to the Short Form CDI, which will be described below.

Given that most of the participants seemed not to complete the CDI due to its length, we decided to shorten it into 100 words using data collected on the initial full form. The development of short CDI forms is increasingly favored (Fenson et al., 2000; short Oxford CDI: Floccia et al., 2018; short forms of the Kilifi CDIs: Alcock et al., 2015; short form of the Spanish CDI: Jackson-Maldonado et al., 2013), to

Table 1. *Pearson correlations between vocabulary knowledge scores in comprehension and production (N = 23) as measured by the Egyptian version of the full form Arabic CDI words only and the Arabic language test (*Rifaie, 1994).

	CDI comp	CDI prod.	Lang Test comp.	Lang Test prod.
CDI comp.	1			
CDI prod.	.574**	1		
Lang Test comp.	.653**	.783**	1	
Lang Test prod.	.511*	.795**	.894**	1

*p < .05, **p < .01

provide greater engagement from parents, which is essential to obtain an accurate measure of the child's achievements.

Egyptian dialect study

In the second part of the study, we focused on the Egyptian dialect, in which we tested test–retest reliability and effect of completion form (paper versus online), before moving on to the main data collection leading to the norms.

Development of the short form of the Arabic CDI words only in all dialects

On the model of the development of the Short Form of the Oxford CDI version (Floccia et al., 2018), we selected 100 words based on their frequency ranges (that is, how often they were reported as being known) as derived from the 50 participants' data collected on the Full Form Arabic CDI Words Only (see Step 1.3 above). The Short Form included the same 19 categories (names of animals, verbs, etc) but with fewer words in each category. The 50 children were binned in 3 age groups: 8.13 to 16.6 months (16 children: 1 Sudanese, 1 Libyan, 7 Syrians, 7 Egyptians), 16.7 to 23.3 months (17 children: 9 Egyptians, 1 Bahraini, 3 Jordanians, 1 Algerian, 3 Syrians) and 23.4 to 30.5 months (17 children: 4 Syrians, 11 Egyptians, 1 Sudanese, 1 Algerian). Then using the middle age bin (17 children) where vocabulary scores were the most informative (e.g., less susceptible to floor or ceiling effects), i.e., all 404 words were rank-ordered according to the number of times parents affirmed their child's knowledge of the word. We divided this list into deciles, and selected 10 words from each decile. In doing so, we kept in each decile the same proportion of nouns, adjectives, verbs, and function words that was present in the full list. Then we verified whether the selected words were suitable choices for the other two age bins. For example, the word "egg" was known by 36% of the children in the middle age range, but a substantially smaller proportion (4%) in the younger age range; thus "egg" was excluded. In contrast, the word "milk", which was known by 30% of the children in the middle age range, was known by 20% of the children in the young age range, a difference more commensurate with the overall effect of age, and therefore it was included.

Test-retest and reliability of online versus paper form completion for the Egyptian version of the short form Arabic CDI words only

After shortening the Full Form Arabic CDI Words Only into a 100-word list translated into 17 dialects (see Step 2.1), the test–retest reliability of this new list was assessed, together with the comparison of the modes of completion (online vs. paper). A new set of participants were asked to fill in the 100-word CDI on paper, and then online (or vice versa), with a time period of 2 to 3 weeks separating the completion of the two versions. Twenty-one parents of typically developing Egyptian healthy children (M = 23.71 months, SD = 5.77, 11 females and 10 males) took part. The mean score of the children's comprehension when the CDI was filled in online was 76.38 (SD = 24.95) as compared to 77.48 (SD = 24.88) on paper, while the mean score of their online production was 49.24 (SD = 26.75) as compared to 50.10 (SD = 26.81) on paper. The children were recruited through the first author's contacts from different Northern Egypt districts and belonged to middle to high social class.

First, a strong correlation was found between the child's age and vocabulary comprehension and production as measured by the CDI paper version (respectively, r (21) = .67, p = .001; r (21) = .71, p = .0001) and by the online version (respectively, r (21) = .62, p = .003; r (21) = .72, p < .001). When filling in the CDI on their second attempt, parents dropped on average 1.76 words in comprehension (SD 1.92) and 1.19 words in production (SD 1.47). At the same time, they also added words which they had not previously reported: 2.10 words in comprehension (SD 2.72) and 1.19 words in production (SD 2.56).

Second, a repeated measure ANOVA was conducted on comprehension scores with two factors: order of completion as a between subject variable (online version first or second) and mode of completion (online vs. paper) as a within participant variable. Analysis revealed no main effect of the mode of completion, $F(1, 19) = 2.18$, $p = .156$, $\eta^2 = .10$, with a mean score for the paper version at 77.48 (SD 24.88) and 76.38 (SD 24.95) for the online version. No main effect of the order of presentation of the CDI was found, $F(1, 19) = 1.28$, $p = .27$, $\eta^2 = .063$, with a score of 76.90 (SD 24.92) on the first completion versus 76.95 (SD 24.93) on the second. No interaction between modality and order was found, $F(1, 19) < 1$. Similar analysis for production revealed again no main effect of the mode of completion, $F(1, 19) = 1.56$, $p = .23$, $\eta^2 = .076$, with a mean score of 50.10 (SD 26.81) for the paper version and 49.24 (SD 26.75) for the online version. No main effect of the order of presentation of the CDI was found, $F(1, 19) = 2.21$, $p = .15$, $\eta^2 = .104$, with a score of 49.67 (SD 26.53) on the first completion versus 49.67 (SD 27.04) on the second. No interaction between modality and order was found, $F(1, 19) < 1$. To sum up, the test–retest results established the reliability of the Short Form Arabic CDI, in both formats of presentation (online and paper). In line with Kristoffersen et al. (2013), we found collecting data online less time- and resource-consuming than on paper, further justifying our decision to collect data online for the norming study.

Standardization of the short form of the Arabic CDI words only in Egyptian Arabic

The bulk of data collection could then start using the Short Form of the Arabic CDI Words Only, using the same online platform as described in Step 1.3, with the aim of standardizing the Arabic CDI in Egyptian Arabic as a starting point, and examining the factors that would predict language development in Arabic-speaking children aged from 8 to 30 months. A simplified website providing online and downloadable copies of all dialect forms, together with Egyptian norms, is now accessible at http://www.psy.plymouth.ac.uk/OpenArabicCDI/.

Participants

Data were collected from 629 Egyptian participants in total, out of whom 436 participants provided usable data. Ninety-seven children belonged to the age bin of 8 to 12 months, 129 children to the age bin of 13 to 20 months, and 210 children to the age bin of 21 to 30 months (see Table 2). Children were excluded either because the parent did not complete the CDI list of words (n = 180) due to not pressing the "submit" button, or because the child had one or more of the exclusion criteria (n = 13) such as hearing problems, developmental delays, prematurity, aged younger than 8 months and older than 30 months at the time of conducting the study, exposed to a non-Arabic language more than 10 hours in total per week (both inside and outside their home), and living in non-Arabic-speaking countries.

The mean age of the 436 participants was 19.82 months (SD = 7.18) and the group comprised 233 girls and 203 boys. Socio-economic status of the family was obtained on a scale from 0 (education at primary school level) to 4 (postgraduate studies) based on the highest level of educational attainment of the parent who filled in the CDI. Given that most of the parents had at least a degree and very few parents were in the two lowest SES levels (n = 4 for SES rating of 0, and n = 5 for SES rating of 1), scores were collapsed into two categories, with high SES being defined as parents with a degree or advanced diploma (rating of 3 and 4), and low SES defined as those whose highest educational attainment was below a degree (rating from 0 to 2). Overall, 81 participants were categorized as low SES versus 355 participants from high SES. The CDI was filled by 271 mothers and 165 fathers. Analyses with four levels of SES (with 0 and 1 collapsed into a single category) are provided in the Appendix.

Procedure

Most children's data were collected through nurseries (for 390 out of 436 children) by sending a link to the website to nurseries, who forwarded it to the parents of children within the age range of 8 to

Table 2. Descriptive statistics for girls and boys per age bin in the Egyptian group (n = 436).

Age (In months)	Female	Male	Total
8	0	3	3
9	16	15	31
10	16	7	23
11	10	11	21
12	13	6	19
13	14	9	23
14	7	8	15
15	12	8	20
16	7	7	14
17	11	5	16
18	5	5	10
19	7	6	13
20	9	9	18
21	5	8	13
22	6	9	15
23	2	5	7
24	8	9	17
25	16	13	29
26	13	9	22
27	11	10	21
28	9	5	14
29	16	15	31
30	20	21	41
Total	233	203	436

30 months. The remaining data were collected through social media by publishing the link on social groups. Parents who filled in the CDI through five nurseries were paid £3 per participant. The Egyptian private nurseries were located in the North of Egypt in major cities, such as Cairo, Mansoura, and Damietta.

Results

SES and gender of the child participant were both contrast coded before analysis. The Egyptian comprehension and production data were modeled with separate quasi-binomial models[1] which took the participant's age in months, gender, and SES status as factors. The dependent variable was the number of words either comprehended, or comprehended and produced. The model output can be seen in Tables 3 and 4 respectively. Model fit to the data was assessed with a half-normal plot of the residuals. Because the models are run with a logistic link function, the estimate can be understood to be the change in log odds from baseline for each parameter. In the case of the comprehension data, as an example, the log odds will change by 0.15 (the estimate) for every unit increase in age. The model outputs indicate an effect of age on child receptive and productive vocabulary, but no evidence of an effect of SES status or child gender. Figure 2 demonstrates clear improvement in comprehension and production score with age in the SES categories with enough points to reliably calculate the slope (see Appendix for further breakdown of this effect).

The model fit to the data can be seen in Figure 2, where the comprehension data occupy the top panel, and the production data occupy the bottom panel. Minimal differences in SES and gender can be observed.

In order to separately assess whether parental gender attitudes to word learning affected how they filled out the CDI, separate models were run that retained the predictors used above, but adding in

[1]The quasi-binomial model in this sense is a binomial model that allows for extra dispersion – necessary due to the variability in the data. The means do not change from those predicted by a binomial model. Thus, these data can also be modeled with binomial models, but the effects of these models would be inflated; so, the quasi-binomial model is the more conservative option with these data.

Table 3. Fixed effects of quasi-binomial model on comprehension scores from the Egyptian data.

	Estimate	Std. Error	*t* value	Pr(>|t|)
(Intercept)	−3.35	0.22	−14.906	<0.001
Age	0.15	0.01	14.793	<0.001
SES	0.34	0.45	0.755	0.451
Gender	0.18	0.45	0.396	0.693
Age:SES	−0.01	0.02	−0.578	0.563
Age:Gender	0.00	0.02	−0.174	0.862
SES:Gender	0.07	0.90	0.076	0.94
Age:SES:Gender	−0.02	0.04	−0.418	0.676

Table 4. Fixed effects of quasi-binomial model on production scores from the Egyptian data.

	Estimate	Std. Error	*t* value	Pr(>|t|)
(Intercept)	−4.95	0.30	−16.283	<0.001
Age	0.18	0.01	14.411	<0.001
SES	0.28	0.61	0.453	0.651
Gender	0.49	0.61	0.809	0.419
Age:SES	−0.01	0.03	−0.346	0.729
Age:Gender	−0.02	0.03	−0.756	0.45
SES:Gender	−1.08	1.22	−0.89	0.374
Age:SES:Gender	0.04	0.05	0.698	0.486

Table 5. Predicted proportions of words from the short form Arabic CDI words only comprehended at each 10th percentile (column) and each age group (row).

Age (months)	10	20	30	40	50	60	70	80	90
8	0.04	0.05	0.06	0.08	0.09	0.09	0.11	0.14	0.16
9	0.04	0.06	0.07	0.09	0.10	0.11	0.13	0.17	0.20
10	0.05	0.07	0.08	0.10	0.12	0.13	0.16	0.20	0.25
11	0.06	0.08	0.09	0.12	0.14	0.15	0.18	0.23	0.30
12	0.07	0.09	0.11	0.14	0.16	0.18	0.21	0.27	0.36
13	0.08	0.10	0.13	0.16	0.18	0.21	0.25	0.31	0.42
14	0.09	0.12	0.15	0.18	0.21	0.24	0.29	0.36	0.48
15	0.10	0.14	0.17	0.21	0.24	0.28	0.33	0.40	0.55
16	0.11	0.16	0.19	0.24	0.27	0.31	0.37	0.45	0.61
17	0.13	0.18	0.22	0.27	0.31	0.36	0.42	0.51	0.67
18	0.15	0.20	0.25	0.31	0.35	0.40	0.47	0.56	0.72
19	0.17	0.23	0.29	0.34	0.39	0.44	0.52	0.61	0.77
20	0.19	0.26	0.32	0.38	0.43	0.49	0.56	0.65	0.82
21	0.21	0.29	0.36	0.42	0.47	0.54	0.61	0.70	0.85
22	0.24	0.32	0.40	0.47	0.52	0.58	0.66	0.74	0.88
23	0.26	0.36	0.44	0.51	0.56	0.63	0.70	0.78	0.91
24	0.29	0.40	0.49	0.55	0.60	0.67	0.74	0.81	0.93
25	0.32	0.44	0.53	0.59	0.64	0.71	0.77	0.84	0.94
26	0.36	0.48	0.57	0.63	0.68	0.75	0.81	0.86	0.95
27	0.39	0.52	0.61	0.67	0.72	0.78	0.83	0.89	0.96
28	0.43	0.55	0.65	0.70	0.75	0.81	0.86	0.91	0.97
29	0.46	0.59	0.69	0.74	0.78	0.84	0.88	0.92	0.98
30	0.50	0.63	0.73	0.77	0.81	0.86	0.90	0.94	0.98

a main effect of parental gender. Parental gender did not improve model fit, having no effect on the number of words comprehended ($\chi^2(1) = 1.713$, $p = .191$) or on the number of words produced ($\chi^2(1) = 2.304$, $p = .129$).

Creating norms

The Egyptian data (n = 436) were then used to generate percentile-based norms. At each age group (month by month from 8 to 30 months), the participants were split into nine quantiles,

Table 6. Predicted proportions of words from the short form Arabic CDI words only produced at each 10th percentile (column) and each age group (row).

Age (months)	10	20	30	40	50	60	70	80	90
8	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.04	0.05
9	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.04	0.06
10	0.01	0.01	0.02	0.02	0.03	0.03	0.04	0.06	0.08
11	0.01	0.02	0.02	0.03	0.03	0.04	0.05	0.07	0.10
12	0.01	0.02	0.03	0.03	0.04	0.05	0.06	0.08	0.12
13	0.01	0.02	0.03	0.04	0.05	0.06	0.08	0.10	0.14
14	0.02	0.03	0.04	0.05	0.06	0.07	0.09	0.12	0.18
15	0.02	0.03	0.05	0.06	0.07	0.08	0.11	0.15	0.21
16	0.03	0.04	0.06	0.07	0.09	0.10	0.14	0.18	0.25
17	0.03	0.05	0.07	0.09	0.11	0.13	0.17	0.22	0.30
18	0.04	0.06	0.09	0.11	0.13	0.15	0.20	0.26	0.35
19	0.04	0.07	0.11	0.13	0.16	0.18	0.23	0.30	0.40
20	0.05	0.09	0.13	0.15	0.19	0.22	0.27	0.35	0.46
21	0.06	0.10	0.15	0.18	0.22	0.26	0.32	0.41	0.52
22	0.08	0.13	0.18	0.21	0.26	0.30	0.37	0.46	0.57
23	0.09	0.15	0.21	0.24	0.30	0.35	0.42	0.52	0.63
24	0.11	0.18	0.25	0.28	0.35	0.40	0.47	0.57	0.68
25	0.13	0.21	0.29	0.32	0.40	0.45	0.52	0.63	0.73
26	0.15	0.24	0.33	0.36	0.45	0.51	0.57	0.68	0.77
27	0.18	0.28	0.38	0.41	0.50	0.56	0.63	0.72	0.81
28	0.21	0.32	0.43	0.46	0.56	0.62	0.67	0.77	0.84
29	0.24	0.37	0.48	0.51	0.61	0.67	0.72	0.80	0.87
30	0.28	0.42	0.53	0.56	0.66	0.71	0.76	0.84	0.89

representing every tenth percentile of word comprehension between 10% and 90%. Each quantile was then individually fitted with a logistic regression (as recommended by Fenson et al., 1994), and the predicted values were used as the norms. The same procedure was then repeated for word production. Figure 3 demonstrates the quantile-based norms from the Egyptian Arabic CDI data, for comprehension and production, respectively. Tables 5 and 6 demonstrate the quantile-based norms as a proportion of the 100 words for comprehension and production, respectively. The tables indicate the predicted quantiles, thus an 18-month-old who comprehends 25 words (proportion of .25) from the Short Form of the Arabic CDI Words Only would be in the 30th percentile.

Mixed Arabic study

A similar data collection procedure was followed for the 16 non-Egyptian dialects, and data were compared to the Egyptian sample, using the Short Form of the Arabic CDI Words Only in the relevant dialects.

Participants

A total of 295 participants were tested, out of whom 168 provided usable data ((10 Algerians, 8 Bahrainis, 22 Emiratis, 15 Iraqis, 8 Jordanians, 8 Lebanese, 8 Libyans, 8 Moroccans, 9 Omanis, 10 Palestinians, 10 Qataris, 15 Saudis, 8 Sudanese, 8 Syrians, 12 Tunisians, 9 Yemenis). Twenty nine children belonged to the age bin of 8 to 12 months, 64 children to the age bin of 13 to 20 months, and 75 children to the age bin of 21 to 30 months (see Table 7). Children were excluded either because the parent did not complete the CDI list of words (n = 63) due to not pressing the "submit" button, or because the child had one or more of the exclusion criteria (n = 64) such as hearing problems, developmental delays, prematurity, aged younger than 8 months and older than 30 months at the time of conducting the study, exposed to a non-Arabic language more than 10 hours in total per week (both inside and outside their home), and living in non-Arabic-speaking countries.

The mean age of the 168 participants was 19.26 months (SD = 6.52) and the group comprised 81 girls and 87 boys. As above, given that most of the parents had at least a degree and very few parents

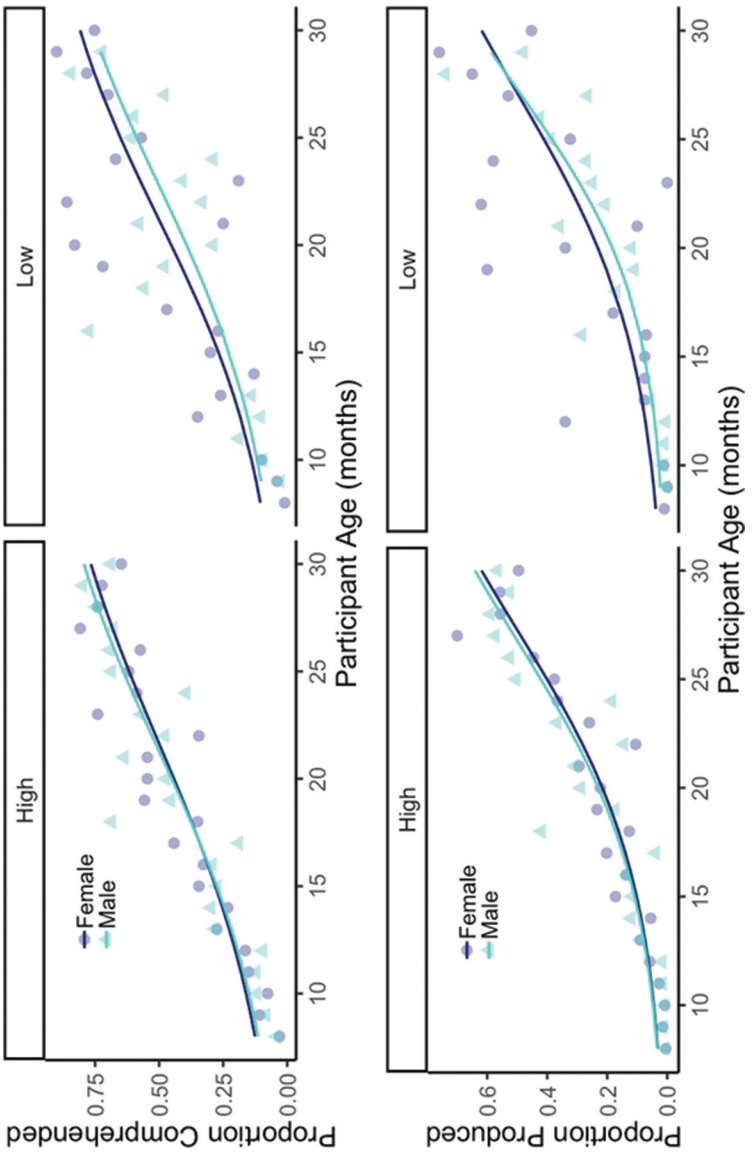


Figure 2. Comprehension and production model fits to raw Egyptian data. Top panel is high and low SES for comprehension data, while bottom panel is high and low SES for production (understands and says) data, with boys and girls represented in blue and purple respectively. Points indicate actual means, lines indicate model fit.

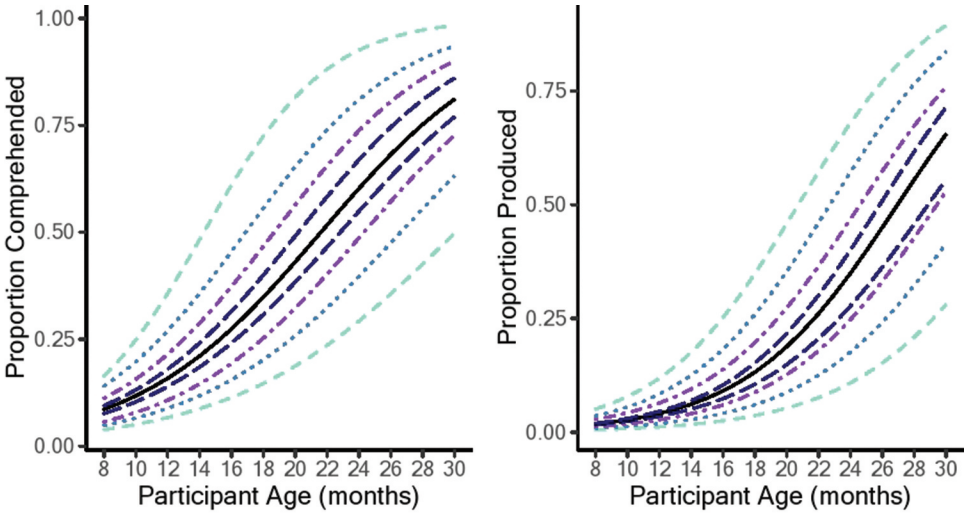


Figure 3. 10th percentile quantiles for word comprehension (left) and word production (right) in the short Arabic CDI Words Only, based on Egyptian data. Black solid line indicates the 50th percentile.

Table 7. Descriptive statistics for girls and boys per age bin and dialect in the other dialects group.

Age bin	Female	Male	Total
8– 12 months	15 (2 Saudi, 1 Algerian, 1 Tunisian, 2 Syrian, 2 Lebanese, 1 Qatari, 2 Omani, 2 Yemeni, 1 Jordanian, 1 Palestinian)	14 (1 Lebanese, 1 Emirati, 1 Qatari, 2 Sudanese, 1 Libyan, 4 Omani, 1 Yemeni, 2 Bahraini, 1 Palestinian)	29
13– 20 months	36 (3 Saudi, 5 Iraqi, 2 Algerian, 4 Tunisian, 2 Moroccan, 2 Lebanese, 5 Emirati, 2 Qatari, 1 Sudanese, 1 Libyan, 1 Omani, 2 Yemeni, 2 Bahraini, 3 Jordanian, 1 Palestinian)	28 (2 Saudi, 2 Iraqi, 4 Algerian, 2 Tunisian, 2 Moroccan, 2 Syrian, 3 Emirati, 3 Qatari, 2 Sudanese, 1 Omani, 1 Yemeni, 2 Bahraini, 1 Jordanian, 1 Palestinian)	64
21– 30 months	30 (1Saudi, 4 Iraqi, 3 Algerian, 3 Tunisian, 2 Moroccan, 2 Syrian, 1 Lebanese, 6 Emirati, 2 Qatari, 3 Libyan, 1 Bahraini, 1 Jordanian, 1 Palestinian)	45 (7 Saudi, 4 Iraqi, 2 Tunisian, 2 Moroccan, 2 Syrian, 1 Lebanese, 8 Emirati, 1 Qatari, 3 Sudanese, 3 Libyan, 1 Omani, 3 Yemeni, 1 Bahraini, 2 Jordanian, 5 Palestinian)	75

were in the two lowest SES levels (n = 4 for SES rating of 0, and n = 2 for SES rating of 1), scores were collapsed into two categories, with high SES being defined as parents with at least a degree or advanced diploma (rating of 3 and 4), and low SES defined as those whose highest educational attainment was below a degree (rating from 0 to 2). Overall, 14 participants were categorized as low SES versus 154 participants from high SES. The CDI was filled by 110 mothers and 58 fathers. Analyses with a 4-point rating scale for SES (with 0 and 1 collapsed into a single category) can be found in the Appendix.

Procedure

Most of the data from non-Egyptians was collected online; however, some nurseries helped to collect data from some non-Egyptian countries (56 out of 168). For the 15 private nurseries contacted in other non-Egyptian countries participant payment was not possible, due to the difficulty of transferring money (Iraq: 4; Algeria: 3; Lebanon: 2; Morocco: 1; UAE: 3; and Qatar: 2).

Table 8. Fixed effects of quasi-binomial model on comprehension scores from other dialects data.

	Estimate	Std. Error	*t* value	Pr(>|t|)
(Intercept)	−3.22	0.54	−5.946	<0.001
Age	0.15	0.03	5.817	<0.001
SES	0.33	1.08	0.305	0.761
Gender	−0.67	1.08	−0.616	0.539
Age:SES	−0.01	0.05	−0.194	0.846
Age:Gender	0.02	0.05	0.323	0.747
SES:Gender	1.31	2.17	0.604	0.547
Age:SES:Gender	−0.03	0.11	−0.285	0.776

Table 9. Fixed effects of quasi-binomial model on production scores from other dialects data.

	Estimate	Std. Error	*t* value	Pr(>|t|)
(Intercept)	−4.38	0.72	−6.108	<0.001
Age	0.15	0.03	4.854	<0.001
SES	−0.04	1.43	−0.028	0.978
Gender	−0.80	1.43	−0.555	0.58
Age:SES	0.01	0.06	0.084	0.934
Age:Gender	0.01	0.06	0.208	0.836
SES:Gender	1.07	2.87	0.374	0.709
Age:SES:Gender	−0.01	0.12	−0.11	0.912

Results

The non-Egyptian data was treated as a single group (Other Dialects) due to the small number of participants in each dialect (maximum 22, minimum 8, mean 10.3). As above, both production and comprehension were modeled separately using quasi-binomial models. The output of these models can be seen in Tables 8 and 9, respectively. As with the Egyptian data, there is strong evidence for an effect of participant age, but no other effects are evident from the data.

The Egyptian and Other Dialects data were compared by re-fitting the models based on the Egyptian data (Tables 5 and 6) to the Other Dialects data. The Root Mean Squared Error (RMSE) was then calculated, and compared to the RMSE of the original model created from the Other Dialects data (Tables 8 and 9). For the comprehension data, the Egyptian model fitted the Other Dialects data reasonably well (RMSE = 0.220), and was only slightly higher than the RMSE of the original, Other Dialects model (RMSE = 0.215). Likewise for the production data, the Egyptian model fitted the Other Dialects data well (RMSE = 0.196), again only slightly higher than the model created from the Other Dialects data (RMSE = 0.189).

Figure 4 demonstrates the model fits to the Other Dialects data, where solid lines indicate the models based on the Other Dialects data itself, while dotted lines indicate the model fit based on refitting the Egyptian model to the Other Dialects data. The top panel shows high and low SES for comprehension, while the bottom panel depicts high and low SES for production. While the fit between the two models appears very close for the high SES group (left panels), a lesser fit is found for the low SES group (right panels), which likely relates to the smaller amount of data points in this sample (1:6 ratio). Overall, differences in trajectory between the two alternative models are minimal, indicating that the Egyptian data may be generalized to the data from the other dialects.

It must be noted for readers unfamiliar with this type of modeling, that the CDI comprehension or production scores analyzed in an ANOVA with age as a covariate, gender (male vs female), SES (high vs low) and dialect group (Egyptian vs Other Dialects) as between participants variables, lead to the same result: a significant effect of age for comprehension (F (1, 595) = 661.6, $p < .001$, $\eta^2 = .525$) but no effect of the other variables (all $Fs < 1$) and no interaction (all $Fs < 1.89$). Similarly, the only significant effect for production was age ($F(1, 595) = 592.8$, $p < .001$, $\eta^2 = .503$) with no other effect (all $Fs < 1$) and no interactions (all $Fs < 1.46$).

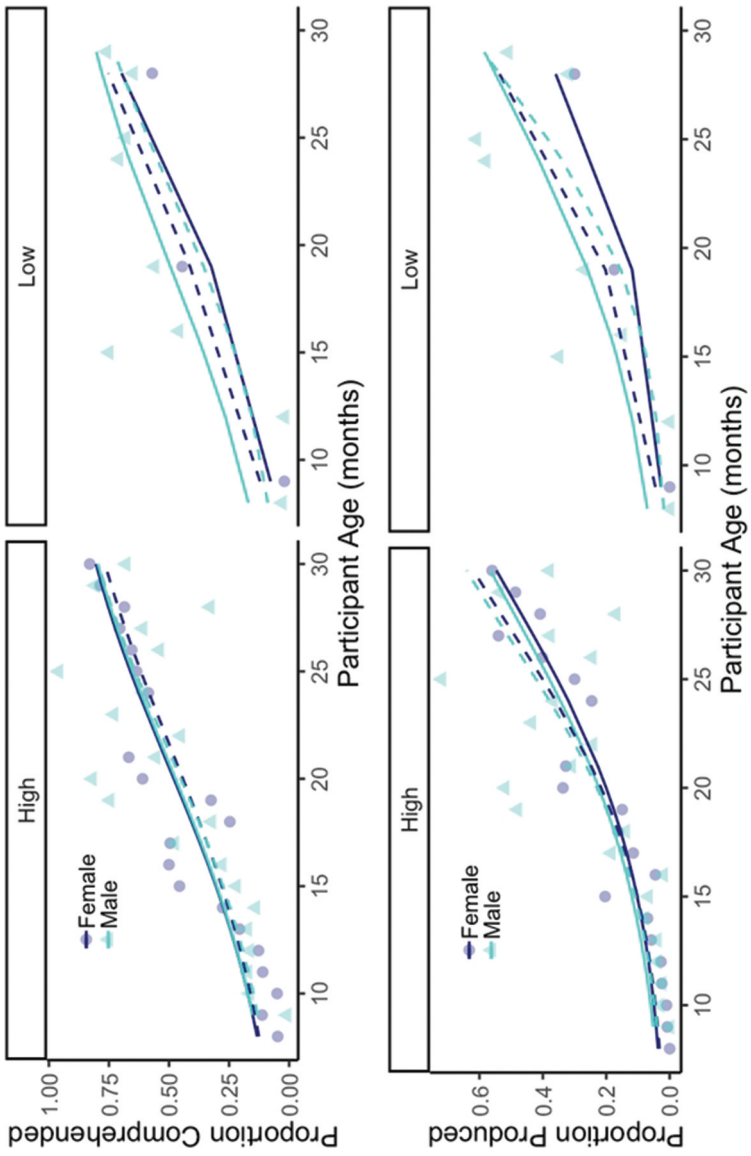


Figure 4. Model fits to data from the non-Egyptian sample. Top panel indicates high and low SES for the comprehension data, bottom panel indicates high and low SES for the production data, with boys and girls represented in blue and purple respectively. Points indicate means, solid lines indicate fit from the models based on the Other Dialects data itself, dotted lines indicate fit from the model based on the Egyptian data, refit to this sample.

Discussion

Following the development and adaptation of the MacArthur CDI in up to 60 languages (see https://mb-cdi.stanford.edu/adaptations.html), we undertook the development of the Arabic CDI Words Only for its use in 17 Arabic countries and corresponding mainstream dialects, for use with infants aged 8 months to 30 months. First, we verified that an initial Full Form Arabic CDI Words Only containing 404 words, culturally adapted from a merging of three English CDIs (MacArthur-Bates CDI: Words and Sentences, Fenson et al., 2007; Oxford CDI; Hamilton et al., 2000; UK-CDI Words and Gestures; Alcock et al., 2017), provided comparable scores of vocabulary knowledge in an initial subset of Arab countries, preliminarily establishing the comparability of the Arabic CDI Words Only for use in the chosen age range and across dialects. Validity was assessed in Egyptian Arabic through a comparison with an Egyptian dialect-specific language test, the Arabic Language Test (Rifaie, 1994). Based on these findings, we concluded that this initial version of the CDI would be valid for assessing comprehension and production in children in the other 16 dialects.

After initiating online data collection in the 17 countries, a high attrition rate alerted us to the fact that we needed a more condensed version the CDI, on the model of the Short Form of the MacArthur Bates CDI (Fenson et al., 2000) or the Oxford CDI (Floccia et al., 2018). The data collected were used to shorten the Arabic CDI Words Only into a 100-word list, leading to the Short Form Arabic CDI Words Only. Test–retest validity and comparison of completion mode (online versus paper) were concurrently assessed, with a strong correlation between the two modes of completion. Finally, the main online data collection was undertaken in Egypt (n = 436) and 16 other countries (n = 168), and norms for the Egyptian dialect calculated.

In the process of standardizing the Short Form Arabic CDI Words Only, we examined the impact of core demographic factors known to affect language development in various children populations, namely SES and child gender. Contrary to our expectations, we did not find any significant impact of SES (assessed based upon parental education) on word comprehension or production in Egyptian nor non-Egyptian groups. It must be noted that in our sample 85% of parents who filled the CDI had at least a degree. A sample bias toward highly educated parents, masking SES effects, is actually often found in CDI studies: in Fenson et al. (1994) who conducted the first standardization of the MacArthur CDI in the US (n = 1130), 77% of parents had reached a level of university education, against 35% in the general American population in those days. No impact of SES on vocabulary skills was found before 3 years of age, with a small effect after that age. Similarly, Hamilton et al. (2000) again found no impact of SES on comprehension nor on production of children when collecting Oxford CDI data in the UK with children (n = 200) aged 1;0 to 2;1, with parents in majority from middle-class background in the affluent and educated area of Oxford (in 2011 43% of the Oxford population had a degree or above, compared to 27% in England; Oxford City Council). In contrast, in the most recent standardization of the MacArthur-Bates CDI (Fenson et al., 2007), 68% of parents had had some university education, against 58% in the general population, and then a modest effect of SES accounted for 0.6% of the variance in production, mainly due to children aged 21 months and older.

It is, therefore, likely that the lack of SES effect observed here is due to a sampling bias, augmented by the fact that in our self-selected sample, parents may have been, irrespective of their own education level, actively engaged in their child's language development, as is probably the case on other CDI norming studies; alternatively, SES effects on the size of the lexicon may appear only later in development (e.g., for Arabic: Aram et al., 2013, in kindergarten; Korat et al., 2013, in 5- to 6-year-olds). This is clearly a limitation to this research, and future development of the Arabic CDI Words Only should involve targeting a wider representation of families across all SES strata.

Our second finding is that we did not observe any impact of gender on the development of vocabulary knowledge in Arabic-speaking children. Large-scale studies in other cultures or countries have reported, rather consistently, an advantage of girls over boys in early language production (e.g., Bornstein & Cote, 2005; Eriksson et al., 2012; Huttenlocher et al., 1991). Before concluding that we

uncovered a cultural difference in the way Arabic-speaking girls and boys acquire language, we need to consider the possibility that the absence of gender effect is due to (1) the use of the CDI, (2) the socio-economic profile of our sample, and/or (3) the age range. Regarding the first point, it is worth mentioning that for CDI norming studies, Fenson et al. (1994) found out that gender only accounted for 1%-2% of the variance in vocabulary scores, with females scoring slightly higher than males in all measures. However, Floccia et al. (2018), using adapted versions of the MacArthur-Bates CDI in various languages, found a larger effect size, with gender effect amounting to 5% of variance in production scores in 24-month-old bilingual children (see also Eriksson et al., 2012). So, it is unlikely that the use of the CDI is solely responsible for the absence of a gender effect with the Arabic CDI Words Only.

Another possibility is that the absence of gender effect is related to the SES profile of the sample. Zambrana et al. (2012) found that between 18 and 36 months of age, gender differences in relation to language comprehension development are attenuated – but not erased – by a high level of maternal education. Furthermore, the distribution of SES in the current study, skewed toward highly educated parents, could have minimized the expected gender effect.

Finally, it is also possible that the children in our sample were too young for any substantial gender effect to be revealed. In favor of this are the findings of Marjanovič-Umek and Fekonja-Peklaj (2017) who conducted a meta-analysis of Slovenian data where language was measured using different tools, including CDIs. They found that the effect size of gender in word production substantially increased with age, with small to null effect sizes in the age range of 8 to 30 months. This contradicts, however, the CDI data analyzed in Eriksson et al.'s study (Eriksson et al., 2012) with multiple languages, where a substantial effect of gender in production was found to emerge as early as between 14 and 16 months.

In sum, the absence of a gender effect in Arabic-learning children tested in this current study could be attributed to the socio-demographic profile of the sample and to its low age range, which could have conspired to mask any emerging gender effect in word production. Further research will be necessary to clarify this issue, before drawing any conclusion related to cultural differences in language development in Arabic-speaking communities and Western/Asian populations, which we alluded to in the introduction.

One important contribution of this paper is the finding from the modeling analyses that vocabulary growth in Egyptian children is very similar to that of children from 16 other Arabic dialects pooled together, including dialects where lexical and phonological variations can hinder mutual understanding (e.g., Morocco versus Gulf dialects). It provides proof of concept showing feasibility of developing a potentially valid tool for measuring spoken language development in multiple dialects of Arabic, and these data show promise for the pan-Arabic approach.

Usually, comparisons across dialects of the same language are difficult because of the use of different CDIs. For example, Mexican Spanish and European Spanish have two different CDIs (respectively, Jackson-Maldonado et al., 2003; López Ornat et al., 2005); inspection of vocabulary curves in the multi-language comparison paper by Bleses et al. (2008) shows that between 8 and 15 months, Mexican children outperform European children in production and comprehension, while for production, the reverse is found between 16 and 30 months. It is not clear whether these differences are due to the use of two different instruments, or whether it relates to more fundamental causes such as demographic, cultural or linguistic variables.

There is a limited amount of CDI development for two dialects or languages where similar or close forms were used. The Oxford CDI (Hamilton et al., 2000), developed from the American MacArthur-Bates CDI, revealed that American toddlers consistently over-performed their British peers between 12 and 24 months in production, and between 12 and 16 months in comprehension, which has been tentatively attributed to cultural differences in parental reporting (Hamilton et al., 2000) or infant-directed speech properties (Floccia et al., 2016). In a similar vein, comparing the vocabulary scores of Cantonese and Mandarin children with highly similar CDIs, Tardif et al. (2009) found a robust advantage of Mandarin children between 8 and 30 months. This was partially attributed to the predominance of monolingualism and first-born children in the Mandarin group, but also to

phonological and phonotactic differences between the two dialects/languages, such as a larger number of syllables and tones in Cantonese.

In sum, it is not clear whether differences, when they emerge, are attributable to the linguistic characteristics of the dialects, or to cultural and demographic differences. The geographical widespread of the 17 Arabic countries where data were collected in the current study necessarily calls for a great many cultural and demographic specificities, and even if our current data suggest that Egyptian norms are not distinguishable from the sample collected in the remaining 16 countries, further research will be necessary to estimate fine-grained comparability of the Arabic CDI use across these varied dialects.

Regarding the clinical use of the Short Form Arabic CDI Words Only, which should be limited to children learning Egyptian Arabic, percentile scores (Tables 5 and 6) computed on the basis of age can be used to identify children at the 15th percentile or lower in comprehension and production, bearing in mind that caution should be exercised when applying these norms for children whose parents have a very low education, given the sampling bias of our data. In addition to genuine SES effects directly related to the nature of the input, parents from low SES may sometimes either over or under estimate their children's vocabulary (Feldman et al., 2000), possibly because they can conflate expressive and receptive vocabulary, leading to providing excessively high scores for their children's vocabulary (Reese & Read, 2000). Following a common practice recommended by Rescorla (1989) known as the Delay 3 cutoff, referring for further assessment any (American English) 2-year-old child who produces fewer than 50 words from the Language Development Survey (which contains 310 words), identifies about 15% of children as being at risk of language delay. In our data, this would mean that 2-year-olds would understand fewer than 40 words of the Short Form Arabic CDI Words Only (20th percentile) and produce fewer than 18 words (20th percentile). However, it must be remembered that a delayed onset and slow growth of vocabulary under the age of 3 is not a reliable predictor of later language disorders – although most children who will be diagnosed with a language disorder will have shown protracted development in early childhood. Therefore using the Short Form Arabic CDI Words Only to clinically assess, a child should lead to a "watchful waiting" approach in case of a low score, rather than to a valid diagnosis.

To conclude, we have developed an easy-to-use, freely available, language tool that researchers, parents and Early Years professionals can use in Egypt and potentially in widespread geography (16 countries), to assess the language skills of children between 8 and 30 months, and to identify those at risk of language delay. This parental report tool does not take more than 10 minutes to complete, and provides at this point reliable estimates in Egyptian Arabic, with a reasonable extrapolation to the other 16 dialects. This unique pan-Arabic approach to language development will need to be consolidated in future research, through the demonstration that norms from Egypt can apply to countries with different demographic (e.g., higher prevalence of multilingualism as in Lebanon or the UAE), cultural (e.g., more traditional upbringing as in Saudi Arabia or Yemen), and linguistic characteristics (e.g., less lexical overlap as in Morocco). The Short Form Arabic CDI Words Only can be freely accessed by parents, practitioners and researchers at http://www.psy.plymouth.ac.uk/OpenArabicCDI/.

Acknowledgments

This paper is dedicated to the memory of Dr Sabah Safi who initiated the Arabic adaptation of the MCDI in Saudi Arabia. We would like to thank Dr Soumaya Bourgou who agreed to let us develop a Tunisian version in parallel to her own. Many thanks to Martyn Atkins, Prof Philip Dale, and to all parents and nurseries in the Middle-East who contributed to data collection.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was funded by an Egyptian Government PhD grant awarded to the first Author.

ORCID

Alshaimaa Gaber Salah Abdelwahab http://orcid.org/0000-0002-7853-223X

References

Abdalla, F., & Shaalan, S. *Kuwaiti Arabic adaptation of the CDI*. Unpublished script. https://mb-cdi.stanford.edu/adaptations.html

Abo Ras, Y., Aref, S., El-Raghy, A., Gaber, O., & El-Maghraby, R. (2010). *Comprehensive Arabic language test (CALT) as a tool for assessing delayed language impaired Egyptian children*. Alexandria University, Alexandria. Retrieved November 27, 2019 from http://www.alexorl.edu.eg/alexorlfiles/pptorl2009/014003.pdf

Al-Akeel, A. I. (1998). *The acquisition of Arabic language comprehension by Saudi children*. PhD Dissertation., Newcastle University.

Alcock, K. J., Meints, K., & Rowland, C. F. (2017). *UK-CDI words and gestures – Preliminary norms and manual*. Unpublished manuscript. http://lucid.ac.uk/ukcdi

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*(4), 763–785. https://doi.org/10.1017/S0305000914000403

Alkhamra, R. A., & Al-Jazi, A. B. (2016). Validity and reliability of the Arabic token test for children. *International Journal of Language & Communication Disorders*, *51*(2), 183–191. https://doi.org/10.1111/1460-6984.12198

Al-Walaie, M. A., & Khan, M. B. (2017). Arabic dialects classification using text mining techniques. *Proceedings of the 2017 International Conference on Computer and Applications (ICCA)*, 325–329. Dubai.

Aram, D., Korat, O., Saiegh-Haddad, E., Arafat, S. H., Khoury, R., & Elhija, J. A. (2013). Early literacy among Arabic-speaking kindergartners: The role of socioeconomic status, home literacy environment and maternal mediation of writing. *Cognitive Development*, *28*(3), 193–208. https://doi.org/10.1016/j.cogdev.2012.10.003

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*(3), 619–650. https://doi.org/10.1017/S0305000908008714

Bornstein, M. H., & Cote, L. R. (2005). Expressive vocabulary in language learners from two ecological settings in three language communities. *Infancy*, *7*(3), 299–316. https://doi.org/10.1207/s15327078in0703_5

Bornstein, M. H., Hahn, C. S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267–304. https://doi.org/10.1177/0142723704045681

Bourgou, S. *Tunisian Arabic adaptation of the CDI*. Unpublished manuscript. https://mb-cdi.stanford.edu/adaptations.html

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldman, H. M., Shriberg, L. D., Sabo, D. L., & Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development*, *74*(2), 346–357. https://doi.org/10.1111/1467-8624.7402002

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant form). *Journal of Child Language*, *30*(1), 213. https://doi.org/10.1017/S0305000902005482

Crais, E. R. (2011). Testing and beyond: Strategies and tools for evaluating and assessing infants and toddlers. *Language, Speech, and Hearing Services in Schools*, *42*(3), 341–364. https://doi.org/10.1044/0161-1461(2010/09-0061)

Dashash, N., & Safi, S. (2008). *Investigating lexical development of Hijazi infants and toddlers using the Arabic version of the MCDI. Paper presented at ASHA Convention. Chicago IL.*

Dwairy, M., Achoui, M., Abouserie, R., Farah, A., Sakhleh, A. A., Fayad, M., & Khan, H. K. (2006). Parenting styles in Arab societies: A first cross-regional research study. *Journal of Cross-Cultural Psychology*, *37*(3), 230–247. https://doi.org/10.1177/0022022106286922

Endendijk, J. J., Groeneveld, M. G., Bakermans-Kranenburg, M. J., Mesman, J., & Christiansen, H. (2016). Gender-differentiated parenting revisited: Meta-analysis reveals very few differences in parental control of boys and girls. *PLoS One*, *11*(7), e0159193. https://doi.org/10.1371/journal.pone.0159193

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., Marjanovič-Umek, L., Gayraud, F., Kovacevic, M., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, *30*(2), 326–343. https://doi.org/10.1111/j.2044-835X.2011.02042.x

Essa, L. H., & Armon-Lotem, S. . *Palestinian Arabic adaptation of the CDI*. Unpublished script. https://mb-cdi.stanford.edu/adaptations.html

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasdy, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Development*, *71*(2), 310–322. https://doi.org/10.1111/1467-8624.00146

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M., Mervis, C. B., & Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, *59*(5), i+iii–v+1-185. https://doi.org/10.2307/1166093

Fenson, L., Marchman, V., Thal, D., Reznick, J., & Bates, E. (2007). *MacArthur- Bates communicative inventories (CDI): Words and gestures; words and sentences*. Paul Brookes.

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, *21*(1), 95–116. https://doi.org/10.1017/S0142716400001053

Floccia, C., Keren-Portnoy, T., DePaolis, R., Duffy, H., Delle Luche, C., Durrant, S., White, L., Goslin, J., & Vihman, M. (2016 Mar). British English infants segment words only with exaggerated infant-directed speech stimuli. *Cognition*, *148*, 1–9. https://doi.org/10.1016/j.cognition.2015.12.004

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., & Plunkett, K. (2018). Vocabulary of 2–year–olds learning English and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*(1), 1–135. http://hdl.handle.net/10026.1/10044

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209

Grigore, G., & Bițună, G. (2016). Arabic varieties: Far and wide. *Proceedings of the 11th International Conference of AIDA – Bucharest, 2015*. Bucharest: Editura Universității din București. ISBN 978-606-16-0709-9.

Hamilton, A., Plunkett, K., & Schafer, G. (2000). Infant vocabulary development assessed with a British Communicative Development Inventory. *Journal of Child Language*, *27*(3), 689–705. https://doi.org/10.1017/S0305000900004414

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H Brookes Publishing.

Heilmann, J., Weismer, S. E., 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. https://doi.org/10.1044/1058-0360(2005/006)

Hoff, E., Laursen, B., Tardif, T., & Bornstein, M. (2002). Socioeconomic status and parenting. In M. H. Bornstein (Ed.), *Handbook of parenting volume 2: Biology and ecology of parenting* (pp. 231–252). Lawrence Erlbaum Associates.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236. https://doi.org/10.1037/0012-1649.27.2.236

Iqbal, F., & Riad, N. (2004). *Increasing girls' school enrolment in the Arab republic of Egypt. World Bank: Social and economic development department, Middle East and North Africa region: Washington, DC*. Retrieved June 12, 2019, from http://documents.worldbank.org/curated/en/159901468770075929/pdf/308340EGY0Girl1ion01see0also0307591.pdf

Jackson-Maldonado, D., Marchman, V. A., & Fernald, L. C. (2013). Short-form versions of the Spanish MacArthur–Bates communicative development inventories. *Applied Psycholinguistics*, *34*(4), 837–868. https://doi.org/10.1017/S0142716412000045

Jackson-Maldonado, D., Thal, D., Marchman, V., Newton, T., Fenson, L., & Conboy, B. (2003). *MacArthur Inventarios del Desarrollo de Habilidades Comunicativas. User's guide and technical manual*. Paul H. Brookes.

Khamis-Dakwar, R., Froud, K., & Gordon, P. (2012). Acquiring diglossia: Mutual influences of formal and colloquial Arabic on children's grammaticality judgments. *Journal of Child Language*, *39*(1), 61–89. https://doi.org/10.1017/S0305000910000784

Kim, S. W., Jeon, H. R., Park, E. J., Kim, H. I., Jung, D. W., & Woo, M. R. (2014). The usefulness of MB CDI-K short form as screening test in children with language developmental delay. *Annals of Rehabilitation Medicine*, *38*(3), 376. https://doi.org/10.5535/arm.2014.38.3.376

Korat, O., Arafat, S. H., Aram, D., & Klein, P. (2013). Book reading mediation, SES, home literacy environment, and children's literacy: Evidence from Arabic-speaking families. *First Language*, *33*(2), 132–154. https://doi.org/10.1177/0142723712455283

Kristoffersen, K. E., Simonsen, H. G., Bleses, D., Wehberg, S., Jørgensen, R. N., Eiesland, E. A., & Henriksen, L. Y. (2013). The use of the Internet in collecting CDI data–an example from Norway. *Journal of Child Language*, *40*(3), 567–585. https://doi.org/10.1017/S0305000912000153

Kuhl, P. K., Tsao, F. M., & Liu, H. M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences*, *100*(15), 9096–9101. https://doi.org/10.1073/pnas.1532872100

Lam, C. C. (2006). Child assessment service in Hong Kong. In R. H. L. Mak, C. C. Y. Ho, & M. M. Y. Wong (Eds.), *A primer in common developmental disabilities: Experience at child assessment service, Hong Kong* (pp. 1-16). Child Assessment Service, Department of Health, Hong Kong SAR Government.

López Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Mart´ınez, M. (2005). Inventario de Desarrollo Comunicativo MacArthur. Adaptaci´on espa˜nola. *[The MacArthur Communicative Development Inventory. Adaptation to Spanish]*. TEA.

Mancilla-Martinez, J., Gámez, P. B., Vagh, S. B., & Lesaux, N. K. (2016). Parent reports of young Spanish–English bilingual children's productive vocabulary: A development and validation study. *Language, Speech, and Hearing Services in Schools*, *47*(1), 1–15. https://doi.org/10.1044/2015_LSHSS-15-0013

Marchman, V. A., Hurtado, N., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish–learning children. *Developmental Science*, *11*(6), F31–F39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

Marjanovič-Umek, L., & Fekonja-Peklaj, U. (2017). The roles of child gender and parental knowledge of child development in parent-child interactive play. *Sex Roles*, *77*(7–8), 496–509. https://doi.org/10.1007/s11199-016-0734-7

Mayor, J., & Mani, N. (2019). A short version of the MacArthur–Bates communicative development inventories with high validity. *Behavior Research Methods*, *51*(5), 2248–2255. https://doi.org/10.3758/s13428-018-1146-0

Moeller, M. P. (2000). Early intervention and language development in children who are deaf and hard of hearing. *Pediatrics*, *106*(3), E43. https://doi.org/10.1542/peds.106.3.e43

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). Developmental dyslexia: A review. *Reading Research Quarterly*, *27*(1), 29–53. https://doi.org/10.2307/747832

Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur Communicative Development Inventory: Words and sentences. *Journal of Child Language*, *27*(2), 255–266. https://doi.org/10.1017/S0305000900004098

Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the national assessment of educational progress (NAEP). *American Psychologist*, *74*(4), 445. https://doi.org/10.1037/amp0000356

Rescorla, L. (1989). The language development survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, *54*(4), 587–599. https://doi.org/10.1044/jshd.5404.587

Rifaie, N. (1994). *The construction of an Arabic test to evaluate child language* [Unpublished PhD thesis]. Ain Shams University.

Rossetti, L. (2001). *Communication intervention: Birth to three* (2nd ed.). Singular Publishing Group.

Smetana, J. G., & Ahmad, I. (2018). Heterogeneity in perceptions of parenting among Arab refugee adolescents in Jordan. *Child Development*, *89*(5), 1786–1802. https://doi.org/10.1111/cdev.12844

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child Language*, *36*(5), 1115–1144. https://doi.org/10.1017/S0305000908009185

Vohr, B., Jodoin-Krauzyk, J., Tucker, R., Johnson, M. J., Topol, D., & Ahlgren, M. (2008). Early language outcomes of early-identified infants with permanent hearing loss at 12 to 16 months of age. *Pediatrics*, *122*(3), 535–544. https://doi.org/10.1542/peds.2007-2028

Yoshinaga-Itano, C., Sedey, A. L., Coulter, D. K., & Mehl, A. L. (1998). Language of early-and later-identified children with hearing loss. *Pediatrics*, *102*(5), 1161–1171. https://doi.org/10.1542/peds.102.5.1161

Zambrana, I. M., Ystrøm, E., & Pons, F. (2012). Impact of gender, maternal education, and birth order on the development of language comprehension: A longitudinal study from 18 to 36 months of age. *Journal of Developmental & Behavioral Pediatrics*, *33*(2), 146–155. https://doi.org/10.1097/DBP.0b013e31823d4f83