SUPPLEMENT S1: MBCDI ADAPTATION & VALIDATION STUDY

To create an adaptation of the MacArthur-Bates Communicative Development Inventories (MBCDI), we started by compiling a list of 113 English words, consisting of the entire MBCDI Short Form I (89 words) and some additional vocabulary (24 words) from the MBCDI Short Forms II-A and II-B (Fenson et al., 2000). Because the MBCDI was administered through face-to-face interviews, we chose to measure vocabulary with shorter lists rather than longer lists of words. Words from Form II were added in order to create a single word-list for all three data collection periods: Form I is designed for infants only between 0;8 and 1;4, whereas Form II is for infants between 1;4 and 2;6.

A direct translation of the list would yield a culturally insensitive tool, so we identified 38 words that were not culturally understandable, and replaced these with words more appropriate to the culture, lifestyle and environment of Mozambican families. All replacement words fulfilled the syntactic-semantic properties of the original item. Examples of items that we replaced include *goat* for *duck*, *ox* for *lion*, *cellphone* for *television*, and also *bring* for *help*. The reasons for replacing these items are that goats are more common than ducks, and most children have encountered neither lions nor a television. In addition, some replacements were based upon problems of explanation. The translation provided by the local informants for *help* did not convey the same meaning as it does in English, and is also not a word as commonly used in Changana/Ronga (at least not with or by children). We therefore decided to change the word to *bring*, keeping it within the same syntactic category and a prominent word in child-directed speech.

This first adaptation, in Portuguese, was verified and further translated into Changana and Ronga with the help of local assistants. We confirmed translations, when possible, with a Ronga–Portuguese dictionary (Sitoe, Mahumana, & Langa, 2008). During the data collection for the research project, we noticed that the research assistants either did not ask certain items in a consistent manner as intended, or that respondents were not able to separate the word from the meaning. We therefore removed five additional items from the list: vocalizations such as *beheh* (sound of a goat), *ow,* and *uh-oh*; other words removed were easily confused with the action they also refer to, such as *patty cake* and *laugh*. The result was a checklist containing 108 items with translations into Mozambican Portuguese, Changana and Ronga.

We trained a total of four local research assistants (two urban and two rural) to administer the MBCDIs with primary caregivers through interviews, which is necessary because of the high illiteracy rate among the studied communities. This training was further finalized during the first meetings with the primary subjects of our research project, which was part of the first data collection of our longitudinal study. The same local research assistants also carried out a norming study in both communities among 260 urban and 378 rural mothers with infants between 0;11 and 2;2, which coincides with the age span of the participants from our main study.

To assess validity of the responses given by the participants of our study, we compared the vocabulary scores with the speech produced by the infants during the same 30 minute fragments from the recordings at 1;6 and 2;1, as described in the main article. Local research assistants transcribed the infants' speech of these 30-minute fragments under continuous direct supervision of one of the first two authors. All intelligible speech was first transcribed in the language spoken, and where necessary, this was translated into Portuguese. All unintelligible speech and vocalizations, such as ’uhm’, laughter and cries, were marked, but left out of the present analysis.

For each video, we counted the number of different word types the infants spoke. Words were considered different if they had completely different meanings. Words that were similar (e.g. “mama” and “ma!” for mother, “avo” and “vovo” for grandmother, “keke” and “makeke” for biscuit, or “nila” and “nilava” for ‘I want’) were counted as one. Also words with relatively complex morphology, such as “nitakuba” (“ni”-I, “ta”-will, “kuba”-hit, ∅-you), were counted as one, because it is unclear whether the infants learned the morphology of the word, or whether it was stored as a holophrase. The number of different types measured at the two age groups was then correlated with the vocabulary measured by the MBCDI at these two age groups. We calculated the Spearman rank correlations, because the speech data in this small sample revealed a skewed distribution.

Table S1. Results from the validation study.

	Urban		Rural	
	1;6	2;1	1;6	2;1
Type frequencies				
Mean	14.07	44.77	5.36	18.43
Median	10.50	27.00	4.00	18.50
Min	0	12	0	6
Max	36	100	22	35
MBCDI score				
Mean	29.00	73.15	17.71	50.86
Median	27.50	77.00	13.5	61.00
Min	4	9	4	13
Max	59	100	45	77
Correlations[a]				
Speech at 1;6	0.668*	0.221	-0.004	0.095
Speech at 2;1	0.517[b]	0.154[b]	0.801**	0.551*

Notes: [a] Spearman correlations between type frequencies of child speech (rows) and expressive MBCDI scores (columns).

[b] Missing transcription for one urban participant (here n = 13).

*p < .05; **p < .001.

Table S1 summarizes the statistics concerning the type frequencies of the child speech recorded at 1;5/1;6 and at 2;1 in both communities, the CDI scores at the same periods and the Spearman correlation coefficients between these measures. In the urban community, the type frequency of child speech recorded at 1;5 correlates significantly with the CDI score at 1;5 (r_{14}=0.668, p=.009), but not significantly with the CDI at 2;1 (r_{14}=0.221, p=.448). The type frequency of urban children’s speech at 2;1 shows positive, but no significant correlations with CDI at 1;5 (r_{13}=0.517, p=.071) and at 2;1 (r_{13}=0.154, p=.615). In the rural community, the type frequencies recorded at 1;6 yield nearly zero correlations with the CDI scores, but the type frequency of speech recorded at 2;1 correlates significantly with the CDI at 1;6 (r_{14}=0.554, p=.040) and at 2;1 (r_{14}=0.801, p<.001).

When comparing the CDI scores of vocabulary for this sample with those found among the same age groups from the norming sample, we get the following results: In the urban community, at 1;5 the CDI score of the small sample (M = 31.83) is not significantly higher than in the norming study (*M* = 25.27, *t*(17.57) = 0.16, *p* = .876), but at 2;1 the small sample score (*M* = 72.93) is significantly higher than in the norming study (*M* = 55.76, *t*(16.10) = 2.41, *p* = .028). The CDI score of the small rural sample at 1;6 (*M* = 17.71) is not significantly different from the norming sample in the same age group (*M* = 20.90, *t*(18.77) = 0.89, *p* = .386). The rural score at 2;1 in the small sample (*M* = 50.86) is not significantly higher than in the norming sample (*M* = 37.98), but it tends towards significance (*t*(15.09) = 1.97, *p* = .068).

To summarize, type frequencies of words produced by children and their MBCDI scores mostly reveal positive correlations, often to a significant degree. This is particularly the case for type frequencies recorded at 1;5 in the urban community, and those recorded at 2;1 in the rural community.

The correlations between type frequencies and MBCDI at 2;1 in the urban community only show small positive correlations. A possible reason for this lower correlation could be the significant difference in MBCDI scores between the small sample and the norming sample. This difference may be an effect of these children having acquired more language as a result of their participation in our research, which could have triggered caregivers to stimulate their infants' language acquisition. However, this difference could also be due to an exaggeration of the respondents due to the presence of one of the authors while the CDI was administered. Nevertheless, although not significantly so, child speech and vocabulary do correlate positively.

The lack of correlations between type frequencies and CDI scores in the rural community at 1;6 can be explained by a flooring effect in the type frequencies: 11 out of 14 infants had a type frequency lower than five, so reliable rank correlations could not be calculated. The positive correlations between the type frequencies of child speech and vocabulary that were obtained indicate that the CDI scores reflect observed speech reasonably well in our small sample. This, thus, provides a positive validation of the CDI. Full details of the development procedure, norming study and validation study are provided in Vogt, Mastin, Aussems and Schots (2015).

References

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,* 95-116.

Sitoe, B., Mahumana, N., & Langa, P. (2008). *Dicionário: Ronga – Português.* Mozambique: Proprimento.

Vogt, P., Mastin, J.D., Aussems, S., & Schots, D.M. (2015) Early vocabulary development in urban and rural Mozambique. *Under review*.

SUPPLEMENT S2: SELECTION OF VIDEO DATA FOR ANALYSIS

In this supplement, we briefly review how we selected the roughly 30-minute video data selections. These selections were taken from the natural observation videos recorded during the second visit of each collection period. Videos could last anywhere from 45 to 75 minutes in length for a variety of reasons – infant falls asleep, infant is in a long period of distress, mother and infant are called away, infant is too far away, the infant or other persons present interact with the researchers, etc. – which would all result in unviable data. For this reason, we took 'live' notes, in 5-minute segments (Table S2-1), on what was occurring during observation recordings.

Table S2-1. Hypothetical example of live field notes.

Time	Comments
0:00 – 5:00	Family adjusting to presence of researcher and camera
5:00 – 10:00	Child playing w/ sibling, mother
10:00 – 15:00	Eating snack
15:00 – 17:00	Running around w/ peers
17:00 – 20:00	Infant off-camera, inside house
20:00 – 25:00	Playing game w/ family
25:00 – 30:00	(cont'd)
30:00 – 35:00	Eating; Change clothes
35:00 – 40:00	Infant is blocked from camera by other child
40:00 – 45:00	Too much wind to hear
45:00 – 48:00	Breastfeeding w/ mother

By taking live notes, we were able to tally the amount of time infants spent during recording sessions in engagement that was visible and code-able, while simultaneously discounting recording time that covered examples such as those in Table S2-1. Once the accumulation of code-able data reached 40 minutes, recording stopped. Videos were later uploaded and coded using ELAN[1] (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). All field notes were used as a general map to what would happen during coding. Videos were coded for engagement levels via the infant's perspective; therefore, if their perspective could not be ascertained, then such segments could not be coded as viable data. Field notes gave warning to extended periods of unviable data or where the normal behavior was frequently interrupted, but shorter periods also occurred and were excluded as need be (i.e., something blocks camera view of infant, infant leaves view momentarily). In addition, we excluded segments where the infant tried to interact at all with the researcher present, or attended to the researcher's presence for longer than 10 seconds. These exclusions should not skew the data we retrieved because technically unless an infant is asleep, they are always interacting with their environment in some way, and we only exclude instances where the infant's perception is unattainable, or unnatural due to our own foreign presence. By using ELAN, we were able to track the amount of actual time coded for viable interactions. Once 30 minutes of coded data had

[1] http://tla.mpi.nl/tools/tla-tools/elan/

accumulated, we stopped coding. In a few occasions, we had to exclude more data than originally anticipated, yielding slightly shorter overall coded data.

References

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In: *Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.*

SUPPLEMENT S3: ENGAGEMENT LEVEL DEFINITIONS

The following nine engagement level categories, and their definitions, were used as our coding scheme for annotating the video data.

1. *Unengaged*. The infant is present, but is not interacting with any specific partner, object or activity. This applies to situations when the infant scans the environment, but the infant's attention is not fixed on anything. Furthermore, when an infant is moving towards a partner, or towards a non-human object, but does not reach said target, then this is also considered unengaged.
2. *Onlooking*. The infant fixes his/her attention on a partner, but makes no effort to engage with said partner. Whatever the infant is focused on, it is animate but is not actively interacting with a target object. A basic example of this is when an infant's attention is drawn towards a person that is moving across their field of vision, but not in any state of interaction with a target.
3. *Objects*. The infant is manipulating or interacting with a specific non-human object(s), but does not include or attend to any possible partners that are present in the immediate environment. Examples of this include playing with toys, eating food by themselves, or slapping a hand against an object.
4. *Observing*. The infant is observing an activity or event that is being undertaken by others within their vicinity. A basic example is when a mother is preoccupied with a given household chore (i.e., laundry), and these actions overtly captivate the infant's attention, sometimes to the point of imitation. This is different from the category of *Onlooking*, because the partner is now attending to or interacting with a target object/event. However, the action being undertaken by the partner is not for the benefit of attracting the infant's attention, nor is he or she necessarily aware of the infant's attention.
5. *Persons*. The infant is involved in a dyadic event with a partner, directly through touch, person play, or reciprocated speech. This applies to times of breast-feeding as well, due to the direct human contact and intimacy between infant and caregiver. In addition, episodes where an infant tries to create a triadic event including a target object, but the partner does not take note of the target at all, are also coded as *Persons* interactions.
6. *Passive Joint Attention*. A partner and the infant share attention to an object or activity, but one individual does not attend to the other's gaze/attention. The prime example of *Passive Joint Attention* usually involves a partner showing an object of interest to the infant and enticing them to play with said object, but the infant's attention does not exceed the object introduced to her/him. The main point here is that the infant does not display overt awareness (e.g., through checking) that the partner is attending to the object as well.
7. *Shared Joint Attention*. Both the infant and a partner attend to the same target object or activity; in addition, both infant and partner are aware that the other's attention is focused on each other and the same target object or activity, but neither coordinate their attention to create a triadic event involving a mutual interaction goal. For example, an infant is playing with a toy; the mother notices and reaches out her hand and asks for the toy; the infant notices the mother, her gaze, and her gesture; the infant then throws the toy in the opposite direction,

away from the mother. At this level of engagement, both parties are aware of the other's attention on a target, and this attention is shared, but there is no understood or shared goal between parties.

8. *Coordinated Joint Attention*. The infant and a partner are jointly involved with an object or activity. Their attention is shared, they are both aware of the other's attention, and this shared attention has been directed towards a mutual interaction goal through the use of explicit cues, gestures and directions. For example, an infant walks up to her/his mother holding an object, the infant looks at the object, then at the mother and extends the object; the mother looks at the object and then to the infant, and either takes the object that is being offered, directs the infant's attention to another partner by a hand gesture or nod, or responds to the child verbally through speech or physically through touching or holding the infant.
9. *Unknown Attention*. The infant is present, but line of visual interest cannot be ascertained. This applies when something or someone obstructs the field of view, the infant hides their face, the infant is too far away from the camera, or the video footage is out of focus. While this category is used for coding purposes, it is not included in the 30-minute criterion, or statistical analyses.