Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1

J. Douglas Mastin and Paul Vogt
Tilburg University

Abstract

This study analyzes how others engage rural and urban Mozambican infants during naturalistic observations, and how the proportion of time spent in different engagements relates to infants' language development over the second year of life. Using an extended version of Bakeman and Adamson's (1984) categorization of infant engagement, we investigated to what extent a detailed analysis of infant engagement can contribute to our understanding of vocabulary development in natural settings. In addition, we explored how the different infant engagements relate to vocabulary size, and how these differ between both communities. Results show that rural infants spend significantly more time in forms of solitary engagement, whereas urban infants spend more time in forms of triadic joint engagement. In regard to correlations with reported productive vocabulary, we find that dyadic *Persons* engagement (i.e. interactions *not* about concrete objects) has positive correlations with vocabulary measures in both rural and urban communities. In addition, we find that triadic *Coordinated Joint Attention* has a positive relationship with vocabulary in the urban community, but a contrasting negative correlation with vocabulary in the rural community. These similarities and differences are explained based upon the parenting beliefs and socialization practices of different prototypical learning environments. Overall, this study concludes that the extended categorization provides a valuable contribution to the analysis of infant engagement and their relation to language acquisition, especially for analyzing naturalistic observations as compared to semi-structured studies. Moreover, with respect to vocabulary development, Mozambican infants appear to benefit strongest from dyadic *Persons* engagement, while they do not necessarily benefit from joint attention, as tends to be the case for children from industrial, developed communities.

PREPRINT: To appear in *Journal of Child Language*

Introduction

A fundamental question in developmental psychology is how infants begin to participate in the social practices of their culture and their language. These shared experiences are realized in forms of joint engagement, where caregivers facilitate symbol learning during goal-oriented interactions (Hobson, 2005; Tomasello, 1995). Infants improve their joint engagement skills around one year of age, and they begin to produce single words not long after, suggesting the two are intertwined.

At the hub of early infant engagement research is Bakeman and Adamson's (1984) study of infants' coordination of attention to people and objects. They analyzed infants' attention states (i.e., *engagement levels*), and showed that triadic joint engagement is the natural culmination of early social development. They proposed six levels of engagement: *Unengaged* with any specific thing or partner*; Onlooking* to another person's activity*; Object* play*; Persons* interaction, face-to-face or through play*; Passive Joint Attention* (*Passive-JA*) between an infant, a partner and an object, but no attention from infant to partner; and *Coordinated Joint Attention* (*Coordinated-JA*) between an infant, a partner and an object, where infant and partner attend to each other. Various studies have focused on individual types and aspects of joint engagement, and how these relate with vocabulary development in middle-class infants from industrialized societies (Adamson, Bakeman & Deckner, 2004; Carpenter, Nagell, Tomasello, Butterworth & Moore 1998; Mundy & Gomes, 1998). However, there are three distinct limitations in such studies.

First, many use semi-structured observation or simulated spontaneous play rather than fully naturalistic observation methods (e.g., Bakeman & Adamson, 1984; Carpenter et al., 1998; Morales et al., 2000), which cannot represent the entirety of infant engagement (Eisenbeiss, 2010). Such methods create a bias towards engagement involving a target object, which could drastically increase triadic interactions. Semi-structured observation can easily omit time infants spend alone, as well as partners other than caregivers. In many cultures adult caregivers do not play with their children, so instructing them to simulate play may be unnatural (Abels, Keller, Mohite, Mankodi, Shastri, Bhargava, Jasrai, & Lakhani, 2005; Lieven & Stoll, 2013). To overcome these limitations, we relied on daily interactions within the home, and did not offer toys to infants or instructions to parents, thus providing natural observations of infant engagement for analysis.

Second, many studies since Bakeman and Adamson (1984) have focused on relations between triadic joint engagement and vocabulary (Carpenter et al., 1998; Morales et al., 2000; Tomasello & Farrar, 1986). While complex types of engagement may be more beneficial to learning, this does not mean that solitary play or observation, for example, bear *no* relation to language acquisition and vocabulary development. We believe that a more complete correlational analysis of engagement levels and vocabulary can uncover aspects of social behavior that have been overlooked. Notice that engagement levels are mutually exclusive, but not necessarily independent. Bakeman and Adamson (1984) showed some distinct patterns in how engagement levels emerged over time, so a broad classification might reveal dependencies between levels when all possible engagements are included.

Third, most studies have been carried out in industrial societies. However, socialization of children and attitudes about child rearing differ greatly across cultures (Greenfield, 2009; Hoff, 2006; Keller, 2012; Schieffelin & Ochs, 1986). For instance, multi-party interactions are more frequent in non-industrial communities, and infants often have secondary caregivers, including siblings (Brown, 2011; Gaskins, 2006; Harkness, 1977; Lieven & Stoll, 2013; Zukow-Goldring, 2002). Families in industrial communities, though, have a more nuclear structure, which may not involve regular exposure to as many communication partners. Furthermore, industrial cultures are usually high on the Human Development Index (HDI), and mothers in high-HDI countries engage in more book reading,

story telling, and object naming and counting, than mothers in low-HDI countries (Bornstein & Putnick, 2012).

In addressing these three limitations we have categorized infant engagement in more naturalistic observations in non-industrial communities. In Mastin, Vogt, Schots & Maes (2015), we presented the design of an extended categorization of engagement levels based on Bakeman and Adamson (1984). By implementing a component-based approach to the construction of engagement categories, we extended their categorization by adding two further engagement levels. In our extended categorization, we included goal-oriented behavior as a necessary component of joint engagement. In the present study, we explore the value of this approach by studying correlations between the proportion of time infants spend in different engagement levels and their reported productive vocabulary (from here referred to as 'vocabulary'), and how these differ in non-industrial rural and urban communities in Mozambique. Our main question is: To what extent can a detailed analysis of infant engagement contribute to our understanding of vocabulary development in natural settings? A second question is: Do correlations between infant engagement and vocabulary size differ between these communities?

In the next section we review how our approach furthers research in the study of infant development. To address our research questions, we first explore how the proportions of infants' engagements differ between the two communities. Second, we investigate infants' vocabulary sizes. Third we explore relations between the proportions of infants' engagements and vocabulary size. Fourth, we compare our approach with two other approaches to early engagement. Finally, we discuss the results, their implications, and what further steps should be considered.

Expanding the Spectrum

Language Socialization in Non-Industrial Communities

"Studies of joint attention and early language need to take account for the real-life and often polyadic contexts in which young children interact with others" (Akhtar & Gernsbacher, 2007, p. 200). We agree: we need to study not only how infants interact but also with whom. This is particularly true for non-industrial cultures, where the extended family and unrelated members of the community play a regular role in the daily life and socialization of infants (Lieven, 1994). However, infant socialization can manifest in different types of interactions in different degrees. For example, Brown (2011) showed that infants from Rossel Island in Papua New Guinea were socialized twice as often as infants from a Mayan community. In particular, many studies have found that the amount of child-directed speech is relatively small in many non-industrial cultures (Gaskins, 2006; Harkness, 1977; LeVine et al., 1994; Rabain-Jamin, Maynard, & Greenfield, 2003; Schieffelin & Ochs, 1986; Shneidman & Goldin-Meadow, 2012). Moreover, the amount of cognitive stimulation infants receive relates to the Human Development Index (HDI), which is low for many non-industrial societies (Bornstein & Putnick, 2012).

Such differences in cognitive stimulation could affect how caregivers engage infants, as well as how infants' vocabulary develops (Hart & Risley, 1995). For instance, in industrial societies, face-to-face cognitive stimulation occurs more frequently than in non-industrial societies, where caregivers are more concerned with children's motor development (Bornstein & Putnick, 2012; Keller, 2007). So, studies of industrial cultures cannot be generalized to non-industrial societies or historical paradigms. Recent research suggests that there are three more or less prototypical learning environments: urban industrial, urban non-industrial, and rural non-industrial communities (Greenfield, 2009; Keller, 2012). Each environment tends to foster children's development based on the daily lifestyles of these

communities. Urban industrial communities foster *individual psychological autonomy*, focusing on cognitive development. Rural non-industrial communities focus on the development of *communal action autonomy* that allows children to participate in a subsistence-based lifestyle from early on. Finally, urban non-industrial communities form a hybrid between the other two, focusing on *communal psychological autonomy* (i.e. on development of cognitive skills and communal responsibilities, Keller, 2012). Due to differences across learning environments, children show different developmental trajectories in these prototypical environments (cf. Abels et al., 2005; Keller, 2007; Keller, 2012). We therefore explore the differences between non-industrial rural and urban communities from Mozambique.

Joint Attention and Vocabulary Development

Although research has focused on aspects of infant engagement and relations to vocabulary, none, to our knowledge, have analyzed correlations between all engagement levels in natural settings and infants' vocabulary development in production. Two studies have come close: Carpenter et al.'s (1998) research on joint attention and communicative competence among English-speaking infants from America, and Childers et al.'s (2007) study of engagement levels and noun versus verb learning in Ngas-speaking children in Nigeria.

Carpenter et al. (1998) analyzed how infants, between 0;9 and 1;3 years old, and their primary caregivers, *share*, *follow* and *direct* each other's attention. Inspired by the theoretical perspective of Tomasello (1995), Carpenter and colleagues expanded Bakeman and Adamson's (1984) definition of joint attention to include infants' understanding of others as intentional agents with goals, choices of how to attain said goals, and what to attend to in pursuing these goals. But their correlational analysis focused only on triadic engagement with objects and people: *Attention Following* (cf. Bakeman and Adamson's *Passive-JA*) and *Joint Engagement* (i.e., *Coordinated-JA*). They showed that the age of onset of different skills in joint attention predicted later vocabulary acquisition, and that the frequencies of these skills were correlated with vocabulary size. However, they excluded categories of solitary engagement, as well as social engagement without objects. But omitting some kinds of engagement could distort the analysis. For example, does time spent alone, observing, or interacting without target objects, affect word learning? Children's solitary engagements, such as symbolic play, can have a great impact on their own development (Rabain-Jamin et al., 2003). Moreover, children who are talked to infrequently may learn from overheard speech (Lieven, 1994; Schieffelin & Ochs, 1986). Carpenter et al. (1998) also instructed parents to simulate normal play using provided toys, chosen to maximize interest and promote triadic engagement. However, providing toys chosen to elicit interactions manipulates the naturalness of the environment.

Childers et al. (2007) provide an example of another semi-structured study, which relied on Bakeman and Adamson's (1984) six-level engagement categorization for their analysis of engagement *distributions* (i.e., time spent in each engagement level). However, for correlating those with vocabulary size, they collapsed the engagement categories into three levels: *Low-level Attention* (*Unengaged* and *Onlooking*), *Mid-level Attention* (*Object* and *Persons*), and *High-level Attention* (*Passive-JA* and *Coordinated-JA*). Childers et al. found that only *Mid-level Attention* correlated with both noun and verb learning, but *Mid-level Attention* combines *Object* and *Person* engagement. This seems inappropriate since object manipulation does not involve joint engagement, whereas engagement with people is both dyadic *and* joint. Their results also showed that *High-level Attention* was more frequent than less complex engagement. Yet, mothers had been instructed to simulate play with their children, which could create a bias towards more *High-level Attention*. Overall, we cannot be sure what affect this had on their observations.

Both Carpenter et al. (1998) and Childers et al. (2007) used parental checklists to assess the infants' vocabulary sizes. Where Carpenter et al. used the MacArthur-Bates Communicative Development Inventories (henceforth MBCDI; Fenson, Dale, Reznick, Bates, Thal, & Pethick, 1994), Childers et al. constructed an adaptation of the MBCDI. Although the use of parental checklists has been criticized on the grounds of unreliability - parents may overestimate or underestimate their children's vocabulary size (Houston-Price, Mather, & Sakkalou, 2007; Law & Roy, 2008) - they are standard for assessing early vocabulary comprehension and production (Bornstein et al., 2004; Fernald, Marchman, & Weisleder, 2012). Moreover, while a parental checklist is not perfect, it is more representative than tokens from selective observations (Pine, Lieven & Rowland, 1996). Since this method has been used in both related studies (i.e., Carpenter et al., 1998; Childers et al., 2007), we used it, with caution, to measure vocabulary size.

Analyzing Infant Engagement by Feature-components

The definition of engagement used in this study is the following:

> *Engagement* involves the increasingly complex ways individuals interact with and within their environment, namely, interaction with themselves, other individuals, events, and objects (both animate and inanimate). Engagement can manifest through either solitary or joint engagement:
>
> - *Solitary engagement* occurs when an individual does not interact with any other individual or group in the environment. The individual may watch others, act with himself alone (in play, for example), or interact with only objects.
> - *Joint engagement* occurs when an individual interacts with another individual or a group in the environment, and the interaction includes only themselves (*social dyadic engagement*) or also some target object or event (*triadic engagement*). At least one individual in the interaction is overtly aware that their focus of attention coincides with that of another individual(s) via verbal and/or non-verbal communication: verbal language, body language, gestures, coordination of eye gaze, or corresponding behaviors.

Engagement, then, is a spectrum of levels that are inter-related yet mutually exclusive. The infant's *coordination of attention* is generally assumed only from checking a partner's eye-gaze (e.g., Bakeman & Adamson, 1984; Carpenter et al., 1998; Childers et al., 2007; Tomasello & Farrar, 1986). We instead broadened this coordination of attention to include all communication, language and behavior, rather than just eye-gaze. This addition was inspired by Barton and Tomasello's (1991) account of joint action (i.e., joint engagement) as including *appropriate* responses. Previous research does not often address the issue of goals within engagement levels (but see Carpenter & Liebal, 2011; Tomasello, 1995, Tomasello, Carpenter, Call, Behne & Moll, 2005), possibly because goals are a unique aspect of human engagement, and harder to identify objectively. Carpenter et al. (1998) included goal-oriented actions within joint attention in their interpretation of intentional agency, while Carpenter and Liebal (2011) argued that for both partners *knowing together* requires simultaneous attention (e.g., Hobson, 2005; Tomasello, 1995), and this *sharing* in mutual knowledge is what changes parallel attention into joint attention.

By including goals as a component of engagement, we derived two new engagement levels by dividing two of Bakeman and Adamson's (1984) engagement level categories (see, Mastin et al., 2015, for more details). Within the *Onlooking* category, we distinguish *Observing* – where an infant focuses their attention to, and sometimes imitates, another individual's goal-oriented actions with a target object/event, from *Onlooking* to an

individual's presence within the infant's field of vision. From the category of *Coordinated-JA*, we distinguish *Shared-JA* – where an infant and partner attend to each other and to a target object, but their goals do not align toward the same outcome, so not allowing for coordination of goal-oriented behavior.

Methods

Participant and Site Selection

We selected Mozambique for our field research. To our knowledge, no previous study on first language acquisition has been reported for Mozambique. We chose an understudied and non-industrial community, because we expected the proportion of time infants spend in particular engagement levels would differ substantially from industrial middle-class urban families. Moreover, we expected to see differences between non-industrial rural and urban communities (Keller, 2012). We therefore selected two field sites: a rural site made up of three adjacent villages just outside the provincial town of Chokwe in Gaza province, about 225 kilometers from the country's capital, Maputo; and an urban site made up of two adjacent residential suburbs in Maputo. The rural and urban communities share some traditions, are both relatively poor, and have low health standards. Daily lifestyle, though, differs considerably: the rural area relies on subsistence farming, whereas the urban areas are market-based.

With mediation from two local community organizations, we asked for volunteers with infants between 1;0 and 1;2 at the start of a longitudinal study with three visits (average ages of 1;1, 1;6, and 2;1). We hired and trained four local research assistants (two in each field site) who explained to caregivers in their native language the purpose of the study and our procedures at each visit. The families were informed that our goal was to investigate how Mozambican infants learn their first words. We also explained that this research offered no immediate benefits to the families who volunteered, that their data would be treated confidentially, and that they could withdraw from the study at any time. All participants gave informed consent. In this paper, we present data and results from 28 participants (Table 1), half each in the rural and urban sites.

The participants from the rural community were all native speakers of Changana– a Southern Bantu language spoken in parts of Mozambique and in South Africa, where it is called Tsonga (Lewis, 2009). This was generally the only language spoken in the household. In the urban community, most children are raised bilingually in Portuguese, the official language, and Ronga, another dialect of Tsonga mutually intelligible with Changana. While there is not a significant difference between family sizes, we believe urban participants have a more dynamic social environment due to population density, industry and technology.

Most rural parents had either no education or only completed the lower levels of education, while all urban parents (except one) have received some education. A nominal logistic regression on education level relating to location and gender revealed a significant effect for location ($\chi^2(3) = 16.415$; $p = .001$), but not for gender ($\chi^2(3) = 4.107$; $p = .250$). More urban parents received a higher education level than rural parents. In addition, most rural fathers worked far from home in South Africa or Maputo. Rural mothers worked as subsistence farmers, whereas urban mothers tended to work in domestic services and fathers had local jobs. Based on these differences in education and employment, we judged the urban site to have a higher socio-economic status (SES) than the rural one.

Table 1. Demographic information of participants in the study (infants and their parents).

Infants	**Rural (n=14)**	**Urban (n=14)**
Female infants	7	5
Male infants	7	9
Average age (SD)	1;1.8 (0;0.26)	1;1.6 (0;0.28)
Average Family size (SD)	8.2 (5.8)	7.4 (4.4)
Average number of siblings (SD)	2.3 (1.5)	3.5 (2.5)
Average birth order (SD)	3.2 (2.4)	2.5 (1.5)
Mother's Average age (SD)	28.4 (7.8)	27.5 (5.3)
Father's Average age (SD)	35.7 (11.6)	33.1 (8.6)

Parents - Education level	**Mother (n=14)**	**Father (n=14)**	**Mother (n=13)**	**Father (n=13)**
No education	6	5	1	0
5-year early primary school	5	7	5	4
Additional 2-year primary school	3	1	6	5
Higher education	0	1	1	4

Parents - Occupation	**Mother (n=14)**	**Father (n=14)**	**Mother (n=14)**	**Father (n=14)**
Paid occupation	0	9	2	10

Note: Parental education for one urban family is missing.

Materials

To measure infants' vocabulary over development, we adapted the short versions of the MBCDI (Fenson, Pethick, Renda, Cox, Dale, & Reznick, 2000) into the three languages of our communities, and administered them in face-to-face interviews, given the level of illiteracy in both communities. Instead of adapting the MBCDI for the three languages (Changana, Tsonga, and Portuguese) separately, we constructed one culturally broad adaptation of the list into Portuguese first and then translated this into the other two languages. Our final adaptation of the MBCDI contained 108 culturally appropriate words. (See Supplement S1 for a detailed description.)

Due to urban bilingualism, we assessed both Portuguese and Ronga simultaneously to assure an accurate comparison. Children in bilingual environments develop language skills similarly to monolingual children when both languages are jointly taken into consideration (Junker & Stockman, 2002); this measure is known as *total conceptual vocabulary* (i.e. the union of both vocabularies - L1∪L2; Patterson, 1998).

The vocabulary scores at 1;6 and 2;1 were validated with the type frequencies of words produced in the transcriptions of the infants' speech from the same video fragments

analyzed here (see, Supplement S1 for details). Table 2 summarizes the results, and gives Spearman correlations between type frequencies and vocabulary sizes.

Table 2. Spearman correlations between type frequencies of child speech (rows) and expressive MBCDI scores (columns).

	Urban MBCDI		Rural MBCDI	
	1;6	2;1	1;6	2;1
Speech at 1;6	**0.668***	0.221	-0.004	0.095
Speech at 2;1	0.517[a]	0.154[a]	**0.801****	**0.551***

Notes: [a]Missing transcription for one urban participant at 2;1 (so n=13). *p<.05; **p<.001.

In the urban community, MBCDI scores at 1;6 correlated significantly for type frequencies of the infants' speech at 1;6 (r_{14} = 0.668, p = .009) and tended towards significance for 2;1 (r_{13} = 0.517, p = .071). The urban 2;1 vocabulary scores revealed positive, but no significant correlations with type frequencies measured at both ages, which may be due to a ceiling effect caused by overestimations of vocabulary at 2;1 (cf. Supplement S1). In the rural area, the correlations between type frequency at 1;6 and vocabulary were virtually zero at both 1;6 and at 2;1, due to a floor effect in the measured type frequencies in the infants' speech: 11 of 14 infants had a type frequency lower than five, which made ranking impossible. Type frequency recorded at 2;1, however, correlated significantly with vocabulary size at 1;6 (r_{14} = 0.801, p<.001) and at 2;1 (r_{14} = 0.551, p = .041). So rural mothers reported their infants' vocabulary fairly accurately at both 1;6 and 2;1 years, compared to the speech the infants produced at 2;1.

The 1;1 vocabulary scores were not validated, but analyses indicate that in the rural area, vocabulary at this age may be underestimated compared to our norming study (Vogt, Mastin, Aussems & Schots, 2015). So results relating to the 1;1 MBCDI should be interpreted with care, which also holds for MBCDI scores at 2;1 from the urban community.

Procedure

All data were collected during visits to the infants' homes. Since most rural daily activities take place outside in open areas and courtyards, filming occurred mostly outside. We placed our camera on a tripod at a distance of between 5 and 15 meters from the participants, depending on the location of shaded areas from which to make recordings. In the urban area, families live in one-floor houses with small courtyards in densely populated suburbs. Due to more confined spaces, urban daily interactions and routines occur inside the home, in the courtyard, and/or in nearby public spaces. Most filming here too occurred outside. Where possible, we followed the same set-up as in the rural area, but in smaller spaces we filmed from 2 to 5 meters away from participants, often by hand.

Video data was collected when infants were on average 1;1, 1;6 and 2;1. The 1;6 data in the urban community were collected two weeks early for logistic reasons, so in effect those infants were 1;5 and 12 days old on average. Each family was visited twice during each collection period. At the first visit, we videotaped the infants' interactions with their families to allow everyone to get used to our presence and the filming procedures. During the second visit, we videotaped the infants from 45 up to 75 minutes for data analysis. On all occasions, caregivers and others present were asked to continue with their daily routines as if we were not present, and not to worry about positioning or moving the infant for our benefit. To ensure natural interactions, and not fabricated ones, we gave no other instructions to caregivers or families. After recording during the second visit, assistants administered the adapted MBCDI through face-to-face interviews in the caregivers' native language under the

supervision of one of the authors. Since parents are likely to underestimate (Houston-Price et al., 2007) and overestimate (Law & Roy, 2008) their child's receptive vocabulary, we relied only on infants' production vocabulary in our analyses.

Data Analysis

Coding Scheme. The videos were coded for approximately 30 minutes (*Mean* 27:57; *SD* 01:52) in segments where the infant displayed 'natural' behavior (i.e., not sleeping, not off camera, not interacting with or disturbed by the researchers; see Supplement S2). We used the following categories in coding as we annotated the video data (see Supplement S3):

1. *Unengaged*: The infant is present, but not interacting with any person or target. This applies, for instance, to situations when the infant scans the environment or moves about without any apparent goal.
2. *Onlooking*: The infant fixes attention on someone, but makes no effort to engage with that person. This person is neither interacting with a target, nor aware of or responding to the infant's attention.
3. *Objects*: The infant is manipulating or interacting (e.g., playing) with a specific object(s) of their own accord, and does not interact with or attend to any person present.
4. *Observing*: The infant is actively observing an activity by someone else close by, sometimes to the point of imitation. This is related to, but different from the category of *Onlooking*, because the observed person is actively manipulating a target object/event.
5. *Persons*: The infant is involved in a dyadic event with a communication partner, through touch, ritualized play, or reciprocated speech, but no target is included in the engagement. This category applies to times of breast-feeding as well.
6. *Passive Joint Attention*: The infant and a communication partner share attention to a target, and only one of them is overtly aware that the attention is shared, while the other appears not to be aware of this. A typical situation is when the infant plays with a toy introduced by the mother, and the mother follows the infant's attention with the toy, but the infant appears not to notice the mother.
7. *Shared Joint Attention*: Both the infant and partner attend to the same target, and both infant and partner are aware that the other's attention is focused on each other and the target. However, neither coordinates their attention to create a triadic event involving an alignment of goals and actions.
8. *Coordinated Joint Attention*: The infant and a partner are mutually involved with a target or event. Their attention is aligned, they are both aware of the other's attention, and this alignment of attention is directed towards a goal via mutual interaction.

Following Bakeman and Adamson (1984), we required a minimum of 2 seconds of fixated attention or interaction for each category of engagement; segments of less than 2 seconds were not differentiated from the surrounding types of interaction. If an infant's point of view could not be ascertained (usually due to technical issues), the engagement was coded as *Unknown*. The *Unknown* category was excluded from all analyses. We calculated the proportion of time infants spend in each category by dividing the total duration for that category by the total duration of all engagement levels together within each video (because total duration did not equal exactly 30 minutes).

Both authors coded half the videos using ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). After coding, we selected 20 videos (10 for each author) at random to be cross-coded for reliability. Ten videos were selected from the 1;1 data from the rural site and for each of these we cross-coded an arbitrarily selected 5-minute segment. The other ten videos were selected from the 1;6 data from both sites, and, for these, we selected

10-minute segments for cross-coding. Cohen's Kappa was calculated on a 100-msec rate and yielded an overall value of 0.73 (0.62 for the 1;1 data and 0.75 for the 1;6 data). The two coders' agreement for individual engagement levels yielded the following Kappa's: 0.30 for *Passive-JA*, 0.34 for *Shared-JA*, 0.57 for *Observing*, 0.60 for *Unengaged*, 0.66 for *Onlooking*, 0.78 for *Coordinated-JA*, 0.81 for *Persons*, 0.84 for *Objects* and 0.85 for *Unknown*. For *Passive-JA* and *Shared-JA* the agreement is rather low, but we believe this does not affect our overall results much for two reasons. First, *Passive-JA* and *Shared-JA* were infrequent (less than 4% in the cross-coded samples), and it is known that Cohen's Kappa reports relatively low scores for disagreements when the category in question occurs infrequently, while actual agreement can be fairly high (Feinstein & Cicchetti, 1990). Second, these two categories were mostly confused with *Objects*, *Persons* and *Coordinated-JA*, all with a high Kappa.

Comparisons With Other Studies. We also assessed differences between correlations with vocabulary using our extended classification of engagement levels compared to the categorizations used by Childers et al. (2007), and by Carpenter et al. (1998). This re-analysis was to assess how informative different engagement level classifications are. To do this, we replicated the 'adjusted' tri-level categorization of Childers et al. and the two triadic engagement categories of Carpenter et al. and applied these to our data. For Childers et al., we summed *Unengaged*, *Onlooking* and *Observing* to create their *Low-Level* category, *Objects* and *Persons* to create their *Mid-Level* category, and *Passive*, *Shared* and *Coordinated-JA* to create their *High-Level* category. For Carpenter et al.'s classification, we summed *Shared* and *Coordinated-JA* to construct their category of *Joint Engagement*, and *Observing* and *Passive-JA* to construct their *Attention Following.*

Results

Engagement Level Proportions and Expressive Vocabulary

Table 3 provides the descriptive statistics for the occurrences and proportions of engagement levels for both sites; these are presented in graphic form in Figure 1. According to the Wilcoxon rank sum test, infants at 1;1 spent significantly more time *Unengaged* in the rural area (*Mdn* = .145) than in the urban area (*Mdn* =.078, $W = 32$, p =.003, r = -.569), and they spent more time *Observing* (*Mdn* = .054) than urban infants (*Mdn* = .023; $W = 49.5$; p = .027; r = -.417). The proportions of *Observing* were also higher in the rural area (*Mdn* = .090) than in the urban area (*Mdn* = .020) at 1;6 ($W = 25$; p<.001, r = -.630). At 2;1, the rural infants (*Mdn* = .129) spent more time *Unengaged* than urban infants (*Mdn* = .073, $W = 48$, p = .023, r = -.430). Urban infants at 1;1 spent more time (*Mdn* = .038) than rural infants (*Mdn* = .021) in *Passive-JA* engagement ($W = 54$, p = .046, r = -.378), and at 2;1 they spent more time (*Mdn* = .036) than rural infants (*Mdn* = .010) in *Shared-JA* ($W = 32$, p = .003, r = -.569).

Results from the MBCDI parental checklist are given in Table 4. These show that urban infants have substantially larger vocabularies than rural infants. A 2 (location) x 3 (age) ANOVA shows a significant main effect of location: urban infants have a larger vocabulary than rural infants ($F(1,78) = 9.349$; $p < .01$) at every collection period. Also there is a main effect of age ($F(2,78) = 81.283$; $p < .001$). A Post-hoc Tukey analysis showed that the MBCDI scores across the three collection periods – 1;1 vs. 1;6; 1;1 vs. 2;1; 1.6 vs. 2.1 – all differ significantly ($p < .001$). There was no interaction of age and location ($F(2,78) = 0.131$; $p = .877$).

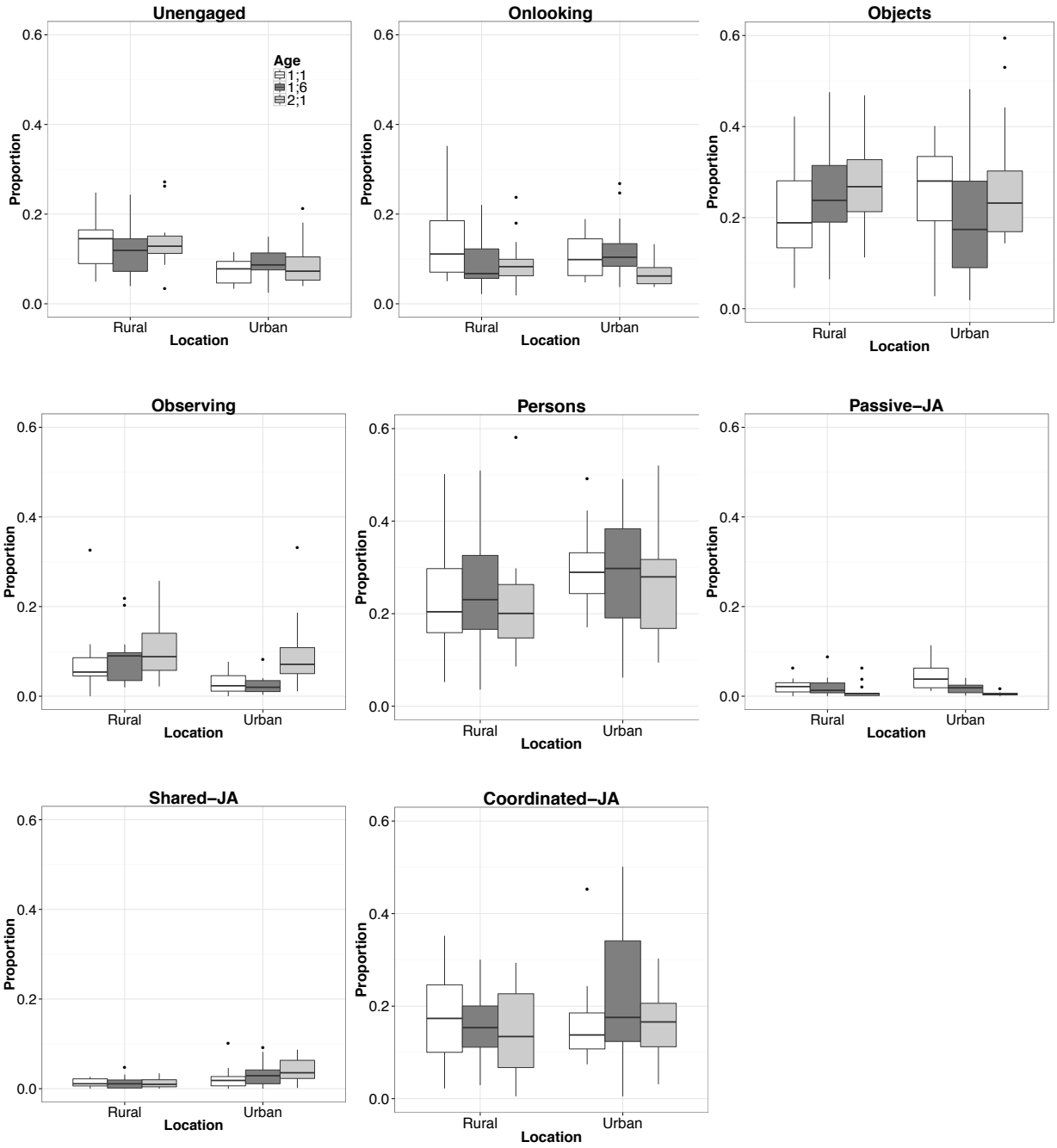

Figure 1. Summary statistics for eight engagement levels at three ages for the two locations. The graphs show the medians, upper and lower quartiles in boxes, and top and bottom 25% in the error-bars. The scales on the y-axes are the same for ease of comparison.

Table 3. Descriptive statistics of infants' engagement levels for ages 1;1, 1;6 and 2;1. The statistics show mean number of occurrences (*N*), and the median (*Mdn*), mean (*M*), minimum (*Min*) and maximum (*Max*) values of the proportion of time infants spent in various engagement levels. The results are distinguished between the rural and urban communities.

	1;1					1;6					2;1				
	N	*Mdn*	*M*	*Min*	*Max*	*N*	*Mdn*	*M*	*Min*	*Max*	*N*	*Mdn*	*M*	*Min*	*Max*
Rural															
Unengaged	28	0.145**	0.141	0.049	0.248	29	0.119	0.115	0.040	0.243	34	0.129*	0.139	0.034	0.271
Onlooking	26	0.111	0.144	0.050	0.352	27	0.067	0.098	0.022	0.220	22	0.083	0.093	0.019	0.237
Objects	33	0.189	0.206	0.046	0.422	42	0.238	0.249	0.065	0.475	43	0.268	0.269	0.113	0.469
Observing	13	0.054*	0.077	0.000	0.327	18	0.090**	0.088	0.020	0.218	21	0.088	0.103	0.022	0.258
Persons	28	0.204	0.226	0.052	0.502	37	0.230	0.259	0.036	0.510	43	0.201	0.223	0.086	0.582
Passive-JA	5	0.021	0.023	0.000	0.063	6	0.013	0.022	0.000	0.087	2	0.005	0.011	0.000	0.063
Shared-JA	2	0.011	0.013	0.000	0.027	3	0.011	0.014	0.000	0.047	3	0.010	0.014	0.000	0.034
Coord-JA	15	0.174	0.171	0.022	0.352	20	0.154	0.155	0.029	0.301	24	0.134	0.149	0.005	0.294
Urban															
Unengaged	21	0.078	0.072	0.033	0.115	25	0.087	0.088	0.025	0.150	30	0.073	0.090	0.039	0.213
Onlooking	31	0.099	0.105	0.048	0.189	29	0.104	0.122	0.037	0.267	23	0.062	0.068	0.037	0.133
Objects	46	0.280	0.260	0.027	0.401	37	0.174	0.206	0.019	0.482	53	0.232	0.279	0.144	0.594
Observing	9	0.023	0.032	0.000	0.077	6	0.020	0.025	0.003	0.082	24	0.071	0.094	0.011	0.331
Persons	47	0.290	0.298	0.171	0.491	40	0.298	0.283	0.062	0.491	57	0.280	0.262	0.094	0.520
Passive-JA	10	0.038*	0.045	0.011	0.114	5	0.019	0.019	0.001	0.041	2	0.004	0.005	0.000	0.016
Shared-JA	5	0.018	0.023	0.000	0.102	6	0.029	0.031	0.000	0.092	8	0.036**	0.041	0.002	0.087
Coord-JA	20	0.138	0.164	0.074	0.453	23	0.176	0.226	0.005	0.501	30	0.166	0.162	0.031	0.303

Note: Comparisons across communities are made via Wilcoxon rank sum tests for each engagement level proportion for each collection period – the proportion that is significantly greater is marked in that site (*p < .05; **p < .01).

Table 4. Expressive vocabulary scores (means and standard deviations) for the rural and urban MBCDI at 1;1, 1;6 and 2;1. Total score possible was 108 at each age.

	At 1;1	At 1;6	At 2;1
Rural	3.35 (1.08)	17.71 (12.23)	50.85 (23.59)
Urban	**10.14 (7.25)****	**29.00 (19.61)***	**72.92 (23.18)****

Note: Significant differences across sites are indicated *$p < .05$; **$p < .01$.

Correlations with Vocabulary

To calculate correlations between the proportions of engagement levels and vocabulary size, we used the Spearman's correlation coefficient, because the data did not reveal a normal distribution. Although multiple regression analysis would be preferable, this was not possible for two reasons. First, the sample size is too small for multiple regression analysis with eight predictors (engagement levels). Second, due to the fact that engagement is part of a spectrum of possibilities, there is a high colinearity of predictors for engagement levels. Since there is also variation within such a small sample, outliers cannot be removed, and multiple regression analysis cannot take these into account.

When proportions of engagement levels are correlated with vocabulary at each age, we see some significant correlations (Table 5). In the rural area, there were no correlations between the proportions of engagement level categories at 1;1 and measured vocabulary at 1;1. The proportion of 1;1 *Coordinated-JA* and 1;6 vocabulary showed a negative correlation ($r_{14} = -0.538$, $p = .050$), while *Persons* engagement reveals a strong positive correlation with 2;1 vocabulary ($r_{14} = 0.723$, $p = .003$). No significant correlations were observed for any 1;6 engagement level with vocabulary at 1;6 or 2;1 in the rural community. Between 2;1 proportions and concurrent vocabulary, *Observing* was positively correlated with vocabulary ($r_{14} = 0.659$, $p = .010$), while *Shared-JA* was negatively correlated ($r_{14} = -0.568$, $p = .034$).

In the urban area, there were also no correlations between 1;1 engagement proportions and concurrent vocabulary. When 1;1 proportions are correlated with 1;6 vocabulary, *Objects* engagement showed a significant negative correlation ($r_{14} = -0.706$, $p = .005$), while *Persons* engagement showed a positive correlation ($r_{14} = 0.772$, $p = .001$). When 1;1 proportions were correlated with 2;1vocabulary, *Persons* engagement remained significant ($r_{14} = 0.598$, $p = .024$). In addition, *Coordinated-JA* engagement now positively correlated with vocabulary size ($r_{14} = 0.660$, $p = .010$). Also, rather than *Objects* engagement, the data was now negatively correlated for *Onlooking* ($r_{14} = -0.552$, $p = .041$) and vocabulary. Correlations between proportions at 1;6 and concurrent vocabulary only showed *Objects* engagement as negatively correlated ($r_{14} = -0.532$, $p < .050$). The urban 1;6 and 2;1 engagement proportions showed no significant correlations with 2;1 vocabulary.

Applying other approaches

We next show how replicated categorizations from previous research correlate with vocabulary to demonstrate how other approaches, with collapsed categories, yield different results. For this, we present only correlations between the 1;1 engagement level proportions with vocabulary at 1;6 and 2;1. Table 6 presents correlations for the Childers et al. (2007) tri-level engagement classification applied to our data.

Table 5. Spearman's correlations between engagement levels' proportions and vocabulary sizes at all collection periods using the categorization set forth in this paper.

Engagement Level	Rural Vocabulary			Urban Vocabulary		
	1;1	1;6	2;1	1;1	1;6	2;1
Unengaged						
1;1	0.134	0.064	-0.324	-0.518	-0.242	-0.374
1;6		0.244	-0.158		-0.206	0.066
2;1			-0.143			-0.096
Onlooking						
1;1	0.139	0.055	-0.181	-0.048	-0.297	**-0.552***
1;6		0.173	-0.147		-0.072	-0.363
2;1			-0.235			-0.325
Objects						
1;1	0.060	0.033	-0.101	-0.160	**-0.706****	-0.459
1;6		-0.046	-0.359		**-0.532***	-0.033
2;1			-0.489			0.193
Observing						
1;1	0.081	0.314	0.187	0.040	0.268	-0.231
1;6		0.099	0.223		-0.015	0.000
2;1			**0.659***			-0.206
Persons						
1;1	0.236	0.200	**0.723****	0.073	**0.772****	**0.598***
1;6		-0.050	0.130		0.510	0.095
2;1			0.097			0.052
Passive-JA						
1;1	-0.406	-0.464	-0.227	-0.351	-0.189	-0.053
1;6		-0.415	-0.187		-0.288	-0.220
2;1			0.190			-0.154
Shared-JA						
1;1	0.012	0.363	0.366	-0.051	-0.287	-0.039
1;6		0.100	0.290		0.046	0.181
2;1			**-0.568***			-0.352
Coordinated-JA						
1;1	-0.307	**-0.538***	-0.474	0.142	0.129	**0.660***
1;6		-0.147	0.157		0.136	0.172
2;1			-0.123			-0.070

Note: *p < .05; **p < .01.

Table 6. Spearman's correlations between the proportions of time spent in 1;1 engagement levels and vocabulary size at 1;6 and 2;1 assessed by the Childers et al. (2007) categorization.

	Vocabulary at 1;6	Vocabulary at 2;1
RURAL		
Low-Level	0.134	-0.351
Mid-Level	0.371	**0.798****
High-Level	**-0.591***	-0.476
URBAN		
Low-Level	-0.249	**-0.695****
Mid-Level	-0.017	0.114
High-Level	0.004	0.457

Note: Low-Level = *Unengaged* + *Onlooking* + *Observing*; Mid-Level = *Objects* + *Persons*; High Level = *Passive-JA* + *Shared-JA* + *Coordinated-JA*. *$p < .05$; **$p < .01$.

Results show that between 1;1 proportions of the tri-level categorization and 1;6 vocabulary, only *High-Level* engagement in the rural area was negatively correlated ($r_{14} = -0.591$, $p = .029$), but there were no significant relations in the urban area. Correlations of the same proportions with 2;1 vocabulary were positively correlated with rural *Mid-Level* engagement ($r_{14} = 0.798$, $p < .001$), and a significant negative correlation with urban *Low-Level* engagement ($r_{14} = -0.695$, $p = .005$).

Table 7. Spearman's correlations between the proportions of time spent in 1;1 engagement levels and vocabulary size at 1;6 and 2;1 assessed for the Carpenter et al. (1998) categories.

	Vocabulary at 1;6	Vocabulary at 2;1
RURAL		
Attention Following	0.187	-0.015
Joint Engagement	**-0.560***	-0.480
URBAN		
Attention Following	0.101	-0.279
Joint Engagement	0.114	**0.623***

Note: Attention Following = *Passive-JA* + *Observing*;
Joint Engagement = *Shared-JA* + *Coordinated-JA*. *$p < .05$; **$p < .01$.

Table 7 provides the results for the Carpenter et al. (1998) engagement level classification. They showed that rural *Joint Engagement* has a significant negative correlation with 1;6 vocabulary ($r_{14} = -0.560$, $p = .040$), while urban *Joint Engagement* had a positive correlation with 2;1 vocabulary ($r_{14} = 0.623$, $p = .017$).

Discussion

Our main research question was: To what extent can an extended, full-spectrum analysis of infant engagement contribute to our understanding of vocabulary development in natural non-industrial settings? In addition, how do the correlations between infant engagement and vocabulary size differ across non-industrial rural and urban communities? To find answers, we first explore how proportions of infants' engagements differ between the two communities. Second, we investigate the vocabulary sizes of the infants. Third, we analyze the cultural differences in correlations between proportions of infant engagements and vocabulary size. Fourth, we compare our approach to two other approaches.

Differences in Infant Engagement

In engagement levels, the results in Table 3 show that infants in both communities appear to have a similar distribution for engagement levels, but there are also significant differences between the two communities. In the rural area, infants spent significantly more time in forms of solitary engagement – *Unengaged* and *Observing* – than in the urban area, where they spent more time in forms of triadic engagement – *Passive-JA* and *Shared-JA*.

Explanations for these differences are based on community lifestyles. The rural area relies on subsistence farming for sustenance and income, whereas the urban area follows a market-economy. Due to the greater demands of subsistence lifestyle, mothers often work in the fields, and the entire community is responsible for household and caregiving chores (Greenfield, 2009; Keller, 2012). This was true in our rural community: most fathers worked in South Africa or Maputo and were away for several months at a time, and siblings take care of many household tasks, including caring for infants. As infants are yet unable to participate in the community, and other individuals have daily tasks, this could result in an environment where infants spend more time in solitary engagement (Hoff, 2006; Keller, 2012), which would explain the significantly higher rural proportions of *Unengaged* and *Observing*.

These findings are also consistent with the view that caregiving in the rural community focuses on developing communal action autonomy (Keller, 2012). The fostering of action autonomy presupposes that infants should engage autonomously, which might be triggered by leaving them to act on their own. In particular, the higher proportion of *Observing* could be the result of this, as it entails that infants attend to other people's activities autonomously. Further research into the motives of caregivers in leaving infants on their own, as well as caregivers' perceptions of their role in infant development, could confirm whether more solitary engagement does actually foster action autonomy.

In a non-industrial urban area, daily life is more focused on individual specialization and intra-community markets, and education levels tend to be higher than in the prototypical rural area (Keller, 2012). The socio-demographics of urban areas could explain why the learning environment there focuses on developing communal psychological autonomy (Greenfield, 2009; Keller, 2012), where others actively involve infants in engagements that focus on cognitive development, all the while learning communal responsibilities. Compared to non-industrial rural communities, urban communities are characterized as focusing more on the interests and goals of children in regard to object stimulation, as well as more face-to-face interactions, and so provide more opportunities for triadic joint engagement (Callaghan et al., 2011; Carpenter & Liebal, 2011; Keller, 2007). This in turn would account for the significantly higher urban proportions of *Passive-JA* and *Shared-JA*. Moreover, the decrease of *Passive-JA* and increase of *Shared-JA* over time could be explained by the increased ability of infants to actively engage in joint attention as a result of developing psychological autonomy. At the same time, infants' overall engagement in joint attention remains fairly constant, so any developmental change is probably in quality, not quantity.

This finding differs from Bakeman and Adamson (1984), and from Childers et al. (2007), who found that the amount of time infants spend in all joint attention categories increased over time for the comparable age groups. To a large extent, our difference with Bakeman and Adamson can be explained by the difference in culture, since an industrial community is known to engage infants in more object-oriented interactions. The difference with the Childers et al. study is more likely due to the semi-structured methods used to elicit simulated play and the introduction of novel toys, both of which may have triggered more joint attention than normal. This also applies to Bakeman and Adamson who also used semi-structured elicitation. As a result, earlier observations may not have yielded a reliable representation of natural interactions (see, Mastin et al., 2015, for an extended discussion).

To summarize our first step, we see that our novel categories *Observing* and *Shared-JA*, as well as one category of solitary engagement (*Unengaged*) and one of joint engagement (*Passive-JA*), play a substantial role in cross-cultural differences. Now, what relationship, if any, is there between engagement level proportions and vocabulary development? Given the results of earlier studies (Adamson et al., 2004; Carpenter et al., 1998; Childers et al., 2007; Morales et al., 2000; Tomasello & Farrar, 1986), urban infants might be expected to gain more from increased interactions relying on joint attention. The higher proportion of *Observing* in the rural area, on the other hand, may provide infants with more opportunities to learn vocabulary from overheard speech.

Higher Expressive Vocabulary Scores in the Urban Area

Results from the adapted MBCDI (Table 4) show that vocabulary size in the urban site was larger than for rural infants at all three ages observed. We discuss four possible explanations for this. First, the adaptation of the MBCDI may have been more culturally appropriate for the urban area. However, the adaptation and piloting of the MBCDI took place with local informants in both sites. We took care to choose appropriate terms in both communities, and when we chose words that could be more appropriate in one community this was counterbalanced by other words that would be more appropriate in the other community.

Second, caregivers have been known to both overestimate and underestimate vocabulary (Houston-Price et al., 2007; Law & Roy, 2008). Urban mothers may have overestimated their infants' vocabularies more than rural mothers did. The urban vocabularies at 2;1 are significantly higher than those in our norming sample (Supplement S1), which suggests that either these mothers overestimate their children’s vocabulary or that participation in this research had a beneficial effect on the children's development. Equally, we found that rural mothers may have underestimated their infants' vocabulary at age 1;1. This could be because rural mothers are away from the house a lot, and leave their children in someone else’s care. De Houwer, Bornstein, & Leach (2005) suggested that, when mothers spend much time away from their child, administrating MBCDIs from multiple reporters might produce a better measure. We observed that some mothers regularly consulted other members of the household during the MBCDI interviews, especially in the rural area, but we did not keep a record of how frequently this occurred. Recall that the validation of the vocabulary with the infants' own speech production yielded good results for the MBCDI scores at 1;6 in both communities, and at 2;1 in the rural community. Since we found no significant correlations with MBCDI scores at 1;1, the rural underestimation for this age group does not affect our findings. The possible overestimation in the urban community at 2;1, however, may affect our results.

Third, it is possible that bilingualism in the urban area caused vocabulary to become overestimated. While infants in bilingual environments tend to have smaller vocabularies for each individual language (Oller & Eilers, 2002), their total conceptual vocabulary size tends to be the same as that of monolingual infants (Junker & Stockman, 2002; Patterson, 1998).

Since the urban MBCDI adaptation was administered to measure total conceptual vocabulary, bilingualism is unlikely to be relevant.

Finally, the difference could be due to differences in the amounts of language socialization in different communities. A different analysis of the same data, in fact, demonstrated that the mean number of infant-directed utterances is six times higher in the urban community than in the rural one (Schots, Vogt & Mastin, 2015), and we found similar differences in the amount of infant-directed co-speech gestures (Vogt & Mastin, 2013). This could be explained by different socio-demographics in these two environments; slightly higher urban SES level, family size, and both urban parents living at home — all could result in greater amounts of and greater variation in infant-directed speech and gesture (Hoff, 2006). This, in turn, could have a cumulative effect on vocabulary development (Fernald et al., 2012; Hart & Risley, 1995; Hoff, 2006).

Although part of the difference in vocabulary may be attributed to one of the first three explanations, we believe that differences in SES and in the rural and urban socio-demographics provide the most likely explanation for the differences in vocabulary size. Moreover, such differences may not only relate to differences in the amount infant-directed speech (Hart & Risley, 1995), but also in other non-verbal aspects of infant socialization and engagement.

Infant Engagement and Vocabulary Development

For the relation between infant engagement and vocabulary development, our results show differences between sites for the relations of solitary and triadic engagements to infants' vocabulary, and also similarities between sites for the relation of dyadic engagement with vocabulary size (Table 5). There was a positive correlation between the amounts of *Observing* at 2;1 and infants' vocabulary at 2;1 in the rural environment. Given that engagements in prototypical rural environments generally involve actions displayed for infants to mimic and master (Greenfield, 2009; Keller, 2012; Schieffelin & Ochs, 1986), it seems appropriate the amount of time infants spend *Observing* others might relate to word learning. In situations where infant-directed speech and other forms of child-centered socialization are scarce, infants would have to rely more on overheard speech (Akhtar & Gernsbacher, 2007; Lieven, 1994), although a recent study from a Mayan village suggests that children may not learn much from overheard speech (Shneidman & Goldin-Meadow, 2012). When infants focus their attention on goal-oriented actions of others', there may be some situations where infants could learn from overheard speech. Rather than *Onlooking* to someone, *Observing* could provide enough contextual information for infants to infer the meaning of some overheard words. That *Observing* has a positive correlation in the rural, but not the urban area could be because at both 1;1 and 1;6 the proportion of time rural infants spent *Observing* was significantly greater than for urban infants (Table 3). Perhaps *Observing* is beneficial for word learning when it occurs often, and in the same contexts, throughout development.

In the urban community, all significant relations between solitary engagements and vocabulary are negative. First, the proportions of *Objects* engagement at 1;1 and 1;6 were negatively related to vocabulary at 1;6. As *Objects* engagement involves no communication partners, there is little likelihood that the proportion of time spent *Onlooking* could be beneficial to word learning. Second, the proportion of *Onlooking* engagement at 1;1 was negatively related to vocabulary at 2;1. *Onlooking* likewise involved no interaction between an infant and a target or partner, so, unlike in *Observing*, any speaker's behavior provides no clear context in goal-oriented behavior, thus making it hard to infer what an unfamiliar word means. The more time infants spend in solitary engagements, except *Observing*, the less time they spend interacting with people, and will have fewer opportunities to learn novel words.

With respect to joint engagements in both communities, we found correlations between *Persons* engagement at 1;1 and vocabulary at 2;1 were positive in both locations. Yet, correlations between *Coordinated-JA* engagement at 1;1 and vocabulary at later ages were negative in the rural community, yet positive in the urban community. Why these two patterns? First, in regard to *Persons* engagement, it may be the case, in non-industrial communities that social joint engagement interactions (excluding target objects or events) provide infants with culturally salient situations that focus on the fostering of communal responsibilities of the infant. Since non-industrial environments consider communal autonomy to be important, socialization tends to focus on the development of social knowledge and skills, with attention to kinship relations, turn taking, communal service, interpersonal responsibilities, etc. (Abels et al., 2005; Greenfield, 2009; Keller, 2012). The acquisition of such knowledge would be better fostered through *Persons* engagements than through triadic joint attention, especially since during *Persons* interactions, any information exchanged should relate more to social relations and interpersonal activities, than to physical targets within an environment. One difference is that the rural community focuses more on action autonomy, so the development of motoric skills might be considered most important (Keller, 2007; Schieffelin & Ochs, 1986), while the urban community focuses more on the acquisition of turn taking skills and interpersonal relationships important to achieving psychological autonomy. This nuanced difference is supported in our analysis of the same data with respect to the gestures addressed to infants (Vogt & Mastin, 2013).

Second, for *Coordinated-JA*, there is a negative relation with rural infants' vocabulary, and a positive relation with urban infants' vocabulary. The positive urban relation is not surprising since urban non-industrial learning environments share characteristics with prototypical *industrial* urban cultures, such as a preference for object stimulation and child-centered interactions to achieve psychological autonomy (Keller, 2012), which could often manifest as *Coordinated-JA*. Moreover, many studies from industrial communities have shown a positive relation between joint attention and vocabulary development (Adamson et al., 2004; Carpenter et al., 1998; Morales et al., 2000; Mundy & Gomes, 1998; Tomasello & Farrar, 1986). Note, however, that we should treat all positive correlations with urban infants' vocabulary size at 2;1 with care, since mothers may have overestimated their infants' vocabulary. All the other correlations between *Coordinated-JA* and urban vocabulary are low, so the urban situation in this respect may be close to the rural community.

The fact that rural *Coordinated-JA* was negatively correlated with vocabulary was unanticipated given that infants appear to master joint attention skills across cultures around the same age (Callaghan et al., 2011; Lieven & Stoll, 2013; Salomo & Liszkowski, 2013). Note that at 2;1, *Shared-JA* also revealed a negative correlation with vocabulary, but due to its infrequent occurrence and low inter-rater reliability, we will focus our discussion on *Coordinated-JA* instead. In view of the data analyzed here, we offer two possible explanations. First, if object stimulation is not characteristic of non-industrial rural environments, then language socialization is unlikely to occur during joint attention with objects. To some extent, this is supported by our analysis of infant-directed speech and gestures. Schots et al. (2015) found that in both Mozambican communities few objects are labeled in infant-directed speech, and even less so in rural Mozambique as there is overall six times less speech addressed to infants. In addition, while nearly 60% of the infant-directed gestures in the urban community were accompanied by speech, only 33% were in our rural sample (Vogt & Mastin, 2014). Moreover, in about 80% of the rural interactions where speech is accompanied by gestures, the gestures convey information not contained in the speech. These results suggest that rural infants' *Coordinated-JA* interactions are often silent, but when speech does occur there is little naming of objects, and when caregivers do name

objects, they often do not use gestures to provide deictic information that could help acquire the appropriate association. So, the more time infants spend in *Coordinated-JA*, the fewer opportunities they have to learn from the utterances addressed to them, since infant-directed utterances rarely contain object labels. For urban infants, the larger numbers of infant-directed utterances result in more object labeling, often supported by gestures indicating the target object, thus providing them with more opportunities to learn object labels.

Second, the time infants spent with specific communication partners may play a crucial role in explaining the negative correlation between *Coordinated-JA* and vocabulary size in the rural community. A deeper exploration into the relation between infant engagement and vocabulary has shown that the amount of time rural infants at 1;1 spent in *Passive-JA* and *Shared-JA* with their mothers correlated positively with vocabulary, but that triadic engagements (including *Coordinated-JA*) with non-caregivers and groups result in negative correlations (Mastin, 2013). Interactions with non-caregivers, then, may not be beneficial. This parallels findings from a study of the Dogon in Mali, where children often have to compete for resources with other household members, especially grandmothers, and this competition is related to a slower growth rate (i.e., stunting), as well as higher infant/child mortality (Strassman, 2011). Stunting is a crucial factor in delaying children's cognitive development (Grantham-McGregor et al., 2007). The negative correlations in non-caregiver and multiparty interactions could be understood by the complexity of navigating attention between multiple communication partners, a target object, and any verbal utterance(s) addressed to the infant (or not addressed to her). Interestingly, however, the time urban infants spend in *Coordinated-JA* with multiple communication partners revealed a positive correlation with vocabulary at 2;1. Although cognitively demanding, multi-party interactions could further explain the negative correlation in the rural community.

In sum, the results suggest that *Coordinated-JA* may not necessarily be the major contributor and scaffold to language acquisition (Akhtar & Gernsbacher, 2007; Mundy & Gomes, 1998; Scofield & Behrend, 2011), at least not for all cultures. Instead, other types of engagement, such as *Observing* and *Persons* engagements, could significantly relate to word learning over early development. Moreover, the shared positive relation with *Persons* engagement in both communities, and the conflicting significant relation with *Coordinated-JA* engagement suggest that urban and rural non-industrial communities do, indeed, represent separate, but not mutually exclusive, learning environments (cf., Greenfield, 2009; Keller, 2012). However, we need to bear in mind that these findings are based on an exploratory study and that more structured research is required to investigate the validity and generalizability of these findings.

Other Approaches

For the fourth step of our analysis, we discuss the differences in the correlations between vocabulary and proportions of engagement levels obtained in our extended categorization compared to those obtained by applying the less extensive engagement categorizations from Childers et al. (2007) and Carpenter et al. (1998).

The correlation analyses using the engagement level categorizations of these two studies resulted in three findings that followed a similar trend. First, in Childers et al.'s (2007) tri-level categorization in Table 6, there were no significant correlations between *Mid-Level* engagement (*Objects* and *Persons*) in the urban area and vocabulary at either 1;6 or 2;1. However, in our results in Table 5, both *Objects* and *Persons* engagements in the urban area were significantly correlated with vocabulary at 1;6, and *Persons* engagement continued to be a significant correlate of vocabulary at 2;1. These two categories' results cancel each other out when combined in Childers et al.'s (2007) *Mid-Level* category since they have

opposite correlations to urban infants' vocabulary. Second, also in Childers et al.'s (2007) tri-level categorization, there were no significant correlations between proportions of urban *High-Level* engagement from 1;1 with vocabulary at 2;1 (Table 6). However, when correlations are computed using either our own categories or Carpenter et al.'s (1998; cf. Table 7), the significant relation of *Coordinated-JA* engagement still remains evident. The third difference relates to solitary engagement. The results from both our own categorization, and Childers et al.'s (2007), show that non-joint engagement behaviors (i.e., the *Low-Level* category that combines *Onlooking, Observing* and *Unengaged*) can be negatively correlated to vocabulary, which Carpenter et al. (1998) did not analyze. These differences make it clear that our extended categorization reveals correlations that would have been overlooked if our analysis were based on the engagement levels applied in earlier studies. These examples illustrate the complexity of measuring the relations between infant engagement and vocabulary development, and show that analysis of extended engagement level categories is more informative.

Conclusions

The main research question we addressed was: To what extent can an extended, full-spectrum analysis of infant engagement contribute to our understanding of vocabulary development in natural settings? In brief, our exploration demonstrates that engagements, which often fall outside the scope of research into the relation between (joint) attention and vocabulary development (e.g., *Onlooking, Objects, Observing, Persons* and *Shared-JA*), can have significant correlations to later vocabulary size and therefore demand attention in future investigations. In addition, our study demonstrates the potential role that non-triadic joint engagements (i.e., *Persons*) may have on vocabulary development. One reason why we found these results was that we observed natural situations without providing any instructions to the participants, as opposed to the semi-structured or experimental methods usually used to study the relations between attention and vocabulary development (Bakeman & Adamson, 1984; Carpenter et al., 1998; Childers et al., 2007). The present study, though, only begins to explore the value of this approach. Due to our small samples, use of parental checklists to assess vocabulary size, use of correlations, and use of an understudied cultural setting, this study lacks the power to provide conclusive evidence. Nevertheless, it provides new questions for further study: What exactly is the role of solitary engagement in language development? To what extent can children learn vocabulary by observing others? To what extent do children learn language via dyadic interactions, and what qualities of such interactions relate best to vocabulary development?

The secondary issue we explored here was: How do correlations between infant engagement and vocabulary size vary in non-industrial rural and urban communities? We identified at least two factors that may play a role in Mozambican language acquisition, factors that are neither mutually exclusive nor exhaustive. First, the positive correlations between *Persons* engagement and vocabulary, and the conflicting correlations between *Coordinated-JA* and vocabulary, indicate that the rural and urban Mozambican communities represent different, non-industrial learning environments (Keller 2012). Second, our results suggest that *Coordinated-JA* may not have to be the primary contributor and scaffold to language acquisition (cf. Akhtar & Gernsbacher, 2007; Mundy & Gomes, 1998; Scofield & Behrend, 2011). In the Mozambique communities we studied, *Persons* interactions related best to language learning, reflected in the acquisition of words for kinship relations, and non-nouns (i.e., pronouns or verbs). This is consistent with the division between urban industrial and non-industrial communities that foster the development of communal responsibilities and action autonomy (Keller, 2012).

To conclude, a full-spectrum analysis of infant engagement, with naturalistic observations in a variety of (non-industrial) cultures, like the one presented here, has the potential to contribute new insights to the relations between different forms of engagement and infants' early vocabulary development. In particular, the present study suggests that *Observing* and dyadic *Persons* engagements may contribute more to vocabulary development than *Coordinated Joint Attention* in at least some non-industrial communities. But since this study was an exploratory one, we need additional - more structured - research before these conclusions can be generalized.

Acknowledgements

This research was funded by the Netherlands Organization for Scientific Research (NWO) with a VIDI grant (number 276-70-018) awarded to PV. We thank Afra Alishahi, Eve Clark and Fons Maes for their invaluable comments on earlier versions of this manuscript. Many thanks to Wona Sanana, Associação Communitário Ambiente da Mafalala, and the local research assistants for their support in Mozambique. Finally, our gratitude goes to all participants involved in this study.

References

Abels, M., Keller, H., Mohite, P., Mankodi, H., Shastri, J., Bhargava, S., Jasrai, S., & Lakhani, A. (2005). Early socialization contexts and social experiences of infants in rural and urban Gujarat, India. *Journal of Cross-Cultural Psychology, 36,* 717-738.

Akhtar, N., & Gernsbacher, M.A. (2007). Joint attention and vocabulary development: A critical look. *Language and Linguistic Compass, 1*(3), 195-207.

Adamson, L.B., Bakeman, R., & Deckner, D.F. (2004). The development of symbol-infused joint engagement. *Child Development, 75*(4), 1171-1187.

Bakeman, R., & L.B. Adamson. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55*(4), 1278-1289.

Barton, M.E., & Tomasello, M. (1991). Joint attention and conversation in mother-infant-sibling triads. *Child Development, 62,* 517-529.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., et al. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development, 75 (4)*, 1115–1139.

Bornstein, M.H., & Putnick, D.L. (2012). Cognitive and socioemotional caregiving in developing countries. *Child Development, 83*(1), 46-61.

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs & B. Schieffelin (Eds.), *The handbook of language socialization.* Oxford: Blackwell.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., & Tomasello, M. (2011). Early social cognition in three cultural contacts. *Monographs of the Society for Research in Child Development, 76*(2), 1-142.

Carpenter, M., & Liebal, K. (2011). Joint attention, communication, and knowing together in infancy. In A. Seemann (Ed.), *Joint Attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 159-181). Cambridge, MA: MIT Press.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), 1-133.

Childers, J.B., Vaughan, J., & Burquest, D.B. (2007). Joint attention and word learning in Ngas-speaking toddlers in Nigeria. *Journal of Child Language, 33*, 199-225.

De Houwer, A., Bornstein, M., & Leach, D. (2005). Assessing Early Communicative Ability: A Cross-Reporter Cumulative Score for the MacArthur CDI. *Journal of Child Language, 32*(4), 735-758.

Eisenbeiss, S. (2010). Production Methods in Language Acquisition Research. In E. Blom & S. Unsworth (eds.) *Experimental Measures in Language Acquisition Research* (pp. 11-34). Amsterdam: John Benjamins.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, *43*(6), 543-549.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D., & Pethick, S. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1-185.

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.

Fernald, A., Marchman, V.A. & Weisleder, A. (2012). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Psychology*, 1-13.

Gaskins, S. (2006) Cultural perspectives on infant-caregiver interaction. In N.J. Enfield & S.C. Levinson (Eds.) *Roots of Human Sociality*. Oxford, UK: Berg.

Grantham-McGregor, S., Cheung, Y. B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. (2007). Developmental potential in the first 5 years for children in developing countries. *The Lancet, 369 (9555)*, 60–70.

Greenfield, P.M. (2009). Linking social change and developmental change: Shifting pathways of human development. *Developmental Psychology, 45*(2), 401-418.

Harkness, S. (1977). Aspects of social environment and first language acquisition in rural Africa. In C. Snow & C. Ferguson (Eds.), *Talking to children: Language input and acquisition* (pp. 309-316). Cambridge: Cambridge University Press.

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

Hobson, R. P. (2005), Social engagement and understanding in chimpanzees and humans. *Monographs of the Society for Research in Child Development*, *70*, 133–152.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26,* 55-88.

Houston-Price, C., Mather, E., & Sakkalou, E. (2007). Discrepancy between parental reports of infants' receptive vocabulary and infants' behavior in a preferential looking task. *Journal of Child Language, 34*(4), 701–724.

Junker, D.A., & Stockman, I.J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology, 11*(4), 381-395.

Keller, H. (2007). *Cultures of infancy*. Mahwah, NJ : Lawrence Erlbaum. .

Keller, H. (2012). Autonomy and relatedness revisited: Cultural manifestations of universal human needs. *Child Development Perspectives, 6*(1), 12-18.

Law, R. & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Developmental Inventories. *Child and Adolescent Mental Health*, 13, 198-206.

LeVine, R.A., Dixon, S., LeVine, S., Richman, A., Leiderman, P.H., Keefer, C.H. & Brazelton, T.B. (1994). *Childcare and culture: Lessons from Africa*. New York: Cambridge University Press.

Lewis, M.P. (2009). Ethnologue: Languages of the world, 16th edition. Dallas, Texas, USA: SIL International. Retrieved from: http://www.ethnologue.com/.

Lieven, E. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition*. Cambridge: Cambridge University Press.

Lieven, E. & Stoll, S. (2013). Early Communicative Development in Two Cultures: A comparison of the communicative environments of children from two cultures. *Human Development*, 56, 178-206.

Mastin, J. D. (2013) *Exploring infant engagement, language socialization & vocabulary development: A study of rural and urban communities in Mozambique*. TiCC Ph.D. Series no. 31. Tilburg University

Mastin, J.D., Vogt, P., Schots, D. & Maes, A.A. (2015). Analyzing infant engagement: Filling in the gaps in research approaches. *Under review.*

Morales, M., Mundy, P., Delgado, C.E.F., Yale, M., Messinger, D., Neal, R., & Schwartz, H.K. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21*(3), 283-298.

Mundy, P., & Gomes, A. (1998). Individual differences in joint attention skill development in the second year. *Infant Behavior and Development, 21*(3), 469-482.

Oller, D.K., & Eilers, R.E. (Eds.). (2002). *Language and literacy in bilingual children*. Clevedon: Multilingual Matters.

Patterson, A.L. (1998). Expressive Vocabulary Development and Word Combinations of Spanish-English Bilingual Toddlers. *American Journal of Speech Language Pathology, 7*, 46-56.

Pine, J.M., Lieven, E.V.M., & Rowland, C. (1996). Observational and checklist measures of vocabulary composition: What do they mean? *Journal of Child Language*, *21*, 573-590.

Rabain-Jamin, J., Maynard, A. E., & Greenfield, P. M. (2003). Implications of sibling caregiving for sibling relations and teaching interactions in two cultures. *Ethos, 31*(2), 204-231.

Salomo, D. & Liszkowski, U. (2013). Sociocultural Settings Influence the Emergence of Prelinguistic Deictic Gestures. *Child Development*, *84*(4), 1296-1307.

Schieffelin, B., & Ochs, E. (1986). Language socialization. *Annual Review of Anthropology, 15*, 163-191.

Schots, D.M., Vogt, P., & Mastin, J.D. (2015) Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands, and rural and urban Mozambique. *First Language*. In press.

Scofield, J., & Behrend, D.A. (2011). Clarifying the role of joint attention in word learning. *First Language, 31*(3), 326-341.

Shneidman, L. & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: how important is directed speech? *Developmental Science, 15*(5), 659-73.

Strassmann, B. I. (2011). Cooperation and competition in a cliff-dwelling people. *Proceedings of the National Academy of Sciences*, *108*, 10894-10901.

Tomasello, M. (1995). Joint attention as social learning. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp.103-130). Hillsdale, NJ: Lawrence Erlbaum.

Tomasello, M., & Farrar, M.J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454-1463.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioural and Brain Sciences, 28,* 675-735.

Vogt, P. & Mastin, J. D. (2013) Rural and urban differences in language socialization and early vocabulary development in Mozambique. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 3687-3692). Austin, TX: Cognitive Science Society.

Vogt, P. & Mastin, J.D. (2014) Cross-cultural differences in three prototypical learning environments. Presented at *International Congress for the Study of Child Language (IASCL)*, Amsterdam, July 2014

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

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.*

Zukow-Goldring, P. (2002). Sibling caregiving. In M.H. Bornstein (Ed.), *Handbook of parenting: Vol. 3- Being and becoming a parent* (2nd ed., pp. 253-286). Mahwah, NJ: Lawrence Erlbaum. .