Published in final edited form as:
J Child Lang. 2012 September ; 39(4): 899–918. doi:10.1017/S0305000911000407.

Basic language comprehension and production in >100,000 young children from sixteen developing nations

MARC H. BORNSTEIN[*] and **CHARLENE HENDRICKS**
Eunice Kennedy Shriver National Institute of Child Health and Human Development

Abstract

Using the Multiple Indicator Cluster Survey, language comprehension and production were compared in a sample of 101,250 children aged 2;00 to 9;11 and a focus subsample of 38,845 children aged 2;00 to 4;11 from sixteen under-researched developing nations. In the whole sample, comprehension slightly exceeded production; correlations between comprehension and production by country were positive and significant, but varied in size, and the average correlation was positive, significant, and small to medium. Mean comprehension and production varied with child age, reaching an asymptote at 5;00, and correlations between comprehension and production by age were positive, significant, and similar at each age. In the focus subsample, comprehension exceeded production; correlations between comprehension and production by country were positive and significant, but varied in size, and the average correlation was positive, significant, and medium in size. Children in countries with lower standards of living were less likely to demonstrate basic language comprehension or production.

INTRODUCTION

This study reports variation in basic language comprehension and production as well as relations between comprehension and production in 101,250 children aged 2;00 to 9;11 from sixteen under-researched developing countries.

Child language comprehension and production

Child language study traditionally distinguishes between comprehension (broadly, understanding) and production (broadly, expression). Notably, relations between these two basic integral facets of language are still somewhat clouded. Comprehension and production feasibly relate to one another in two main ways. If comprehension and production are measured comparably in the same children, first, mean absolute levels of comprehension and production can be compared and, second, relative levels of individual variation between comprehension and production can be determined. It is important to understand both relations because, ultimately, comprehension and production need to be coordinated;

Address for correspondence: Dr Marc H. Bornstein, Child and Family Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Suite 8030, 6705 Rockledge Drive, Bethesda, MD 20892-7971, USA. tel: 301-496-6832; fax: 301-496-2766; Marc_H_ Bornstein@nih.gov.

[*]We thank A. Blaine and A. Bradley. This research was supported by the Intramural Research Program of the NIH, NICHD.

without such coordination, “speakers would be unable to use language to communicate and to infer intentions” (Clark & Hecht, 1983: 326).

Two relations between comprehension and production: absolute and relative

In absolute terms, it is usually the case that comprehension precedes production developmentally and exceeds production substantively, pointing to a degree of independence in the two. Comprehension and production follow different timetables with respect to onset, milestone achievement, and trajectory and rate of development, for example (Bates, 1993; Clark & Hecht, 1983; Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994). Benedict (1979) studied the first 50 words children comprehended and produced. Children first understood words at 0;9 but first produced words at 1;0, on average, and children reached a 50-word milestone in comprehension at around 1;1, whereas they reached 50 words in production at around 1;6. Bauer, Goldfield and Reznick (2002) identified distinct trajectories of expressive and receptive vocabulary from ages 0;8, 0:9 or 0;10 to 1;2 and then to projected performance at 1;9. Campbell, Macdonald and Dockrell (1982) pointed to distinct lexical representations in the two domains in preschool-age children learning new words. In addition, children learning different languages almost invariably have larger concurrent receptive than productive vocabularies (Kern, 2007; Zink & Lejaegere, 2002). Finally, comprehension does not always predict production (Bauer *et al.*, 2002). In these several respects, comprehension and production appear to be somewhat distinct aspects of language, and some authors have posited that comprehension and production are ‘dissociated’ psycholinguistic processes that draw on different skills and neurological bases (Bates, 1993; Fenson *et al.*, 1994; see also Dale & Goodman, 2005).

In relative terms, however, empirical work shows that associations between comprehension and production vary from moderate to strong, depending on the ages of the children studied, evaluation procedures, and the like, and so the two are thought to be related. Thus, in different languages, young children who comprehend more words are also likely to produce more words (Bornstein & Haynes, 1998; Reznick & Goldfield, 1992). MCDI norming data for Dutch (N-CDI; Zink & Lejaegere, 2002) show correlations between (word) comprehension and (word) production based on 1,024 children 0;8 to 1;4 (r=0.62) and 958 children 1;4 to 2;6 (r=0.88). MCDI norming data for European French (Kern, 2007), based on 548 children 0;8 to 1;4, show a slightly smaller, but still respectable, correlation between (word) comprehension and (word) production (r=0.45). Using the (UK version) CDI-WS, Tan and Schafer (2005) asked parents of children between 1;4 and 1;8 to indicate which items their children produced and which items they understood but did not produce; the two were correlated (r=0.80). Predictively, Harris, Yeeles, Chasin and Oakley (1995) hypothesized that early comprehension should underpin later production, and contrary to studies cited earlier many reports document a predictive relation between early receptive and later expressive language in children with language skills ranging from delayed to typical (Lyytinen, Poikkeus, Laakso, Eklund & Lyytinen, 2001; Tamis-LeMonda, Bornstein, Kahana-Kalman, Baumwell & Cyphers, 1998; Zink & Lejaegere, 2002). Moreover, just as children overextend words in comprehension (Behrend, 1988), they overextend them in production (Clark, 1993). In most respects, comprehension and production call on many of

the same constituent skills (Harris *et al.*, 1995). These several kinds of associations and similarities between comprehension and production implicate their intimate connection.

The variation in comprehension–production relations could be accounted for by any of several reasons. Many studies compare comprehension and production among children who are very young (in infancy, both comprehension and production are immature and restricted in range, thereby attenuating potential correlation); other studies include older children in whom difficulties in assessing comprehension place a different kind of methodological limit on understanding putative comprehension–production relations. Small sample size (and therefore limited power to detect effects) is another factor determining significance in this literature. Much, too, depends on the ways in which comprehension and production are measured: observations, reports and testing child language can yield different measures and results (Bornstein & Haynes, 1998). In short, age of child, sampling, procedure and language index each may influence comprehension–production relations. Anyway, it is likely that dissociation is never complete and that dissociation and association positions converge and are not mutually exclusive (Reznick & Goldfield, 1992).

The present study aimed to learn more about absolute and relative relations between basic language comprehension and production by overcoming some prior limitations. For this purpose, comprehension and production data were collected from a substantial number of children over a wide age range using a common measure of child language assessment: parent report. The samples in this study differed widely, too, in terms of language (and cultural) backgrounds; therefore, results ought also to transcend local variation in norms and patterns of development. In addition, we studied comprehension and production in nations barely represented (if at all) in contemporary developmental and language research.

International developmental language science

There is a dearth of population-based multinational data on child language, especially from developing countries. Most contemporary research in child language is Western (European or North American) in origin. Less than 10% of the literature in developmental science generally emanates from regions of the world that account for more than 90% of the world's population (Arnett, 2008), and critics have long wisely rejected broad generalizations derived from contextually restricted findings (Bornstein, 2010). As a corollary, the societies typically included in developmental research are usually sociodemographically and linguistically similar. Therefore, much less is currently known scientifically than is commonly acknowledged about child language generally or the majority of ecological contexts of child language development specifically. This sampling restriction is as limiting in terms of understanding idiosyncrasies of child language as it is generalizations and universals about it. Child language science can only benefit from an enlarged empirical representation of the world's children, and studies that use such a contextual lens promise deeper insights into how children comprehend and produce language (Bornstein, 2010). This situation is especially true for developing countries, which are virtually unexplored in the extant child language literature and, if at all, usually contribute small samples from single locales.

The countries we report about here represent the majority developing world, contra the minority developed world (National Center for Children in Poverty, 1999; UNICEF, 2006), and vary widely in terms of language as well as history and ideology, social situations, and other factors. Notably, however, all are low in socioeconomic status, which is thought generally to influence (later) child language development (Hoff, 2006). The nature of the social interactions that expose children to language and the properties of language itself are shaped by larger social and cultural considerations. The amount of speech addressed to children, the richness of the vocabulary, the rate of question asking, and the length of utterances are positive environmental input predictors of child language development (Bauer *et al.*, 2002; Bornstein, Haynes & Painter, 1998; Hart & Risley, 1995; Weizman & Snow, 2001). Beginning in the second year of life and continuing through the school years, children from lower socioeconomic strata tend to show slower rates of development than do children from higher socioeconomic strata (Dollaghan, Campbell, Paradise, Feldman, Janosky & Pitcairn, 1999; Huttenlocher, Vasilyeva, Cymerman & Levine, 2002; Rescorla & Alley, 2001). Indeed, children from lower socioeconomic strata are over-represented in samples identified as having poor communication skills (Tomblin, Hardy & Hein, 1991).

This study

The complex relations between comprehension and production allow for both association and dissociation. In this study, variation and concurrent associations in comprehension and production were evaluated in more than 100,000 children aged 2;00 to 9;11 from sixteen under-researched developing countries learning more than a dozen languages (including, for example, Arabic, Ashante, Bengali, English, French, Macedonian, Sango, Serbian, Spanish, Thai and Uzbek) that are typologically and genetically varied (representing, for example, Indo-Aryan, Indo-European, Mongolic, Niger-Congo, Romance, Semitic, Slavic, Thai-Kadai and Turk). The wider range of nations included in this study opens inquiry into understudied contexts of child language acquisition. Multinational–linguistic developmental inquiry can contribute to identifying, distinguishing and understanding general as well as specific patterns of child language development (Bornstein, 2010). Comprehension and production are normally studied in infancy and toddlerhood, yet both comprehension and production continue to develop at least until early adolescence (see, e.g., Wassenberg *et al.*, 2008, for comprehension, and Berman, 2004, for production). The wide age range included in this study opens inquiry into an understudied period of childhood as well.

METHOD

The MICS3

In collaboration with selected developing countries around the world, UNICEF (2006) developed the Multiple Indicator Cluster Survey (MICS), a household questionnaire designed to collect internationally comparable data to evaluate country-level progress in child development and family life (Bornstein, Britto, Nonoyama-Tarumi, Ota, Petrovic and Putnick, 2012). The main purposes of the MICS are to support evidence-based policy formulation, assess trends, and identify and measure disparities in progress towards international goals, such as the Millennium Declaration and the Millennium Development Goals. The present report uses data from the third MICS (MICS3) conducted 2005–2007.

The items used for this study came from an optional module of the Household Questionnaire in the MICS3 which asked mothers (or primary caregivers) about basic language skills in children aged 2;00 to 9;11. The respondent for each child answered *No* (=0) or *Yes* (=1) to the following two questions about the target child's basic language comprehension and production, respectively: “When you tell (name) to do something, does he/she seem to understand what you are saying?” and “Does (name) speak at all; can (he/she) make (him or herself) understood in words; can (name) say any recognizable words?” Other contemporarily prominent parent-questionnaire child-language measures, such as CDI: Words and Gestures, tap into associated processes by asking similar kinds of nominal questions (Fenson, Marchman, Thal, Dale, Reznick & Bates, 2007).

The Human Development Index

Child language comprehension and production data in the sixteen countries reported here are related to key country-level indicators from the Human Development Index (HDI; United Nations Development Programme, 2007). The HDI was developed by the United Nations as a measure of the social and economic status of a country as a preferred way to represent general standards of living. The HDI has three major indices: life expectancy (in years), education (adult literacy rate and the percentage of school-aged children enrolled in primary, secondary and tertiary school), and gross domestic product (GDP; in purchasing power parity [PPP] in US dollars). Each country's life expectancy, education and GDP were scaled from 0 to 1 (based on minimum and maximum values of 25–85 years for life expectancy; 0–100% for literacy and school enrollment; and $100–40,000 PPP for GDP) and then averaged to compute the country's HDI. The HDI offers a reasonable proxy for the level of support that is generally available for promoting human development in poor nations. As such, it likely connects to many aspects of the family and home environment with known relations to child development. The HDI provides a set of universally accepted standards for human development and represents a global shift in thinking about development from purely economic progress to human well-being (Mankiw, Romer & Weil, 1992). The HDI is more about the population that resides in the country and less about country indicators per se. We therefore selected the HDI over other global social and economic national indices as a means of grouping countries in a meaningful way. Countries with an HDI of 0.80 or greater are considered high, 0.50 to 0.79 medium, and 0.00 to 0.49 low. MICS3 data generally draw from the high, medium and low regions of the HDI, and this tripartite division is used to organize countries. (The HDI was not available for Iraq because of missing GDP data; however, the life expectancy index and the education index were available.)

Analytic plan

First, we examined the whole sample of children aged 2;0 to 9;11. We computed, for the entire age range and by year, the means and standard deviations, by country, for the proportions of children reported as comprehending and producing. We then computed correlations between language comprehension and production by country. Because of the disparity in sample sizes, a mean weighted and normalized correlation coefficient was computed for the whole sample. The correlations to be combined were transformed into Fisher *Z* values, which are approximately normally distributed and are numbers on a ratio scale and can thus be directly compared. *Z* values were then weighted and subjected to a

linear combination. The result is a weighted and normalized average correlation (Hedges & Olkin, 1985). In describing effect sizes, we follow Cohen's (1988: 223) terminology regarding coefficients: estimate of population correlation for a small effect size, $r_\Phi \approx 0.10$, medium effect size: $r_\Phi \approx 0.30$, and large effect size: $r_\Phi \approx 0.50$. The primary measures used in this study are dichotomous, and appropriate non-parametric analyses are used throughout. Given the wide age range it is possible that the correlations are larger than what might be expected for more age-restricted samples of younger children alone. Next, we examined linear or non-linear growth functions across age groups and, as a result, restricted the age range for a focus subsample to 2;00 to 4;11. Analyses of the focus sample were expanded to include potential covariates and a comparison of each country for each of the language measures to determine in which countries children performed better or worse than the overall effect of country.

WHOLE SAMPLE: 2;00–9;11

Participants

All sixteen developing nations provided data. If there was more than one child between the ages of 2;00 and 9;11 in a family, we randomly selected a target child in the household. Table 1 presents the countries, number of families providing data, average age of the selected child, and percentage of female children for the whole sample (ages 2;00 to 9;11). Also included is the predominant language(s) spoken in each country. Comprehension and production data are provided for 101,250 children of an average age of 5;5 (*SD*=2;3) with approximately equal numbers of girls and boys in each national sample. *N*s vary slightly by analysis as data were not complete for all variables (less than 1% of data was missing for any measure).

Comprehension, production and their relations

Descriptive statistics and the mean proportions of children comprehending and producing language and the correlation between language comprehension and production by country are reported in Table 1. Mean proportions of children comprehending and producing were high. However, the mean levels for comprehension and production marginally differ. A MANOVA comparing mean comprehension and production showed ($F(1,101189)=293.99$, $p<0.001$, $\eta^2_p=0.003$), with the difference between estimated marginal means=0.008. Not unexpectedly, comprehension is reportedly higher than production. Zero-order phi (r_Φ) correlations for the dichotomous measures of comprehension and production were uniformly positive, significant in all countries (except Montenegro), but varied from small to large. The weighted and normalized average correlation for the entire sample was positive, significant, and small to medium in size (r=0.21, standard error of the mean Z coefficient=0.003; 95% confidence interval=(0.202, 0.214)).

Comprehension and production by age

For each age for each country separately (Table 2), we computed mean proportions of children comprehending and producing and correlations between the two measures. The mean proportions of children comprehending and producing varied with age. Correlations

between comprehension and production by age were uniformly positive, significant at all ages, and varied only slightly.

We plotted the mean proportions of children comprehending and producing by age to investigate the nature of their respective growth functions (Figure 1). Examination of the plots suggested linear relations from ages 2;00 to 4;11, after which trend lines appeared to reach an asymptote. We therefore performed a series of logistic spline regressions to identify non-linear relations. Spline regression is a method for fitting separate regression lines in segments of a sample (Marsh & Cormier, 2001). The segments are defined by 'knots', the points on the independent variable (child age) where the regression line changes slope. Spline regression can be used to estimate the location of knots as well as test whether there is a significant change in slope before and after a knot. In this study, logistic spline regression was used to determine the point on the child age continuum where the slope changed (i.e. the location of a single knot in the regression line) and whether there was a significant difference in slopes when the distribution was segmented at the knot age. Start values for the intercept and slopes were set at 0, and the start value for the knot age was set at 4;0 based on visual examination of the plots.

Table 3 presents the estimated knot age and 95% confidence intervals. Because child age was collected only in years, we rounded up to 4;0 and tested whether the slope changed at age 4;0. The results showed significant changes in the slopes of regression lines at 4;0 for both comprehension and production. Furthermore, the estimation of the slopes from 2;00 to 4;11 was significant, with increasingly older children in that age range reporting higher mean proportions comprehending and producing language. The slope from 5;00 to 9;11 was not significant for production, and, although the Wald estimate was significant for comprehension, the odds ratio shows a negligible effect of child age.

Based on these results, we restricted the range of the next analyses to a focus subsample of children aged 2;00 to 4;11. The analyses and tables that follow are based on this focus sample.

FOCUS SAMPLE: 2;00–4;11

Participants

All sixteen developing nations provided data. Table 4 presents the countries, number of families providing data, the average age of the selected child, the percentage who were female for the focus sample ages 2;00 to 4;11, and the average number of children aged 2;00 to 9;11 in the household. Also included are the mean proportions of children comprehending and producing language and the correlation between language comprehension and production by country. Comprehension and production data are provided for 38,845 children of an average age of 3;0 (*SD*=0;9) and for approximately equal numbers of girls and boys. *N*s vary slightly by analysis as data were not complete for all variables (less than 1% of data was missing for any measure). We used pairwise deletion when computing the analyses.

Covariates

We considered child age and gender and number of children in the family aged 2;00–9;11 as potential covariates. Mean child age varied across countries (F(15,38849)=5.12, p<0.001, η^2_p=0.002), and, as might be expected, was associated with language comprehension (r_Φ=0.01, p<0.001) and production (r_Φ=0.07, p<0.001), but the magnitudes (effect sizes) are small. Given this pattern and the restricted age range of 2;00 to 4;11, we rejected child age as a covariate. Child gender was not used as a covariate either because similar percentages of girls and boys were selected in each country and across countries (χ^2(15, N=38865)=6.84, n.s.), and child gender in this age range for these samples for these variables was unrelated to either measure of language comprehension or production (r_Φs=0.01 and 0.01). We also compared language comprehension and production by gender by country, and only Iraq showed a gender difference in language production (χ^2(1, N=4204)=13.41, p<0.001), with boys greater than girls. Number of children 2;00–9;11 in the family varied across country (F(15,38849)=329.47, p<0.001, η^2_p=0.11); the point-biserial correlations with language comprehension and production were (r_{pb}=–0.05, p<0.001, r_{pb}=–0.04, p<0.001), respectively, and therefore we retained number of children 2;00–9;11 in the family as a covariate.

Language measures were explored using logistic regression, with country as the predictor, controlling for number of children aged 2;00 to 9;11. We used a deviation contrast to compare each country to the overall effect to investigate the general ordering of these sixteen countries on a continuum. We report Cox and Snell's and Nagelkerke's pseudo-R^2 values as estimates of the percentages of variance accounted for by country; Wald estimates for significance of the logistic regression coefficients (which correspond to significance testing of b coefficients in OLS regression); and odds ratios for effect size.

Comprehension versus production

Table 4 shows the mean proportions of children reported as comprehending and producing language. As can be seen, both were high (uniformly in the 0.90s). However, the means for comprehension and production differ (χ^2(1, N=38845)=2787.27, p<0.001). Overall, as expected, reported comprehension was higher than reported production.

Comprehension–production correlations

Correlations between comprehension and production were uniformly positive, significant in all countries (except Montenegro and Serbia), and varied from small to large. The weighted and normalized average correlation for the focus sample was positive, significant, and small to medium in size (r=0.22; standard error of the mean Z coefficient=0.005; 95% confidence interval=(0.206, 0.225)).

Child language: deviations from the overall effect

Comprehension—Controlling for number of children aged 2;00 to 9;11 in the family, the majority of countries (12 of 16) differed significantly from the overall effect of country. Country explained between 1.5% (Cox & Snell R^2) and 6.3% (Nagelkerke R^2) of the variance. Results for high- and medium-HDI countries varied in whether young children

were more or less likely to understand when their mother (or caregiver) asked them to do something (ORs=0.46–4.48). Young children in both countries with low HDI and in Iraq were less likely to understand when their mother (or caregiver) asked them to do something (ORs=0.42–0.62) (see Table 5).

Production—Controlling for number of children aged 2;00 to 9;11 in the family, the majority of countries (10 of 16) differed significantly from the overall effect of country. Country explained between 3.1% (Cox & Snell R^2) and 8.8% (Nagelkerke R^2) of the variance. Results for high- and medium-HDI countries varied in whether young children were more or less likely to speak and say recognizable words (ORs=0.54–4.59). Children in both countries with low HDI and in Iraq were less likely to speak and say recognizable words (ORs=0.26–0.44) (see Table 5).

DISCUSSION

Language is a requisite for successful functioning in society, and, although complex and dynamic in its ontogeny, it is a fundamental skill and a cornerstone of human development. Children's language has been linked to a variety of other important areas of growth from emotional maturity and social interaction to reading and academic outcomes.

Language comprehension and production are usually studied in infants or very young children, but this age range imposes limitations and both aspects of language continue to develop beyond this early stage of the life course (Berman, 2004; Wassenberg *et al.*, 2008). Our study compared comprehension and production to examine their mean absolute and relative associations. We approached the assessment of child comprehension and production of language using a common parental report measure, following Bates's (1993) admonition that different measures of comprehension and production can yield different results (in mean level, though perhaps not rank order; see Bornstein & Haynes, 1998). We also used the Multiple Indicator Cluster Survey, one of the main tools applied in developing nations in the world today to measure progress in child development (see Bornstein *et al.*, 2012). The MICS plays a role in planning and reporting on children and families, being a reliable source of data for many indicators which for developing nations are difficult to find otherwise. It thus affords unique and useful insights on child language from usually wholly neglected sources. The MICS3 is admittedly a blunt instrument not designed to assess language development in children in depth, as it trades depth for extensive breadth.

Based on the MICS3 we report about language comprehension and production in a sample of 101,250 children age 2;00 to 9;11 and a focus subsample of 38,845 children age 2;00 to 4;11 from sixteen under-researched developing nations. In the whole sample of children in these developing nations, average comprehension and production were reportedly high, but still not perfect, even by 9;11.

In both samples, many of the central findings paralleled those previously reported for English and other more commonly studied languages (Dale & Goodman, 2005). Comprehension slightly exceeded production. The correlations between comprehension and production by country were positive and significant, but varied in size, and the weighted and

normalized average correlation was positive, significant, and small to medium in size. Thus, MICS3 data show that basic language comprehension and production are moderately related at each age from 2;00 to 9;11, but share less than 10% of common variance. At the same time, language and culture moderated comprehension–production relations as the relatively wide range in their shared variances (00–28%) amply attests. (Of marginal note, in different countries that speak the same dominant language, comprehension–production correlations are comparable; see English in Belize and Jamaica and Serbian in Montenegro and Serbia.) These small but significant correlational results are in line with the extant literature on comprehension–production relations reviewed at the outset. Of course, the finding that comprehension and production are minimally associated with one another suggests, but does not prove, that these two basic language faculties spring from the same source(s).

Comprehension and production varied with child age, and each grew from 2;00 to 5;00 after which each reached an asymptote. In addition, analyses comparing countries to the overall effect of country showed a general relation between language in the 2;00 to 4;11 age range and national standard of living as measured by the Human Development Index: young children in countries with lower standards of living were less likely to demonstrate basic language comprehension or production to mother (or caregiver). In representative samples in developed nations, children from lower socioeconomic strata tend to show slower rates of development than do children from higher socioeconomic strata, as here beginning within the second year of life and continuing through the school years (Dollaghan *et al.*, 1999; Huttenlocher *et al.*, 2002; Rescorla & Alley, 2001). A variety of input factors (amount of speech addressed to children, richness of the vocabulary used, rate of question asking, and length of utterances) has been identified as positive predictors of language development (Bornstein *et al.*, 1998; Weizman & Snow, 2001). Hart and Risley (1995) reported significant associations between family socioeconomic status on the one hand and both vocabulary exposure and vocabulary growth on the other. Children from low-income families in the United States were exposed to significantly fewer words than children from high-income families and subsequently followed slower rates of vocabulary development.

Comprehension and production may be culturally common patterns in language development, but their acquisition and the degree of their association appear to be community specific and express flexible adaptations to the language setting in which the child develops. Evolutionary psychology seeks evidence for human universals, and cultural psychology seeks evidence for human diversity. These complementary approaches converge to define conditions under which culturally universal mechanisms are expressed in community-specific ways. In essence, language development is common AND specific. An evolutionary model posits a language instinct from the perspective of an inborn and universal acquisition device, but specificity of environmental input plays a demonstrable role in the acquisition of each specific language (Pinker, 2007). In these terms, language comprehension and production might be best described by the interplay of evolutionarily based, inborn biases that are culture-common with environmental input that stimulates community-specific patterns of relations and development. Cross-national linguistic research, such as the study presented here, exposes the ways by which ecological context may moderate universal tendencies in child language. More generally, it is imperative to

learn more about children in widely diverse language communities so that linguists and developmental scientists as well as educators, practitioners and policy-makers can better understand human development and advantageously promote children's language and cognitive functioning and their well-being.

REFERENCES

Arnett JJ. The neglected 95%: Why American psychology needs to become less American. American Psychologist. 2008; 63:602–614. [PubMed: 18855491]

Bates E. Comprehension and production in early language development. Monographs of the Society for Research in Child Development. 1993; 58:222–42. Comments on Savage-Rumbaugh *et al.* [PubMed: 8366873]

Bauer DJ, Goldfield BA, Reznick JS. Alternative approaches to analyzing individual differences in the rate of early vocabulary development. Applied Psycholinguistics. 2002; 23:313–35.

Behrend DA. Overextensions in early language comprehension: Evidence from a signal detection approach. Journal of Child Language. 1988; 15:63–75. [PubMed: 3350877]

Benedict H. Early lexical development: Comprehension and production. Journal of Child Language. 1979; 6:183–200. [PubMed: 468932]

Berman, RA. Language development across childhood and adolescence. John Benjamins; Philadelphia, PA: 2004.

Bornstein, MH., editor. The handbook of cultural developmental science: Vol. 1. Domains of development across cultures; Vol. 2. Development in different places on earth. Routledge; New York: 2010.

Bornstein MH, Britto PB, Nonoyama-Tarumi Y, Ota Y, Petrovic O, Putnick DL. Child development in developing countries: Introduction and methods. Child Development. 2012

Bornstein MH, Haynes OM. Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. Child Development. 1998; 69:654–71. [PubMed: 9680678]

Bornstein MH, Haynes OM, Painter K. Sources of child vocabulary competence: A multivariate model. Journal of Child Language. 1998; 25:367–93. [PubMed: 9770912]

Campbell, R.; Macdonald, TB.; Dockrell, J. The relationship between comprehension and production and its ontogenesis.. In: Lowenthal, F.; Vandamme, J.; Cordier, J., editors. Language and language acquisition. Plenum Press; New York: 1982. p. 141-50.

Clark, EV. The lexicon in acquisition. Cambridge University Press; Cambridge: 1993.

Clark EV, Hecht BF. Comprehension, production, and language acquisition. Annual Review of Psychology. 1983; 34:325–49.

Cohen, J. Statistical power analysis for the behavioral sciences. 2nd edn.. Erlbaum; Hillsdale, NJ: 1988.

Dale, P.; Goodman, J. Commonality and individual differences in vocabulary growth.. In: Tomasello, M.; Slobin, DI., editors. Beyond nature–nurture. Essays in honor of Elizabeth Bates. Lawrence Erlbaum Associates; London: 2005. p. 41-80.

Dollaghan CA, Campbell TF, Paradise JL, Feldman HM, Janosky JE, Pitcairn DN. Maternal education and measures of early speech and language. Journal of Speech, Language, and Hearing Research. 1999; 42:1432–43.

Fenson L, Dale P, Reznick JS, Bates E, Thal D, Pethick S. Variability in early communicative development. Monographs of the Society for Research in Child Development. 1994; 59(5, Serial No. 242)

Fenson, L.; Marchman, VA.; Thal, D.; Dale, P.; Reznick, S.; Bates, E. The MacArthur Communicative Development Inventories: User's guide and technical manual. 2nd edn.. Paul. H. Brookes Publishing Co.; Baltimore, MD: 2007.

Harris M, Yeeles C, Chasin J, Oakley Y. Symmetries and asymmetries in early lexical comprehension and production. Journal of Child Language. 1995; 22:1–18. [PubMed: 7759573]

Hart, B.; Risley, TR. Meaningful differences in the everyday experience of young American children. Paul H. Brookes Publishing; Baltimore, MD: 1995.

Hedges, LV.; Olkin, I. Statistical methods for meta-analysis. Academic Press; San Diego, CA: 1985.

Hoff E. How social contexts support and shape language. Developmental Review. 2006; 26:55–88.

Huttenlocher J, Vasilyeva M, Cymerman E, Levine S. Language input and child syntax. Cognitive Psychology. 2002; 45:337–74. [PubMed: 12480478]

Kern S. Lexicon development in French-speaking infants. First Language. 2007; 27:227–50.

Lyytinen P, Poikkeus A-M, Laakso M-L, Eklund K, Lyytinen H. Language development and symbolic play in children with and without familial risk for dyslexia. Journal of Speech, Language, and Hearing Research. 2001; 44:873–885.

Mankiw NG, Romer G, Weil DN. A contribution to the empirics of economic growth. The Quarterly Journal of Economics. 1992; 107:407–437.

Marsh, LC.; Cormier, DR. Spline regression models. Sage Publications; Thousand Oaks, CA: 2001.

National Center for Children in Poverty. Poverty and brain development in early childhood. National Center for Children in Poverty; New York: 1999.

Pinker, S. The language instinct. Harper; New York: 2007.

Rescorla L, Alley A. Validation of the Language Development Survey (LDS). Journal of Speech, Language, and Hearing Research. 2001; 44:434–45.

Reznick JS, Goldfield BA. Rapid change in lexical development in comprehension and production. Developmental Psychology. 1992; 28:406–413.

Tamis-LeMonda CS, Bornstein MH, Kahana-Kalman R, Baumwell L, Cyphers L. Predicting variation in timing of language milestones in the second year: An events history approach. Journal of Child Language. 1998; 25:675–700. [PubMed: 10095330]

Tan SH, Schafer G. Toddlers' novel word learning: Effects of phonological representation, vocabulary size and parents' ostensive behaviour. First Language. 2005; 25:131–55.

Tomblin JB, Hardy JC, Hein HA. Predicting poor-communication status in preschool children using risk factors present at birth. Journal of Speech and Hearing Research. 1991; 34:1096–1105. [PubMed: 1749241]

UNICEF Division of Policy and Planning. Multiple indicator cluster survey manual 2005: Monitoring the situation of children and women. UNICEF; New York: 2006.

United Nations Development Programme. [10 July 2009] Human Development Reports. 2007. from: http://hdr.undp.org/en/statistics/hdi/

Wassenberg R, Hurks PPM, Hendriksen JGM, Feron FJM, Meijs CJC, Vles JSH, Jolles J. Age-related improvement in complex language comprehension: Results of a cross-sectional study with 361 children aged 5 to 15. Journal of Clinical and Experimental Neuropsychology. 2008; 30:435–48. [PubMed: 18938680]

Weizman ZO, Snow CE. Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. Developmental Psychology. 2001; 37:265–79. [PubMed: 11269394]

Zink I, Lejaegere M. N-CDIs. Lijsten voor Communicatieve Ontwikkeling. 2002Acco.Leuven Aanpassing en hernormering van de MacArthur CDIs van Fenson *et al*.

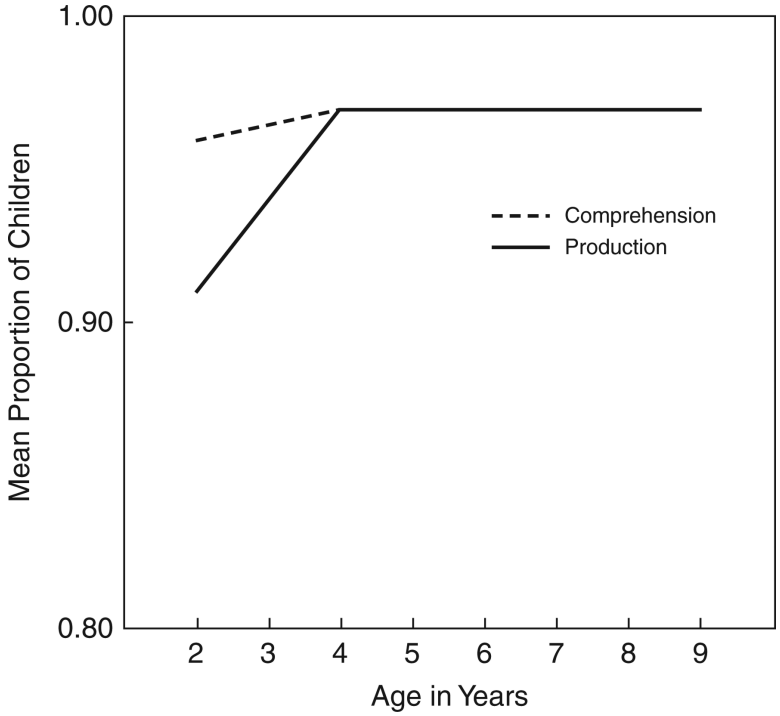

Fig. 1.

Mean proportions of children in sixteen developing countries who comprehend and produce basic language by age.

TABLE 1

Sociodemographic statistics, proportions of children comprehending and producing, and correlations between comprehension and production by country by HDI for whole sample (2 ; 00–9 ; 11)

		Child age		Child gender		Comprehension		Production		Comprehension–production correlation
Country	***n***	***M***	***SD***	**% Female**	**Predominant spoken language(s)**	***M***	***SD***	***M***	***SD***	r_{Φ}
High HDI										
Montenegro	940	5;1	2;3	48.5	Serbian	0.94	0.25	0.97	0.26	0.06 *ns*
Serbia	3065	4;10	2;3	48.1	Serbian	0.95	0.22	0.93	0.24	0.06
Macedonia	3266	4;6	1;11	50.7	Macedonian	0.97	0.18	0.96	0.20	0.34
Albania	1636	5;11	2;2	45.1	Albanian	0.96	0.21	0.94	0.24	0.52
Bosnia and Herzegovina	2526	4;7	2;2	50.6	Bosnian, Serbian	0.99	0.11	0.98	0.14	0.25
Medium HDI										
Thailand	12911	5;9	2;3	48.3	Thai	0.99	0.12	0.95	0.21	0.18
Belize	830	5;8	2;2	52.0	Spanish, Kriol, English	0.95	0.22	0.95	0.21	0.24
Jamaica	1630	5;7	2;3	48.1	English	0.96	0.20	0.97	0.18	0.34
Mongolia	3478	5;2	2;3	47.6	Mongolian	0.98	0.14	0.96	0.20	0.26
Uzbekistan	4933	5;7	2;3	47.6	Uzbek	1.00	0.07	0.99	0.08	0.53
Yemen	2512	5;6	2;3	48.8	Arabic	0.93	0.25	0.93	0.25	0.34
Ghana	3240	5;7	2;2	49.1	Ashante, Ewe, Fanti	0.96	0.19	0.97	0.16	0.23
Bangladesh	36987	5;6	2;3	48.6	Bengali	0.98	0.13	0.98	0.13	0.24
Low HDI										
Central African Republic	6825	5;3	2;2	49.7	Sango, French	0.94	0.23	0.92	0.28	0.27
Sierra Leone	5308	5;5	2;2	49.3	Kiro	0.95	0.22	0.90	0.30	0.31
HDI n/a										
Iraq	11 163	5;4	2;3	47.9	Arabic	0.96	0.20	0.94	0.24	0.30
Total	101 250	5;5	2;3	48.6		0.97	0.16	0.96	0.20	0.21[a]

NOTES : Unless otherwise noted, all phi correlations are $p<0.001$.

[a]Weighted and normalized average correlation.

TABLE 2

Proportions of children comprehending and producing and correlations between comprehension and production by age by country by HDI

Country		2;00–2;11	3;00–3;11	4;00–4;11	5;00–5;11	6;00–6;11	7;00–7;11	8;00–8;11	9;00–9;11
High HDI									
Montenegro	Comprehension	0.97 (0.16)	0.92 (0.27)	0.95 (0.22)	0.91 (0.29)	0.94 (0.24)	0.93 (0.25)	0.92 (0.27)	0.92 (0.27)
	Production	0.08 (0.13)	0.97 (0.16)	0.96 (0.19)	0.97 (0.17)	0.99 (0.10)	0.98 (0.15)	0.97 (0.17)	0.96 (0.21)
	r_Φ	– 0.02 n.s.	–0.05 n.s.	0.11 n.s.	0.11 n.s.	–0.03 n.s.	–0.4 n.s.	–0.05 n.s.	0.34***
Serbia	Comprehension	0.94 (0.23)	0.96 (0.21)	0.96 (0.21)	0.95 (0.22)	0.94 (0.24)	0.94 (0.24)	0.95 (0.21)	0.95 (0.22)
	Production	0.95 (0.22)	0.94 (0.25)	0.96 (0.19)	0.92 (0.27)	0.92 (0.27)	0.93 (0.25)	0.90 (0.30)	0.91 (0.28)
	r_Φ	–0.02 n.s.	0.02 n.s.	0.10*	0.13*	0.06 n.s.	–0.01 n.s.	0.11 n.s.	0.12 n.s.
Macedonia	Comprehension	0.97 (0.17)	0.96 (0.19)	0.97 (0.17)	0.97 (0.17)	0.97 (0.17)	0.95 (0.22)	0.95 (0.22)	0.97 (0.17)
	Production	0.91 (0.28)	0.96 (0.20)	0.97 (0.16)	0.97 (0.16)	0.97 (0.17)	0.96 (0.19)	0.98 (0.16)	0.97 (0.17)
	r_Φ	0.44***	0.28***	0.31***	0.49***	0.27***	0.66***	0.26***	0.38***
Albania	Comprehension	0.96 (0.19)	0.97 (0.17)	0.98 (0.13)	0.96 (0.20)	0.98 (0.16)	0.96 (0.19)	0.99 (0.11)	0.86 (0.34)
	Production	0.96 (0.21)	1.0(0.00)	0.95 (0.22)	0.95 (0.23)	0.93 (0.26)	0.95 (0.23)	0.97 (0.18)	0.85 (0.35)
	r_Φ	0.44***	—[a]	0.37***	0.56***	0.32***	0.17*	0.18**	0.80***
Bosnia & Herzegovina	Comprehension	0.99 (0.10)	0.99 (0.08)	0.98 (0.10)	0.97 (0.14)	0.98 (0.13)	0.98 (0.14)	1.0 (0.07)	0.98 (0.14)
	Production	0.98 (0.15)	0.98 (0.13)	0.98 (0.12)	0.97 (0.16)	0.99 (0.12)	0.98 (0.14)	0.99 (0.10)	0.97 (0.16)
	r_Φ	0.12***	–0.01 n.s.	0.47***	0.35***	0.28***	0.49***	–0.01 n.s.	0.21**
Medium HDI									
Thailand	Comprehension	0.98 (0.13)	0.99 (0.12)	0.99 (0.11)	0.98 (0.15)	0.99 (0.10)	0.99 (0.11)	0.98 (0.12)	0.99 (0.12)
	Production	0.92 (0.27)	0.95 (0.21)	0.97 (0.18)	0.95 (0.20)	0.97 (0.19)	0.96 (0.19)	0.95 (0.22)	0.96 (0.20)
	r_Φ	0.16***	0.16***	0.24***	0.27***	0.13***	0.18***	0.14***	0.16***
Belize	Comprehension	0.91 (0.29)	0.93 (0.25)	0.99 (0.11)	0.94 (0.25)	0.95 (0.21)	0.97 (0.18)	0.92 (0.27)	0.99 (0.10)
	Production	0.85 (0.36)	0.97 (0.18)	0.99 (0.11)	0.95 (0.23)	0.94 (0.24)	0.97 (0.18)	0.98 (0.13)	0.99 (0.10)
	r_Φ	0.20 n.s.	– 0.05 n.s.	1.0***	0.27**	0.40***	0.22*	0.21*	–0.01 n.s.
Jamaica	Comprehension	0.95 (0.22)	0.95 (0.22)	0.97 (0.16)	0.98 (0.15)	0.94 (0.24)	0.96 (0.21)	0.96 (0.20)	0.96 (0.19)
	Production	0.96 (0.21)	0.97 (0.16)	0.98 (0.14)	0.97 (0.18)	0.94 (0.23)	0.97 (0.18)	0.98 (0.13)	0.96 (0.20)
	r_Φ	0.20**	0.40***	0.43***	0.32***	0.40***	0.48***	0.30***	0.22***

Country		2;00–2;11	3;00–3;11	4;00–4;11	5;00–5;11	6;00–6;11	7;00–7;11	8;00–8;11	9;00–9;11
Mongolia	Comprehension	0.97 (0.17)	0.98 (0.15)	0.99 (0.11)	0.99 (0.11)	0.99 (0.11)	0.97 (0.16)	0.97 (0.17)	0.98 (0.16)
	Production	0.87 (0.34)	0.96 (0.20)	0.98 (0.15)	0.98 (0.15)	0.98 (0.12)	0.98 (0.16)	0.97 (0.16)	0.98 (0.13)
	r_{Φ}	0.10*	0.23***	0.23***	0.34***	0.36***	0.62***	0.46***	0.26***
Uzbekistan	Comprehension	0.99 (0.10)	0.99 (0.10)	1.0 (0.04)	1.0 (0.04)	1.0 (0.06)	0.99 (0.09)	1.0 (0.04)	1.0 (0.05)
	Production	0.98 (0.14)	0.99 (0.11)	1.0 (0.04)	1.0 (0.00)	1.0 (0.06)	1.0 (0.07)	0.99 (0.09)	0.99 (0.08)
	r_{Φ}	0.49***	0.77***	– 0.00 n.s.	— *a*	0.50***	0.51***	0.45***	0.71***
Yemen	Comprehension	0.92 (0.28)	0.93 (0.26)	0.94 (0.24)	0.94 (0.25)	0.94 (0.23)	0.93 (0.25)	0.92 (0.27)	0.95 (0.21)
	Production	0.89 (0.31)	0.91 (0.29)	0.94 (0.24)	0.96 (0.19)	0.92 (0.27)	0.96 (0.20)	0.95 (0.22)	0.94 (0.24)
	r_{Φ}	0.24***	0.38***	0.28***	0.35***	0.56***	0.30***	0.22***	0.44***
Ghana	Comprehension	0.95 (0.22)	0.97 (0.18)	0.97 (0.16)	0.95 (0.21)	0.97 (0.17)	0.97 (0.16)	0.98 (0.15)	0.94 (0.23)
	Production	0.90 (0.30)	0.99 (0.10)	0.97 (0.16)	0.98 (0.14)	0.99 (0.11)	0.99 (0.10)	0.98 (0.12)	0.98 (0.15)
	r_{Φ}	0.38***	0.27***	0.09 n.s.	0.20***	0.09 n.s.	– 0.02 n.s.	0.11*	0.41***
Bangladesh	Comprehension	0.98 (0.14)	0.98 (0.13)	0.99 (0.12)	0.98 (0.12)	0.99 (0.12)	0.98 (0.12)	0.98 (0.13)	0.98 (0.13)
	Production	0.97 (0.17)	0.98 (0.14)	0.98 (0.14)	0.98 (0.14)	0.99 (0.11)	0.99 (0.12)	0.99 (0.12)	0.98 (0.12)
	r_{Φ}	0.24***	0.18***	0.25***	0.23***	0.14***	0.31***	0.20***	0.33***
Low HDI									
Central African Republic	Comprehension	0.90 (0.30)	0.93 (0.26)	0.95 (0.22)	0.96 (0.20)	0.95 (0.23)	0.96 (0.20)	0.97 (0.18)	0.95 (0.22)
	Production	0.77 (0.42)	0.89 (0.32)	0.95 (0.22)	0.94 (0.25)	0.95 (0.21)	0.96 (0.21)	0.97 (0.18)	0.95 (0.22)
	r_{Φ}	0.37***	0.19***	0.29***	0.15***	0.24***	0.27***	0.27***	0.09*
Sierra Leone	Comprehension	0.88 (0.32)	0.94 (0.23)	0.95 (0.22)	0.96 (0.18)	0.96 (0.19)	0.97 (0.18)	0.95 (0.21)	0.97 (0.17)
	Production	0.76 (0.43)	0.85 (0.36)	0.90 (0.30)	0.91 (0.29)	0.93 (0.26)	0.94 (0.24)	0.94 (0.23)	0.97 (0.17)
	r_{Φ}	0.39***	0.22***	0.27***	0.19***	0.32***	0.26***	0.43***	0.25***
HDI N/A									
Iraq	Comprehension	0.93 (0.26)	0.96 (0.19)	0.95 (0.22)	0.96 (0.19)	0.97 (0.17)	0.96 (0.19)	0.97 (0.17)	0.96 (0.18)
	Production	0.84 (0.37)	0.93 (0.26)	0.93 (0.25)	0.96 (0.20)	0.96 (0.19)	0.96 (0.20)	0.97 (0.18)	0.96 (0.19)
	r_{Φ}	0.29***	0.20***	0.37***	0.31***	0.20***	0.37***	0.36***	0.24***
Total	Comprehension	0.96 (0.20)	0.97 (0.17)	0.98 (0.15)	0.97 (0.16)	0.98 (0.15)	0.98 (0.15)	0.98 (0.15)	0.97 (0.16)
	Production	0.91 (0.28)	0.95 (0.21)	0.97 (0.18)	0.96 (0.19)	0.97 (0.17)	0.97 (0.16)	0.97 (0.17)	0.97 (0.17)
	r_{Φ}	0.30***	0.21***	0.28***	0.25***	0.23***	0.29***	0.24***	0.31***

Country	2;00–2;11	3;00–3;11	4;00–4;11	5;00–5;11	6;00–6;11	7;00–7;11	8;00–8;11	9;00–9;11
Total *N*	12 743	13 404	12 698	12 724	12 360	12 876	13 210	11 191

NOTES : *M* (*SD*) Correlations reported are phi.

[a] No statistics are computed because at least one variable is a constant.

* $p<0.05$

** $p<0.01$

*** $p<0.001$

TABLE 3

Logistic spline regressions for language comprehension and production with child age

		Estimation of knot age			Estimation of 1st slope 2;00–4;11		Estimation of 2nd slope 5;00–9;11	
	N	**Knot**	**95% CI**	***Wald***	***Wald***	***Odds Ratio***	***Wald***	***Odds Ratio***
Comprehension	101 231	3;7	3;0–3;1	48.13***	326.92***	1.69	7.50**	1.04
Production	101 221	3;8	3;3–3;9	204.02***	58.93***	1.32	0.03 n.s.	0.99

NOTES:

** $p < 0.01$

*** $p < 0.001$.

TABLE 4

Sociodemographic statistics, proportions of children comprehending and producing, and correlations between comprehension and production by country by HDI for the focus sample (2;00–4;11)

		Child age		Child gender	Number of children 2;00–9;11	Comprehension		Production		Comprehension–production correlation
Country	***n***	***M***	***SD***	**% female**	***M* (*SD*)**	***M***	***SD***	***M***	***SD***	r_{Φ}
High HDI										
Montenegro	424	3;1	0;9	47.6	1.63 (0.75)	0.95	0.22	0.97	0.17	0.02 *ns*
Serbia	1569	2;11	0;9	48.8	1.50 (0.73)	0.95	0.21	0.95	0.22	0.03 *ns*
Macedonia	1810	3;1	0;9	49.5	1.81 (0.90)	0.97	0.17	0.95	0.22	0.28
Albania	477	3;0	0;9	46.5	1.44 (0.61)	0.97	0.16	0.97	0.18	0.33
Bosnia and Herzegovina	1476	3;0	0;9	50.4	1.42 (0.62)	0.99	0.09	0.98	0.13	0.20
Medium HDI										
Thailand	4338	3;0	0;9	47.9	1.28 (0.55)	0.99	0.12	0.95	0.23	0.18
Belize	263	3;0	0;9	50.6	1.84 (1.00)	0.94	0.23	0.94	0.25	0.20*
Jamaica	569	3;0	0;9	48.0	1.52 (0.78)	0.96	0.20	0.97	0.17	0.32
Mongolia	1500	2;11	0;9	48.5	1.41 (0.64)	0.98	0.15	0.93	0.25	0.16
Uzbekistan	1752	3;0	0;9	48.6	1.63 (0.78)	0.99	0.08	0.99	0.10	0.57
Yemen	932	2;11	0;9	49.0	2.46 (1.37)	0.93	0.26	0.91	0.28	0.30
Ghana	1127	2;11	0;9	48.7	1.83 (1.00)	0.96	0.19	0.96	0.20	0.28
Bangladesh	13 740	3;0	0;9	48.7	1.56 (0.75)	0.98	0.13	0.98	0.15	0.23
Low HDI										
Central African Republic	2768	2;11	0;9	49.2	2.05 (1.15)	0.92	0.27	0.87	0.34	0.30
Sierra Leone	1896	3;0	0;9	50.3	2.05 (1.14)	0.93	0.26	0.84	0.37	0.32
HDI n/a										
Iraq	4204	2;11	0;9	48.6	2.31 (1.20)	0.95	0.23	0.90	0.30	0.29
Total	38845	3;0	0;9	48.8	1.70 (0.93)	0.97	0.18	0.94	0.23	0.22[a]

NOTES : Except when noted otherwise, for all correlations $p<0.001$.

*$p<0.05$.

[a]Weighted and normalized average correlation.

TABLE 5

Odds ratios (*ORs*) for country deviation from overall effect for language comprehension and production by country by HDI for focus sample (2;00–4;11)

Country	*n*	Odds ratio comprehension	Odds ratio production
High HDI			
Montenegro	424	0.61*	1.72*
Serbia	1571	0.65***	0.93 n.s.
Macedonia	1810	1.03 n.s.	0.97 n.s.
Albania	477	1.17 n.s.	1.45 n.s.
Bosnia and Herzegovina	1476	3.64***	2.69***
Medium HDI			
Thailand	4338	2.27***	0.88 n.s.
Belize	263	0.56*	0.73 n.s.
Jamaica	570	0.75 n.s.	1.63*
Mongolia	1500	1.45*	0.69***
Uzbekistan	1752	4.48***	4.59***
Yemen	934	0.46***	0.54***
Ghana	1127	0.90 n.s.	1.11 n.s.
Bangladesh	13 740	1.98***	2.16***
Low HDI			
Central African Republic	2774	0.42***	0.32***
Sierra Leone	1896	0.44***	0.26***
HDI N/A			
Iraq	4201	0.62***	0.44***

NOTES:

* $p < 0.05$

*** $p < 0.001$.