# Assessing child development scores among minority and Indigenous language versus dominant language speakers: a cross-sectional analysis of national Multiple Indicator Cluster Surveys

*Ann C Miller, David Flood, Scott Tschida, Katherine Douglas, Peter Rohloff*


## Summary

**Background** Multiple studies have highlighted the inequities minority and Indigenous children face when accessing health care. Health and wellbeing are positively impacted when Indigenous children are educated and receive care in their maternal language. However, less is known about the association between minority or Indigenous language use and child development risks and outcomes. In this study, we provide global estimates of development risks and assess the associations between minority or Indigenous language status and early child development using the ten-item Early Child Development Index (ECDI), a tool widely used for global population assessments in children aged 3–4 years.

**Methods** We did a secondary analysis of cross-sectional data from 65 UNICEF Multiple Indicator Cluster Surveys (MICS) containing the ECDI from 2009–19 (waves 4–6). We included individual-level data for children aged 2–4 years (23–60 months) from datasets with ECDI modules, for surveys that captured the language of the respondent, interview, or head of household. The Expanded Graded Intergenerational Disruption Scale was used to classify household languages as dominant versus minority or Indigenous at the country level. Our primary outcome was on-track overall development, defined per UNICEF's guidelines as development being on track for at least three of the four ECDI domains (literacy–numeracy, learning, physical, and socioemotional). We performed logistic regression of pooled, weighted ECDI scores, aggregated by language status and adjusting for the covariables of child sex, child nutritional status (stunting), household wealth, maternal education, developmental support by an adult caregiver, and country-level early child education proportion. Regression analyses were done for all children aged 3–4 years with ECDI results, and separately for children with functional disabilities and ECDI results.

**Findings** 65 MICS datasets were included. 186 393 children aged 3–4 years had ECDI and language data, corresponding to an estimated represented population of 34 714 992 individuals. Estimated prevalence of on-track overall development as measured by ECDI scores was 65·7% (95% CI 64·2–67·2) for children from a minority or Indigenous language-speaking household, and 76·6% (75·7–77·4) for those from a dominant language-speaking household. After adjustment, dominant language status was associated with increased odds of on-track overall development (adjusted OR 1·54, 95% CI 1·40–1·71), which appeared to be largely driven by significantly increased odds of on-track development in the literacy–numeracy and socioemotional domains. For the represented population aged 2–4 years (n=11 465 601), the estimated prevalence of family-reported functional disability was 3·6% (95% CI 3·0–4·4). For the represented population aged 3–4 years with a functional disability (n=292 691), language status was not associated with on-track overall development (adjusted OR 1·02, 95% CI 0·43–2·45).

**Interpretation** In a global dataset, children speaking a minority or Indigenous language were less likely to have on-track ECDI scores than those speaking a dominant language. Given the strong positive benefits of speaking an Indigenous language on the health and development of Indigenous children, this disparity is likely to reflect the sociolinguistic marginalisation faced by speakers of minority or Indigenous languages as well as differences in the performance of ECDI in these languages. Global efforts should consider performance of measures and monitor developmental data disaggregated by language status to stimulate efforts to address this disparity.


**Funding** None.



*Lancet Glob Health* 2023

Published **Online** November 10, 2023 https://doi.org/10.1016/S2214-109X(23)00456-4

See **Online/Comment** https://doi.org/10.1016/S2214-109X(23)00511-9

For the Spanish translation of the abstract see **Online** for appendix 1

For the Kaqchikel translation of the abstract see **Online** for appendix 2

For the K'iche' translation of the abstract see **Online** for appendix 3

**Department of Global Health and Social Medicine, Blavatnik Institute, Harvard Medical School, Boston, MA, USA** (A C Miller PhD)**; Center for Indigenous Health Research, Wuqu' Kawoq – Maya Health Alliance, Tecpán, Guatemala** (D Flood MD, S Tschida MPhil, P Rohloff MD)**; Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA** (D Flood)**; Boston Medical Center, Boston, MA, USA** (K Douglas MD)**; Department of Global Health Equity, Brigham and Women's Hospital, Boston MA, USA** (P Rohloff)

Correspondence to: Dr Ann C Miller, Department of Global Health and Social Medicine, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA **ann_miller@hms.harvard.edu**

## Introduction

More than half of languages spoken globally are Indigenous languages.[1] Incontrovertible evidence has shown that speaking one's language, and educational systems that incorporate Indigenous wisdom and heritage, have positive benefits for Indigenous children's

## Research in context

### Evidence before this study

Language status is rarely analysed separately from ethnicity in large-scale surveys, and little is known specifically about the association between minority or Indigenous language use and child development outcomes, as a metric for the magnitude of inequity in early life. We did a systematic search of the PubMed and Web of Science databases for articles available from Jan 1, 2020, to Jan 14, 2023, using the terms "Indigenous language" or "minority language" or "Indigenous" and "child development". Numerous studies have highlighted inequities in child development outcomes for Indigenous children, although these studies did not focus on language. Other studies have highlighted the importance for equity of cultural adaptation of child development assessment tools and approaches to evaluating child development in minority language or multilingual contexts. These include multiple studies showing how dominant-language measures of development underestimate the language capabilities of Indigenous children, thereby contributing to stigmatisation. Several studies have demonstrated how minority or Indigenous language use in health and education settings directly improves health-care and educational quality, or how maternal language-centric programmes for minority or Indigenous children improve such outcomes. We identified one study that used large national datasets from Bolivia, Guatemala, Mexico, and Peru to examine the association between Indigenous language use and maternal health-care outcomes, and another study that used Australian census data to investigate the association between Indigenous language and child development outcomes.

### Added value of this study

In this study, we used a systematic strategy to extract household language data from 65 Multiple Indicator Cluster Survey study datasets across multiple countries and years. We classified household languages as minority or Indigenous versus dominant using a novel approach applying Expanded Graded Intergenerational Disruption Scale codes from the Ethnologue catalogue of human languages. To estimate the effect of discrimination against minority and Indigenous languages on child development, we disaggregated Early Childhood Development Index (ECDI) scores in more than 186 000 children from 40 countries by minority or Indigenous language status. We provide weighted population estimates, controlling for household wealth, nutritional status and sex of the child, maternal education, level of developmental support provided by adults, and country-level proportion of children attending early child education. We also provide estimates stratified by World Bank income group, WHO geographical region, and proportion of minority or Indigenous language-speaking households in each country. Additionally, we provide development estimates separately for children with family-identified functional disabilities in vision, hearing, or mobility.

### Implications of all the available evidence

In this analysis, being from a minority or Indigenous language-speaking household versus a dominant language-speaking household was associated with lower odds of on-track development scores when controlling for household wealth, maternal education, child's sex and nutritional status, adult support for development, and country-level early child education. Given the well known health and educational benefits of Indigenous language use for Indigenous children, our findings provide quantitative estimates from population-representative datasets of the impact of discrimination against and inequities among children speaking minority or Indigenous languages. Disaggregating population-level child development estimates by language status, in addition to other social determinants, should be considered in future monitoring efforts to track this inequity and stimulate advocacy to improve opportunities for children who speak these languages.

learning and health.[2–4] However, multiple studies have documented health and early learning inequities among Indigenous children globally.[5–7] Possible reasons for this disparity include cultural barriers, insufficient inclusion in national policy agendas, and explicit or implicit discriminatory policies, including structural institutional legacies of colonialism in some settings.[8,9]

Although these inequities for Indigenous children are well documented, an important yet understudied factor is the languages children speak. A child's primary language is of central importance to positive wellbeing and educational outcomes.[2] A child's primary language (the language they first learn from caregivers) is of central importance, both in terms of measures chosen to assess child development and also as a determinant of the availability of opportunities for early learning and education, given that many services are only offered in dominant, government-endorsed languages.[10] Of the more than 6000 living human languages worldwide, more than 50% are highly endangered from encroachment largely by official languages promoted by governments for use in educational systems.[1,11] Paucity of adequate early learning services in Indigenous languages is a human rights concern. The right of Indigenous children to receive learning opportunities in their own language is set out in the UN Convention on the Rights of the Child and Declaration on the Rights of Indigenous Peoples.[12]

Although numerous studies discuss child development in Indigenous and multilingual contexts, few quantitatively estimate the mediating status of these languages as a fundamental determinant on child development. In this study, we provide descriptive estimates of global child development scores disaggregated by minority or

Indigenous language status, adjusting for wealth and other important covariates. We use the Early Childhood Development Index (ECDI), a ten-item scale that assesses global development in children aged 3–4 years, which has been extensively used in the Multiple Indicator Cluster Surveys (MICS).[13] The ECDI shows heterogeneity across different sociocultural contexts, and is being replaced by an expanded measure, the ECDI2030, as of 2022. Nevertheless, for the foreseeable future, few ECDI2030 datasets will be available, and so the original ECDI remains a valuable comparative cross-country measure.

## Methods

### Study design and data sources

We conducted a secondary, cross-sectional analysis of ECDI data from the MICS. The UNICEF-sponsored MICS programme includes cross-sectional household surveys of maternal and child health indicators, which have been conducted in more than 115 participating countries to date. Sampling and data collection methods are standardised across countries and over time. The sampling methods involve a multistage cluster design in which primary sampling units (PSUs) are chosen according to probability-proportional-to-size sampling that is based on an already available sampling frame for each country, such as a census. Once selected, all households within the PSUs are enumerated. Households within the PSUs are randomly selected to participate. Starting with the fourth round of data collection (MICS-4), the ECDI child development measure for children aged 3–4 years was included. We used publicly available anonymised datasets from the MICS website from three survey rounds that included the ECDI: MICS-4 (2009–12), MICS-5 (2013–17), and MICS-6 (2017–19).

The GATHER checklist was used in preparing this manuscript (appendix 4 pp 2–3). Ethics approval was not sought as the study involved anonymised, publicly available data that could not be linked to a specific individual.

### Procedures

All MICS survey datasets from waves 4, 5, and 6 available for download during the month of September, 2019, were evaluated. We included datasets in our analysis if they had ECDI modules and had captured the language of the respondent, interview, or the head of household (in order of priority). Surveys with none of the aforementioned language variables were excluded.

We merged child and household datasets for each country, to ensure all important variables were captured. We limited inclusion of individual-level data for children with a recorded age between 23 and 60 months and any recorded height-for-age z-score. We included variables capturing information on ECDI scores, parent involvement, height-for-age z-score, maternal education, age of child, sex of child, relative wealth score and quintile of the household, language variables, household stratum, cluster, PSU, and survey sampling weights. Sex of child was captured on the MICS forms by the interviewer at the in-person interview, and cannot be left unanswered. Household strata were the subgroups used within MICS sampling frames (eg, urban *vs* rural). When available, we also included variables on family-identified functional disability. Of note, the ECDI was designed for children aged 3–4 years, whereas questions on disability were asked about all children aged 2 years and older; thus, in our association analyses, we selected an age range of 2–4 years (23–60 months) to capture both populations, allowing for one month either side given that birthdates are sometimes imprecise or imputed in the MICS. The MICS surveys collect three items related to functional disability assessments: vision (child needs glasses or has caregiver-reported moderate to severe difficulty seeing); hearing (child needs a hearing aid or has moderate to severe difficulty hearing); and mobility (child uses mobility assistive devices or has moderate to severe difficulty with mobility). We defined any disability if the child screened positive in any of these three areas and we defined separate variables specific for visual, auditory, and mobility disabilities. Other methodological details are tabulated by country and MICS round in appendix 4 (pp 5, 7–34).

For more on the **ECDI2030** see https://data.unicef.org/resources/early-childhood-development-index-2030-ecdi2030/

We used a standardised workflow to establish the primary language of households (appendix 4 p 37). Given a wide range of non-standard spelling in language fields, we manually compared all primary household language data to country-specific language monographs in Ethnologue (24th edition, released Feb 22, 2021), a comprehensive language reference tool. For each primary household language, the standard spelling used in Ethnologue and International Organization for Standardization (ISO) 639-3 codes[14] were extracted.

For **Ethnologue** see https://www.sil.org/about/endangered-languages/languages-of-the-world

For the **MICS website** see https://mics.unicef.org/surveys

See **Online** for appendix 4

Finally, we established whether a household language was likely to be a dominant language or a minority or Indigenous language. For each language we extracted the corresponding Expanded Graded Intergenerational Disruption Scale (EGIDS) code[15] from Ethnologue. The EGIDS is an ordinal scale that measures a language's official political support and endangerment status, ranging from 0 ("the language is widely used between nations in trade, knowledge exchange, and international policy") to 10 (extinct—"the language is no longer used and no one retains a sense of ethnic identity associated with the language"; appendix 4 p 38). All languages were assigned country-specific EGIDS codes, allowing for language status to vary at the country level for languages spoken in multiple countries. We defined dominant language as any language with evidence of widespread political support or use in formal educational settings and mass media (EGIDS categories 0–4); languages without such support (EGDIS categories 5–10) were considered minority or Indigenous languages. The main analyses exclude Pakistan, which represented an important language outlier, given that although 39% of the population in Pakistan were reported to speak Punjabi in the 2017 census,[16] it is not an official language of Pakistan and therefore has an EGIDS of a minority language.

For the **UN World Population Prospects database** see https://population.un.org/wpp/DataQuery/

## Primary outcome

The ten-item ECDI is a caregiver-reported measure of on-track development in physical, learning, socioemotional, and literacy–numeracy domains. Data are collected from each child's primary caregiver during the MICS face-to-face interviews in participants' homes. Interviewers in all locations are trained in a standard fashion, which includes a focus on interview privacy and skills to reduce bias and increase completion. Interviewers are also expected to be of the same gender as interviewees. Our primary outcome variable was on-track development, as defined per UNICEF's ECDI guidelines;[17] overall development is on track if at least three of the four ECDI domains (literacy–numeracy, learning, physical, and socioemotional) are on track. The complete ECDI instrument, including definitions of on-track development for each domain, is in appendix 4 (p 35).

For the **MICS instructions for interviewers** see https://mics.unicef.org/tools?round=mics6#data-collection

## Statistical analysis

We created pooled, weighted survey data using sampling weights from each survey. Samples were weighted for country, survey wave (MICS-4, MICS-5, or MICS-6), and household stratum. Stratum weights were calculated for each survey by the country's institute of statistics, and were extracted from each survey. A country–survey weighting factor was created by taking the population estimates for each country at the time of each included survey from the UN World Population Prospects database (2022 revision), summing them and dividing the population of children younger than 5 years in that country at the time of the survey by the total under-5 population of all of the surveys in the present study. Each child's individual sampling weight, calculated by each country's institute of statistics and extracted for this study, was multiplied by the inverse of the country–survey factor. We used Stata's SVY commands to apply these weights and calculated estimated means, prevalences (reported as percentages), and unadjusted and adjusted odds ratios (ORs) with 95% CIs. Stata's package of SVY commands fit models for complex survey data, incorporating the effects of weighting, clustering, and stratification.[18,19] Using Stata's SVY tools, we estimated weighted characteristics of the base population (children aged <5 years) from all included surveys, and separately for subpopulations with ECDI results. We provided weighted estimates of ECDI and disability on subpopulations with these data available. We used the Stata user-defined command ADJRR to estimate unadjusted and adjusted risk differences.[18] The ADJRR command calculates risk differences after running a logit or probit model, controlling for complex survey designs. The risk difference was averaged over the entire dataset and expressed in percentage points with 95% CIs. We performed logistic regression of pooled data to assess the impact of dominant language status on children's ECDI scores for the subpopulation of children with these data available, adjusting for household wealth, child sex, child nutritional status (stunting), maternal education, developmental support by an adult caregiver, and country-level early child education. We conceptualised criteria for model adjustment using a directed acyclic graph (DAG) to identify potential confounders, and to elucidate plausible direct and indirect pathways from minority or Indigenous language status to child development (appendix 4 p 36). We did not adjust for age as the ECDI is collected only among children aged 3–4 years. Stunting was defined as a height-for-age z-score lower than –2 SD when compared with the WHO child growth reference standards.[20] Wealth quintiles were taken from each MICS dataset, and represent relative wealth at the country level compared within the same survey. Wealth quintiles are calculated by each country's institute of statistics for each survey on the basis of principal component analyses of household assets, construction materials, and water and sanitation variables; specific components vary from survey to survey.[21] We defined maternal educational status as a three-level categorical variable (no formal education, primary education, or secondary education or higher). We defined country-level early child education as the proportion of children aged 3–4 years in each country and wave of data collection who attended an early childhood education programme, and categorised into quartiles.

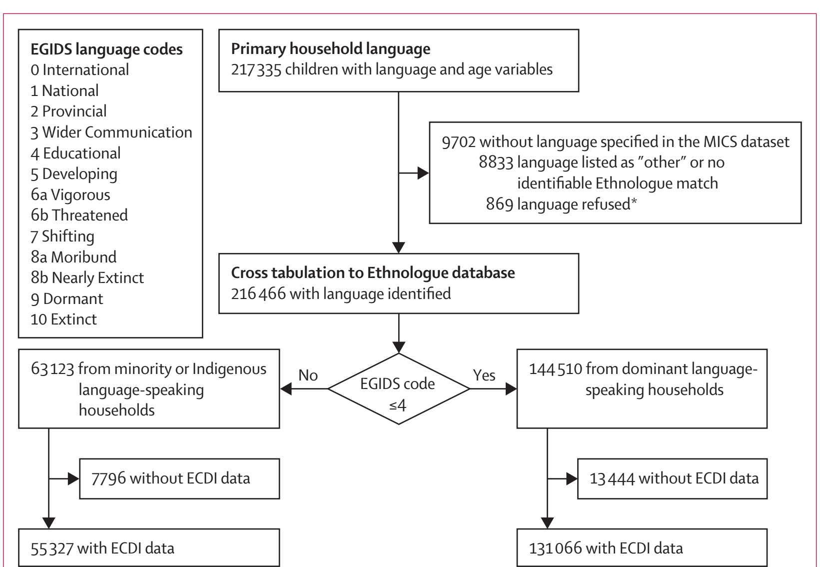


*Figure 1:* **Classification of household language as dominant versus minority or Indigenous language**
Country-specific language monographs published in Ethnologue (24th edition) were used to classify languages by ISO 639-3 code and endangerment or minority status with use of the EGIDS scale. Languages were classified as dominant for EGIDS codes 0–4 and as minority or Indigenous for EGIDS codes 5–10. The inset provides the labels for EGIDS codes; full descriptions are in appendix 2 (p 38). EGIDS=Expanded Graded Intergenerational Disruption Scale. ISO=International Organization for Standardization. MICS=Multiple Indicator Cluster Surveys. *Active refusal to specify either household head's primary language or respondent's primary language.

Adult development support was defined per the UNICEF indicator (coded in the MICS as "engagement of household members age 15 or over")as an adult caregiver completing at least four of the following activities with a child in the last 3 days: reading books, telling stories, singing songs, going outside the family house, playing together, or naming and counting things.

We conducted the regression analyses for all children aged 3–4 years with ECDI results and separately for the subgroup of children aged 3–4 years with ECDI results and functional disabilities when disability data were available. We stratified analyses by World Bank income group for fiscal year 2021 (low-income countries [LICs] and lower-middle-income countries [LMICs] *vs* upper-middle-income countries [UMICs]), WHO geographical region, and proportion of minority or Indigenous language-speaking households in each country (using a cutoff of ≤25% *vs* >25%). Code to replicate the ORs and risk differences are available in a public repository. To visualise data, we created forest plots of the ORs and risk differences for each of the stratified analyses, and for the ECDI overall and ECDI subdomains. We used complete case analysis and did not correct for multiplicity.

As sensitivity analyses, first, we re-estimated prevalences with Pakistan included; second, we calculated of country-specific crude and adjusted ORs; and third, we calculated E-value to assess the strength of unmeasured confounding necessary to negate the relationships between exposure and outcome, conditional on the included covariates.[22]

Data were analysed with Stata (version 15). Statistical significance was defined on the basis of 95% CIs, with a null value of 1 for odds ratios and 0 for risk differences.

## Role of the funding source

There was no funding source for this study.

## Results

107 MICS from 69 countries were considered for inclusion; 55 surveys from MICS wave 4, 44 from wave 5, and eight from wave 6 (appendix 4 p 4). After excluding the surveys without identifiable language status variables, we included 65 surveys from 40 countries with data on 217 335 children younger than 5 years (figure 1), corresponding to an estimated population of 39 833 646 individuals. 186 393 children aged 3–4 years (35–60 months) had both ECDI data and data for one of the language variables (interview language, respondent language, or head of household language), corresponding to an estimated population of 34 714 992 individuals. Language distribution and ISO and EGIDS codes in the final dataset are in appendix 4 (pp 39–49). We summarised descriptive characteristics for the estimated represented populations (table, appendix 4 p 57). These data and analyses exclude Pakistan, which represented a language outlier. Further details on the present sample are tabulated by country and MICS round in appendix 4 (pp 5, 7–34). A flow diagram and a visualisation of included MICS waves and countries is presented in appendix 4 (pp 4, 6).

For all children younger than 5 years, the estimated prevalence of minority or Indigenous language status was 28·1% (95% CI 27·0–29·2). As expected on the basis of the known societal marginalisation of and frequent discrimination against minority or Indigenous language groups, we observed that household wealth, maternal education status, and the prevalence of a developmentally supportive home environment were all significantly

| | Children from dominant language-speaking households | Children from minority or Indigenous language-speaking households |
|---|---|---|
| Overall sample | 71·9% (70·8–73·1) | 28·1% (27·0–29·2) |
| Sex | | |
| Female | 49·9% (49·0–50·8) | 49·3% (47·8–50·7) |
| Male | 50·1% (49·2–51·0) | 50·7% (49·3–52·2) |
| Mean age, months | 46·1 (45·9–46·3) | 44·5 (44·2–44·7) |
| Stunting prevalence (height-for-age <–2 SD) | 18·2% (17·5–19·0) | 20·0% (18·8–21·3) |
| Household wealth quintile | | |
| Q5 (highest wealth) | 17·2% (16·3–18·0) | 10·3% (9·3–11·4) |
| Q4 | 20·4% (19·4–21·3) | 13·4% (12·2–14·6) |
| Q3 | 20·3% (19·5–21·2) | 17·5% (16·1–18·9) |
| Q2 | 21·5% (20·6–22·3) | 22·0% (20·5–23·7) |
| Q1 (lowest wealth) | 20·7% (19·8–21·7) | 36·8% (34·6–39·2) |
| Mother's education | | |
| None | 4·1% (3·8–4·4) | 10·5% (9·4–11·9) |
| Primary | 49·9% (48·8–51·0) | 66·9% (65·1–68·7) |
| Secondary or higher | 46·1% (45·0–47·2) | 22·5% (21·0–24·2) |
| Supportive developmental environment* | 70·1% (69·1–71·0) | 54·6% (52·7–56·6) |
| ECDI† | | |
| On-track development in ≥3 of 4 domains | 76·6% (75·7–77·4) | 65·7% (64·2–67·2) |
| On-track development in literacy–numeracy | 31·6% (30·6–32·7) | 18·8% (17·4–20·2) |
| On-track development in physical | 97·9% (97·6–98·1) | 96·6% (96·2–97·0) |
| On-track development in learning | 93·2% (92·7–93·7) | 91·3% (90·6–92·0) |
| On-track development in socioemotional | 76·1% (75·3–77·0) | 66·9% (65·4–68·5) |
| Family-identified functional disability‡ | | |
| Any disability | 3·5% (2·9–4·2) | 3·8% (2·7–5·5) |
| Vison disability | 1·1% (0·8–1·4) | 0·8% (0·5–1·4) |
| Hearing disability | 1·4% (1·1–1·8) | 1·1% (0·7–1·8) |
| Mobility disability | 2·9% (1·8–4·6) | 1·7% (1·2–2·4) |
| Among children with any family-identified disability, on-track development in ≥3 of 4 ECDI domains§ | 75·3% (59·0–86·5) | 71·7% (62·1–79·7) |

Demographic factors are weighted to correspond to the estimated represented population of 39 833 646 children aged <5 years (see footnotes for populations for the ECDI and disability modules); data are percentage (prevalence) with 95% CIs or mean with 95% CIs. Data on race and ethnicity were not extracted as race is generally not collected in the MICS and ethnicity data collection in the MICS varies substantially from survey to survey. ECDI=Early Child Development Index. MICS=Multiple Indicator Cluster Surveys. *Defined as an adult doing four or more developmentally supportive activities (reading books, telling stories, singing songs, going outside the family house, playing together, or naming and counting things) in the past 3 days. †For 186 393 children aged 35–60 months (3–4 years) with ECDI data; estimated represented population was 34 714 992 individuals. ‡For 66 723 children aged 23–60 months (2–4 years) with available data; estimated represented population was 11 465 601 individuals. §For 1544 children with a family-identified disability and ECDI results; estimated represented population was 292 691 individuals.

***Table:* Weighted characteristics of the study population**

For the **MICS6 tools** see https://mics.unicef.org/tools?round=mics6#analysis

For **World Bank income groups** see https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups

For **WHO geographical regions** see https://www.who.int/countries

For the **public repository** see https://doi.org/10.7910/DVN/X1UAU8

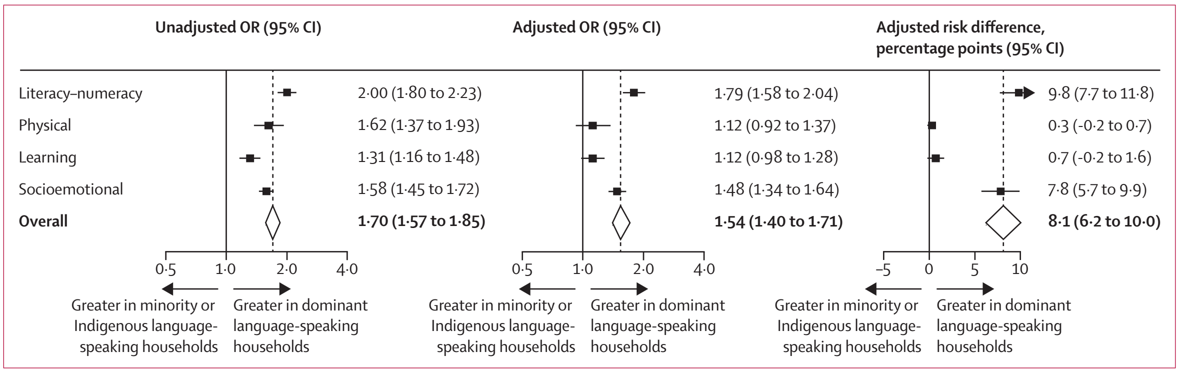


*Figure 2:* **Effect estimates of on-track child development among children from dominant language-speaking households versus minority or Indigenous language-speaking households**
On-track child development was assessed by the overall and domain scores of the Early Childhood Development Index. Each unadjusted or adjusted estimate reflects a separate logistic regression model for the outcome of interest. Only the coefficient for language status is reported. Diamond symbols show the overall estimate. Unadjusted risk differences are provided in appendix 4 (p 50). OR=odds ratio.

lower for children from households speaking a minority or Indigenous language (table).

In children aged 3–4 years, the estimated prevalence of on-track overall development as measured by the ECDI was 73·7% (95% CI 72·9–74·4); this rate was 65·7% (64·2–67·2) for children from households speaking a minority or Indigenous language and 76·6% (75·7–77·4) for those from dominant language-speaking households (table). The largest differences in on-track development for dominant versus minority or Indigenous language status were noted for the literacy–numeracy domain (difference 12·9 percentage points, 95% CI 11·1–14·7) and socioemotional domain (difference 9·2 percentage points, 7·4–11·0).

The disability module was available for 66 723 children aged 2–4 years (23–60 months) with language data. The module was not used in MICS-4 and rarely used in MICS-5, thus most data (n=61 559, 92·3%) were from MICS-6. The estimated represented population was 11 465 601 children. The estimated prevalence of family-reported functional disability was 3·6% (95% CI 3·0–4·4); this rate was 3·8% (95% CI 2·7–5·5) for children from minority or Indigenous language households and 3·5% (95% CI 2·9–4·2) for children from dominant-language households.

The sample of 1544 children with a family-identified disability and ECDI and language data was used to estimate rates of on-track development (estimated represented population n=292 691). The estimated prevalence of on-track overall development among children with a family-identified disability was 73·3% (64·1–80·9); this rate was 71·7% (62·1–79·7) for children from minority or Indigenous language-speaking households and 75·3% (59·0–86·5) for dominant language-speaking households.

We derived adjusted ORs and risk differences for on-track development by language status (figure 2, appendix 4 p 50). Being from a dominant language-speaking household was significantly associated with increased odds of on-track overall development in children, when adjusting for child sex, household wealth, child nutritional status, maternal education, country-level early child education, and a developmentally supportive home environment (adjusted OR 1·54, 95% CI 1·40–1·71). This association appeared to be largely driven by significantly increased odds of on-track development in the literacy–numeracy domain (adjusted OR 1·79, 1·58–2·04) and socioemotional domain (adjusted OR 1·48, 1·34–1·64). Results for the physical and learning domains were not significant.

Adjusted ORs and risk differences were calculated for children with family-identified disabilities (figure 3, appendix 4 p 51). Language status was not associated with differences in the odds of on-track status on the overall ECDI among children with any functional disability (adjusted OR 1·02, 95% CI 0·43–2·45). For individual domains, dominant language status was associated with lower odds of on-track physical development for children with any disability (adjusted OR 0·25, 95% CI 0·07–0·88), for those with vision disability (adjusted OR 0·05, 0·00–0·70), and for those with hearing disability (adjusted OR 0·02, 0·00–0·68; figure 3, appendix 4 p 51).

Adjusted ORs and risk differences for on-track development were stratified by country-level characteristics (World Bank income group, WHO geographical region, and proportion of minority or Indigenous language-speaking households; figure 4, appendix 4 pp 52–53). In the stratified analyses, dominant language status was associated with on-track overall development and literacy–numeracy and socioemotional development for the LIC and LMIC group and for the UMIC group, except for socioemotional development in LICs and LMICs. When stratified by high proportion (>25%) versus low proportion (≤25%) of minority or Indigenous language-speaking households, dominant language status was associated with on-track overall development and literacy–numeracy development regardless of the

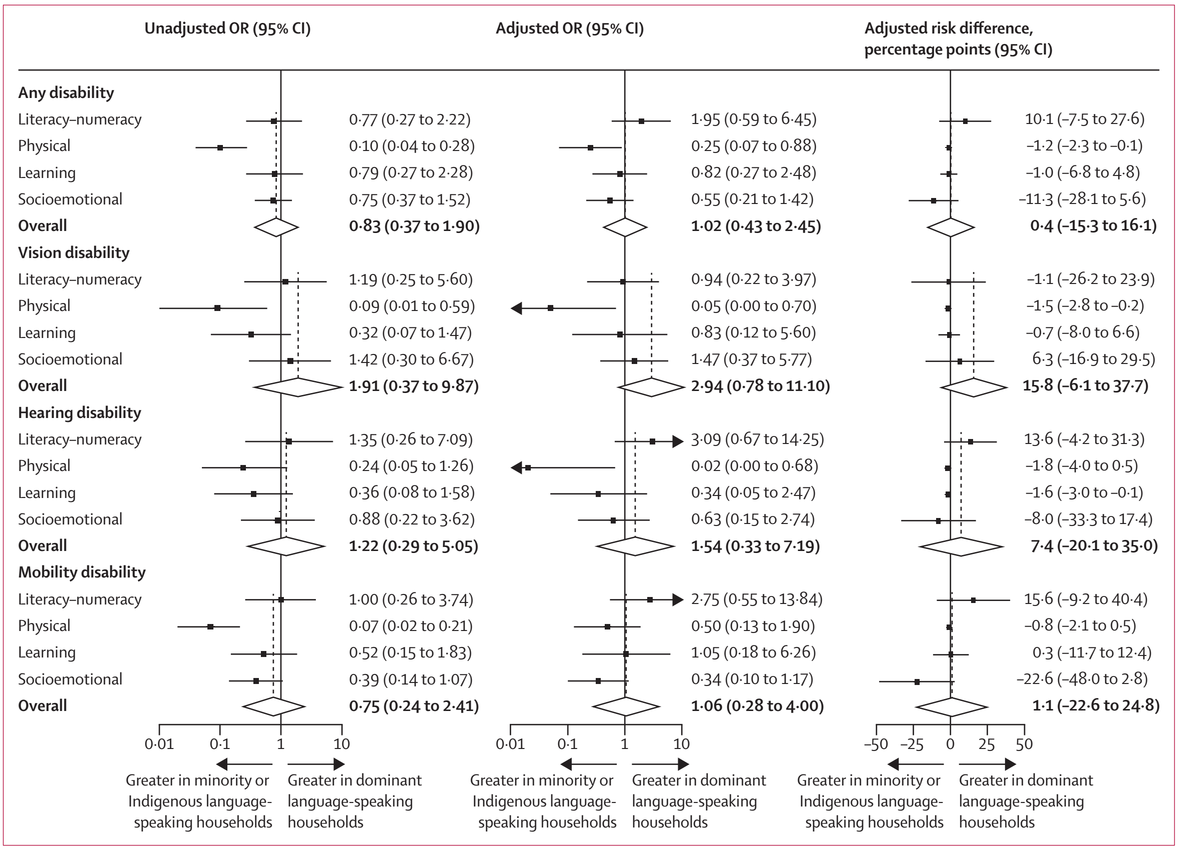


***Figure 3:* Effect estimates of on-track child development among children with family-identified disability from dominant language-speaking households versus minority or Indigenous language-speaking households**

On-track child development was assessed by the overall and domain scores of the Early Childhood Development Index. Each unadjusted or adjusted estimate reflects a separate logistic regression model for the outcome of interest. Only the coefficient for language status is reported. Forest plots ending in arrows go beyond the scale shown. Diamond symbols show the overall estimate. Unadjusted risk differences are provided in appendix 4 (p 51). OR=odds ratio.

stratification factor, but not with socioemotional development. There was substantial heterogeneity of effect by WHO geographical region.

We derived country-specific crude and adjusted models to examine the association between dominant language status and overall on-track development (appendix 4 pp 54–55). We also explored how the inclusion of Pakistan affected associations with on-track development overall and for individual development domains (appendix 4 p 56). Results with the inclusion of Pakistan were broadly consistent with the primary findings. Variability was observed for country-specific estimates.

We calculated an E-value to assess the likelihood that an unmeasured confounder not included in our DAG could influence findings enough to render associations between language status and ECDI scores non-significant;[22] an unmeasured confounder would need a minimum strength of association (conditional on the measured covariates) with both language and ECDI scores of an OR of 1·7 for overall development or 2·0 for literacy–numeracy development to render the findings non-significant.

## Discussion

In this descriptive analysis, minority or Indigenous language was associated with lower odds of on-track development, especially in the literacy–numeracy and socioemotional domains, after adjustment for multiple covariables. This large, pooled analysis of population-representative data from more than 186 000 children aged 3–4 years from 40 countries provides important estimates that were missing from the global literature. Our findings provide important estimates missing from the global literature on language status as a possible social determinant of child developmental health. We also provide a new methodology for classifying language status in large datasets.

Given the history of oppression against and differential opportunities for minority or Indigenous language speakers worldwide, structural explanations for the observed disparities by language status include various forms of sociolinguistic marginalisation. Discriminatory national language policies, such as not producing early reading materials in minority or Indigenous languages or not funding Indigenous-language early learning centres,

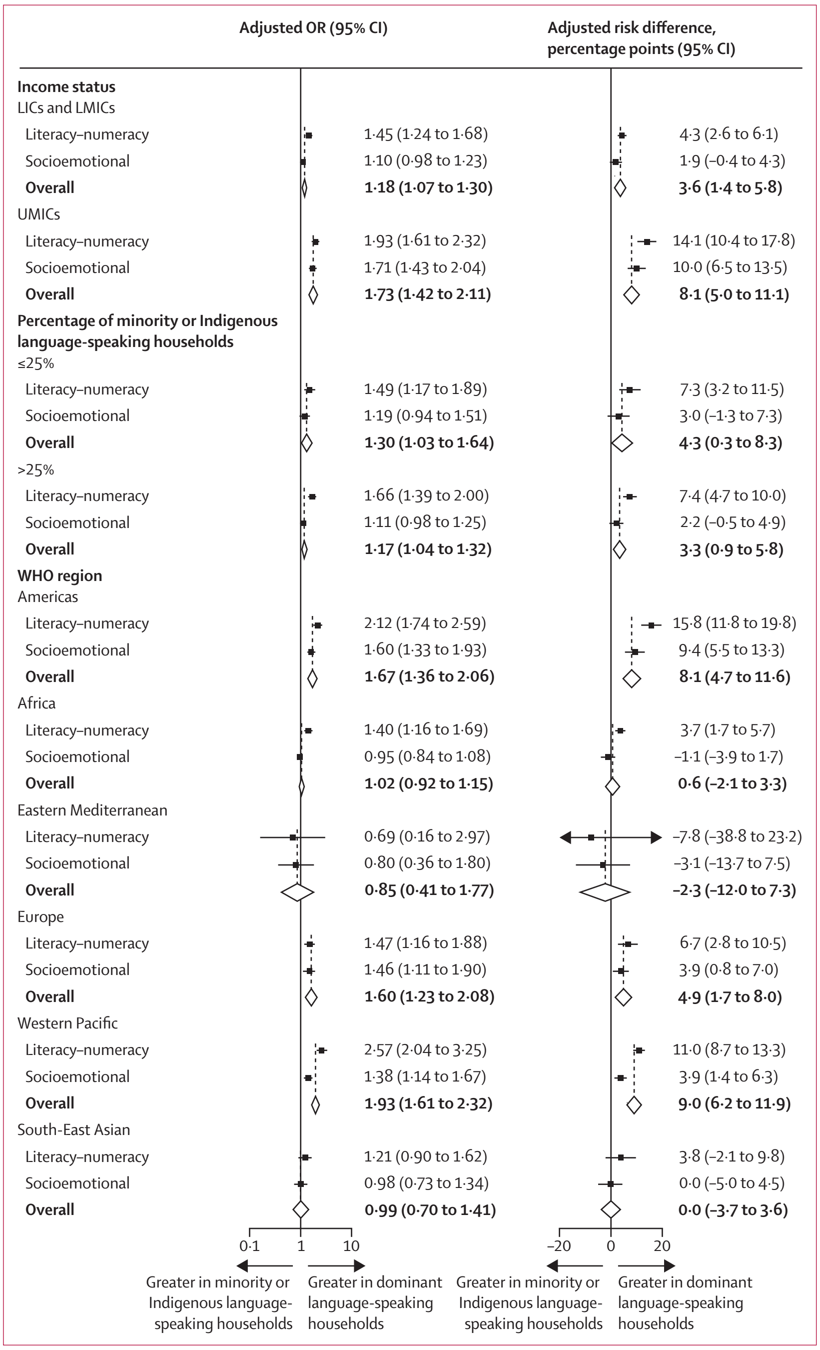


*Figure 4:* **Effect estimates of on-track child development among children from dominant language-speaking households versus minority or Indigenous language-speaking households, stratified by country-level characteristics**

On-track child development was assessed by the overall and domain scores of the Early Childhood Development Index. Physical and learning domains are not presented because when adjusted for covariables (figure 2), these were not statistically significantly associated with language status. Results for all four development domains are presented in appendix 4 (pp 52–53). Each adjusted estimate reflects a separate stratified logistic regression model for the outcome of interest. Only the coefficient for language status is reported. Forest plots ending in arrows go beyond the scale shown. Diamond symbols show the overall estimate. OR=odds ratio.

are common.[23,24] In our stratified analyses, the disparity in development for children from minority or Indigenous language-speaking households was conserved except for the socioemotional domain across World Bank income group categories and regardless of minority language population proportion. Recently, Bornstein and colleagues showed that the human development and educational indices of countries were associated with higher ECDI scores.[25] However, these indices only accounted for 25% of the variance in child development in the Bornstein study. The additional disparity by language status that we describe in this study suggests that some of the residual country-level difference might be due to discrimination by language status.

A second consideration when interpreting our main finding is that the ECDI might overestimate developmental risk in minority or Indigenous language children. Research has shown that the age of attainment of developmental milestones in healthy, well nourished children is consistent across cultures.[26,27] Nevertheless, some functional domains might be less well conserved cross-culturally and cross-linguistically. In particular, for ECDI questions in the literacy–numeracy domain, McCoy and colleagues reflect that "observed differences... are more likely to reflect differences in countries' social/cultural norms around early education than they are likely to reflect children's cognitive capacity."[28] In addition, a large body of literature has shown that instruments applied in the dominant language routinely underestimate language capabilities when administered to children who speak a minority or Indigenous language.[29–31] As both previous literature and our own study show, parent-reported concerns about children's capabilities are similar regardless of language status, suggesting measurement bias when using a test which might privilege dominant languages and culture.[32,33]

In this study, we also disaggregated our findings by family-reported disability status of children. Our prevalence estimates of vision and hearing disabilities in the overall population are similar to those recently identified in an analysis of the Global Burden of Disease Study 2017.[34] We found few differences by language status, although dominant language was associated with lower odds of on-track physical development in children with disabilities, except for those with mobility disabilities. This finding might be related to unmeasured differences in severity of disability, survival of children with complex illness, or differing thresholds for reporting disability among caregivers.

Our study has several notable strengths. The MICS data are an exceptional resource with standardised global methodologies. This analysis provides weighted and representative data from diverse global settings collected in each country by local agencies. Our study represents a novel effort to quantify disparities in measured child development by language status, which reflects the burden of discrimination and marginalisation of


www.thelancet.com/lancetgh **Published online November 10, 2023** **https://doi.org/10.1016/S2214-109X(23)00456-4**

minority or Indigenous languages on child health. Additionally, a major methodological contribution of our study is the standardised approach to extracting and coding language data, and determining dominant versus minority or Indigenous language status, in large population-based datasets.

Our study has several important limitations. The Ethnologue database used to categorise languages is updated annually. We categorised languages on the basis of the most recent database available at the time of the study (released in 2021), which might not accurately reflect language status at the time of a specific MICS in all cases. In addition, the MICS have some inherent biases. They are more likely to include the stably housed; our population therefore might be biased away from vulnerable unhoused populations. MICS also use cluster sampling with anonymised data designed to be nationally representative; it is therefore not possible to assign individual households to regions to assess within-country variation. Language status was determined with an algorithm that used the available information, starting with the primary language of the respondent, then the language of the interview, and finally the primary language of the head of household. The assumption that this is the language that the child hears most might not always be correct. Missing data are also a concern, as the MICS datasets did not always capture language in the same way over time. Some MICS datasets included only the dominant language and "other", which were unable to be categorised. Less than half of MICS-4 (2009–12) datasets were included in our analysis because of missing language variables, whereas all MICS-6 (2017–19) datasets were included (appendix 4 p 4). We provided missing language estimates for each survey in appendix 4 (pp 39–49). Additionally, Ethnologue did not provide language assessments for regions with contested diplomatic status; Kosovo, specifically, could not be included in analyses. Furthermore, we were only able to control for variables that were collected in MICS datasets; missing data might have led to residual confounding. However, we used authors' expert knowledge to create a DAG to guide our models, and we controlled for plausible confounders in the DAG. In addition, in a sensitivity analysis, we calculated an E-value to estimate the effect of residual confounding. One final limitation is our use of the ten-question ECDI, which has been replaced by the expanded ECDI2030. However, few ECDI2030 datasets are publicly available, and so the original ECDI remains one of the largest and most valuable repositories of developmental data in children worldwide. The ten-question ECDI was used only in children aged 3–4 years, and therefore we were unable to assess language relationships outside that age range, and the results cannot be assumed to apply to all under-5 children.

In conclusion, our analysis of a large multicountry MICS dataset shows notable disparities in on-track development for children from minority or Indigenous language-speaking households compared with those from dominant language-speaking households. The observed disparity could partly be due to mismeasurement resulting from the ECDI instrument's cross-linguistic limitations, and the new expanded ECDI2030 has the potential to improve comparative measures as it gradually rolls out. However, our results emphasise the degree to which marginalisation of minority and Indigenous languages impacts children's developmental health, and represent a call for vigorous efforts to address this disparity.

#### Contributors

ACM co-developed the research question, developed the analysis plan, conducted the literature search, merged and cleaned the data, reviewed language data, conducted analyses, and co-wrote the first draft and all subsequent drafts of the paper. DF created the visual displays of data, and reviewed, edited, and approved all manuscript versions. ST conducted data analyses, prepared tables and the visual displays of data, and reviewed and approved the final manuscript. KD accessed the Multiple Indicator Cluster Surveys (MICS) data, applied exclusion criteria, collected and entered language data, and reviewed, edited, and approved all manuscript versions. PR co-developed the research question, conducted the literature search, collected and entered language data, reviewed analytical code, conducted analyses, and co-wrote the first and all subsequent drafts of the manuscript. ST and PR accessed and verified the data. All authors had full access to all data in the study and had final responsibility for the decision to submit for publication.

#### Declaration of interests

DF serves as an unpaid staff physician for Wuqu' Kawoq – Maya Health Alliance, which is a non-governmental health organisation in Guatemala that conducts advocacy at the intersection of health care and Indigenous language revitalisation. ST is employed by Wuqu' Kawoq – Maya Health Alliance. PR is Chief Science Officer for Wuqu' Kawoq – Maya Health Alliance. ACM and KD declare no competing interests.

#### Data sharing

MICS study data are the property of the UNICEF MICS programme and the data use agreement is that the data may not be redistributed or passed on to others in any form. However, the MICS study datasets are available upon request to the MICS programme for legitimate research purposes at the website: https://mics.unicef.org/surveys. The statistical analysis plan and Stata code are available on ACM's Dataverse repository at https://doi.org/10.7910/DVN/X1UAU8.

#### Acknowledgments

This study had no outside funding. We acknowledge our respective universities for allowing us the time to work on this project. DF was supported by the US National Institutes of Health (NIH; award number K23HL161271) and PR was partially funded by a grant from the NIH National Institute of Child Health and Human Development (award number R01HD100984) during the course of this work.

Editorial note: The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

#### References

1 Austin PK, Sallabank J. Cambridge handbook of endangered languages. Cambridge, UK: Cambridge University Press, 2011.

2 Elek C, Gubhaju L, Lloyd-Johnsen C, Eades S, Goldfeld S. Can early childhood education programs support positive outcomes for Indigenous children? A systematic review of the international literature. *Educ Res Rev* 2020; **31:** 100363.

3 Beauchamp MLH, Amorim K, Wunderlich SN, Lai J, Scorah J, Elsabbagh M. Barriers to access and utilization of healthcare services for minority-language speakers with neurodevelopmental disorders: a scoping review. *Front Psychiatry* 2022; **13:** 915999.

4 Faruk MO, Rosenbaum S. The mental health consequences of Indigenous language loss. *Lancet Psychiatry* 2022; **9:** e46.

5 Brown SJ. Disparities in outcomes for Indigenous children. *Lancet Child Adolesc Health* 2020; **4:** 5–6.

6 Anderson I, Robson B, Connolly M, et al. Indigenous and tribal peoples' health (The Lancet–Lowitja Institute Global Collaboration): a population study. *Lancet* 2016; **388:** 131–57.

7 Paulino NA, Vazquez MS, Bolumar F. Indigenous language and inequitable maternal health care, Guatemala, Mexico, Peru and the Plurinational State of Bolivia. *Bull World Health Organ* 2019; **97:** 59–67.

8 Stephens C, Porter J, Nettleton C, Willis R. Disappearing, displaced, and undervalued: a call to action for Indigenous health worldwide. *Lancet* 2006; **367:** 2019–28.

9 Horton R. Indigenous peoples: time to act now for equity and health. *Lancet* 2006; **367:** 1705–07.

10 Morelli G, Quinn N, Chaudhary N, et al. Ethical challenges of parenting interventions in low- to middle-income countries. *J Cross Cult Psychol* 2018; **49:** 5–24.

11 Flood D, Rohloff P. Indigenous languages and global health. *Lancet Glob Health* 2018; **6:** e134–35.

12 UN Human Rights Office of the High Commissioner. A/HRC/43/47: education, language and the human rights of minorities—report of the Special Rapporteur on minority issues. Jan 9, 2020. https://www.ohchr.org/en/documents/thematic-reports/ahrc4347-education-language-and-human-rights-minorities-report-special (accessed Oct 20, 2023).

13 UNICEF. Development status. June, 2023. http://data.unicef.org/ecd/development-status.html (accessed Sept 18, 2023).

14 SIL International. ISO 639-3 change requests series 2022: summary of outcomes. Jan 23, 2023. https://iso639-3.sil.org/sites/iso639-3/files/reports/639-3_ChangeRequests_2022_Summary.pdf (accessed Sept 18, 2023).

15 Lewis MP, Simons GF. Assessing endangerment: expanding Fishman's GIDS. *Revue Roumaine Linguistique*. 2010; **55:** 103–20.

16 Government of Pakistan Population Census Organization. Population by mother tongue. 2017. https://www.pbs.gov.pk/sites/default/files/population/2017/tables/pakistan/Table11n.pdf (accessed Oct 19, 2023).

17 Loizillon A, Petrowski N, Britto P, Cappa C. Development of the Early Childhood Development Index in MICS surveys. New York, NY: UNICEF, 2017.

18 Norton EC, Miller MM, Kleinman LC. Computing adjusted risk ratios and risk differences in Stata. *Stata J* 2013; **13:** 492–509.

19 StataCorp. Stata survey data reference manual, release 18. 2023. https://www.stata.com/manuals/svy.pdf (accessed Oct 20, 2023).

20 WHO, UNICEF. WHO child growth standards and the identification of severe acute malnutrition in infants and children: A joint statement by the World Health Organization and the United Nations Children's Fund. 2009. https://iris.who.int/bitstream/handle/10665/44129/9789241598163_eng.pdf?sequence=1 (accessed Oct 16, 2023).

21 Rutstein SO, Rojas G. Guide to DHS statistics. Calverton, MD: MEASURE DHS/ICF International, 2006.

22 VanderWeele TJ, Ding P. Sensitivity analysis in observational research: introducing the E-value. *Ann Intern Med* 2017; **167:** 268–74.

23 McIvor O, Ball J. Language-in-education policies and Indigenous language revitalization efforts in Canada: considerations for non-dominant language education in the Global South. *Forum Int Res Educ* 2019; **5:** 12–28.

24 Becerra-Lubies R, Mayo S, Fones A. Revitalization of Indigenous languages and cultures: critical review of preschool bilingual educational policies in Chile (2007–2016). *Int J Biling Educ Biling* 2021; **24:** 1147–62.

25 Bornstein MH, Rothenberg WA, Lansford JE, et al. Child development in low- and middle-income countries. *Pediatrics* 2021; **148:** e2021053180.

26 Ertem IO, Krishnamurthy V, Mulaudzi MC, et al. Similarities and differences in child development from birth to age 3 years by sex and across four countries: a cross-sectional, observational study. *Lancet Glob Health* 2018; **6:** e279–91.

27 Fernandes M, Stein A, Newton CR, et al. The INTERGROWTH-21st Project Neurodevelopment Package: a novel method for the multi-dimensional assessment of neurodevelopment in pre-school age children. *PLoS One* 2014; **9:** e113360.

28 McCoy DC, Peet ED, Ezzati M, et al. Early childhood developmental status in low- and middle-income countries: national, regional, and global prevalence estimates using predictive modeling. *PLoS Med* 2016; **13:** e1002034.

29 Makgamatha MM, Heugh K, Prinsloo CH, Winnaar L. Equitable language practices in large-scale assessment: possibilities and limitations in South Africa. *South Afr Linguist Appl Lang Stud* 2013; **31:** 251–69.

30 Pearce WM, Flanagan K. Language abilities of Indigenous and non-Indigenous Australian children from low socioeconomic backgrounds in their first year of school. *Int J Speech Lang Pathol* 2019; **21:** 212–23.

31 Zupan B, Campbell-Woods N, Thompson H. Scoping review: language assessment practices for Aboriginal and Torres Strait Islander children in Australia and guidelines for clinical practice. *Aust J Rural Health* 2021; **29:** 879–95.

32 McLeod S, Verdon S, Bennetts Kneebone L. Celebrating young Indigenous Australian children's speech and language competence. *Early Child Res Q* 2014; **29:** 118–31.

33 Knauer HA, Kariger P, Jakiela P, Ozier O, Fernald LCH. Multilingual assessment of early child development: analyses from repeated observations of children in Kenya. *Dev Sci* 2019; **22:** e12875.

34 Olusanya BO, Wright SM, Nair MKC, et al. Global burden of childhood epilepsy, intellectual disability, and sensory impairments. *Pediatrics* 2020; **146:** e20192623.