NET WORK LISER

Discussion Paper Series

IZA DP No. 19079

November 2026

Language Match Is Fixed, Language Understanding Can Be Built: Rethinking Home Language Instruction in Sub-Saharan Africa

Ernesta Orlovaitė
Royal Holloway, University of London

Arnaud Chevalier
Royal Holloway, University of London,
CESifo, IFS, IZA@LISER and RFBerlin

Kathleen Rastle
Royal Holloway, University of London

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.

Language Match Is Fixed, Language Understanding Can Be Built: Rethinking Home Language Instruction in Sub-Saharan Africa*

Abstract

Nine in ten children in Sub-Saharan Africa cannot read a simple text by the age of ten, and most are taught in a language they do not speak at home. The dominant policy response is home language instruction, yet the evidence that it improves reading is weaker than its adoption suggests. Using a unique dataset of 1,718 formerly out-of-school children in an accelerated learning programme in Ghana and Liberia, we compared matched pupils, who spoke the language of instruction at home, with mismatched pupils, who did not. Speaking the language of instruction at home provided a clear advantage in Ghana but not in Liberia—a puzzle we resolved with three findings. First, understanding of the language of instruction is a better predictor of reading progress across countries than language match. Second, reading is shaped by classroom characteristics: the more classmates shared a mismatched pupil's home language, the worse that pupil read. Third, language understanding can be built: most mismatched pupils developed oral proficiency in the language of instruction within a single programme cycle, the classroom's language mix predicted whether they did, and those who did read better. No language policy can match the home language of every child, but every child should be taught to understand the language of the classroom—and once they do, they can learn to read in it, whatever language they speak at home.

JEL classification

I21

Keywords

reading, mother tongue, language of instruction, oral language, peer effects, Sub-Saharan Africa

Corresponding author

Ernesta Orlovaitė
Ernesta.Orlovaite@rhul.ac.uk

* *Funding*: This work was supported by the Economic and Social Research Council (ESRC) through the SEDarc Doctoral Training Partnership (grant number ES/Y001656/1).

Acknowledgements: The authors gratefully acknowledge the Luminos Fund for their invaluable in-kind support, including expert advice and the provision of proprietary data from their accelerated learning programmes. The views and opinions expressed in this article are those of the authors and do not necessarily reflect the position of the Luminos Fund.

Declaration of competing interests: Ernesta Orlovaitė was employed by the Luminos Fund from 2021 to 2025 and currently provides paid consultancy to the organisation. Kathleen Rastle is a member of the Luminos Fund's Advisory Board. As required by the data sharing agreement, the Luminos Fund provided comments on the manuscript before submission; the authors retained full control over the study design, data analysis, interpretation of results, and the decision to publish.

Data availability: The individual pupil data analysed in this study were provided by the Luminos Fund and are not publicly available due to their proprietary nature and to protect participant confidentiality. The analysis code and outputs that support the findings of this study are available at https://anonymous.4open.science/r/phd-study1-luminos-egra.

1. Introduction

Nine in ten children in Sub-Saharan Africa cannot read and understand a simple text by the age of ten (World Bank et al., 2022). The crisis persists despite decades of progress in school access: primary enrolment is now near-universal across much of the region (World Bank, 2026), yet it has not translated into learning. The central challenge for education systems is thus no longer getting children into school, but ensuring that those who attend learn.

Four in five children in Sub-Saharan Africa are taught in a language they do not speak at home (World Bank, 2021). Former colonial languages, native to almost no one, dominate education systems throughout the region (UNESCO, 2025). African languages are used as well, but over 2,000 are spoken on the continent (Eberhard et al., 2026), and few are widely shared as second languages of early childhood. For most children in the region, then, learning to read means learning to read in a language they may not understand.

The dominant policy answer is to teach children in the language they speak at home. The World Bank (2021) made this the first principle of its literacy policy in *Loud and Clear*, and UNESCO has advocated the same for decades. The intuition is hard to argue with: after all, '[i]f you don't understand, how can you learn?' (UNESCO, 2016). Home language instruction policies have since been adopted across much of the region, although they are rarely implemented in practice (Trudell, 2016; UNESCO, 2025).

In this study, we examine how the language of instruction shapes reading in Ghana and Liberia, for matched pupils, who speak it at home, and mismatched pupils, who do not. The two countries produce a puzzle: speaking the language of instruction at home is a clear advantage in Ghana, but not in Liberia. We report three findings that resolve this puzzle. First, the focus on language match is misleading: it predicts reading differently across settings, while language understanding predicts it consistently. Second, reading is shaped by classroom characteristics: the more classmates share a mismatched pupil's home language, the worse the pupil reads. Third, while language match is fixed, language understanding can be developed: the classroom predicts whether mismatched pupils develop it, and pupils who do read better. Together, these findings offer a more

nuanced view of home language instruction: children can learn to read in a language they do not speak at home if they are supported to understand it at school.

1.1. Evidence for home language policy

The home language policy is rooted in the science of reading. Under the Simple View of Reading, reading is the product of two components: decoding the written word, and understanding the spoken language it represents (Gough & Tunmer, 1986; Hoover & Gough, 1990). This relationship is multiplicative: a child who can sound out a story but does not understand its language will not read with comprehension (Hjetland et al., 2020). This is the logic behind *Loud and Clear*: teaching in a child's home language makes reading possible, because the understanding is already there (World Bank, 2021).

Yet the evidence that home language instruction improves reading is weaker than the policy's adoption suggests. Earlier literature reviews were encouraging, finding consistent gains of home language instruction interventions across Africa (Evans & Mendez Acosta, 2021; Trudell, 2016). The most recent review is more cautious: drawing on 45 studies from Sub-Saharan Africa and South and Southeast Asia, Nakamura et al. (2023) concluded that home language interventions *might* improve reading, but possibly only if bundled with structured pedagogy. Indeed, home language instruction is almost never delivered on its own, so the evidence behind the policy consensus is confounded. For example, a study in Uganda found significant reading gains from teaching children in their home language (Brunette et al., 2019), but the programme also included teacher training, structured materials, and ongoing coaching[1]—interventions that improve learning outcomes on their own (Snilstveit et al., 2015). When Kim et al. (2020) meta-analysed 67 studies from 32 low- and middle-income countries and adjusted for these confounds, first- and second-language instruction no longer differed on any

[1]PRIMR-MT in Kenya is another study that is often cited as evidence that home language instruction improves reading (e.g. Evans & Mendez Acosta, 2021; UNESCO, 2025). It comes closer than most experiments to isolating the language effect, adding literacy instruction in the home languages, Lubukusu and Kikamba, to an otherwise identical base programme (Piper et al., 2016). But all pupils, including those who did not receive instruction in the home languages, were assessed in either Lubukusu or Kikamba. Reading was never assessed in the languages of the base programme, so the experiment cannot show which approach produced better readers.

reading outcome.

1.2. The present study

Perhaps the home language advantage disappears not because it is absent, but because the language match captures the wrong construct. Research and policy sort children into two groups: matched and mismatched (e.g. Nakamura et al., 2023; UNESCO, 2025). But the Simple View of Reading is not about which language a child speaks at home; it is about how much of the language of the text they understand. How far these two constructs diverge depends on the setting. With over 80 languages spoken, Ghana is among the most linguistically diverse countries in the region (Eberhard et al., 2026). Policy mandates instruction in one of the eleven home languages in the early grades (Appiah & Ardila, 2020), but the language is often set regionally and imposed on communities that may not speak it (Anyidoho, 2018). There, a mismatched child may rarely hear the language of instruction outside the classroom. Liberia is also diverse (Eberhard et al., 2026), but the language of instruction is its lingua franca: most Liberians speak some variety of English, and many acquire it outside school (Singler, 2012). The same label may thus describe a child who understands little of the language of the school in one country, and a child who understands much of it in another. If the home language advantage comes from language *understanding*, language *match* might be too crude to detect it.

Unlike language match, language understanding can be developed (Snowling & Hulme, 2025). A child who does not speak the language of instruction at home might develop it at school. Teaching quality is among the strongest determinants of learning outcomes (Conn, 2017). But children also learn from each other, through group work and play (Quinn et al., 2018; Sacerdote, 2011; Tenenbaum et al., 2020). A mismatched child may spend much of the day talking to classmates who share their home language, and every such conversation is a missed opportunity to practise the language of the school. But when few classmates share their language, the same child may have to use the language of instruction to be understood at all. How much of the language a mismatched child hears may therefore depend on who else is in the classroom. The few studies that address this directly, all in high-income countries, show that classmates shape how

much of the language pupils understand and how well they read (Schmerse, 2021; Seuring et al., 2020). To our knowledge, no study has examined how the classroom's language mix shapes reading in low- and middle-income countries.

We address both gaps in the literature with a unique dataset that allows us to model the child and the classroom. We follow 1,718 formerly out-of-school children enrolled in the Luminos Fund's accelerated learning programme in Ghana and Liberia, shown to produce large learning gains (McManus et al., 2025). It is only because these children learned to read—unlike most of their peers in government schools—that language-related differences can be detected at all. For every child, we measure their reading comprehension and listening comprehension in the language of instruction, at two points in time. For every classroom, we measure its language mix. We model reading comprehension using mixed-effects models to answer three questions. First, does language understanding, rather than language match, explain how well children read? Second, does the classroom's language mix shape how well children read? Third, does it shape reading by building their understanding of the language? Our results suggest a new way of thinking about the language of instruction policy: not as a choice between two languages, but as a question of how children come to understand the language in which they are taught.

2. Methods

2.1. Setting and sample

The data cover three cohorts of the Luminos accelerated learning programme: Ghana 2023—2024 (n = 508), Ghana 2024-–2025 (n = 465), and Liberia 2024-–2025 (n = 745). Luminos is a foundational literacy and numeracy programme delivered by community teachers to out-of-school children aged 8—14 who have never attended school or who dropped out before acquiring basic reading skills (Baron, 2025). It condenses the first three grades of literacy instruction into a single year, pairing systematic phonics and oral language development with small classes, structured pedagogy, and continuous teacher coaching. The model is effective: a randomised controlled trial in Liberia found large gains in reading comprehension (1.33 SDs) over out-of-school children

not enrolled in the programme (McManus et al., 2025). Its pupils are not, however, representative of broader pupil populations in either country, in three respects: Luminos serves some of the most disadvantaged communities; the instruction is more intensive and more faithfully implemented; and children learn to read, while most of their peers in government schools do not.

Between 2023 and 2025, Luminos operated in Ghana's Ashanti region and in eight counties in Liberia (see *Figure 1*). The two Ghana cohorts differed primarily in programme maturity, while the Ghana and Liberia cohorts differed further in curricula, contact hours, and national contexts. Luminos has operated in Ghana since 2021 and in Liberia since 2016; the Liberia programme is thus more established. Both follow their national curricula and teach reading through phonics, but Ghana devotes just over two hours a day to literacy, compared to close to six in Liberia; the Liberia programme is thus more intensive. Cohort fixed effects absorb these differences.

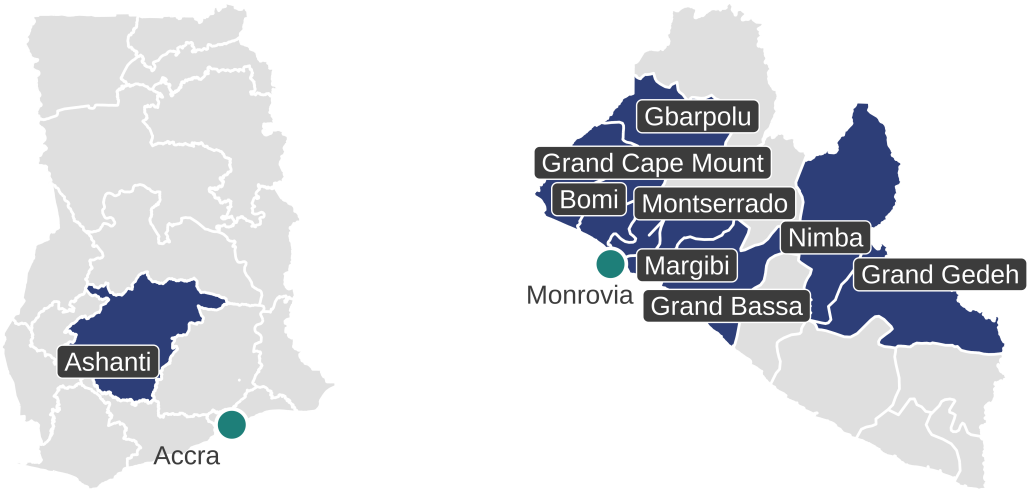


Figure 1. Region in Ghana (left) and counties in Liberia (right) where the Luminos programme operated during the 2023—2024 and 2024–-2025 academic years. Teal dots indicate the country capitals. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

Both programmes teach in the government-mandated language of instruction: Asante Twi in Ghana, and English in Liberia. Many pupils do not speak it at home, especially in Ghana, where the programme serves a region shaped by migration. One in seven Ashanti residents was born in another region, and likely more in rural areas (Ghana Statistical Service, 2023), where Luminos operates. A language mismatch is thus a

feature of the communities the programme serves.

Data for the three cohorts were collected by independent external evaluators: the Education Assessment and Research Centre in Ghana, and The Khana Group in Liberia. Baseline was administered roughly within the first two weeks of the programme, and endline within the last two. In each classroom, the evaluators randomly sampled 6 children in 2023–2024, and 7 in 2024–2025 from pupils present on the baseline assessment day. Of the 2,202 pupils sampled, 1,726 were reassessed at endline (78% retention); a further 8 pupils were dropped due to missing covariate data for a final sample size of 1,718.

Retained and attrited pupils were similar on the characteristics central to our analysis: language match, language peer density, baseline listening comprehension, and baseline reading comprehension (see *Appendix A*). Attrited pupils were concentrated in slightly larger classrooms with slightly lower mean baseline reading comprehension. However, because attrition was unrelated to language match, it is unlikely to bias our estimates.

2.2. Measures

Our analysis used four sets of variables: oral language and reading outcomes, language variables, individual-level covariates, and classroom-level covariates. We describe them below; full summary statistics are reported in *Appendix B*.

2.2.1. Oral language and reading outcomes

Listening comprehension and reading comprehension were measured using the Grade 2 Early Grade Reading Assessment (EGRA), a standardised literacy assessment tool designed for low-resource contexts (Dubeck & Gove, 2015). Subtasks were administered in the language of instruction. In the listening comprehension subtask, an evaluator read a passage aloud and asked three open-ended questions about it. Listening comprehension measures oral proficiency: how much of the language of instruction a pupil understands. In the reading comprehension subtask, the pupil read a passage aloud and answered up to five open-ended questions about it. Both subtasks were scored as the percentage of questions answered correctly.

Assessments were administered one-on-one at baseline and endline. Evaluators completed three to five days of training before each data collection round, including field practice and inter-rater reliability testing; reliability was 98% for the Liberia 2024—2025 cohort[2] (The Khana Group, 2025). Data were recorded in Tangerine, a mobile EGRA data collection platform designed for low-connectivity settings. The same subtask scripts and administration protocols were used across countries, and the same subtask items were used within countries, enabling both cross-cohort and within-pupil comparisons.

2.2.2. Language variables

Pupil language match indicates whether a pupil's home language is the same as the language of instruction: Asante Twi in Ghana[3] or English in Liberia. Pupils self-reported their home language during the baseline interview. Language names were standardised following Ethnologue (Eberhard et al., 2026). Across the three cohorts, 53% of pupils were classified as matched, with a substantial country difference (39% in Ghana, 72% in Liberia). Throughout, we refer to the difference in reading outcomes between matched and mismatched pupils as the mismatch gap.

Language peer density captures the classroom's linguistic environment and is defined for each pupil as the proportion of classmates whose home language matches their own. The measure was computed using Luminos programme records. Language peer density averaged 62% across the sample (standard deviation 40%), with substantial variation by language match status. The measure carries opposite linguistic implications across the pupil language match variable: for matched pupils, same-language peers speak the language of instruction; for mismatched pupils, they do not.

[2]Inter-rater reliability was not reported externally for Ghana 2023—2024 and 2024—2025 cohorts.

[3]While Akwapem Twi and Fante are closely related to Asante Twi (Eberhard et al., 2026), only one pupil reported speaking Fante at home, and no pupils reported speaking Akwapem Twi at home. Broadening the definition of a language match to include these two languages would thus not change the results of the analysis.

2.2.3. Individual-level covariates

The analysis controlled for six individual-level features. Sex (female or male) was recorded by evaluators. Age was self-reported and is likely noisy given low birth registration rates in both countries (World Bank, 2022). Prior schooling, an indicator of whether the pupil had ever attended a formal school, was also self-reported. The baseline value of the outcome of interest captured the pupil's initial ability.

The home literacy support index and the socioeconomic status index, both modest but reliable predictors of literacy outcomes in low- and middle-income countries (Nag et al., 2024; Zuilkowski et al., 2019), were derived from the baseline survey. The home literacy support index is the mean of three domestic literacy practice indicators: whether the pupil read at home, whether someone read to them at home, and whether someone helped them learn at home. The socioeconomic status index is the first principal component of seven household asset ownership indicators: electricity, television, radio, computer or tablet, motorcycle or scooter, bicycle, and phone.[4] We standardised the home literacy support index across the full sample (assuming these practices are comparable across settings), and the socioeconomic status index within each cohort (since the same assets reflect different relative standings across space and time). Both indices were centred at zero, with negative values indicating below-average levels.

2.2.4. Classroom-level covariates

The analysis further controlled for five classroom-level features. Teacher higher education captured whether the teacher held a post-secondary qualification. Teacher experience has three levels: none, one year, or more than one year of prior teaching experience with Luminos. Classroom size recorded the number of pupils enrolled in the classroom, and classroom baseline achievement captured the mean baseline score on the outcome of interest among the assessed pupils.

The teaching quality index used data from a standardised classroom observation tool

[4] Estimating socioeconomic status using consumer durable asset ownership as an alternative to per capita household expenditure has been popularised by (Filmer & Pritchett, 2001).

administered by Luminos partner supervisors every two to three weeks throughout the academic year. The tool covered four domains (classroom setup, lesson delivery, formative assessment and feedback, and classroom management and atmosphere) across 22 items, scored as a percentage of the maximum. The score was averaged per classroom and standardised within cohort to account for systematic variations in observer behaviours.

2.2.5. *Baseline differences between matched and mismatched pupils*

Pupils who spoke the language of instruction at home and those who did not differed systematically across almost every characteristic (see *Figure 2*). Mismatched pupils were significantly more disadvantaged at both individual and classroom levels: they had lower baseline achievement, were less likely to have attended school before, and came from poorer and less supportive households; and they were clustered in classrooms with lower baseline achievement, weaker teaching quality, and less experienced teachers.

These systematic differences informed how we estimated the mismatch gap. We adjusted for a pupil's individual background so that the gap reflected language match and not prior schooling or poverty. When the classroom itself was our subject, we further adjusted for its characteristics so that we could examine how language peer density shaped reading of matched and mismatched pupils. Existing literature routinely controls for individual background but rarely models the classroom.[5]

2.3. *Analytic methods*

We estimated the relationship between language match and reading using a linear mixed-effects model with classroom random intercepts. The full specification is:

$$\begin{aligned} Y_{i(jk)} = \beta_0 + \beta_1 \text{Cohort}_k + \beta_2 \text{Match}_{i(jk)} + \beta_3 \text{Match}_{i(jk)} \times \text{Country}_k + \gamma \mathbf{X}_{i(jk)} \\ + \delta \mathbf{C}_{j(k)} + \beta_4 \text{PeerDensity}_{i(jk)} + \beta_5 \text{Match}_{i(jk)} \times \text{PeerDensity}_{i(jk)} + v_{j(k)} + \epsilon_{i(jk)} \end{aligned} \quad (1)$$

[5]For example, only a handful of studies included in the most recent systematic review of the impact of language of instruction choices in low- and middle-income countries explicitly modelled the classroom (Nakamura et al., 2023).

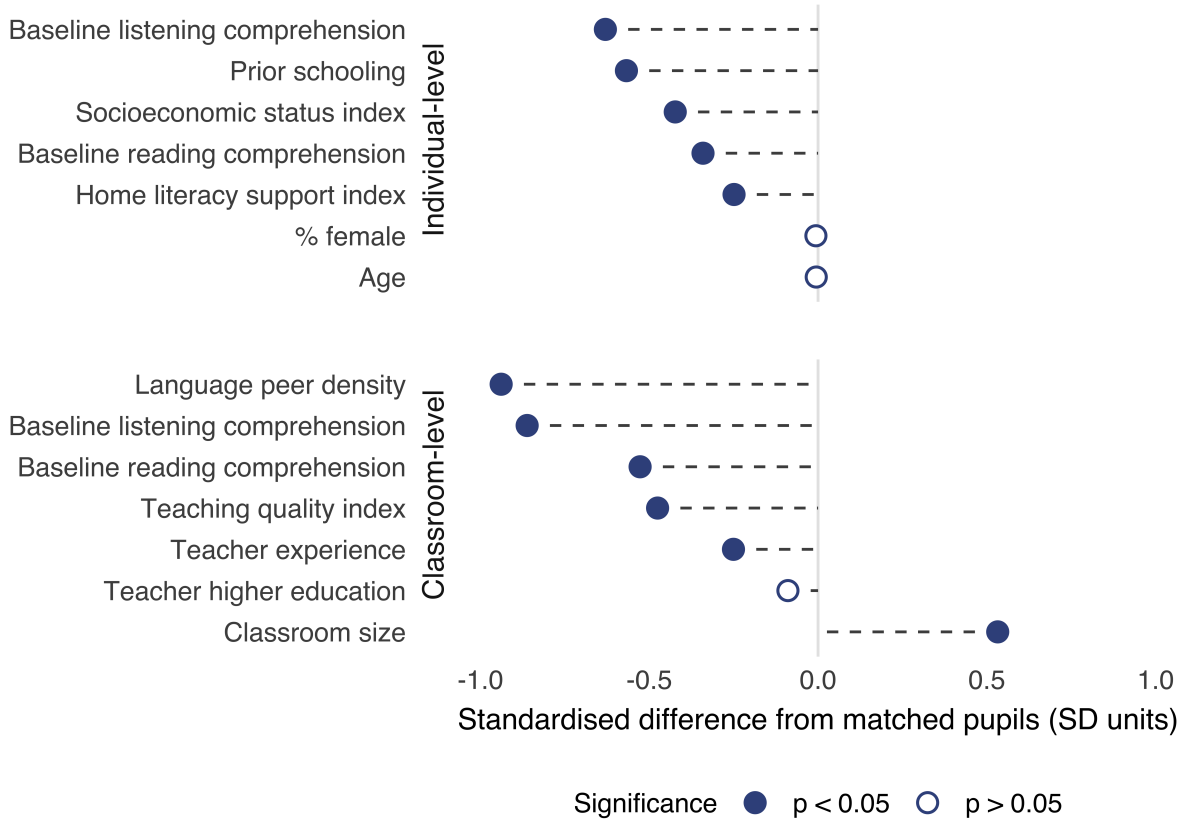


Figure 2. Mismatched pupils were disadvantaged across nearly every baseline characteristic. Each point shows the difference between mismatched and matched pupils on a baseline characteristic, with matched pupils as the reference (vertical line at zero) and differences in pooled standard deviation units. Within each panel, characteristics are ordered by the size of the difference. For most characteristics, a leftward point indicates a disadvantage for mismatched pupils (note that sex is shown for balance and has no advantageous direction). Mismatched pupils were only advantaged on a single characteristic: they were in smaller classes. Hollow points denote differences that are not statistically significant at $p < 0.05$; all others are significant. Standard errors of the underlying matched and mismatched means are reported in *Table B1*.

Here, $Y_{i(jk)}$ is the endline outcome score for pupil i in classroom j in cohort k, and Cohort_k are cohort fixed effects. $\text{Match}_{i(jk)}$ is an indicator of the pupil's language match, and the interaction allows its coefficient to vary by country. $\mathbf{X}_{i(jk)}$ and $\mathbf{C}_{j(k)}$ are the individual- and classroom-level covariates described previously. $\text{PeerDensity}_{i(jk)}$ is the proportion of same-language peers in the classroom, and the interaction allows this relationship to vary by the pupil's language match status. $v_{j(k)}$ is the classroom random intercept, and $\epsilon_{i(jk)}$ is the individual residual.

We report two forms of this specification. The individual model omits the classroom covariates and the language peer density terms. It isolates how a pupil's own characteristics, including language match, predict reading, and underlies the first part of the

Results. The full model adds the classroom covariates and the language peer density term; we use it whenever the classroom itself is the subject of inquiry, in the second and third parts of the Results. Our primary predictor of reading comprehension is language match, which we then replace with listening comprehension. All other models are variations of this specification. Where variables enter interactions, we combine the relevant coefficients and report the resulting associations (simple slopes) for each group of pupils; all models are reported in *Appendix C*.

We chose a mixed-effects model because pupils are nested within classrooms and share observed and unobserved features that violate the independence assumption of standard linear regression. A single-level model would understate standard errors and overstate statistical significance (Luke, 2017). The mixed-effects specification partitions the residual variance between individual-level and classroom-level components to produce standard errors that reflect the hierarchical structure of the data. This adjustment is particularly relevant because pupils with and without language match are unevenly distributed across classrooms.

To assess the robustness of the findings, we conducted three further checks of the primary specification (see *Appendix D*). First, we ran standard model diagnostics, assessing residual normality, variance homogeneity, and random effect normality. Second, we recomputed standard errors using cluster-robust variance estimation at the classroom level (Pustejovsky & Tipton, 2018). Third, we reestimated the model using a classroom fixed-effects specification to relax the assumption that individual-level covariates are not correlated with the unobserved classroom effect (Wooldridge, 2010, Chapter 10). However, with just over 5 pupils sampled per classroom, classroom intercepts were imprecisely estimated, and we thus retained the mixed-effects model as our primary specification.[6]

All analyses were performed in R 4.6.0 (R Core Team, 2026) with the *lme4* (Bates et al.,

[6]We favour the mixed-effects model for two further reasons. First, a fixed-effects approach absorbs all classroom-level variation, preventing the estimation of environmental predictors, such as language peer density. Second, with 327 classrooms in the sample, estimating fixed effects would consume a substantial number of degrees of freedom, resulting in inflated standard errors for individual-level covariates.

2015) and *lmerTest* (Kuznetsova et al., 2017) packages, using the Satterthwaite method to approximate degrees of freedom and with statistical significance assessed at $p < 0.05$. Simple slopes and probability estimates were computed with the *emmeans* package (Lenth & Piaskowski, 2026). These are observational cohorts, not an experiment; throughout, we report associations, and we test the most plausible competing explanations for each.

3. Results

In multilingual settings, mismatched pupils read worse not because of the language they speak at home, but because of how much of the language they understand and develop at school. We make this argument in three steps. First, the mismatch gap depends on the setting; how much of the language of instruction pupils understand predicts their reading in both countries. Second, how well they read varies with the classroom: surrounded by classmates who share their language, mismatched pupils read worse. Third, the understanding they lack can be built: most mismatched pupils improve their oral proficiency within a single programme cycle, and whether they do so varies with the classroom's language mix. Throughout, we report focal parameters in the text and the full model estimates in *Appendix C*, *Table C1* to *Table C9*.

3.1. Mismatched pupils read worse, and oral proficiency is the divide

3.1.1. Mismatched pupils read worse, but only in some settings

Mismatched pupils fell far behind their matched peers in reading. Across both countries, the mismatch gap—the difference in reading comprehension between pupils who spoke the language of instruction and those who didn't—widened from 4.5 percentage points (p.p.) at baseline to 21.4 p.p. at endline. The disparity in average scores was largely a disparity in zero scores: at the end of the programme, 45% of mismatched pupils could not answer a single reading comprehension question, compared to 18% of matched pupils. For many children, then, language mismatch was not a continuous impediment to reading acquisition but rather a barrier to learning to read at all.

The mismatch gap, however, was concentrated in one country. Adjusting for individual background characteristics, matched pupils scored 12.5 p.p. higher than mismatched

pupils in Ghana ($p < 0.001$), but only 2.9 p.p. higher in Liberia, where the difference was statistically indistinguishable from zero (see *Table C1*). *Figure 3* illustrates estimated endline reading comprehension by language match in each country: a wide gap in Ghana, and no gap in Liberia.

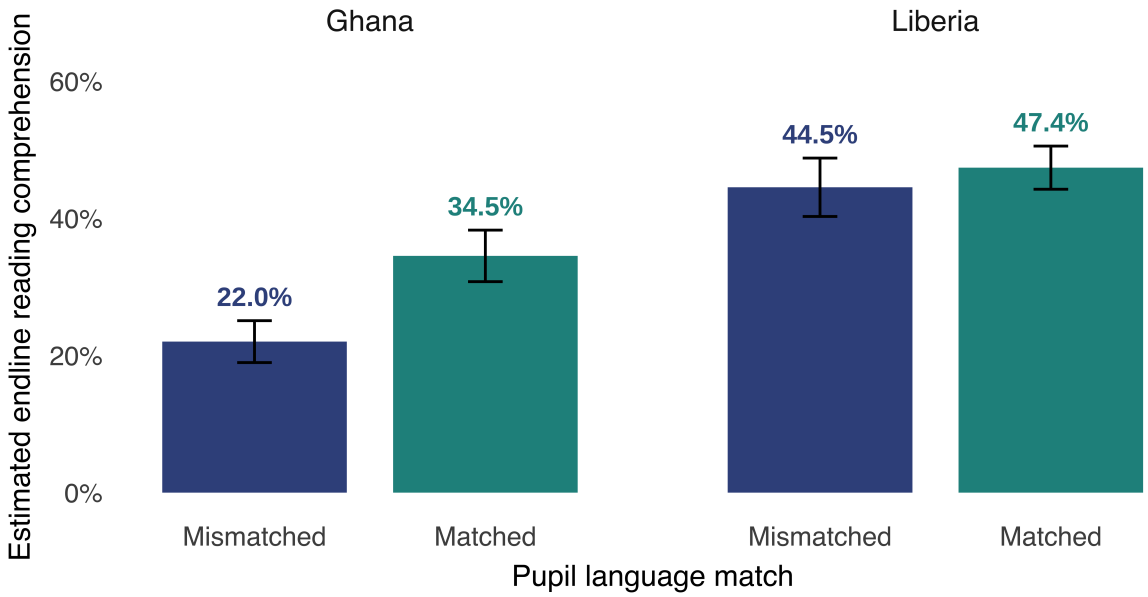


Figure 3. The mismatch gap was wide in Ghana and statistically indistinguishable from zero in Liberia. Estimated endline reading comprehension scores by pupil language match and country, adjusted for individual background characteristics and averaged over the analytic sample, with 95% confidence intervals.

This Ghana–Liberia asymmetry is consistent with the linguistic contexts of the two countries. In Ghana, Asante Twi is a regional language that mismatched pupils may rarely encounter outside the classroom. In Liberia, English serves as a lingua franca, so mismatched pupils may hear and speak the language of instruction daily. The gap thus appears where the language of instruction is confined to the classroom, and vanishes where it is also heard outside school. This is the first indication that language understanding, rather than match, underlies the mismatch gap.

Neither attendance nor the disadvantages that often accompany migration explain the mismatch gap. Mismatched pupils attended slightly less (by 2.1 p.p. averaged across the sample and after adjusting for individual background characteristics, $p < 0.01$; see *Table C2*), and attendance predicted reading comprehension in its own right: each additional 10 p.p. of attendance was associated with 2.0 p.p. higher endline reading

comprehension ($p < 0.001$; see *Table C3*). Yet attendance accounted for less than a percentage point of the gap: adjusting for it moved the Ghana mismatch estimate from 13.0 p.p. to 12.3 p.p. (both $p < 0.001$). The model also adjusted for migration-related disadvantages (UNESCO, 2018): socioeconomic status, prior schooling, and home literacy support. We could not, however, observe migrant status itself.

3.1.2. Language match is a poor proxy for oral proficiency

The binary language match concealed wide variation in oral proficiency. Matched pupils were more proficient on average, scoring 60% on baseline listening comprehension compared to 38% for mismatched pupils. However, oral proficiency varied widely within the two groups: some mismatched pupils could answer all listening comprehension questions, and some matched pupils could answer none. If the mismatch gap reflects oral proficiency rather than the home language label, this within-group variation should predict reading more precisely than the binary match.

The match also concealed a difference between the two countries. In Ghana, as *Figure 4* shows, mismatched pupils clustered at the bottom of the listening comprehension scale, while matched pupils were concentrated in the upper ranges. In Liberia, the two groups looked far more alike: both were spread across the range, with matched pupils only somewhat more concentrated at the top. The binary match thus stood for a wide oral proficiency difference in Ghana and a much narrower one in Liberia. If the mismatch gap reflects oral proficiency, then measuring it directly should dissolve the country difference.

Measured directly, oral proficiency predicted reading the same way in both countries. We reestimated the model with baseline listening comprehension in place of language match: 10 p.p. higher baseline listening comprehension was associated with 1.1 p.p. higher endline reading comprehension ($p < 0.001$; see *Table C4*). This association did not differ between countries: the interaction between listening comprehension and country was not significant. The Ghana–Liberia difference the binary match implied was largely a difference in oral proficiency; it did not reflect differences in teaching or programme implementation.

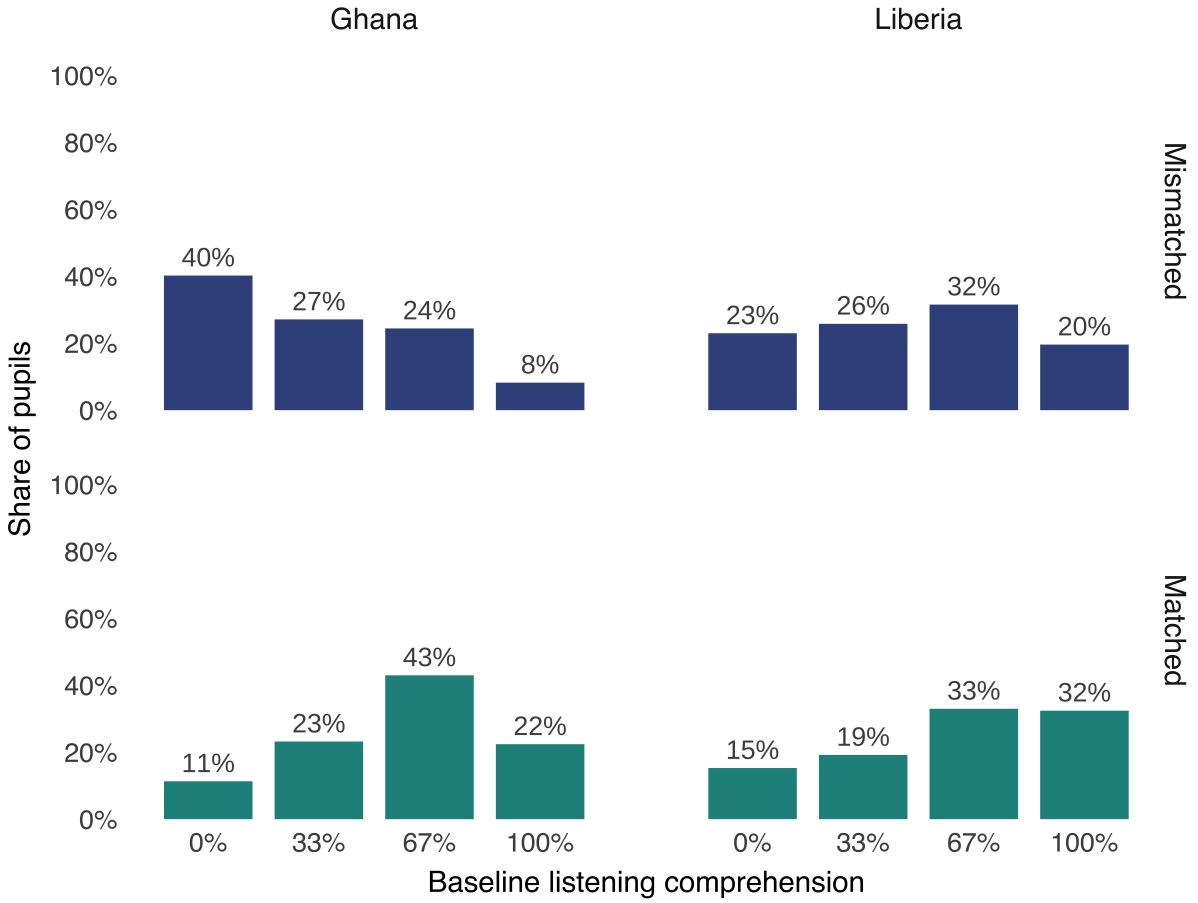


Figure 4. The language match captured a wide oral proficiency difference in Ghana but a much smaller one in Liberia. Share of mismatched and matched pupils at each level of baseline listening comprehension, by country.

Oral proficiency thus captures directly what the binary match could only approximate. As *Figure 5* illustrates, the match assigns every pupil one of two estimated scores; oral proficiency differentiates them across its full range, so that a mismatched pupil who understood much of the language at baseline is estimated to read better than a matched pupil who understood little. The mismatch gap, then, depends on the setting; the association between understanding the language spoken at school and reading in it does not.

Entered together, neither measure subsumed the other. We modelled endline reading comprehension on language match and baseline listening comprehension (see *Table C5*). Oral proficiency again predicted reading the same way in both countries, with the interaction between listening comprehension and Liberia close to zero. The language match, by contrast, retained its association in Ghana, where matched pupils scored

11.5 p.p. higher than mismatched pupils ($p < 0.001$). The two measures are only modestly correlated ($r = 0.31$), and language match may capture something beyond oral proficiency, such as migration or other disadvantages that come with speaking a minority language. But because language match predicted reading in some settings only, it cannot be a proximal cause of reading; from here on, we model the two measures separately.

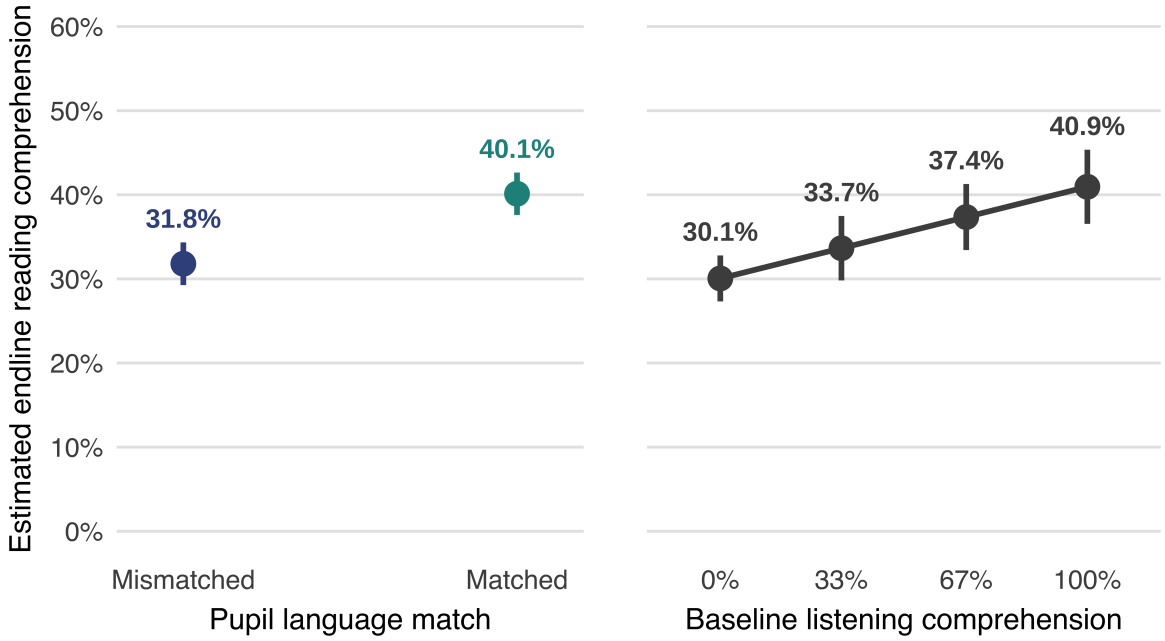


Figure 5. The binary language match assigns every pupil one of two estimated scores; oral proficiency differentiates them more precisely. Estimated endline reading comprehension from two otherwise identical individual models: with the binary language match (left), and with baseline listening comprehension (right), averaged over the analytic sample, with 95% confidence intervals.

3.2. Reading depends on the classroom, not just the pupil

Pupils hear the language of instruction from the teacher, but whether they hear it from anyone else depends on their classmates. For a mismatched pupil, a classroom where most children speak their home language offers few opportunities to practise the language of instruction; for a matched pupil, it offers many. On average, a mismatched pupil shared their home language with 42% of classmates—nearly half the classroom they could talk to without using the language of instruction. How much of the language a pupil understands predicts their reading; how much of it they hear in the classroom should predict it, too.

The classroom's language mix predicted reading in opposite directions for matched and mismatched pupils. For mismatched pupils, reading comprehension was 8.1 p.p. lower when all their classmates spoke their home language than when none did ($p < 0.01$; see *Table C1*). For matched pupils, reading comprehension was 10.2 p.p. higher when all their classmates spoke their home language than when none did ($p < 0.01$). As *Figure 6* illustrates, the more classmates spoke a pupil's home language, the worse mismatched pupils read—and the better matched pupils read.

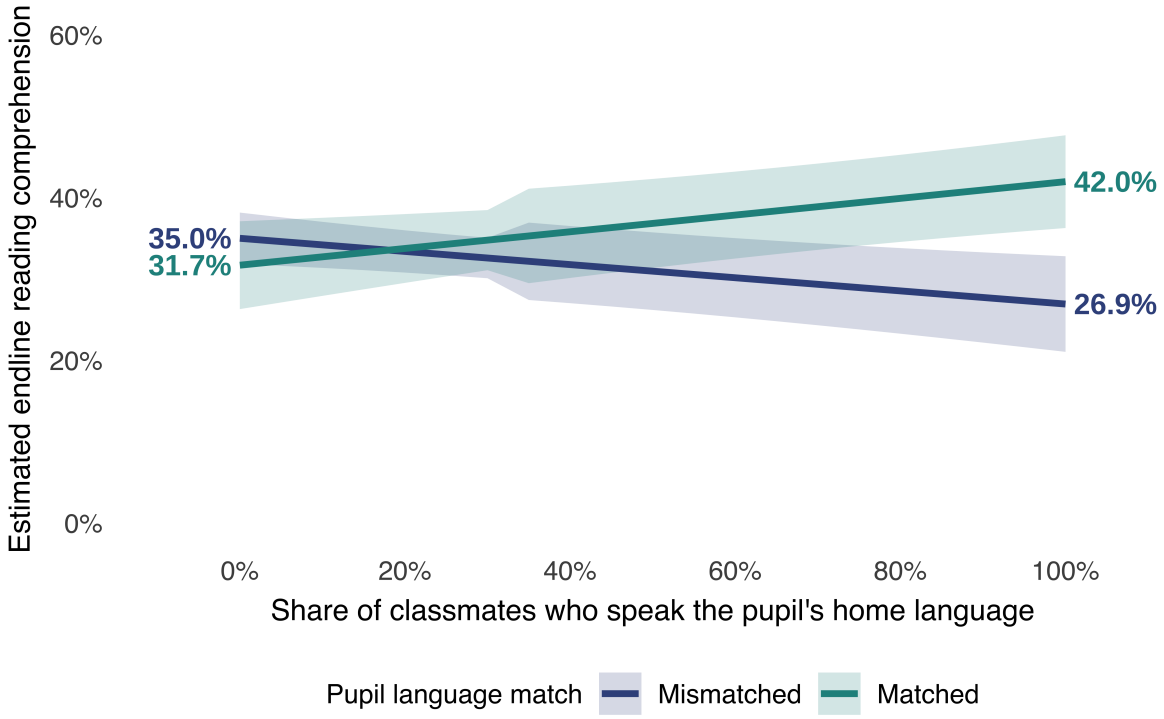


Figure 6. The classroom's language mix predicted reading in opposite directions for mismatched and matched pupils. As more classmates spoke a pupil's home language, estimated reading comprehension fell for mismatched pupils and rose for matched pupils. Estimated endline reading comprehension scores by pupil language match and the share of classmates who speak the pupil's home language, with 95% confidence intervals.

Exposure to the language of instruction is the simplest explanation for this pattern: for a mismatched pupil, a classroom full of same-language classmates is a classroom with fewer opportunities to hear and speak the language of instruction. Other explanations might produce the same association, but those we could test did not account for it. Teachers in classrooms where few children understand the language of instruction might slow the pace or depart from the intended lesson; but the model adjusted for teaching quality, and the pattern held. Teachers in such classrooms may be more likely to share

the pupils' home language—and, perhaps, to lean on it for instruction; but adjusting for the match between the teacher's home language and the language of instruction left the pattern unchanged (see *Table C6*). Mismatched pupils with many same-language classmates may also hear less of the language of instruction outside school; we could not test this explanation. We also cannot rule out that classrooms of different language compositions differ in ways we could not measure—but the classroom characteristics we could control for left the exposure account intact.

3.3. Oral proficiency can be built—and the classroom predicts whether it is

3.3.1. Most mismatched pupils improve their oral proficiency within one programme cycle

Oral proficiency is not a fixed trait but a skill that children can build, and most mismatched pupils did so within a single programme cycle. Among the mismatched pupils who began the programme below full oral proficiency, 58% improved by endline, and those who did gained 53 p.p. on average. Those who started with the least gained the most: as *Figure 7* shows, among pupils who began with no oral proficiency at all, 66% improved, and those who did gained 64 p.p. on average. Oral proficiency, then, was malleable across the range—including the pupils who started at zero.

Improvement, however, was not universal: 42% of mismatched pupils understood the language of instruction no better at the end of the programme than at the beginning. The classroom's language mix predicted how well mismatched pupils read; the same mix may explain why some improved their oral proficiency and others didn't.

3.3.2. The classroom predicts whether mismatched pupils improve their oral proficiency

The classrooms of mismatched pupils varied enormously in their language mix. On average, 42% of a mismatched pupil's classmates shared their home language. But few pupils were 'average': a quarter of them had almost no classmates who shared their language, while another quarter were surrounded by them almost entirely.[7] Mismatched

[7] Among mismatched pupils, the share of classmates who spoke their home language was 5% or lower for the bottom quarter, and 83% or higher for the top quarter—roughly one or no classmates, and 22 or more, in an average classroom of around 27 pupils.

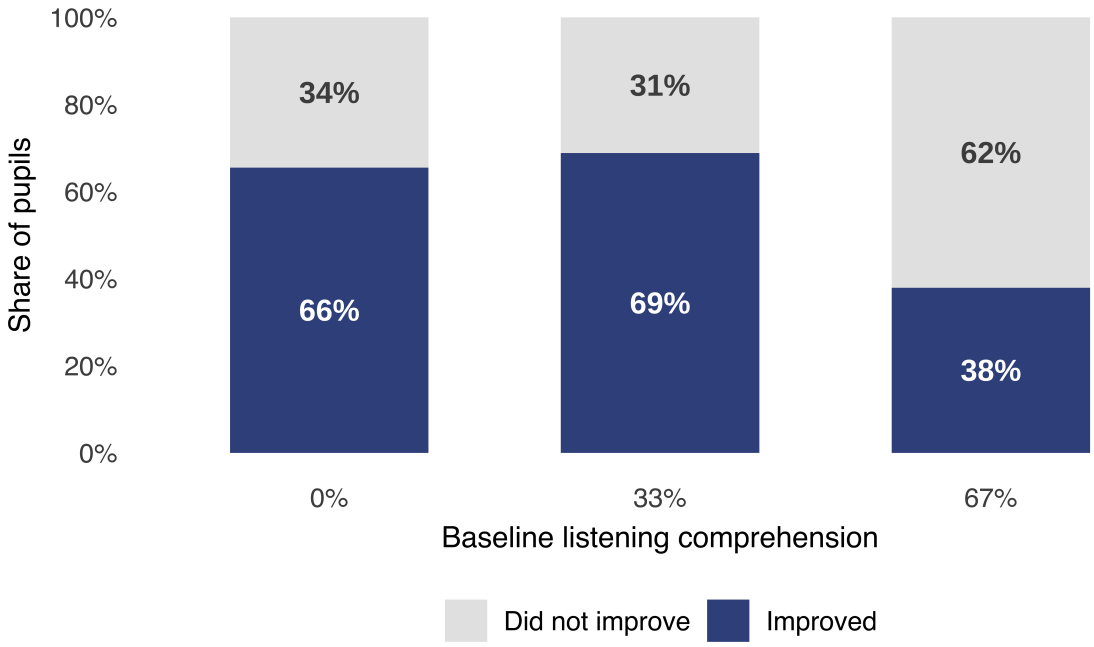


Figure 7. Most mismatched pupils improved their oral proficiency, including those who started with none. Share of mismatched pupils who improved their oral proficiency between baseline and endline, by baseline level, among those who began below full oral proficiency.

pupils therefore differed sharply in how much of the language of instruction they heard from their classmates. The classroom's language mix predicts how well mismatched pupils read. If it works through exposure, it should also predict how much of the language of instruction they learn to understand.

The classroom's language mix predicted whether mismatched pupils improved their oral proficiency, but only for those who began with none. Among mismatched pupils who began the programme understanding none of the language of instruction, the chance of improving their oral proficiency was 28 p.p. lower when all their classmates spoke their home language than when none did ($p < 0.001$; see *Table C7*).[8] Among those who understood some of it, the classroom's language mix made no difference. *Figure 8* plots the estimated probability of improvement at three baseline levels of listening comprehension: none, one-third, and two-thirds. For pupils who started with none, the line falls steeply: the more classmates shared their home language, the less likely they were to improve. For pupils who started with some, the lines are nearly flat: the

[8]These results are robust to a logistic mixed-effects specification (see *Table C8*).

classroom's language mix no longer made a difference.

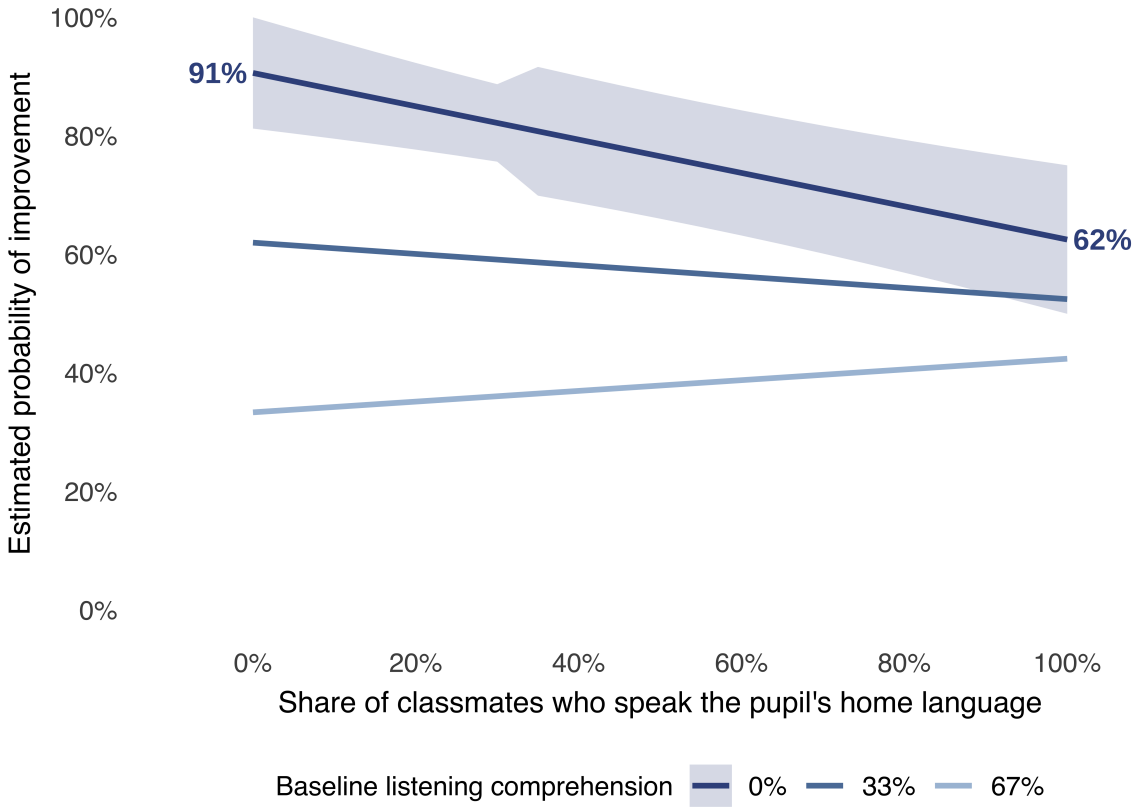


Figure 8. The classroom's language mix predicted whether mismatched pupils improved their oral proficiency, but only for those who started with none. Estimated probability of improving oral proficiency across the share of classmates who spoke the pupil's home language, at three baseline levels (0%, 33%, and 67% of the listening comprehension questions answered correctly), for mismatched pupils who began below full oral proficiency. The association is significant only for pupils starting at zero oral proficiency ($p < 0.001$). Shaded bands show 95% confidence intervals (omitted for non-significant slopes to aid legibility).

3.3.3. Mismatched pupils who improve their oral proficiency read better

Without oral proficiency, reading was out of reach. Among mismatched pupils who started with no oral proficiency, 91% of those who remained at zero by the end of the programme could not answer a single reading comprehension question either. *Figure 9* illustrates the contrast: mismatched pupils who improved their oral proficiency read across the scale, while those who did not were concentrated at zero.

Since reading depends on oral proficiency, a pupil who improves it should read better, too. Across all mismatched pupils who had not yet mastered the language of instruction, those who improved their listening comprehension scored 10.5 p.p. higher on endline

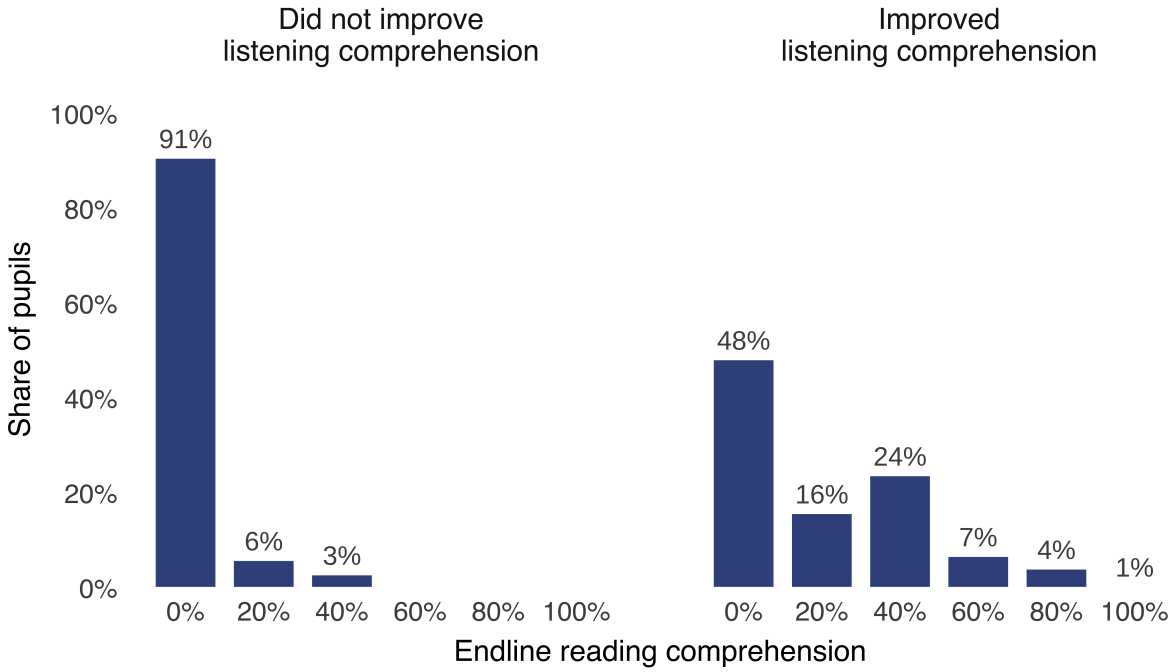


Figure 9. Almost no pupil read with comprehension without oral proficiency. Distribution of endline reading comprehension scores for mismatched pupils who began the programme with zero listening comprehension and either improved or did not.

reading comprehension than those who did not, adjusting for individual characteristics ($p < 0.001$; see *Table C9*). The pupils who developed their oral proficiency, then, not only understood more—they also read better.

4. Discussion

In this study, we argued that reading depends not on the language children speak at home, but on how much of the language they understand and develop at school. We draw four implications for policy. First, reading requires language understanding, and language match is a poor proxy for that. Second, home language policy, though attractive, cannot reach every child. Third, to reach all children, the language of instruction should be taught. Fourth, no language policy can help where teaching produces no learning.

4.1. Language match is a poor measure of understanding

Reading requires understanding the language of the text. Under the Simple View of Reading, reading comprehension is the product of two components: decoding the words on the page, and knowing what they mean (Gough & Tunmer, 1986; Hoover & Gough,

1990). How much of the language of instruction a child understands therefore shapes how well they read. Research and policy capture this through the language match, sorting children into those who speak the language of the school at home and those who do not. Yet the match is only a crude proxy for understanding. This, we argued, might explain why the evidence on home language instruction has been mixed.

The language match can mean different things in different settings. Defined the same way across our sample, it produced a wide reading gap in Ghana and no gap in Liberia. The linguistic contexts of the two countries explain the divergence. In Ghana's Ashanti region, where Luminos classrooms are located, a mismatched pupil, who does not speak the language of instruction at home, may hear Asante Twi in the classroom and nowhere else. In Liberia, English is the lingua franca, and a mismatched pupil may hear and speak it daily. The label 'mismatched', then, describes a child with daily exposure to the language of instruction in one setting, and a child with almost none in the other. Measured directly, oral proficiency resolved the contradiction: a 10 p.p. increase in baseline listening comprehension was associated with 1.1 p.p. higher endline reading comprehension, in both countries. Research and policy should therefore measure how much of the language of instruction a child understands, rather than which language they speak at home.

4.2. Home language instruction cannot reach every child

Home language instruction is the most direct route to understanding, yet no set of languages can match every child in a diverse classroom. As a thought experiment, we used Luminos programme enrolment records in Ashanti to simulate how many pupils home language instruction could reach if languages were added to the curriculum one at a time (see *Appendix E*). At each step, we added the next most widely spoken language, and let every classroom either keep its current language of instruction or switch to the new one. Starting from Asante Twi, which matched 31%[9] of pupils, Konkomba was the most valuable addition, raising the match to 55%. Fifteen languages would be needed to

[9]This differs from the match rate reported in the *Results* because the simulation uses a larger sample: all enrolled pupils across four academic years, rather than the externally assessed pupils across two academic years.

reach 73%, beyond which linguistic diversity within classrooms allows no further gains.

Literacy is a challenge even in high-income countries that teach in a single language (Seidenberg, 2017). In low- and middle-income countries, teaching in multiple languages, each requiring its own trained workforce and effective curriculum (Trudell et al., 2023), is infeasible. Partly for that reason, the language of instruction is often set regionally, not by classroom (Anyidoho, 2018). Under that constraint, the optimal choice in Ashanti, Asante Twi, matched only a third of pupils. Across Ghana, only 46% of primary schools had a close match between the language of the teacher, the language of the pupils, and the mandated language of instruction (Hatch et al., 2018). Even under the most generous assumptions, then, universal home language instruction might be both theoretically impossible and economically unachievable.

Where home language instruction is possible, it is often not delivered (Trudell, 2016). In Ghana, where eleven home languages are mandated (Appiah & Ardila, 2020), the actual language of the classroom is usually English, and neither pupils nor teachers are equipped for it. In Grades 1 to 3, 84% of pupils reported that their teachers mostly used English (UNICEF & Ghana Statistical Service, 2018), a language spoken well by a small, educated elite (Anyidoho, 2018). Every child in an English-speaking classroom is mismatched, because almost no Ghanaian child speaks English at home. Every teacher is unsupported, because the curriculum, materials, and training were developed for the Ghanaian languages. The result is a paradox: in a country with a home language policy, most children are taught to read in a language they do not speak. By the end of Grade 2, most cannot read with comprehension either (Ghana Education Service et al., 2016).

4.3. Oral proficiency should be taught

Where a language match cannot be delivered, our work suggests that children can acquire the language of instruction quickly enough to learn to read in it. In the Luminos programme, most mismatched pupils who began below full oral proficiency improved within a single programme cycle, and those who began with no proficiency in the language of instruction improved the most. These gains translated into reading outcomes: pupils who improved their oral proficiency scored 10.5 p.p. higher on endline

reading comprehension than those who did not. A language mismatch therefore does not have to be resolved before a child can begin to read.

Teaching oral proficiency explicitly, as part of a structured approach to literacy, may be the most direct way to build it. This is not a new demand—the World Bank already advocates instruction in oral language as a foundation for literacy (Crawford et al., 2022), and programmes such as Luminos already provide it—but the recommendation has drawn far less attention than the choice of language of instruction. We argue that the balance is off: most pupils in Sub-Saharan Africa are taught in a language they do not speak at home (World Bank, 2021), and across much of the region a full match is neither theoretically nor practically possible. Teaching the language of instruction is therefore the more urgent priority for most children. Learning the language is inseparable from learning to read in it, and oral language instruction belongs at the centre of the literacy curriculum, not at its margins.

Beyond the curriculum, oral proficiency likely develops through everyday talk between classmates. A mismatched pupil with no same-language classmates may have little choice but to communicate and play in the language of instruction, whereas a mismatched pupil surrounded by same-language classmates barely needs it. Our estimates are consistent with this: mismatched pupils with more same-language classmates were less likely to develop oral proficiency, and read worse. These are, to our knowledge, the first such estimates in low- and middle-income countries, and they align with studies of minority-language pupils in high-income countries (Schmerse, 2021; Seuring et al., 2020). Peer exposure of this nature is incidental, but education systems need not leave it to chance. Classroom composition is one route, where school size and the local language mix allow it. Classroom arrangement is another: grouping, pairing, and play may shape how much of the language a child uses.

4.4. Better teaching matters more than the right language

A language policy matters only where teaching produces learning. *Loud and Clear* ranks mother tongue instruction as the cheapest route to learning, and better teaching as the most expensive (World Bank, 2021), which invites decision-makers to treat the two

as alternatives. They are not: a language policy changes how much of the teaching a child understands, but where teaching produces no learning, it has nothing to change. In Ghana's Ashanti region, 77% of Grade 2 pupils in government schools spoke the mandated language of instruction at home. Yet pupils in Asante Twi schools read just 2.9 words per minute and scored 5.2% on reading comprehension on average (Ghana Education Service et al., 2016). The Luminos programme in Ashanti served some of the most disadvantaged children in the region: most did not speak the language of instruction at home, and all were taught by community members. After a single school year, they were reading 24.9 words per minute[10] and scoring 28.5% on reading comprehension. A high language match did not produce reading, and a low one did not prevent it. Where few children are learning, teaching quality may matter more than the language of instruction.

4.5. Limitations and future research

These are observational cohorts, and we report associations, not causal effects. The competing explanations we could test did not account for the patterns we observed, but unmeasured differences between pupils and classrooms remain possible. Several measures are also imprecise. Listening comprehension is based on three items and reading comprehension on up to five, as is standard for EGRA instruments (RTI International, 2016), and a number of variables are self-reported. When a predictor is noisily measured, classical measurement error attenuates its coefficient towards zero; when an outcome is noisily measured, it widens standard errors without biasing the coefficients (Wooldridge, 2013, Chapter 9). These sources of error make our findings more likely to be understated, but not more likely to be spurious.

Our sample both enables our analysis and limits how far it generalises. These are children who learned to read, unlike most of their peers in government classrooms, which gives the analysis something to explain. These are also formerly out-of-school children in a high-fidelity programme, so the estimates apply most directly to similar

[10]For comparison, pupils across 47 low- and middle-income countries read 22 words per minute on average in their third year of reading instruction (Crawford et al., 2025). Luminos pupils in Ashanti reached this after nine months, with just over two hours of literacy instruction a day.

programmes. The mechanisms we discuss, however, are not specific to this programme, since reading requires understanding the language of the text in any context.

Three questions for low- and middle-income countries follow from this work. The first is whether explicit oral language instruction improves reading, which could be tested by randomising an oral language component within an existing literacy programme. The second is how language peer effects operate, and how a classroom can be arranged so that these effects help mismatched children rather than hold them back. Descriptive studies should come first, since we know too little about these effects to design an experiment. The third is whether our findings hold in government schools, where teaching might be weaker and learning gains smaller—the study the policy debate most needs.

4.6. Conclusion

Home language policies are guided by a worthy goal: that every child learns to read in the language they speak at home. Yet no set of languages can reach every child: across much of Sub-Saharan Africa a full match is all but impossible, and even modest policies rarely deliver on their promises. In this study, we showed that a home language match predicts reading in one setting but not the other; understanding the language of the school predicts it in both. A child's home language is fixed, but their language understanding is not—and it can be built within a single school year. Language of instruction matters only where teaching produces learning, and where it does, children can be taught to understand it. With that, they can learn to read, whatever language they speak at home.

References

Anyidoho, A. (2018). Shifting sands: Language policies in Education in Ghana and implementation challenges. *Ghana Journal of Linguistics*, *7*(2), 225–243. https://doi.org/10.4314/gjl.v7i2.10

Appiah, S. O., & Ardila, A. (2020). The question of school language in multilingual societies: The example of Ghana. *RUDN Journal of Psychology and Pedagogics*, *17*(2), 263–272. https://doi.org/10.22363/2313-1683-2020-17-2-263-272

Baron, C. (Ed.). (2025). *The Luminos Method: The secret to unlocking the light of learning in every child*. Marigold Ink Press.

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brunette, T., Piper, B., King, S., & Nabacwa, R. (2019). The impact of mother tongue reading instruction in twelve Ugandan languages and the role of language complexity, socioeconomic factors, and program implementation. *Comparative Education Review*, *63*(4). https://doi.org/10.1086/705426

Conn, K. M. (2017). Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research*, *87*(5), 863–898. https://doi.org/10.3102/0034654317712025

Crawford, M., Oviedo Buitrago, M. E., Faquir, N. F., & Raheel, N. (2022). *Handbook for literacy lesson planning: Early-grade classroom activities for teaching children to read*. World Bank. Washington, D.C. Retrieved July 20, 2026, from https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099010004212217887

Crawford, M., Raheel, N., Korochkina, M., & Rastle, K. (2025). Inadequate foundational decoding skills constrain global literacy goals for pupils in low- and middle-income countries. *Nature Human Behaviour*, *9*(1), 74–83. https://doi.org/10.1038/s41562-024-02028-x

Dubeck, M. M., & Gove, A. (2015). The Early Grade Reading Assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, *40*, 315–322. https://doi.org/10.1016/j.ijedudev.2014.11.004

Eberhard, D. M., Simons, G. F., & Robinson, A. J. (Eds.). (2026). *Ethnologue: Languages of the world* (29th ed.). SIL Global. https://www.ethnologue.com

Evans, D. K., & Mendez Acosta, A. (2021). Education in Africa: What are we learning? *Journal of African Economies*, *30*(1), 13–54. https://doi.org/10.1093/jae/ejaa009

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography*, *38*(1), 115–132. https://doi.org/10.1353/dem.2001.0003

Ghana Education Service, RTI International & Education Assessment and Research Centre. (2016). *Ghana 2015 Early Grade Reading Assessment and Early Grade Mathematics Assessment: Report of findings*.

Ghana Statistical Service. (2023). *Ghana 2021 population and housing census: Thematic report on migration*. https://statsghana.gov.gh/gssmain/fileUpload/pressrelease/Thematic%20Report%20on%20Migration_09032022.pdf

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology: General*, *150*(4), 700–709. https://doi.org/10.1037/xge0000920

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Hatch, R., Cao, Y., Omoeva, C., Frazier, J., Moussa, W., Atintono, S. A., Nsoh, A. E., & Owu-Ewie, C. (2018). *Language mapping study: Analysis report*. USAID Partnership for Education: Learning.

Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, *30*, 100323. https://doi.org/10.1016/j.edurev.2020.100323

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/bf00401799

Kim, Y. S. G., Lee, H., & Zuilkowski, S. S. (2020). Impact of literacy interventions on reading skills in low- and middle-income countries: A meta-analysis. *Child Development*, *91*(2), 638–660. https://doi.org/10.1111/cdev.13204

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Lenth, R. V., & Piaskowski, J. (2026). *Emmeans: Estimated marginal means, aka least-squares means*. https://rvlenth.github.io/emmeans/

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

McManus, J., Rudasingwa, M., Nijhof, E., Mokobi, K., Deme, S. F., Kiawoin, J., Garay, F. A., Mwai, L., & Pignon, C. (2025). Improving learning outcomes for out-of-school children: Evidence from a randomized evaluation of an accelerated learning program in Liberia. *Education Economics*, *33*(1), 19–38. https://doi.org/10.1080/09645292.2024.2410773

Nag, S., Vagh, S. B., Dulay, K. M., Snowling, M., Donolato, E., & Melby-Lervåg, M. (2024). Home learning environments and children's language and literacy skills: A meta-analytic review of studies conducted in low- and middle-income countries. *Psychological Bulletin*, *150*(2), 132–153. https://doi.org/10.1037/bul0000417

Nakamura, P., Molotsky, A., Zarzur, R. C., Ranjit, V., Haddad, Y., & Hoop, T. D. (2023). Language of instruction in schools in low- and middle-income countries: A systematic review. *Campbell Systematic Reviews*, *19*(4), e1351. https://doi.org/10.1002/cl2.1351

Piper, B., Zuilkowski, S. S., & Ong'Ele, S. (2016). Implementing mother tongue instruction in the real world: Results from a medium-scale randomized controlled trial in Kenya. *Comparative Education Review*, *60*(4), 776–807. https://doi.org/10.1086/688493

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

Quinn, S., Donnelly, S., & Kidd, E. (2018). The relationship between symbolic play and language acquisition: A meta-analytic review. *Developmental Review*, *49*, 121–135. https://doi.org/10.1016/j.dr.2018.05.005

R Core Team. (2026). *R: A language and environment for statistical computing*. Vienna, Austria. https://doi.org/10.32614/R.manuals

RTI International. (2016). *Early Grade Reading Assessment (EGRA) toolkit: Second edition*. RTI International. Washington, D.C.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In E. A. Hanushek, S. Machin & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 249–277, Vol. 3). Elsevier. https://doi.org/10.1016/B978-0-444-53429-3.00004-1

Schmerse, D. (2021). Peer effects on early language development in dual language learners. *Child Development*, *92*(5), 2153–2169. https://doi.org/10.1111/cdev.13588

Seidenberg, M. (2017). *Language at the speed of sight: How we read, why so many can't, and what can be done about it*. Basic Books.

Seuring, J., Rjosk, C., & Stanat, P. (2020). Ethnic classroom composition and minority language use among classmates: Do peers matter for students' language achievement? *European Sociological Review*, *36*(6), 920–936. https://doi.org/10.1093/esr/jcaa022

Singler, J. V. (2012). Vernacular Liberian English. In *The Mouton world atlas of variation in English* (pp. 369–381). De Gruyter Mouton.

Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geleen, M., Pastorello, M. G., & Eyers, J. (2015). *Interventions for improving learning outcomes and access to education in low- and middle-income countries: A systematic review*. International Initiative for Impact Evaluation (3ie). London, England. Retrieved June 2, 2026, from http://3ieimpact.org/evidence-hub/publications/systematic-reviews/interventions-improving-learning-outcomes-and-access

Snowling, M. J., & Hulme, C. (2025). The reading is language model: A theoretical framework for language and reading development and intervention. *Annual Review of Developmental Psychology*, *7*, 195–218. https://doi.org/10.1146/annurev-devpsych-111323-084821

Tenenbaum, H. R., Winstone, N. E., Leman, P. J., & Avery, R. E. (2020). How effective is peer interaction in facilitating learning? A meta-analysis. *Journal of Educational Psychology*, *112*(7), 1303–1319. https://doi.org/10.1037/edu0000436

The Khana Group. (2025). *2024–25 external evaluation report for The Luminos Fund in Liberia: Summary*. https://luminosfund.org/wp-content/uploads/2025/12/Luminos-Fund-Liberia-2024-25-External-Endline-Evaluation-Report.pdf

Trudell, B. (2016). Language choice and education quality in Eastern and Southern Africa: A review. *Comparative Education*, *52*(3), 281–293. https://doi.org/10.1080/03050068.2016.1185252

Trudell, B., Piper, B., & Ralaingita, W. (2023). Language of instruction in the African classroom: Key issues, challenges and solutions. In R. M. Joshi, C. A. McBride, B. Kaani & G. Elbeheri (Eds.), *Handbook of literacy in Africa* (pp. 79–102). Springer International Publishing. https://doi.org/10.1007/978-3-031-26250-0_5

UNESCO. (2016). *If you don't understand, how can you learn?* UNESCO. Paris, France. Retrieved December 7, 2025, from https://unesdoc.unesco.org/ark:/48223/pf0000243713

UNESCO. (2018). *Migration, displacement and education: Building bridges, not walls*. UNESCO. Paris, France. https://doi.org/10.54676/XDZD4287

UNESCO. (2025). *Languages matter: Global guidance on multilingual education*. Paris, France. https://doi.org/10.54675/MLIO7101

UNICEF & Ghana Statistical Service. (2018). *Ghana Multiple Indicator Cluster Survey 2017-2018*. Retrieved November 1, 2025, from https://mics.unicef.org/surveys

Wooldridge, J. M. (2010). Basic linear unobserved effects panel data models. In *Econometric analysis of cross section and panel data*. MIT Press.

Wooldridge, J. M. (2013). *Introductory econometrics: A modern approach* (5th ed.). South-Western.

World Bank. (2021). *Loud and clear: Effective language of instruction policies for learning*. World Bank. Washington, D.C.

World Bank. (2022). *Completeness of birth registration (%)*. Retrieved May 17, 2026, from https://data.worldbank.org/indicator/SP.REG.BRTH.ZS

World Bank. (2026). *School enrollment, primary (% gross)*. Retrieved May 17, 2026, from https://data.worldbank.org/indicator/SE.PRM.ENRR

World Bank, UNICEF, FCDO, USAID & Bill & Melinda Gates Foundation. (2022). *The state of global learning poverty: 2022 update*. World Bank.

Zuilkowski, S. S., McCoy, D. C., Jonason, C., & Dowd, A. J. (2019). Relationships among home literacy behaviors, materials, socioeconomic status, and early literacy outcomes across 14 low- and middle-income countries. *Journal of Cross-Cultural Psychology*, *50*(4), 539–555. https://doi.org/10.1177/0022022119837363

Appendix A. Attrition statistics

Table A1. Baseline characteristics of retained and attrited pupils. Baseline characteristics of pupils reassessed at endline (retained, $n = 1,726$) and those sampled at baseline but not reassessed (attrited, $n = 476$). A two-sample t-test was used to compare numeric variables, and Pearson's chi-squared test to compare categorical variables. Standard deviations are reported for numeric variables in parentheses, and the proportion of the most frequent category for categorical variables. Standard errors of each mean or proportion are reported in square brackets.

Variable	Retained ($n = 1,726$)	Attrited ($n = 476$)	Difference p value
Individual-level			
Language Match	Matched (53.2%) [1.2%]	Matched (55.4%) [2.3%]	0.429
Sex	Male (50.6%) [1.2%]	Male (52.7%) [2.3%]	0.436
Age (Years)	10.46 (1.74) [0.04]	10.66 (1.91) [0.09]	0.046
Prior Schooling	Yes (54.7%) [1.2%]	Yes (54.3%) [2.3%]	0.915
Home Literacy Support Index (Raw)	0.40 (0.37) [0.01]	0.35 (0.35) [0.02]	0.006
Socioeconomic Status Index (Raw)	0.01 (1.32) [0.03]	-0.05 (1.30) [0.06]	0.365
Baseline Listening Comprehension (%)	49.59 (35.54) [0.86]	48.39 (37.46) [1.72]	0.530
Baseline Reading Comprehension (%)	3.89 (13.02) [0.31]	2.86 (12.99) [0.60]	0.124
Classroom-level			
Language Peer Density (%)	61.57 (39.63) [0.95]	60.88 (40.32) [1.85]	0.738
Teacher Higher Education	No (61.1%) [1.2%]	No (65.3%) [2.2%]	0.105
Teacher Experience	None (40.6%) [1.2%]	> 1 Year (43.7%) [2.3%]	0.001
Classroom Size (Number of Pupils)	27.71 (3.72) [0.09]	28.25 (4.10) [0.19]	0.010
Teaching Quality Index (Raw)	0.87 (0.10) [0.00]	0.87 (0.10) [0.00]	0.208
Baseline Listening Comprehension (%)	49.37 (23.00) [0.55]	49.19 (23.28) [1.07]	0.877
Baseline Reading Comprehension (%)	3.96 (7.32) [0.18]	2.63 (5.46) [0.25]	< 0.001

Appendix B. Descriptive statistics

Table B1. Descriptive statistics, overall and by pupil language match ($n = 1,718$). Means for numeric variables and the most frequent category for categorical variables, for the full analytic sample and separately for matched and mismatched pupils. A two-sample t-test was used to compare numeric variables, and Pearson's chi-squared test to compare categorical variables. Standard deviations are reported for numeric variables in parentheses, and the proportion of the most frequent category for categorical variables. Standard errors of each mean or proportion are reported in square brackets.

Variable	Pooled ($n = 1,718$)	Matched ($n = 915$)	Mismatched ($n = 803$)	Difference p value
Individual-level				
Sex	Male (50.6%) [1.2%]	Male (50.5%) [1.7%]	Male (50.8%) [1.8%]	0.934
Age (Years)	10.46 (1.74) [0.04]	10.47 (1.70) [0.06]	10.46 (1.78) [0.06]	0.912
Prior Schooling	Yes (54.7%) [1.2%]	Yes (67.9%) [1.5%]	No (60.4%) [1.7%]	< 0.001
Home Literacy Support Index ($M = 0$, $SD = 1$)	0.00 (1.00) [0.02]	0.12 (0.97) [0.03]	-0.13 (1.01) [0.04]	< 0.001
Socioeconomic Status Index ($M = 0$, $SD = 1$ within cohort)	0.00 (1.00) [0.02]	0.20 (1.02) [0.03]	-0.23 (0.92) [0.03]	< 0.001
Baseline Listening Comprehension (%)	49.57 (35.53) [0.86]	60.04 (33.33) [1.10]	37.65 (34.21) [1.21]	< 0.001
Baseline Reading Comprehension (%)	3.91 (13.04) [0.31]	5.99 (16.31) [0.54]	1.54 (7.10) [0.25]	< 0.001
Classroom-level				
Language Peer Density (%)	61.73 (39.56) [0.95]	79.09 (32.09) [1.06]	41.95 (37.98) [1.34]	< 0.001
Teacher Higher Education	No (60.9%) [1.2%]	No (58.9%) [1.6%]	No (63.3%) [1.7%]	0.072
Teacher Experience	None (40.6%) [1.2%]	> 1 Year (40.3%) [1.6%]	None (45.3%) [1.8%]	< 0.001
Classroom Size (Number of Pupils)	27.70 (3.72) [0.09]	28.63 (3.16) [0.10]	26.65 (4.02) [0.14]	< 0.001
Teaching Quality Index ($M = 0$, $SD = 1$ within cohort)	0.00 (1.00) [0.02]	0.22 (0.91) [0.03]	-0.25 (1.04) [0.04]	< 0.001
Baseline Listening Comprehension (%)	49.34 (22.99) [0.55]	58.60 (18.80) [0.62]	38.79 (22.79) [0.80]	< 0.001
Baseline Reading Comprehension (%)	3.98 (7.33) [0.18]	5.78 (8.76) [0.29]	1.92 (4.45) [0.16]	< 0.001

Appendix C. Model estimates

Throughout this appendix, models are built cumulatively so that the reader can see how the coefficients of interest change as controls are introduced. *Model 1* to *Model 3* correspond to the individual model, and *Model 4* and *Model 5* correspond to the full model. Where a model number is split (e.g., *Model 3.1* and *Model 3.2*), the two share the same controls and differ only in the variable of interest, so the effect of its addition can be read directly.

C.1. Reading comprehension vs. language match and language peer density

Mismatched pupils might read worse than matched pupils, and how well they read might vary with the classroom's language mix. To test this, we modelled reading comprehension as an outcome of language match and its interaction with the classroom's language mix (see *Table C1*).

Table C1. Reading comprehension on pupil language match and language peer density ($n = 1,718$). Estimated coefficients from linear mixed-effects models on endline reading comprehension (%), with pupil language match and language peer density as the key predictors. *Model 1* includes cohort and language match; *Model 2* adds the language match × country (Liberia) interaction; *Model 3* adds individual-level covariates; *Model 4* adds classroom-level covariates; *Model 5* adds language peer density and its interaction with language match. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept (Reference Cohort)	19.812***	17.224***	0.017	-7.814	-1.842
	(2.057)	(2.064)	(4.199)	(9.831)	(9.891)
Cohort: Ghana 2024–2025	7.446*	7.078*	8.328**	8.185**	7.894**
	(2.917)	(2.802)	(2.639)	(2.545)	(2.503)
Cohort: Liberia 2024–2025	17.527***	25.000***	25.711***	26.425***	24.889***
	(2.627)	(3.100)	(2.925)	(3.782)	(3.807)
Pupil Language Match (Yes)	11.168***	18.542***	12.505***	7.251**	-1.181
	(1.682)	(2.327)	(2.229)	(2.381)	(3.302)
Pupil Language Match (Yes) × Liberia		-14.213***	-9.644**	-3.868	-4.901
		(3.326)	(3.122)	(3.229)	(3.327)
Pupil Sex (Male)			0.615	0.749	0.919
			(1.148)	(1.142)	(1.148)
Pupil Age (Years)			1.296***	1.260***	1.255***
			(0.360)	(0.357)	(0.357)
Pupil Prior Schooling (Yes)			3.105*	2.615+	2.352+
			(1.417)	(1.403)	(1.400)
Pupil Baseline Reading Comprehension (%)			0.719***	0.676***	0.679***
			(0.050)	(0.052)	(0.052)

Table C1. Reading comprehension on pupil language match and language peer density ($n = 1,718$) *(continued)*.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.998	0.783	0.634
			(0.673)	(0.667)	(0.667)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.439	0.139	-0.099
			(0.690)	(0.683)	(0.684)
Teacher Higher Education (Yes)				1.268	1.946
				(2.033)	(2.011)
Teacher Experience: 1 Year				-0.808	0.107
				(2.520)	(2.496)
Teacher Experience: > 1 Year				-7.568**	-6.237*
				(2.795)	(2.771)
Classroom Size (Number of Pupils)				0.381	0.261
				(0.354)	(0.349)
Classroom Baseline Reading Comprehension (%)				0.292+	0.138
				(0.153)	(0.157)
Teaching Quality Index (M = 0, SD = 1 within cohort)				5.465***	4.973***
				(1.026)	(1.018)
Pupil Language Peer Density (%)					-0.081**
					(0.030)
Pupil Language Match (Yes) × Pupil Language Peer Density (%)					0.184***
					(0.049)
R^2 Marginal	**0.115**	**0.136**	**0.250**	**0.312**	**0.323**
R^2 Conditional	**0.383**	**0.374**	**0.459**	**0.483**	**0.485**

C.2. Attendance vs. language match

Mismatched pupils might read worse simply because they attend less. To test this, we first modelled attendance as an outcome of language match (see *Table C2*).

The model was fitted on the smaller sample of 1,615 pupils whose external evaluation records could be linked to internal programme records. The two systems share no common pupil identifier, so records were linked by name. Because pupil names were entered separately into each system, often with slightly different spellings, they rarely matched exactly, and thus required fuzzy linking. Within each classroom, pupil names from the two records were scored on their similarity—ignoring case, punctuation, spacing, and word order, and comparing phonetic spellings—with a small added weight for agreement on pupil sex. The highest scoring match for each pupil was kept, as long as it exceeded a predefined threshold. The matching was validated against a gold standard sample of 571 Ghana 2024–2025 pupils matched by hand, resulting in 83% recall and 98% precision. Across the three cohorts, 95% of external records were successfully linked to internal programme records.

Attendance, calculated from classroom registers as days in class divided by the cohort's expected contact days, captures the pupil's exposure to teaching (which requires both pupil and teacher presence).

Table C2. Attendance on pupil language match ($n = 1,615$). Estimated coefficients from linear mixed-effects models on attendance (%), with pupil language match as the key predictor. *Model 1* includes cohort and language match; *Model 2* adds the language match × country (Liberia) interaction; *Model 3* adds individual-level covariates. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3
Intercept (Reference Cohort)	73.401***	72.684***	80.237***
	(0.942)	(0.963)	(2.054)

Table C2. Attendance on pupil language match ($n = 1,615$) *(continued)*.

Variable	Model 1	Model 2	Model 3
Cohort: Ghana 2024–2025	1.464	1.350	1.312
	(1.332)	(1.308)	(1.312)
Cohort: Liberia 2024–2025	15.071***	17.224***	17.198***
	(1.203)	(1.440)	(1.448)
Pupil Language Match (Yes)	2.290**	4.338***	3.756***
	(0.769)	(1.074)	(1.096)
Pupil Language Match (Yes) × Liberia		-4.050**	-3.739*
		(1.532)	(1.532)
Pupil Sex (Male)			-1.218*
			(0.559)
Pupil Age (Years)			-0.691***
			(0.175)
Pupil Prior Schooling (Yes)			0.765
			(0.698)
Pupil Baseline Reading Comprehension (%)			0.032
			(0.024)
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.289
			(0.329)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.268
			(0.339)
R^2 Marginal	**0.251**	**0.257**	**0.268**
R^2 Conditional	**0.487**	**0.482**	**0.490**

C.3. Reading comprehension vs. language match and attendance

If mismatched pupils attend less, part of the mismatch gap might be an attendance effect. To test this, we added attendance as a predictor to the model of reading comprehension as an outcome of language match (see *Table C3*). The sample and variables are as before (see *Table C2*).

Table C3. Reading comprehension on pupil language match, adjusting for attendance ($n = 1,615$). Estimated coefficients from linear mixed-effects models on endline reading comprehension (%), with pupil language match as the key predictor and pupil attendance rate as an additional predictor. *Model 1* includes cohort and language match; *Model 2* adds the language match × country (Liberia) interaction; *Model 3.1* adds individual-level covariates; *Model 3.2* adds pupil attendance rate. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3.1	Model 3.2
Intercept (Reference Cohort)	20.037***	17.546***	0.920	-15.068*
	(2.104)	(2.115)	(4.326)	(6.020)
Cohort: Ghana 2024–2025	7.265*	6.852*	8.025**	7.760**
	(2.969)	(2.860)	(2.700)	(2.676)
Cohort: Liberia 2024–2025	16.641***	23.979***	24.533***	21.108***
	(2.685)	(3.185)	(3.008)	(3.121)
Pupil Language Match (Yes)	12.252***	19.426***	12.990***	12.267***
	(1.737)	(2.397)	(2.296)	(2.292)
Pupil Language Match (Yes) × Liberia		-13.952***	-9.001**	-8.281*
		(3.438)	(3.221)	(3.211)
Pupil Sex (Male)			0.588	0.827
			(1.187)	(1.185)
Pupil Age (Years)			1.259***	1.396***
			(0.372)	(0.372)
Pupil Prior Schooling (Yes)			2.846+	2.700+
			(1.475)	(1.470)
Pupil Baseline Reading Comprehension (%)			0.733***	0.727***
			(0.051)	(0.051)

Table C3. Reading comprehension on pupil language match, adjusting for attendance ($n = 1,615$) *(continued)*.

Variable	Model 1	Model 2	Model 3.1	Model 3.2
Pupil Home Literacy Support Index (M = 0, SD = 1)			1.316+	1.255+
			(0.695)	(0.693)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.523	0.479
			(0.716)	(0.713)
Pupil Attendance Rate (%)				0.199***
				(0.052)
R^2 Marginal	**0.116**	**0.136**	**0.255**	**0.265**
R^2 Conditional	**0.384**	**0.376**	**0.466**	**0.471**

C.4. Reading comprehension vs. baseline listening comprehension

The mismatch gap might reflect how much of the language of instruction a pupil understands, not which language they speak at home. To test this, we modelled reading comprehension as an outcome of baseline listening comprehension (see *Table C4*).

Table C4. Reading comprehension on baseline listening comprehension ($n = 1,718$). Estimated coefficients from linear mixed-effects models on endline reading comprehension (%), with baseline listening comprehension as the key predictor in place of language match. *Model 1* includes cohort and baseline listening comprehension; *Model 2* adds the listening comprehension × country (Liberia) interaction; *Model 3* adds individual-level covariates. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3
Intercept (Reference Cohort)	16.678***	14.798***	3.752
	(2.127)	(2.236)	(4.147)
Cohort: Ghana 2024–2025	6.516*	6.119*	7.818**
	(2.936)	(2.915)	(2.716)
Cohort: Liberia 2024–2025	18.242***	23.162***	22.071***
	(2.594)	(3.231)	(3.054)
Pupil Baseline Listening Comprehension (%)	0.181***	0.230***	0.116***
	(0.020)	(0.028)	(0.027)
Pupil Baseline Listening Comprehension (%) × Liberia		-0.101*	-0.015
		(0.040)	(0.038)
Pupil Sex (Male)			0.709
			(1.143)
Pupil Age (Years)			0.934*
			(0.363)
Pupil Prior Schooling (Yes)			3.138*
			(1.416)
Pupil Baseline Reading Comprehension (%)			0.708***
			(0.051)
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.812
			(0.674)

Table C4. Reading comprehension on baseline listening comprehension ($n = 1,718$) *(continued)*.

Variable	Model 1	Model 2	Model 3
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.554
			(0.688)
R^2 Marginal	**0.128**	**0.132**	**0.233**
R^2 Conditional	**0.402**	**0.401**	**0.461**

C.5. Reading comprehension vs. language match and baseline listening comprehension

If the mismatch gap reflects how much of the language of instruction a pupil understands, language match might carry no association with reading once oral proficiency is accounted for. To test this, we modelled reading comprehension as an outcome of both language match and baseline listening comprehension, each with its interaction with country (see *Table C5*).

Table C5. Reading comprehension on pupil language match and baseline listening comprehension ($n = 1,718$). Estimated coefficients from linear mixed-effects models on endline reading comprehension (%), with pupil language match and baseline listening comprehension entered jointly. *Model 1* includes cohort, language match, and baseline listening comprehension; *Model 2* adds the language match × country (Liberia) and listening comprehension × country (Liberia) interactions; *Model 3* adds individual-level covariates. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3
Intercept (Reference Cohort)	14.005***	10.539***	-0.229
	(2.087)	(2.186)	(4.176)
Cohort: Ghana 2024–2025	6.177*	5.601*	7.436**
	(2.790)	(2.698)	(2.604)
Cohort: Liberia 2024–2025	15.138***	25.329***	24.010***
	(2.528)	(3.459)	(3.323)
Pupil Language Match (Yes)	9.204***	15.413***	11.450***
	(1.665)	(2.315)	(2.236)
Pupil Baseline Listening Comprehension (%)	0.167***	0.200***	0.096***
	(0.020)	(0.028)	(0.028)
Pupil Language Match (Yes) × Liberia		-12.380***	-9.457**
		(3.298)	(3.133)
Pupil Baseline Listening Comprehension (%) × Liberia		-0.073+	0.003
		(0.040)	(0.038)
Pupil Sex (Male)			0.683
			(1.143)
Pupil Age (Years)			1.034**
			(0.363)

Table C5. Reading comprehension on pupil language match and baseline listening comprehension ($n = 1,718$) *(continued)*.

Variable	Model 1	Model 2	Model 3
Pupil Prior Schooling (Yes)			2.630+
			(1.414)
Pupil Baseline Reading Comprehension (%)			0.682***
			(0.051)
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.766
			(0.671)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.243
			(0.688)
R^2 Marginal	**0.155**	**0.178**	**0.265**
R^2 Conditional	**0.394**	**0.392**	**0.464**

C.6. Reading comprehension vs. pupil language match and teacher language match

The association between reading and the classroom's language mix might reflect some other characteristic of the classroom, such as the language the teacher speaks at home. To test this, we added teacher language match to the model of reading comprehension as an outcome of language match and its interaction with the classroom's language mix (see *Table C6*). The model was fitted on the 1,393 pupils whose teacher's home language was recorded. Teacher language match indicates whether the teacher's home language is the language of instruction. 87% of matched pupils and 63% of mismatched pupils had a teacher who spoke the language of instruction at home.

Table C6. Reading comprehension on pupil language match, adjusting for teacher language match ($n = 1,393$). Estimated coefficients from linear mixed-effects models on endline reading comprehension (%), with pupil language match as the key predictor and teacher language match as an additional predictor. *Model 1* includes cohort and pupil language match; *Model 2* adds the pupil language match × country (Liberia) interaction; *Model 3* adds individual-level covariates; *Model 4* adds classroom-level covariates, including teacher language match; *Model 5* adds language peer density and its interaction with pupil language match. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept (Reference Cohort)	22.004***	17.196***	-1.049	-13.307	-3.395
	(3.392)	(3.380)	(5.383)	(11.465)	(11.833)
Cohort: Ghana 2024–2025	5.432	6.046	8.242*	9.345**	9.679**
	(3.922)	(3.717)	(3.594)	(3.534)	(3.470)
Cohort: Liberia 2024–2025	15.599***	25.028***	26.312***	25.927***	26.985***
	(3.694)	(4.101)	(3.958)	(5.176)	(5.160)
Pupil Language Match (Yes)	10.790***	21.242***	14.528***	8.494**	-0.321
	(1.882)	(2.876)	(2.808)	(3.056)	(4.139)
Pupil Language Match (Yes) × Liberia		-16.927***	-11.686**	-5.169	-7.161+
		(3.760)	(3.591)	(3.790)	(3.944)
Pupil Sex (Male)			1.345	1.417	1.597
			(1.321)	(1.315)	(1.321)

Table C6. Reading comprehension on pupil language match, adjusting for teacher language match ($n = 1,393$) *(continued)*.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Pupil Age (Years)			1.282**	1.312**	1.308**
			(0.416)	(0.413)	(0.412)
Pupil Prior Schooling (Yes)			3.454*	2.999+	2.647
			(1.629)	(1.614)	(1.613)
Pupil Baseline Reading Comprehension (%)			0.718***	0.684***	0.690***
			(0.061)	(0.063)	(0.063)
Pupil Home Literacy Support Index (M = 0, SD = 1)			1.115	0.950	0.723
			(0.784)	(0.778)	(0.780)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.143	-0.100	-0.293
			(0.801)	(0.795)	(0.796)
Teacher Higher Education (Yes)				-0.084	0.969
				(2.335)	(2.318)
Teacher Experience: 1 Year				1.076	1.869
				(3.067)	(3.020)
Teacher Experience: > 1 Year				-6.511+	-5.765+
				(3.331)	(3.278)
Teacher Language Match (Yes)				3.890	-0.885
				(2.689)	(3.049)
Classroom Size (Number of Pupils)				0.414	0.228
				(0.402)	(0.399)
Classroom Baseline Reading Comprehension (%)				0.195	0.083
				(0.184)	(0.185)
Teaching Quality Index (M = 0, SD = 1 within cohort)				5.376***	4.991***
				(1.176)	(1.160)
Pupil Language Peer Density (%)					-0.094*
					(0.040)
Pupil Language Match (Yes) × Pupil Language Peer Density (%)					0.211**
					(0.066)
R^2 Marginal	**0.085**	**0.112**	**0.214**	**0.269**	**0.281**
R^2 Conditional	**0.347**	**0.335**	**0.421**	**0.444**	**0.445**

C.7. Improvement in listening comprehension vs. language peer density

The classroom's language mix might predict whether mismatched pupils improve their oral proficiency. To test this, we fit a linear probability model[11] of listening comprehension improvement on baseline listening comprehension and its interaction with the classroom's language mix (see *Table C7*). The model was fitted on the 713 mismatched pupils who began the programme below full oral proficiency. Listening comprehension improvement is an indicator equal to one if a pupil's endline listening comprehension score exceeded their baseline score.

Table C7. Improvement in listening comprehension on language peer density (n = 713). Estimated coefficients from linear mixed-effects linear probability models on whether a pupil's listening comprehension improved from baseline to endline, fit on mismatched pupils who began below full oral proficiency. *Model 1* includes cohort; *Model 2* adds baseline listening comprehension; *Model 3* adds individual-level covariates; *Model 4* adds classroom-level covariates; *Model 5* adds language peer density and its interaction with baseline listening comprehension. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept (Reference Cohort)	0.660***	0.634***	0.292**	0.445*	0.640**
	(0.040)	(0.041)	(0.107)	(0.208)	(0.210)
Cohort: Ghana 2024–2025	0.112*	0.106+	0.124*	0.127*	0.120*
	(0.055)	(0.054)	(0.055)	(0.054)	(0.053)
Cohort: Liberia 2024–2025	0.161**	0.303***	0.272***	0.285**	0.218*
	(0.056)	(0.078)	(0.078)	(0.092)	(0.091)
Pupil Baseline Listening Comprehension (%)	-0.005***	-0.004***	-0.005***	-0.005***	-0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

[11]The coefficients of a linear probability model are directly interpretable as differences in the probability of the outcome (here, of developing listening comprehension), expressed as percentage points. We prefer it to a logistic specification because linear models remain interpretable in the presence of interaction terms, which are central to our analysis, while logistic coefficients are expressed as log-odds and require further transformation to yield probabilities (Gomila, 2021). The results are robust to a logistic mixed-effects specification (see *Table C8*).

Table C7. Improvement in listening comprehension on language peer density (n = 713) *(continued)*.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Pupil Baseline Listening Comprehension (%) × Liberia		-0.004**	-0.003+	-0.003+	-0.002
		(0.002)	(0.002)	(0.002)	(0.002)
Pupil Sex (Male)			0.050	0.051	0.067*
			(0.033)	(0.032)	(0.033)
Pupil Age (Years)			0.032**	0.032**	0.031**
			(0.010)	(0.010)	(0.010)
Pupil Prior Schooling (Yes)			0.022	0.015	0.004
			(0.038)	(0.038)	(0.038)
Pupil Baseline Reading Comprehension (%)			0.011***	0.011**	0.011**
			(0.003)	(0.003)	(0.003)
Pupil Home Literacy Support Index (M = 0, SD = 1)			-0.003	-0.004	-0.006
			(0.018)	(0.019)	(0.019)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.045*	0.039+	0.035
			(0.021)	(0.021)	(0.021)
Teacher Higher Education (Yes)				0.042	0.050
				(0.050)	(0.048)
Teacher Experience: 1 Year				-0.050	-0.042
				(0.057)	(0.056)
Teacher Experience: > 1 Year				-0.045	-0.029
				(0.062)	(0.060)
Classroom Size (Number of Pupils)				-0.005	-0.006
				(0.007)	(0.007)
Classroom Baseline Reading Comprehension (%)				-0.003	-0.004
				(0.005)	(0.005)
Teaching Quality Index (M = 0, SD = 1 within cohort)				0.082***	0.079***
				(0.024)	(0.023)
Pupil Language Peer Density (%)					-0.003***
					(0.001)
Pupil Baseline Listening Comprehension (%) × Pupil Language Peer Density (%)					0.000**
					(0.000)

Table C7. Improvement in listening comprehension on language peer density (n = 713) *(continued)*.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
R^2 Marginal	**0.086**	**0.095**	**0.128**	**0.152**	**0.169**
R^2 Conditional	**0.289**	**0.291**	**0.320**	**0.326**	**0.325**

C.8. Improvement in listening comprehension vs. language peer density (logistic specification)

A linear probability model might misrepresent the association between listening comprehension improvement and its interaction with the classroom's language mix by forcing it into a linear form. To test this, we refit the previous model as a logistic mixed-effects specification (see *Table C8*). The sample and variables are as before (see *Table C7*).

Table C8. Improvement in listening comprehension on language peer density, logistic specification ($n = 713$). Estimated odds ratios from logistic mixed-effects models on whether a pupil's listening comprehension improved from baseline to endline, fit on mismatched pupils who began below full oral proficiency, as a robustness check on the linear probability specification in *Table C7*. *Model 1* includes cohort; *Model 2* adds baseline listening comprehension; *Model 3* adds individual-level covariates; *Model 4* adds classroom-level covariates; *Model 5* adds language peer density and its interaction with baseline listening comprehension. 95% confidence intervals of odds ratios are reported in brackets. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept (Reference Cohort)	2.352***	2.005**	0.278*	0.700	2.056
	[1.523, 3.634]	[1.293, 3.109]	[0.080, 0.972]	[0.070, 7.030]	[0.220, 19.221]
Cohort: Ghana 2024–2025	1.850*	1.767+	2.071*	2.095*	1.853*
	[1.040, 3.291]	[0.999, 3.125]	[1.125, 3.812]	[1.150, 3.816]	[1.067, 3.217]
Cohort: Liberia 2024–2025	2.490**	8.190***	6.990***	7.554***	3.985*
	[1.364, 4.547]	[2.907, 23.070]	[2.413, 20.248]	[2.326, 24.531]	[1.316, 12.063]
Pupil Baseline Listening Comprehension (%)	0.973***	0.979***	0.971***	0.970***	0.957***
	[0.965, 0.981]	[0.970, 0.987]	[0.962, 0.981]	[0.961, 0.980]	[0.944, 0.970]
Pupil Baseline Listening Comprehension (%) × Liberia		0.970**	0.977*	0.978*	0.988
		[0.951, 0.990]	[0.958, 0.997]	[0.959, 0.998]	[0.969, 1.007]
Pupil Sex (Male)			1.338	1.342	1.418+
			[0.915, 1.955]	[0.919, 1.959]	[0.987, 2.037]
Pupil Age (Years)			1.207**	1.214**	1.181**
			[1.070, 1.361]	[1.076, 1.369]	[1.055, 1.321]
Pupil Prior Schooling (Yes)			1.141	1.095	1.022
			[0.730, 1.783]	[0.702, 1.709]	[0.670, 1.560]

Table C8. Improvement in listening comprehension on language peer density, logistic specification (n = 713) *(continued)*.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Pupil Baseline Reading Comprehension (%)			1.086**	1.086**	1.077**
			[1.031, 1.145]	[1.028, 1.148]	[1.022, 1.134]
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.978	0.972	0.963
			[0.792, 1.207]	[0.787, 1.202]	[0.787, 1.179]
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			1.332*	1.283+	1.212
			[1.034, 1.717]	[0.999, 1.647]	[0.955, 1.537]
Teacher Higher Education (Yes)				1.295	1.302
				[0.745, 2.252]	[0.780, 2.174]
Teacher Experience: 1 Year				0.759	0.807
				[0.402, 1.433]	[0.448, 1.454]
Teacher Experience: > 1 Year				0.751	0.827
				[0.376, 1.500]	[0.434, 1.576]
Classroom Size (Number of Pupils)				0.972	0.973
				[0.898, 1.052]	[0.904, 1.047]
Classroom Baseline Reading Comprehension (%)				0.975	0.971
				[0.916, 1.038]	[0.915, 1.030]
Teaching Quality Index (M = 0, SD = 1 within cohort)				1.611***	1.507**
				[1.228, 2.114]	[1.173, 1.937]
Pupil Language Peer Density (%)					0.984***
					[0.976, 0.993]
Pupil Baseline Listening Comprehension (%) × Pupil Language Peer Density (%)					1.000**
					[1.000, 1.001]
R^2 Marginal	**0.122**	**0.153**	**0.212**	**0.245**	**0.239**
R^2 Conditional	**0.364**	**0.384**	**0.443**	**0.442**	**0.410**

C.9. Reading comprehension vs. improvement in listening comprehension

Mismatched pupils who improve their oral proficiency might read better. To test this, we modelled endline reading comprehension as an outcome of improvement in listening comprehension (see *Table C9*). The sample and variables are as before (see *Table C7*).

Table C9. Reading comprehension on improvement in listening comprehension (n = 713). Estimated coefficients from linear mixed-effects models on endline reading comprehension (%), with improvement in listening comprehension as the key predictor. *Model 1* includes cohort and baseline listening comprehension; *Model 2* adds the listening comprehension × country (Liberia) interaction; *Model 3.1* adds individual-level covariates; *Model 3.2* adds the listening comprehension improvement indicator. Standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3.1	Model 3.2
Intercept (Reference Cohort)	10.928***	10.058***	-5.192	-8.055
	(2.168)	(2.218)	(5.105)	(5.003)
Cohort: Ghana 2024–2025	4.270	4.079	4.861	3.531
	(3.052)	(3.055)	(2.967)	(2.777)
Cohort: Liberia 2024–2025	24.684***	29.705***	29.439***	26.527***
	(3.015)	(4.016)	(3.928)	(3.781)
Pupil Baseline Listening Comprehension (%)	0.143***	0.176***	0.125***	0.186***
	(0.032)	(0.036)	(0.036)	(0.037)
Pupil Baseline Listening Comprehension (%) × Liberia		-0.142+	-0.093	-0.061
		(0.075)	(0.073)	(0.071)
Pupil Sex (Male)			-0.089	-0.704
			(1.491)	(1.474)
Pupil Age (Years)			1.558***	1.186*
			(0.467)	(0.462)
Pupil Prior Schooling (Yes)			-1.375	-1.434
			(1.808)	(1.770)
Pupil Baseline Reading Comprehension (%)			0.914***	0.811***
			(0.145)	(0.144)

Table C9. Reading comprehension on improvement in listening comprehension (n = 713) *(continued)*.

Variable	Model 1	Model 2	Model 3.1	Model 3.2
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.433	0.423
			(0.872)	(0.853)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			1.770+	1.502
			(1.030)	(1.004)
Pupil Listening Comprehension Improved (Yes)				10.549***
				(1.739)
R^2 Marginal	**0.180**	**0.185**	**0.241**	**0.288**
R^2 Conditional	**0.466**	**0.471**	**0.503**	**0.504**

Appendix D. Further checks

D.1. Standard model diagnostics

We examined three assumptions of the mixed-effects model.

The normality of individual-level residuals requires that error term $\epsilon_{i(jk)}$ is normally distributed. The pupil-level residuals fall close to the reference line across the distribution, lifting slightly above it at the upper tail (see *Figure D1*). The normality assumption holds.

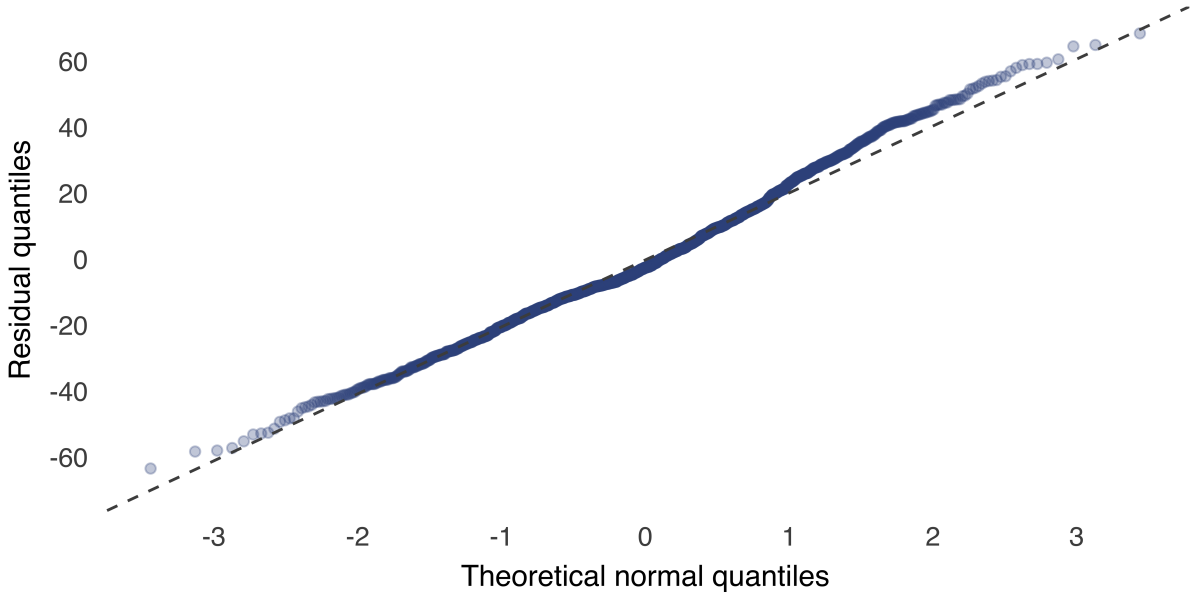


Figure D1. Normal quantile-quantile plot of the individual-level residuals. Estimates are derived from the full linear mixed-effects model (*Model 5*).

The homogeneity of individual-level residual variance requires that the variance of the error term $\epsilon_{i(jk)}$ is constant across the fitted values. The variance is stable across the central range. The reduced spread near 0% and 100% reflects floor and ceiling effects in the bounded outcome, and the arc-shaped banding reflects its discreteness, rather than heteroscedasticity (see *Figure D2*). The assumption is satisfied.

The normality of random effects requires that the random effects $v_{j(k)}$ are normally distributed. The classroom random intercepts fall close to the reference line through the centre, with a slight upward departure at the upper tail (see *Figure D3*). They are approximately normal, and the assumption holds.

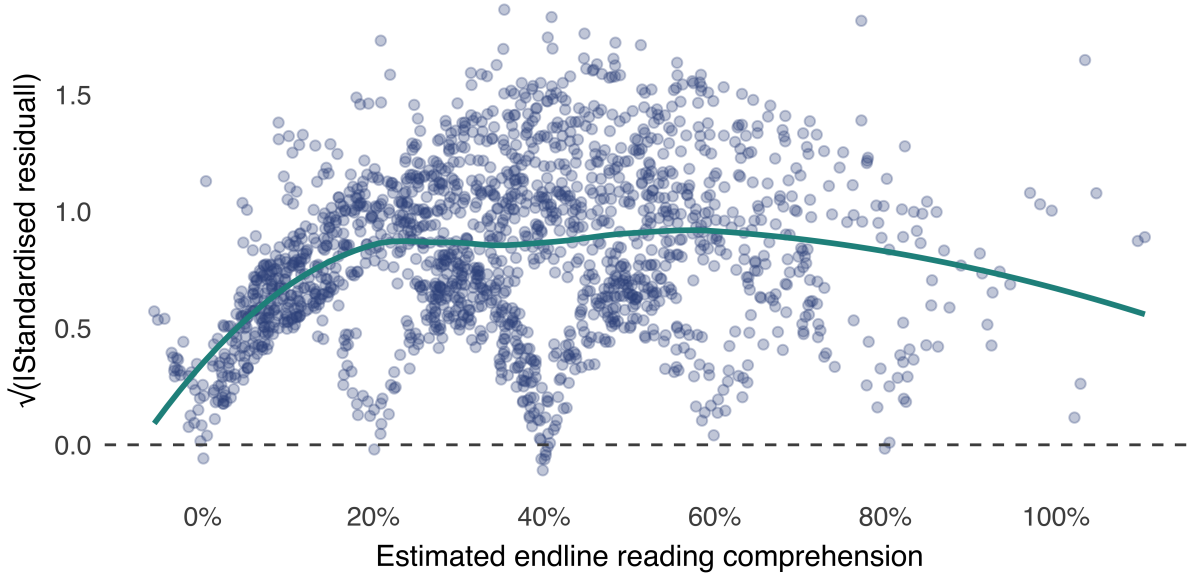


Figure D2. Square root of the absolute standardised residual against estimated endline reading comprehension. Estimates are derived from the full linear mixed-effects model (*Model 5*).

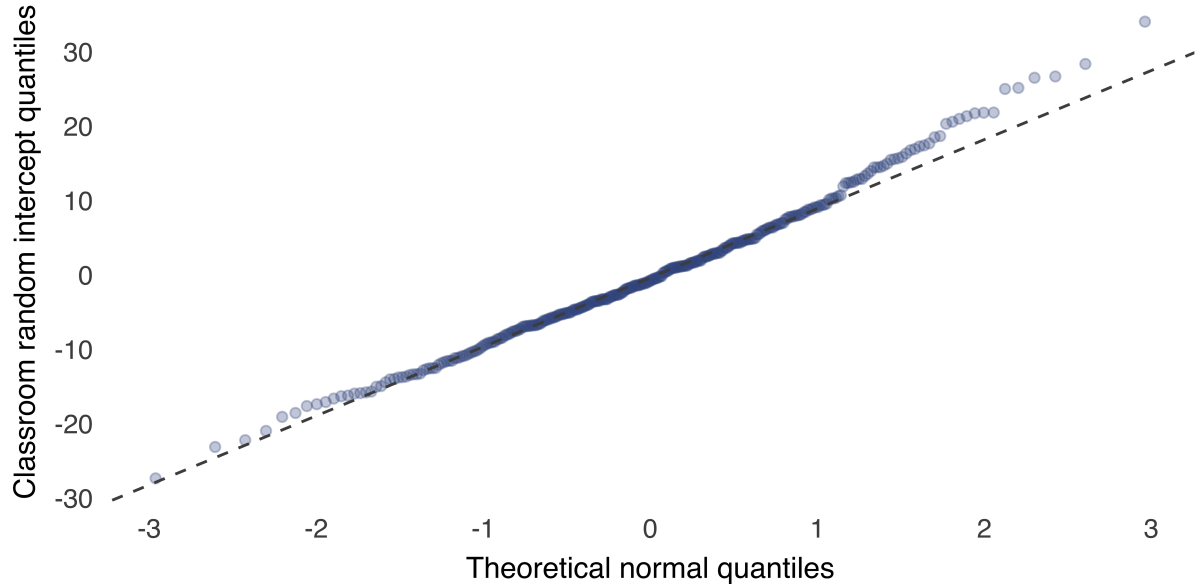


Figure D3. Normal quantile-quantile plot of the classroom random intercepts. Estimates are derived from the full linear mixed-effects model (*Model 5*).

D.2. Reading comprehension vs. language match and language peer density (cluster-robust standard errors)

The mixed-effects model captures clustering of pupils within classrooms through random intercepts, but its standard errors assume a fixed within-classroom correlation. To relax that assumption, we recomputed standard errors for the full model using cluster-robust variance estimation at the classroom level (see *Table D1*). The estimates are identical to those in *Table C1* by construction. The standard errors for classroom-level predictors increased slightly, as expected; the significance of associations of language match and language peer density with reading comprehension did not change.

Table D1. Reading comprehension on pupil language match, with cluster-robust standard errors ($n = 1,718$). The full specification of *Table C1*, reestimated with cluster-robust standard errors at the classroom level. *Model 1* includes cohort and language match; *Model 2* adds the language match × country (Liberia) interaction; *Model 3* adds individual-level covariates; *Model 4* adds classroom-level covariates; *Model 5* adds language peer density and its interaction with language match. Cluster-robust standard errors are reported in parentheses. Models include random intercepts for the classroom. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept (Reference Cohort)	19.812***	17.224***	0.017	-7.814	-1.842
	(1.962)	(1.899)	(3.858)	(8.415)	(8.625)
Cohort: Ghana 2024–2025	7.446*	7.078*	8.328**	8.185**	7.894**
	(3.014)	(2.820)	(2.617)	(2.487)	(2.469)
Cohort: Liberia 2024–2025	17.527***	25.000***	25.711***	26.425***	24.889***
	(2.673)	(3.023)	(2.886)	(3.335)	(3.426)
Pupil Language Match (Yes)	11.168***	18.542***	12.505***	7.251**	-1.181
	(1.744)	(2.545)	(2.373)	(2.437)	(3.091)
Pupil Language Match (Yes) × Liberia		-14.213***	-9.644**	-3.868	-4.901
		(3.452)	(3.247)	(3.333)	(3.492)
Pupil Sex (Male)			0.615	0.749	0.919
			(1.169)	(1.156)	(1.153)
Pupil Age (Years)			1.296***	1.260***	1.255***
			(0.330)	(0.328)	(0.327)

Table D1. Reading comprehension on pupil language match, with cluster-robust standard errors ($n = 1,718$) *(continued)*.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
Pupil Prior Schooling (Yes)			3.105*	2.615+	2.352
			(1.464)	(1.464)	(1.461)
Pupil Baseline Reading Comprehension (%)			0.719***	0.676***	0.679***
			(0.053)	(0.058)	(0.058)
Pupil Home Literacy Support Index (M = 0, SD = 1)			0.998	0.783	0.634
			(0.685)	(0.678)	(0.677)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			0.439	0.139	-0.099
			(0.716)	(0.700)	(0.707)
Teacher Higher Education (Yes)				1.268	1.946
				(2.028)	(2.009)
Teacher Experience: 1 Year				-0.808	0.107
				(2.524)	(2.496)
Teacher Experience: > 1 Year				-7.568**	-6.237*
				(2.765)	(2.793)
Classroom Size (Number of Pupils)				0.381	0.261
				(0.311)	(0.313)
Classroom Baseline Reading Comprehension (%)				0.292+	0.138
				(0.153)	(0.160)
Teaching Quality Index (M = 0, SD = 1 within cohort)				5.465***	4.973***
				(0.870)	(0.842)
Pupil Language Peer Density (%)					-0.081**
					(0.025)
Pupil Language Match (Yes) × Pupil Language Peer Density (%)					0.184***
					(0.046)
R^2 Marginal	**0.115**	**0.136**	**0.250**	**0.312**	**0.323**
R^2 Conditional	**0.383**	**0.374**	**0.459**	**0.483**	**0.485**

D.3. Reading comprehension vs. language match (fixed-effects specification)

Individual-level features might be correlated with unobserved classroom-level features, thus biasing the mismatch gap. To test this, we reestimated the model with classroom fixed effects, which absorb classroom differences (see *Table D2*). The mismatch gap was smaller and not significant, and the country difference disappeared. This specification identifies the gap from within-classroom comparisons alone. With just over 5 pupils per classroom, and pupils clustered by language match across classrooms, the mismatch gap cannot be estimated precisely.

Table D2. Reading comprehension on pupil language match, classroom fixed-effects specification ($n = 1,718$). Estimated coefficients from linear models on endline reading comprehension (%) with classroom fixed-effects in place of random intercepts, relaxing the assumption that pupil-level covariates are uncorrelated with unobserved classroom features. *Model 1* includes language match; *Model 2* adds the language match × country (Liberia) interaction; *Model 3* adds individual-level covariates. Classroom-level covariates are absorbed by the fixed-effects. Standard errors are reported in parentheses. R^2 and adjusted R^2 are reported in place of marginal and conditional R^2. *** – $p < 0.001$, ** – $p < 0.01$, * – $p < 0.05$, + – $p < 0.1$.

Variable	Model 1	Model 2	Model 3
Intercept (Reference Classroom)	55.758***	55.975***	16.111
	(10.234)	(10.540)	(10.860)
Pupil Language Match (Yes)	4.242*	4.025	3.076
	(2.055)	(3.244)	(3.030)
Pupil Language Match (Yes) × Liberia		0.362	-0.002
		(4.193)	(3.916)
Pupil Sex (Male)			0.993
			(1.163)
Pupil Age (Years)			1.343***
			(0.378)
Pupil Prior Schooling (Yes)			1.810
			(1.554)
Pupil Baseline Reading Comprehension (%)			0.680***
			(0.053)

Table D2. Reading comprehension on pupil language match, classroom fixed-effects specification ($n = 1,718$) *(continued)*.

Variable	Model 1	Model 2	Model 3
Pupil Home Literacy Support Index (M = 0, SD = 1)			1.167
			(0.731)
Pupil Socioeconomic Status Index (M = 0, SD = 1 within cohort)			-0.632
			(0.766)
R^2	**0.516**	**0.516**	**0.580**
R^2 Adjusted	**0.402**	**0.401**	**0.479**

Appendix E. Language matching simulation

Home language instruction cannot match every pupil in linguistically diverse settings. To ask how many pupils it could reach, we simulated adding languages of instruction to Ghana's Ashanti region one at a time (see *Table E1*).

The simulation used the Luminos internal programme records for Ashanti. The simulation sample differed from the sample used in the main analysis in two ways. First, it covered four academic years (2022–2023 to 2025–2026) rather than two. Second, it included every enrolled pupil whose home language was known, rather than the externally assessed subsample. The language match rates reported here therefore differ from those in the *Results*.

We report the results of the simulation at the classroom level. Candidate languages were the home languages not already used for instruction, ranked by the number of pupils who spoke them—an ordering that approximates how an education system might prioritise languages in practice. Beginning from Asante Twi, the language of instruction already in use, we added each candidate step-by-step. At each step, every classroom either kept its current language of instruction or adopted the newly available language, if it matched more pupils. A language switch could 'displace' some pupils: those who were matched under the previous language of instruction, but became mismatched after the new language was adopted. We stopped the simulation once no new language additions could produce a net gain.

Table E1. Stepwise introduction of languages of instruction in Ghana's Ashanti region, with each classroom free to choose from the languages available ($n = 7,584$). The first row is the status quo, Asante Twi. At each step, we report the language added; the share of pupils newly matched by its addition; the share of pupils displaced by its addition; the net gain, defined as the difference between the share of pupils matched and displaced; and the cumulative share of pupils matched. Languages that produced no net gain were omitted.

Language Added	Pupils Matched	Pupils Displaced	Net Gain	Total Match
Status Quo (Asante Twi)	0%	0%	0%	31%
Konkomba	25%	1%	24%	55%
Dagaare	11%	2%	9%	63%
Kusaal	2%	1%	2%	65%
Mampruli	2%	0%	1%	66%
Farefare	1%	0%	1%	67%
Dagbani	2%	0%	2%	69%
Gourmanchema	1%	0%	1%	70%

Table E1. Stepwise introduction of languages of instruction in Ghana's Ashanti region, with each classroom free to choose from the languages available ($n = 7,584$) *(continued)*.

Language Added	Pupils Matched	Pupils Displaced	Net Gain	Total Match
Moore	1%	0%	1%	70%
Anufo	1%	0%	1%	72%
Ewe	0%	0%	0%	72%
Pulaar	1%	0%	0%	72%
Wali	1%	0%	0%	73%
Grusi	0%	0%	0%	73%
Moba	0%	0%	0%	73%