

Silenced by an Unknown Language? Exploring Language Matching during Transitions from
Complementary Education into Government Schools in Ghana

Kieran Daly, Emma Carter, Ricardo Sabates*

May 2021

Manuscript ID CCOM-2020-0189.R1

Kieran Daly

University of Cambridge, Faculty of Education, 184 Hills Road, Cambridge.

kierandaly12@gmail.com

Twitter: @KieranSCDaly

Emma Carter

University of Cambridge, Faculty of Education, 184 Hills Road, Cambridge. ejc69@cam.ac.uk

ORCID: [0000-0001-6169-3806](https://orcid.org/0000-0001-6169-3806)

Twitter: @EmmaCar13341799

Ricardo Sabates (* Corresponding author)

University of Cambridge, Faculty of Education, 184 Hills Road, Cambridge.

Rs867@cam.ac.uk

ORCID: 0000-0002-1433-5667

Twitter: @RSabates72

Abstract:

Mother tongue-based education has been central to the promotion of early literacy skills in many multilingual contexts of the Global South. However, learners in such environments may face significant linguistic challenges when changing language of instruction during schooling. In particular, the linguistic distance between mother tongue and official language may be a significant barrier to learners. This paper provides an empirical approach to this issue by employing language matching based on linguistic distance between languages to explore changes in literacy scores for learners who change language of instruction. Findings show that the greater the linguistic distance between two languages the larger the loss in foundational literacy. We conclude that language matching could be introduced as a tool to identify at-risk learners during transitions and, if possible, as a tool for linguistic allocation of students who have the possibility of selecting between schools with different languages for instruction.

Keywords: Mother Tongue Education; Language Matching; Multilingualism; Literacy; Ghana.

Word count: 6,824 (incl. references)
Number of Tables 4 and Number of Figures 3

Acknowledgements: This project was commissioned under the Complementary Basic Education (CBE) Programme, funded by FCDO and USAID and managed by the Management Unit (MU) at Crown Agents, in partnership with the Ministry of Education and Ghana Education Service. The authors are grateful for the support received from RTI International, JEAVCO and PAB Development Consultants and IMC Worldwide. The authors are grateful for support received from Professor Kwame Akyeampong and for the insights into Gruni language provided by Dr Helen Atipoka Adongo from the Faculty of Ghanaian Languages, University of Education Winneba, Ghana.

1. Introduction

Many multilingual educational systems are promoting the use of mother tongue as language of instruction, particularly during early years (Brock-Utne, 2010; Trudell, 2009). Yet, in these complex linguistic environments many children experience learning in different languages as they progress through education. For instance, some children may learn in mother tongue during early years, but then change to an official national language for instruction from around mid-primary education onwards. Other children may only receive informal mother tongue education in their communities, due to formal schooling only being available in a different local language of instruction. Even within the classroom, some children learn in the language utilised by the teacher, which may not be the same as the child's mother tongue. This is the case even when the school's official language of instruction may be the same as the child's mother tongue. Given the complexity of multilingual environments, and the desire to promote mother tongue education in a sustainable way, this raises an important question: are children able to retain foundational literacy skills previously gained in mother tongue as they transition into a different linguistic educational environment?

Previous research using the case of children who participated in the Complementary Basic Education (CBE) programme in Ghana have found that children who change to a different linguistic environment fare worse in foundational literacy acquisition than their peers who continue to learn in mother tongue (AUTHORS). Yet, it is unknown whether differences in foundational literacy acquisition could relate to the local language of instruction, particularly as there are linguistic differences between languages. In order to analyse these linguistic differences we utilise linguistic typology as a methodology. We assess and cross-compare the phonological and orthographic features within the Niger-Congo phylum to match students by 'linguistic similarity'¹. All languages spoken by the Ghanaian children in the study were part of the Niger-Congo phylum.

Empirically, we assess whether there are differences in the proportion of children who are unable to retain foundational literacy skills previously achieved in mother tongue during the CBE programme, and whether these differences relate to the linguistic distance between languages. In the context of this paper, this language matching relates to changes from one language indigenous to Ghana to another, as opposed to changes to English which are not explored in this paper. We have undertaken this work exclusively focusing on Ghanaian indigenous languages as this is an area that has received less attention than research focusing on transitions to English or other western dominant languages. To our knowledge, this is one of the first papers to empirically test the concept of language matching with these Ghanaian languages and therefore contributes to a deeper understanding of the challenges faced by children who move to linguistic environments which are distant from their own mother tongue. As a limitation of this paper, we only focus here on linguistic differences between languages and not the sociolinguistic environment of these learners. We also assess this from a quantitative perspective, and therefore are unable to capture the process of learning these different languages which can be impacted by factors including learners' disposition or the teachers' linguistic skills. Yet, we assess foundational literacy acquisition over time by their linguistic distance and we highlight this as the main contribution of this paper.

¹ We acknowledge that sociolinguistic factors are equally as important linguistic factors in 'matching' language groups. Sociolinguistic frameworks recognise the importance of linguistic enregisterment (Kerswill, 2013a, 2013b, and 2014), contact linguistics (Campbell, 2003; Daniel, 2010), sociolinguistic scales (Blommaert, 2010) and linguistic markets (Bourdieu, 1977, Sherris *et al*, 2014) which can complement the linguistic aspect of language matching. These aspects are beyond the scope of this paper.

2. Transitioning Literacy Acquisition to Distant Languages

Learning to read is inextricably linked to the spoken mode. Children begin to learn the sounds and rules of their language from birth, and start to segment the speech stream into words, syllables, and finally individual phonemes (Reder *et al*, 2013). To segment words into letters and letters into sounds, children must have acute awareness of the phonological properties of their language (Cheung *et al*, 2001). When children realise that each letter matches a sound in their spoken language, they have acquired metalinguistic awareness – or awareness of how language and literacy combine. This involves transitioning from unconscious to conscious knowledge of the mechanisms of literacy acquisition (Reder *et al*, 2013).

Children who acquire metalinguistic awareness from their mother tongue are more able to apply the concepts of sound-letter matching and literacy conventions when learning a new language. Research suggests that children are able to read for the first time in an unfamiliar language once a sufficient amount of exposure to mother tongue is reached (Cummins 1979, 1981). This is further evidenced by Dubeck and Gove's (2015:317) study that found "alphabetic knowledge is a stronger predictor for non-native speakers...predict(ing) reading achievement two years later". Presumably, the closer the linguistic distance between mother tongue and the new language, the easier it will be for children to make the transition to acquire literacy in a new language.

Children without awareness of spoken elements of their own language find metalinguistic awareness from another language difficult to acquire. Consequently, when a teacher points at a letter and produces a sound that is not present in the child's first language, the child does not associate the sound with their language. The child then finds it harder to grasp the concept that each symbol represents a sound in their language. This process is further complicated when the phonological properties of the child's mother tongue differ to those in the new language in which the child is learning. The closer the new language is to the child's mother tongue in terms of how sound is produced and how the child hears this sound, the more able is the child to categorise sounds to words that have been already obtained from mother tongue. The more distant the languages are, the harder it is for the child to categorise sounds in the new language. These phonological similarities and sound-symbol correspondences constitute the basis from which we use linguistic typology between the different Ghanaian languages learned during the CBE programme to explore their linguistic distance. This is done in order to empirically explore the sustainability of foundational literacy acquisition experienced by children who changed to a different local language of instruction.

3. Linguistic Typology and Distance Scale Adapted to the CBE Programme

The main aim of a typological methodology is to classify languages by their features, in order to show how 'related' or 'close' two or more languages are to one another. Popular methods of displaying the relatedness of languages tend to use aspects of phonology, syntax, semantics and pragmatics as basis for classification (Aikhenvald & Dixon, 2017). Based on these aspects, languages can be delineated in order to investigate cross-linguistic diversity (Aikhenvald & Dixon, 2017:1). In order to understand the relationship between the languages within this research, we focussed on the typological features that apply prominently to formative literacy acquisition that is taught through the phonics-based method: phonological and orthographical similarities.

In order to identify phonological and orthographical similarities between languages, we adopted the ‘structures’ and ‘systems’ approaches of the ‘intra-language’ typological method (Aikhenveld & Dixon, 2017:6-7). This method compares individual phonological structures across different languages and whole linguistic systems. Considering both phonological structures and whole linguistic systems ensures that languages are not only classified by their proximity in terms of language family, but by the phonological and orthographical properties which are used by the learner when learning in a different language. By analysing the Niger-Congo language phylum, which contains all mother tongue languages and the languages experienced by children during the CBE programme (Brifo, Dagaare, Dagbani, Ewe, Gonja, Gruni, Kasem, Konkomba, Kusaal, Mampruli, Sisaala, and Twi) we created phonological tables that displays phonemes and their corresponding orthographic symbols for each language. We then identified the phonological properties that are shared by the languages, and those that differ. This process utilised phonological information gathered by Ghanaian language experts on Brifo (Kuch 1993), Dagaare (Bodomo 1997) Dagbani (Crothers *et al*, 1979), Gruni (Adongo, 2018) and other languages associated with the CBE project (Hartell, 1993; Omniglot, 2018; Ire-Mwinlaaru, 2017).

The result of this method for the linguistic differences between languages using their phonological and orthographic properties is shown in Table 1. Our analyses show that languages that are closer to one another in the Niger-Congo phylum have less significant phonological and orthographic differences, or higher levels of mutual intelligibility based on other factors (contact, sociolinguistic factors of power and amelioration) than those further away from one another. This provides evidence that the structural (phonological and orthographic similarities) and systematic (distance between one another in Niger-Congo language phylum) aspects of our typological analyses align with one another, thus triangulating the decisions behind a possible categorisation of ‘close’ and ‘distant’ language transition. Those that are distanced by several branches within the phylum and have significant phonological differences were identified as ‘distant’. Those sharing several branches within the phylum with few phonological differences (sometimes referred to as ‘sister’ languages) were identified as ‘close’.² We will explore the phylum and the phonological and orthographical similarities in more detail below.

==Table 1 about here==

The Niger-Congo phylum, or classification tree, containing the languages involved in the CBE programme is summarised in Figure 1 – this is the authors’ figure, and is based on a similar phylum produced by Adongo (2018), and information from Ethnologue (2018). Language phyla are used to show similarity between languages. The more language classification groups that languages share, the smaller the linguistic differences between them, and consequently the more similar the languages become. Therefore, languages that split at the ‘Kwa’ and ‘North’ classifications represent linguistically distant languages and contain more distinct phonological and syntactic properties. The linguistic ‘distance’ between languages that differ at a ‘higher’ language classification group differ more. There are common properties shared by all the languages within the CBE programme, due to their geographical proximity to one another, and the consequent interaction and exposure between their speakers. All languages sit within the geographically defined group of Volta-Congo of

² It is worth noting that all languages in this study were situated within a close area of the Niger-Congo language phylum. Transitioning to a national language in a different part of the phylum (Twi, Gonja, Ewe), or a language in a different phylum (English) could imply more significant differences.

the Niger-Congo phylum. However, as evidenced by the classification tree and Table 1 (which compares phonological and orthographical features between the languages to triangulate with the phylum), there is still linguistic difference between many of these languages. Closer languages such as Mampruli and Dagbani, and Dagaare and Birifor contain considerable phonological, syntactic and lexical overlap, and others, such as Kusaal and Kasem and Sisaala and Dagaare shared less features. This allowed us to create language ‘close’ and language ‘distant’ groups. We use this classification of close and distant languages, based on phonological and orthographical similarity, as well as the language phylum, to empirically test foundational literacy achievements of groups of children who changed language of instruction between the CBE programme and the first year in government schools.

==Figure 1 about here ==

4. Methodology

4.1 Description of the Sample

Data from this paper comes from an evaluation of the CBE programme in Ghana. For evaluation purposes, an initial sample of 2,360 students was drawn from an approximate 40,000 CBE children who took part in the 2016-17 academic year. The sampling method followed a stratified random sample (stratification by implementing partners who worked within specific linguistic environments). Considerable attrition was experienced overtime due to discontinuation of schooling following the CBE programme, dropout from formal school, migration and absence at the time of data collection (irregular attendance at school is common due to household chores and seasonality) (See AUTHORS, for a more in-depth attrition analysis).

From the initial sample, data with full information across two academic years was available for 1,166 children. Given our interest in examining transitions into linguistically similar or different indigenous languages, we focused our analysis on the 382 students who transitioned into government schools in grade 3 and below, where local language is maintained as the dominant language of instruction. Children were further classified according to whether the transition implied continuation of the mother tongue into official language of instruction or whether the language changed. In case the language changed, language matching was used to identify children in two groups: (1) close or (2) distant from their mother tongue following the linguistic typology in Figure 1.

Table 2 shows descriptive statistics for the full sample of 2,360 children, as well as for the 382 children with full information on literacy assessments and who were placed into grade 3 and below in government schools after completing the CBE programme. We estimate that of all 2,360 children, 52% made a transition into the same language when moving into government schools. For our estimation sample, this proportion is 46%, indicating that significantly more children moved into a different language when transitioning into primary grade 3 and below. Compared with the full sample, children from our estimation sample were also significantly more likely to be low achievers in both numeracy and literacy (as indicated by the higher proportion of zero scores in both cases) and were found to be statistically younger. These findings were unsurprising, given that students in our estimation sample were transitioning into lower grades (i.e. grades 3 and below). As shown, our estimation sample was also statistically more likely to have students from mid-low and mid-high wealth

backgrounds, suggesting that they were relatively better off in terms of assets compared to children from the full sample.

In comparing students from language same, language close and language distant groups who were part of our estimation sample, it can be seen that language close students were significantly more likely to have learning materials at home, engage in reading at home and to have more family members within their household compared with language same students. Language distant students were also significantly more likely to engage in reading activity at home compared to language same students.

==Table 2 about here==

Understanding patterns of sample bias is important for two reasons, which we address in this paper. First, transitioning into a linguistically different school may be associated with indicators of disadvantage, such as poverty, remoteness, and therefore we need to consider these factors as potential controls in our statistical analyses. Second, since we are working with children for whom we have full information, it is possible that the sample is biased towards those who are relatively higher achievers or live in households who prioritise schooling over other chores or needs. Therefore, our results are unlikely to be representative of all children who undertook the CBE programme. Fortunately, this issue is likely to affect the lower bias of our estimates. In other words, for children who may be less likely to continue to government schooling or remain once in there, we can expect an even greater learning challenge due to linguistic distance.

4.2 Assessment of Literacy Skills and Language Matching

Literacy skills were measured throughout the nine-month CBE programme using an adaptation of the Early Grade Reading Assessment (EGRA) generated by the Directorate of Research Innovation and Consultancy (DRIC) of the University of Cape Coast. After the transition to formal school, literacy skills were measured using the EGRA tool developed for Ghana in 2013. Both assessments (EGRA-type & EGRA) measure skills based on the assumption that learners acquire literacy via the phonics-based method, using an alphabetic system. Dubeck and Gove (2015:316) found that “literacy skills in alphabetic languages develop along predictable patterns”. As facilitators supporting the CBE programme and teachers in public schools teach literacy acquisition in this way, these assessment tools provide an appropriate measure of literacy skills.

Due to minor differences between the subtask constitution of the EGRA-type and EGRA instruments, only letter-sound identification and reading comprehension tasks could be selected for comparison over time.³ The letter-sound identification test is the most basic test of metalinguistic awareness and is therefore useful to measure when exploring transfer of language independent properties when students transition to a new language. Reading comprehension is useful to measure, as it explores whether students who transition to a linguistically ‘closer’ language are able to acquire vocabulary, semantic/pragmatic (cultural), and syntactic knowledge faster than those transitioning to a distant language.

³ While the Oral Passage Reading Fluency subtask also was comparable within the test instruments used in CBE and in formal school, complications in the data entry for this component of the assessment meant that no data could be reported during the CBE programme. This prevented us from using results from this subtask for the present study.

Children in northern Ghana who undertook the CBE programme and transitioned into primary grade 3 and below made a total of 16 different linguistic transitions. Table 3 shows each transition group, how many learners made each transition, and the degree of linguistic difference between the two languages involved in the transition based on the point at which they differ in classification shown in Figure 1. For this paper, we selected children where there was a significant number represented in the language transition group, as indicated in Table 3. For the measure of linguistic distance, language close denotes Mampruli and Konkomba to Dagbani as well as Birifor and Gruni to Dagaare. Language distant denotes Kusaal to Kasem and Sisaala to Dagaare. As indicated previously, there are more differences in phonological and orthographical features between the languages in the ‘distant’ category than there are in the ‘close’ category.

== Table 3 about here ==

4.3 Estimation Method

For each of the 3 groups of children identified by language matching (same, close and distant), we estimate the proportion of children with zero scores in the two literacy subscales at three points in time. First, we estimate changes in the proportion of children with zero scores in literacy subscales during the CBE programme when mother tongue was their language of instruction. Then, we estimate changes in these proportions as children make the transition into government schools where language of instruction is one of the 6 official Ghanaian languages and some, therefore, change language of instruction. Finally, we estimate the proportion of children with zero scores in literacy subscales during one year in government schools.

We measure changes in literacy over time using the proportion of zero scores. Given differences in assessment subtasks implemented during CBE and formal school, zero scores would give the most accurate measurement of changes in learning overtime. Additionally, zero scores are informative in that they indicate the proportion of students who could not answer a single item correctly, therefore demonstrating the number of students performing at critically low levels.⁴ An absolute change in the proportion of zero scores is equivalent to within-group change (in other words, whether there is a smaller proportion of children achieving zero scores in the literacy subtasks within the specified period). A relative change of zero scores is equivalent to between-group changes over time (in other words, whether the change in the proportion of zero scores for one group is greater than that for the other group over the period)⁵. The relative change is equivalent to a difference-in-difference estimation. We estimate the statistical significance of absolute and relative changes of the proportion of children with zero scores conditional on the covariates included in Table 2.

We hypothesise that:

- a) During the CBE programme, reductions in the proportion of zero scores in the sample are comparable between the three groups of children identified by language matching as these children are all learning in mother tongue.

⁴ While zero scores constitute the focus of our analysis in the presented paper, we also tested changes using mean scores. Results using this measure demonstrated achievement patterns that were highly comparable to those found when applying zero scores.

⁵ Unfortunately, zero scores are just measure of the distribution of literacy scores which limits our ability to provide a more nuanced picture relating to learning changes over time.

- b) As children moved into government schools, the proportion of children with zero scores is higher for children who move to a more distant language. The more distant the language, the harder it is for children to recognise the new local language thus increasing the chances that these children will not be able to respond to the literacy test.
- c) As children spend one year in government schools, the proportion of zero scores declines more rapidly for children who transition to a language of instruction which is closer to their mother tongue and slower for those who transition into a distant language. The more distant the language, the harder it is for children to transfer previous foundational literacy learning from mother tongue into a different local language. Hence a higher proportion of children in the language distant group remain without any foundational skills (i.e. zero scores).

5 Results

Table 4 shows results from the estimated parameters of absolute and relative changes in the proportion of children unable to perform the literacy subtasks by groups of children identified by language matching. We present results according to each of the 3 hypotheses.

Hypothesis A: During the CBE programme, reductions in the proportion of zero scores in the sample are comparable between the three groups of children identified by language matching as these children are all learning in mother tongue.

During the CBE programme, children in the language close and distant categories made significant progress in reducing the proportion of zeros in letter-sound identification, while this was not as marked for children who were in the same language. For instance, the proportion of children with zero scores in letter-sound identification changed from 14 to 11% during the 9 months of the CBE programme for children classified as language same, but reduced from 21 to 6% for children in classified as language close and from 22 to 4% for children classified as language distant. For reading comprehension, we find that the proportion of children with zero scores reduced by 35 percentage points for children in the language same group, by 30 percentage points for children in the language close group, and by 39 percentage points for children in the language distant group. Recall that during the CBE programme, these children were learning in their own mother tongue.

When measuring the relative change in the proportion of children with zero scores in the literacy subtasks between groups, we found that the above differences are not statistically significant. Yet, we remain cautious as some of the estimated differences are relatively large, and given our small sample sizes, we may be incorrectly assuming that progress for all children was similar. Take for instance the reduction in the proportion of zero scores for letter sound identification. For children in the language same group, the proportion unable to complete any of this task during the CBE programme changed by 3 percentage points (from 14 to 11%). However, for children in the language close and language distance groups, the proportions unable to complete this task changed by 15 and 19 percentage points, respectively. These changes are five times larger than for children in the language same group, yet not statistically significant. For reading comprehension, the differences between groups are less marked, with differences between language close and distant and that of language same equivalent to around 5 percentage points (see Table 4). Therefore, we conclude comparable progress for children in reading comprehension according to language matching groups but potentially a larger improve in letter sound identification for children in

the language close and distant groups relative to children in the language same group. In terms of our next hypotheses, this result implies that children in the language close and distant groups manage at least the same progress in these literacy subtasks, if not higher in letter sound identification, than children in the language same group during their time in the CBE programme.

Hypothesis B: As children moved into government schools, the proportion of children with zero scores is higher for children who move to a more distant language.

The second hypothesis states that at the point of transition, children who change into a distant language will underperform in the literacy test as a result of this being a new distant local language. Table 4 shows that at the start of government school the proportion of children achieving zero scores in the literacy subtasks increased for all children, and particularly for those taking the test in a different local language (close or distant from their mother tongue). The proportion of children achieving zero scores in letter sound identification increased from 4% at the end of CBE to 89% for those in the language distant group. For children in the language close group, the proportion who were unable to identify letters or sounds increased from 6% at the end of the CBE to 35% at the start of government school. Similar results are obtained for reading comprehension, where we find that the proportion of children not being able to undertake this task increased from 46% to 62% for the language same group, from 42% to 75% for the language close and from 36% to 96% for the language distant group as they transitioned into government schools.

The relative change shows that compared with language same students, the proportion of children with zero scores in letter sound identification was 24 and 80 percentage points higher for those in language close and language distant groups, respectively. In reading comprehension, the proportion of children with zero scores was 18 and 45 percentage points higher for those in language close and distant groups relative to those in the language same group. In both language close and language distant cases, differences were found to be statistically significant, with larger discrepancies observed for language distant students. Overall, these findings suggest that the second hypothesis is not rejected and that children underperform in these literacy subtasks after nearly 3 months out of school. In particular, those taking a test in a distant language are likely to struggle to complete the test.

Hypothesis C: As children spend one year in government schools, the proportion of zero scores declines more rapidly for children who transition to a language of instruction which is closer to their mother tongue and slower for those who transition into a distant language.

The third hypothesis proposes that children who transition into a distant language will show far slower progress in the literacy subtasks during the course of the academic year than children who transition into a school where the language is the same as the one previously used for teaching in the CBE programme. Table 4 shows that the proportion of children who were able to identify letters and sounds increased by 7 percentage points during one year in government schools for those in the language distant group. More striking, we found no improvement in reading comprehension during one year in government schools for children in the language distant group. Yet, children in the language close and language same groups did show improvements both in letter sound identification and in reading comprehension as shown in Table 4. For example, the proportion of zero scores in letter sound identification

declined by 32 percentage points and in reading comprehension by 49 percentage points for children in the language close group.

In terms of relative changes, there was a difference of 23 and 17 percentage points between the proportion of children from the language close group to have reductions in zero scores in letter sound identification and reading comprehension and the reductions achieved by children from the language same group. For children in the language distant group relative to children in the language same group, we only found a statistically significant increase in the proportion of children achieving zero scores in reading comprehension by the end of the first year in government schools. This result suggests that the lowest performers from the language distant group are unable to make progress in the new local language after transition. Therefore, a higher proportion of children in the language distant group are likely to remain without foundational literacy skills.

These results are illustrated in Figures 2 and 3. These figures show the proportion of children who could not answer a single item on the literacy assessments of letter sound identification and reading comprehension across four time periods, by language category (i.e. language same, language close and language distant). Here we demonstrate visually learning progress in literacy subtasks during the CBE programme (i.e. when students were still learning in their mother tongue) which is comparable for all children. Second, the proportion of non-performers is higher at the commencement of government schooling for those children who transitioned into either a close or distant language, with the latter group of students demonstrating the biggest gap in learning relative to language same students. Lastly, the percentage of non-performers reduces by the end of government school to comparably low proportions for both language same and language close students. For language distant students, however, a considerable gap in performance remains, with minimal reductions in zero-score performers observed. In other words, while the proportion of language close students without any skills in these subtasks appear to follow the same trajectory as that of students from the language same group, this is not the case for the proportion of students in the language distant group, particularly once learning in a distant language takes place in government schools.

== Figures 2 & 3 here ==

6 Conclusions

The study was motivated by the learning challenges faced by millions of children who live in multilingual environments of the Global South who, even when they have access to mother tongue education, over the course of their schooling are likely to change language of instruction. Previous studies show that these children are likely to fall behind in foundational literacy acquisition (Akyeampong *et al*, 2007; Ball, 2011; AUTHORS). We were also motivated by the fact that not many studies focus on transitions between indigenous local languages due to a greater research focus on transitions between local languages and dominant western or national languages. Finally, we were further motivated by the linguistic diversity that many children experience while in school.

Using data from the evaluation of the CBE programme in Ghana, we highlight three main findings supported by the quantitative evidence. First, during the CBE programme, we showed that the programme was supporting children's progress in letter sound identification and reading comprehension regardless of the linguistic categorisation using language

matching. Second, the proportion of children who are then unable to identify letters and sounds or unable to read and comprehend increases when children are tested in a new language which is different to the one previously utilised for learning. The key finding to highlight here is that when the test is done in a language that is more distant to the one previously experienced by children, the proportion of children unable to achieve any score in these foundational literacy skills is higher than when the linguistic distance between languages is close. Third, once in government schools, and once learning is taking place in a different local language, children learning in a distant language are unable to catch up in reading comprehension. For these children, we are unable to find evidence that some were able to master reading comprehension during their first year in government schools.

Identification and targeting of students who transition into linguistically distant learning environments is an important practical implication of our study. In many multilingual contexts of the Global South, there are classes in which children of numerous different language groups are present (Mackenzie & Walker, 2013; Owalabi, 2013). In such classes, the language of instruction is likely to be different from the mother tongue of many learners. Understanding the linguistic distance between different languages is important to identify learners who are more likely to struggle and therefore require additional resources for literacy learning.

Although our study focused on linguistic distance, there are other important aspects of language not included in this paper, which are areas for further research. For instance, colonial languages may be viewed as apolitical tools that unify countries or as responsible for structural inequality, discrimination, and self-hatred among disadvantaged groups (Benson, 2002). Sociolinguistic factors could also influence a successful transition into a diverse linguistic environment. These should be considered within a broader definition of language matching which is beyond only linguistic distance, as sociolinguistic aspects impact on a child's culture and identity.

It is important to note several limitations of this study. One of these concerns our inability to understand the linguistic make-up of the classroom, whether children transition into a class where the majority of students are using the same language or a distant language and/or the linguistic competencies of the teachers. Teachers living and working within a geographical area around a school are likely to be familiar with the languages spoken even if they may not be fluent in all of them. Therefore, they are likely to know the main differences between all languages of the region (close or distant), and may even use students themselves to translate between languages in the classroom (code-switching). However, utilising code-switching in multilingual classrooms can also favour particular languages in which the teacher may be more fluent. This practice, and the favouring of some languages over others, may also be due to sociolinguistic factors.

There are also important limitations of the assessment tools utilised here that further research should consider. First, we are unable to capture whether children may react differently to a test, meaning we are unable to identify any extraneous variables that affected test performance. Such contextual elements could include their relationship to the tester, and the quality and salience of the story read. Hence, our indicator of zero scores may reflect these biases. In addition, the letter-sound identification task should place sound-letter correspondences found in both languages (child's mother tongue and language of instruction) on the first line, and those that are not found in both languages on the following lines (this was not the case with the EGRA-type and EGRA tools). This would allow researchers to

track improvement, or indeed a decline in alphabetic and metalinguistic awareness once the learner has transitioned into another language. It would also allow researchers to track whether students are acquiring the second language sound-letter awareness, or whether their acquisition rate has plateaued. If this were the case, it would show that exposure and phonological closeness are vital and should be considered prominently when matching languages. Unfortunately, we are unable to investigate these important issues in this paper.

The study also suffers from limitations in terms of external validity due to sample attrition and the focus on those at the lower end of the learning distribution. We are unable here to suggest that our results will hold for all learners of the CBE programme. We have only included learners for whom we have full information, remained in the programme for two consecutive years and were willing to participate in the study. It is likely that these learners have different motivations from other learners, which we are unable to capture here. We also acknowledge that we are testing the idea of language matching to track progress using the proportion of zero scores. Children who achieve a zero score are usually the lowest performers in such a task and we focus on them for reasons of equity. Our results are unchanged when we perform the models using the average results in the literacy tests, but we are unable to test for whole distributional changes in these scales for children who move to distant languages relative to those who continue to learn in the same language. These are important issues which remain to be assessed empirically when samples sizes are large enough to enable these types of analyses.

Regardless of these limitations, to our knowledge this paper is among the first to use linguistic distance to test whether children will be able to maintain literacy skills previously acquired while learning in mother tongue. Language matching could serve as a tool for identification and targeting, and also as an ‘intermediary’ step to assign children to the school where they will perform best whilst resources are being produced in all languages, when school choice is available. It could also be used to prioritise mother tongue resource development by identifying the most distant languages for immediate action. In the post-COVID-19 situation, it could be a tool to provide support to linguistic minorities who are likely to re-enter into different linguistic environments once schools reopen.

References

- Adongo, HA. (2018). *Aspects of Gurene Phonology*. PhD Thesis in Linguistics, University of Ghana.
- Aikhenvald, AY., Dixon, RMW. (2017). *Introduction: Linguistic Typology – Setting the Scene*. In Aikhenvald, AY., Dixon, RMW. (Eds.) *The Cambridge Handbook of Linguistic Typology*. Cambridge University Press. (pp.1-35). DOI: <https://doi.org/10.1017/9781316135716>
- Akyeampong, K., Carter, E., Higgins, S., Rose, P., & Sabates, R. (2019). *Understanding Complementary Basic Education in Ghana: Investigation of the Experiences and Achievements of Children after Transitioning into Public Schools*. Report for Department of International Development (DFID) Ghana office (November 2018). REAL Centre, University of Cambridge. Retrieved from https://www.educ.cam.ac.uk/centres/real/downloads/Policy%20papers/CBE%20-%20Final%20Impact%20Evaluation%20-%20REAL%20RP_V2.pdf
- Akyeampong, K., Djangmah, J., Oduro, A., Seidu, A. and Hunt, F. (2007). *Access to Basic Education in Ghana: The Evidence and the Issues*. CREATE Country Analytical Report. University of Sussex: Consortium for Research on Educational Access, Transitions and Equity. <http://www.create-rpc.org>
- Ball, J. (2011). *Enhancing Learning of Children from Diverse Language Backgrounds: Mother tongue-based bilingual or multilingual education in the early years*. Paris: UNESCO. Retrieved from <http://unesdoc.unesco.org/images/0021/002122/212270e.pdf>
- Benson, C. (2002). *Real and Potential Benefits of Bilingual Programmes in Developing Countries*. *International Journal of Bilingual Education and Bilingualism*, 5 (6), (pp.303-317). DOI: <https://doi.org/10.1080/13670050208667764>
- Blommaert, J. (2010). *The Sociolinguistics of Globalisation*. Cambridge University Press.
- Bodomo, A. (1997). *The Structure of Dagaare*. Stanford Monographs in African Languages. Stanford: Center for the Study of Language and Information.
- Bourdieu, P. (1977). *Outline of a Theory of Practice*. Cambridge University Press.
- Brock-Utne, B. (2010). Research and policy on the language of instruction issue in Africa. *International Journal of Educational Development* 30: 636–645. [doi:10.1016/j.ijedudev.2010.03.004](https://doi.org/10.1016/j.ijedudev.2010.03.004)
- Campbell, L. (2003). In Joseph, BD., Janda, RD (Eds.). *The Handbook of Historical Linguistics*. Blackwell
- Cheung, H., Chen, HC., Lai, CY., Wong, OC., Hills, M. (2001). *The Development of Phonological Awareness: Effects of Spoken Language Experience and Orthography*. *Cognition*, 81(3) (pp.227–241) DOI: [https://doi.org/10.1016/S0010-0277\(01\)00136-6](https://doi.org/10.1016/S0010-0277(01)00136-6)
- Crothers, JH., Lorentz, JP., Sherman, DA., Vihman, MM. (1979). *Handbook of Phonological Data From a Sample of the World's Languages*. Stanford

Cummins, J. (1979). *Linguistic Interdependence and the Educational Development of Bilingual Children*. Review of Educational Research, 49, (pp.222–251). DOI: <http://dx.doi.org/10.3102/00346543049002222>

Cummins, J. (1981). *The Role of Primary Language Development in Promoting Educational Success for Language Minority Students*. In California State Department of Education. (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3–49). Los Angeles, CA: Evaluation, Dissemination and Assessment Center, California State University, Los Angeles.

Daniel, M. (2010). *Linguistic Typology and the Study of Language*. In Song, JJ (Ed.). *The Oxford Handbook of Linguistic Typology*. Oxford University Press: Oxford. DOI: <https://doi.org/10.1093/oxfordhb/9780199281251.013.0004>

Dubeck, MM., Gove, A. (2015). *The Early Grade Reading Assessment (EGRA): Its Theoretical Foundation, Purpose, and Limitations*. International Journal of Educational Development 40 (pp.315-322). DOI: <https://doi.org/10.1016/j.ijedudev.2014.11.004>

Ethnologue: Ghana. (2018). <https://www.ethnologue.com/country/GH>

Hartell, RL. (1993) *Alphabets of Africa*. Dakar, Senegal: UNESCO Regional Office

Kerswill, P., Cheshire, J., Fox, S., Torgersen, E. (2013a). *English As A Contact Language: The Role of Children and Adolescents*. In Hundt, M., Schreier, D. (eds.). (2013). *English As A Contact Language*. CUP (pp.258-282)

Kerswill, P. (2013b). *Identity, Ethnicity and Place: The Construction of Youth Language in London*. In Auer, P., Hilpert, M., Stukenbrock, A., Szmerecsanyi, B. (eds.). (2013). *Space in Language and Linguistics*. De Gruyter (pp.128-164)

Kerswill, P. (2014). *The Objectification of ‘Jafaican’: The Discoursal Embedding of Multicultural London English in the British Media*. In Androutsopoulos, J. (ed.). (2014). *The Media and Sociolinguistic Change*. Walter de Gruyter. Berlin.

Kuch, L. (1993) *Collected Language Notes, No. 21: The Phonology of Birifor*. Institute of African Studies, University of Ghana

Mackenzie, PJ., Walker, J. (2013). *Global Campaign For Education Policy Brief: Mother Tongue Education: Policy Lessons For Quality and Inclusion*. Global Campaign for Education. http://www.campaignforeducation.org/docs/reports/GCE%20Mother%20Tongue_EN.pdf

Omniglot. (2018). www.omniglot.com Accessed 12/06/2018

Owalabi, K. (2013). *Using Africa’s Indigenous Languages as Tools for Sustainable Development*. In *Language, Literature and Culture in a Multilingual Society (A Festschrift for Abubakar Rasheed)*. Eds by Ndimele, A., Ahmed, M., Yakasai, H. Port Harcourt: M&J Communications (pp.5-32).

Reder, F., Marec-Breton, N., Gombert, JE., Demont, E. (2013). *Second-language Learners’ Advantage in Metalinguistic Awareness: A Question of Languages’ Characteristics*. British Journal of Educational Psychology, 83(4) DOI: <https://doi.org/10.1111/bjep.12003>

Sherris, A., Sulemana, OS., Alhassan A., Abudu, G., Karim AR. (2014). *School for Life in Ghana: Promoting Literate Opportunities for Rural Youth*. Journal of Multilingual and Multicultural Development, 35:7 (pp.692-708) DOI: <https://doi.org/10.1080/01434632.2014.908891>

Trudell, B. (2009). Local-language literacy and sustainable development in Africa. *International Journal of Educational Development*, 29: 73-79. DOI: <https://doi.org/10.1016/j.ijedudev.2008.07.002>

Table 1: Phonological, orthographic, and sociolinguistic differences between transition groups selected for analysis. The first four transitions below were grouped together as ‘close’ language transitions, and the last two transitions below were grouped together as ‘distant’ language transitions.

Transition	Key Differences Between the Languages
Mampruli → Dagbani	<p>Phonology: Mampruli has plosive palatal pair (Dagbani nil), Mampruli has ejective /gʼ/ and Dagbani has /g/. Dagbani has an extra pair of affricates /ch/ and /dz/. Dagbani has velar fricative (Mampruli nil). Mampruli has alveolar trill /r/ (Dagbani nil). Mampruli has palatal approximant. Dagbani has /y/, (Mampruli nil). Dagbani has complex fricatives /tʃ/ and /dʒ/ (Mampruli nil). Dagbani has 3 unmarked tones (Mampruli nil). Dagbani has /i/ (Mampruli nil)</p> <p>Orthography: ‘r’ can represent /d/ in Mampruli as well as /r/. /ŋ/ is represented as ‘ng’ in Mampruli and ‘ŋ’ in Dagbani. /o/ is represented as ‘o’ in Mampruli and ‘o’ in Dagbani. Dagbani uses ‘i’ and ‘e’ to also represent vowel /i/, as well as /i/ and /e/ respectively – Mampruli only has /i/ ‘i’ and /e/ ‘e’.</p> <p>Despite phonological differences, they are considered sister languages and dialects, sharing 95% of the vocabulary. High contact.</p>
Brifo → Dagaare	<p>Phonology: Dagaare has voiced alveolar fricative /z/ (Brifo nil). Brifo has /y/ (Dagaare nil). Brifo has vowels /o/ and /u/ (Dagaare nil)</p> <p>Orthography: The only orthographic differences are related to the phonological differences above.</p> <p>Brifo is seen as a dialect of Dagaare, which is evidenced through the very small differences in phonology and orthography</p>
Gruni → Dagaare	<p>Phonology: Gruni has a glottal stop (Dagaare nil). Gruni has a voiced velar fricative /ŋ/ (Dagaare nil). Gruni has /y/ (Dagaare nil). Gruni has /u/ and /o/ (Dagaare nil). Gruni has complex nasal approximant /mw/ (Dagaare nil). Gruni has /y/ (Dagaare nil). Dagaare has palatal plosive pair /c/ and /j/ (Brifo nil). Dagaare has palatal approximant /j/ (Brifo nil). Dagaare has 3 tones and Gruni has 2 tones.</p> <p>Orthography: /ŋ/ is represented as ‘ŋ’ in Gruni and ‘ng’ in Dagaare. /ŋm/ is represented by ‘ŋm’ in Gruni and ‘ngm’ in Dagaare.</p> <p>Gruni is slightly less related than Mampruli→Dagbani and Brifo→Dagaare, but the similarities in overall phonology and orthography show significant similarities.</p>
Konkomba → Dagbani	<p>Phonology: Dagbani has labiodental voiced /v/ (Konkomba nil). Dagbani has /z/ (Konkoma nil). Dagbani has /i/ (Konkomba nil). Konkomba has /o/ (Dagbani nil).</p> <p>Orthography: /g/ is represented as ‘g’ in Konkomba and ‘g’ or ‘y’ is Dagbani. /s/ is represented as ‘s’ in Konkomba and ‘s’ or ‘j’ in Dagbani. /tʃ/ is represented as ‘sh’ in Konkomba and ‘ch’ in Dagbani. /dʒ/ is represented as ‘j’ in Konkomba and ‘dz’ in Dagbani. Dagbani uses ‘i’ and ‘e’ to also represent vowel /i/, as well as /i/ and /e/ respectively – Konkomba only has /i/ ‘i’ and /e/ ‘e’.</p> <p>Konkomba is slightly less related than Mampruli→Dagbani and Brifo→Dagaare, but the similarities in overall phonology and orthography show significant similarities.</p>
Kusaal → Kasem	<p>Phonology: Kusaal has glottal stop (Kasem nil). Kusaal has labiodental voiced fricative /v/ (Kasem nil). Kusaal has glottal fricative /h/ (Kasem nil). Kusaal and Kasem has no crossover in complex consonants – Kusaal /kp/, /gb/, /w/ and Kasem /tʃ/, /dʒ/. Kusaal has vowel /u/ (Kasem nil). Kusaal has a tense/lax vowel distinction but is not orthographically marked (Kasem nil). Kusaal has tones two tones that distinguish lexical and semantic constructions.</p> <p>Orthography: ‘u’ represents /u/ in Kusaal and /u/ in Kasem – ‘u’ is represented by /u/ in Kusaal. ‘e’ represents /e/ and /e/ in Kusaal and ‘e’ represents /e/ in Kasem. ‘o’ represents /o/ in Kasem and /o/ and /o/ in Kusaal. /ŋ/ can be represented by ‘ŋ’ or ‘ny’ in Kusaal and just ‘ŋ’ in Kasem</p> <p>There are more structural and significant differences in the phonology of these two languages (complex consonant inventory, tone and lax/tense vowels), as well as some disparities in orthography.</p>
Sisaala → Dagaare	<p>Phonology: Dagaare has a palatal plosive pair /c/ and /j/ (Sisaala nil). Dagaare has palatal approximant /j/ (Sisaala nil). Differences in complex consonants – Sisaala has /tʃ/, /dʒ/. Sisaala has /y/ (Dagaare nil). Sisaala has /i/ and /u/ (Dagaare nil). Dagaare has 3 tones with high and low marked (Sisaala nil), Dagaare has nasal vowel contrast (Sisaala nil).</p> <p>Orthography: ‘u’ represents /u/ in Sisaala and /u/ in Dagaare – ‘u’ is represented by /u/ in Sisaala. ‘e’ represents /e/ and /e/ in Sisaala and ‘e’ represents /e/ in Dagaare. ‘o’ represents /o/ in Dagaare and /o/ and /o/ in Sisaala.</p> <p>There are more structural and significant differences in the phonology of these two languages (tone and nasality), as well as some disparities in orthography.</p>

Table 2: Descriptive statistics of key variables for analysis (full and estimation samples)

	Full Sample Start of CBE (1)	Estimation Sample (2)	Language Same (3)	Language Close (4)	Language Distant (5)
No. Available	2360	380	175	177	28
Sample for descriptive statistics	2360	380	175	177	28
% Language same transition at government school	52	46.1*	NA	NA	NA
% Female	47	42.9	42.3	41.9	53.6
% Zeros-overall literacy score	14	18.4*	20.6	17.5	10.7
% Zeros- overall numeracy score	6.3	10.0**	12.6	9.0	0.0
% Home learning materials (e.g.books)	81.6	84.0	74.3	92.7*	89.3
% Engaged in reading activity at home	71.2	70.9	62.3	77.7**	82.1*
% Work outside of home	47	43.5	29.1	54.8	60.7
Average missed days at school ^a	0.9(1.2)	0.8 (1.1)	0.8 (1.1)	0.9 (1.0)	0.8(0.9)
Average age	10.1(2.2)	9.4 (1.9)*	9.3 (1.9)	9.4 (2.2)	9.8(1.5)
Average household size	10.1(6.0)	10.6(6.1)	9.5(5.4)	11.7(6.8)**	9.9(4.26)
% Low wealth	27.4	19.4	19.1	20.6	14.3
%Mid low wealth	25.8	30.7**	27.0	33.7	28.6
% Mid high wealth	24.3	28.6**	32.1	25.1	28.6
% High wealth	22.5	21.3	20.8	20.1	28.6

Source: Authors' calculations based on CBE Monitoring and Evaluation 2016-2018.

Notes: Standard deviations shown in parentheses. Binominal logistic regression was used to estimate differences between the full (1) and estimation samples (2). Multinomial logistic regression was used to estimate differences between language same (3), language close (4) and language distant (5) subsamples. Within these tests, 'language same' was used as the referent group. In both binominal and multinomial regression examples, the reference category for the categorical variable 'wealth' was 'low wealth'. Asterisks *, **, *** indicate statistical significance at 5, 1 and 0.1% level.

^a This figure represents the average number of days missed in the five days preceding data collection. It was determined by the question: *In the last five days of school, how many days have you missed school?*⁶

Table 3: Information on all linguistic transitions to grade 3 and below made by CBE students to start 2017-18 academic year in government schools.

Transition Group	Linguistic Distance	Number of Students
Dagaare → Dagaare	Same Language	27
Dagbani → Dagbani	Same Language	74
Ewe → Ewe	Same Language	13
Gonja → Gonja	Same Language	37
Kasem → Kasem	Same Language	6
Twi → Twi	Same Language	18
Mampruli → Dagbani	Language Close	46
Birifor → Dagaare	Language Close	67
Gruni → Dagaare	Language Close	6
Konkomba → Dagbani	Language Close	60
Birifor → Dagbani	Language Close	1
Dagaare → Dagbani	Language Close	1
Kusaal → Kasem	Language Distant	23
Dagaare → Kasem	Language Distant	1
Sisaala → Dagaare	Language Distant	5
Dagaare → Twi	Language Distant	1

Note: Transitions are ranked on the linguistic difference between the two languages involved in each transition, as described in Table 1 and Figure 1. The transitions highlighted in red are those that will be taken forward for analysis, based on number of learners who made the transition.

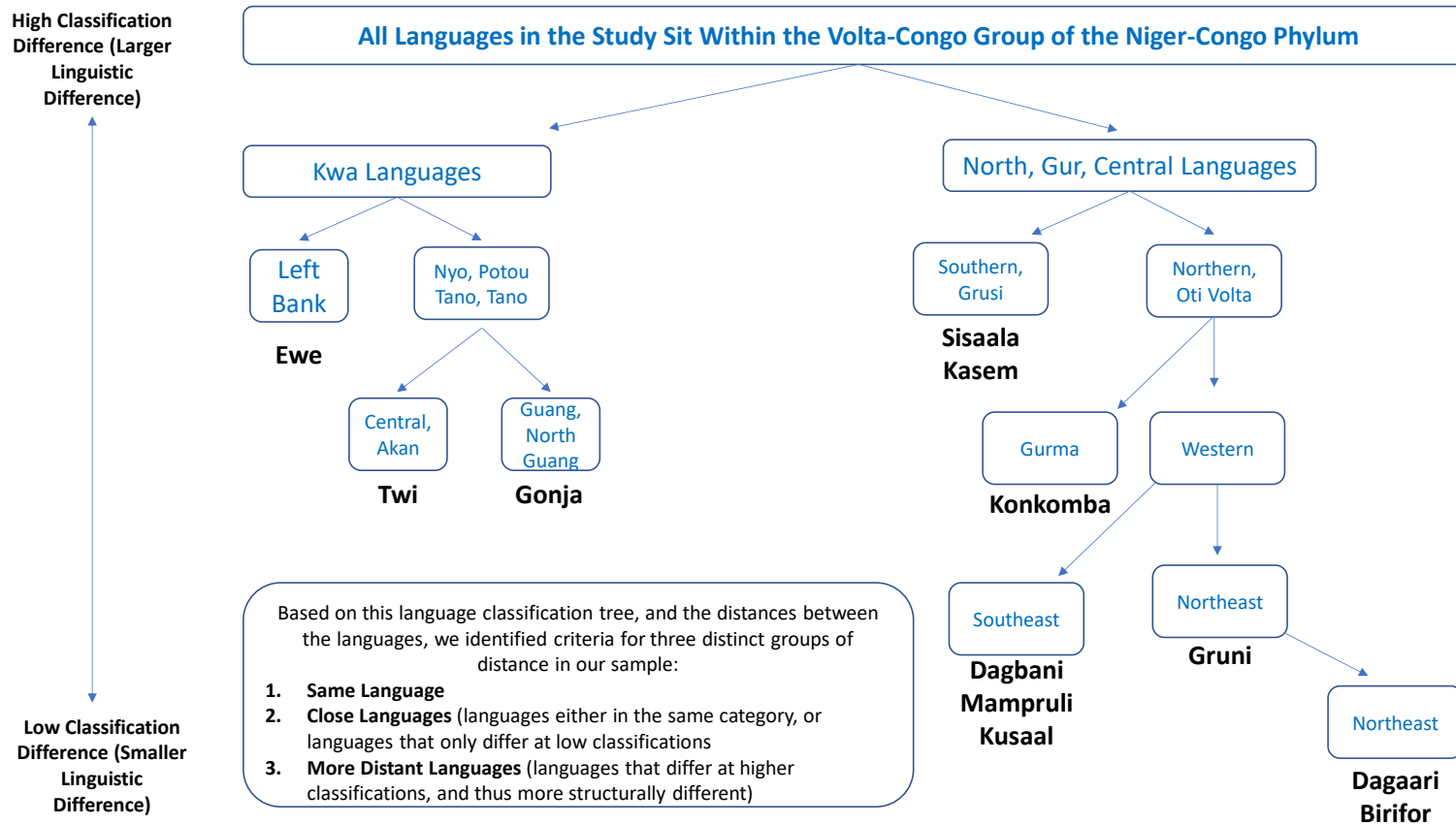
Table 4: Estimated change (absolute and relative) in zero scores in literacy subtasks by language matching groups

CBE Programme (2016-17 academic year)						
	Start of CBE	End of CBE	Absolute % change	Sig	Relative % change (1)	Sig
Letter sound id (Same)	14.29	11.43	-2.86			
Letter sound id (Close)	20.67	5.59	-15.08 **		-12.22	
Letter sound id (Distant)	22.29	3.57	-18.72 **		-15.86	
Reading comp (Same)	81.14	46.29	-34.85 **			
Reading comp (Close)	71.91	42.13	-29.78 **		5.07	
Reading comp(Distant)	75.00	35.71	-39.29 **		-4.44	
Transition to Government School (June 2017 - September 2018)						
	End of CBE	Start of gov. school	Absolute % change	Sig	Relative % change (1)	Sig
Letter sound id (Same)	11.43	17.14	5.71			
Letter sound id (Close)	5.59	35.20	29.61 **		23.90 **	
Letter sound id (Distant)	3.57	89.29	85.72 **		80.01 **	
Reading comp (Same)	46.29	61.71	15.42 **			
Reading comp (Close)	42.13	75.28	33.15 **		17.73 **	
Reading comp(Distant)	35.71	96.43	60.72 **		45.30 **	
Government school (2017-18 academic year)						
	Start of gov. school	End of gov. school (Yr1)	Absolute % change	Sig	Relative % change (1)	Sig
Letter sound id (Same)	17.14	8.57	-8.57			
Letter sound id (Close)	35.20	3.35	-31.85 **		-23.28 **	
Letter sound id (Distant)	89.29	82.14	-7.15		1.42	
Reading comp (Same)	61.71	29.70	-32.01 **			
Reading comp (Close)	75.28	26.40	-48.88 **		-16.87 *	
Reading comp(Distant)	96.43	96.43	0.00		32.01 **	

Source: CBE Evaluation Data. Notes: (1) Relative change is measure against language same.

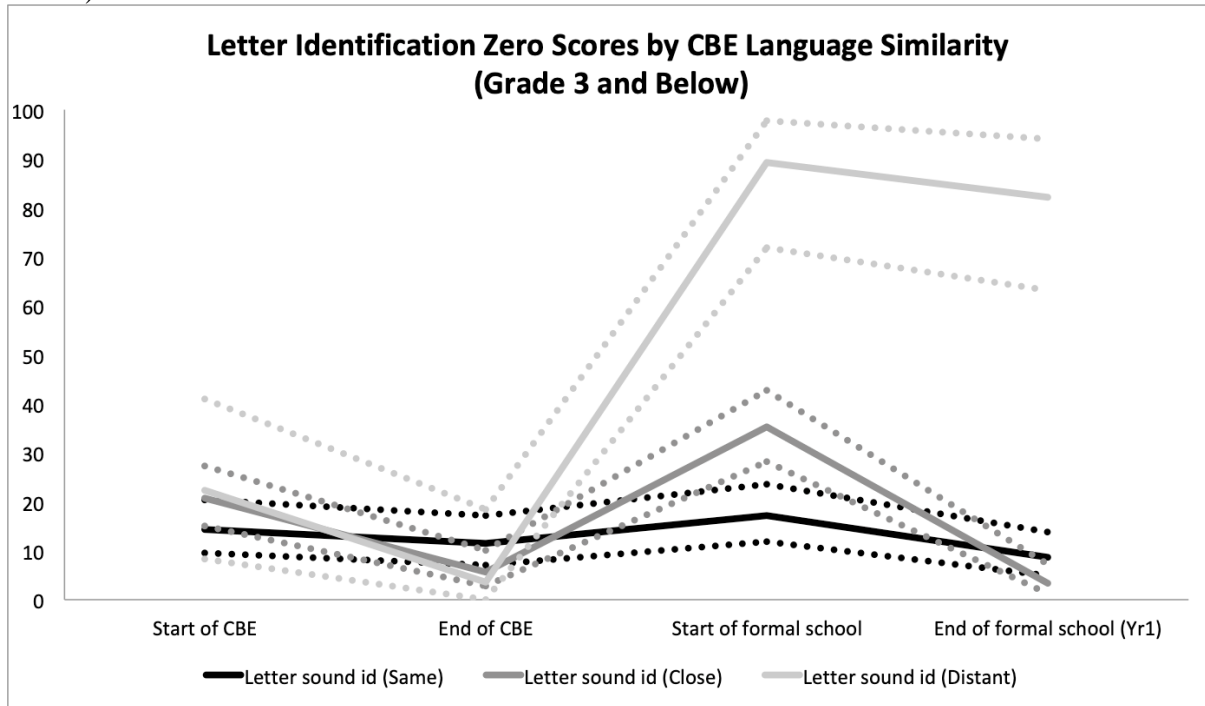
Asterisks (*, **) indicate statistical significance at 5 & 1% level, respectively. Models condition for child and household level factors identified in Table 2.

Figure 1: Language classification tree containing the languages involved in the CBE programme



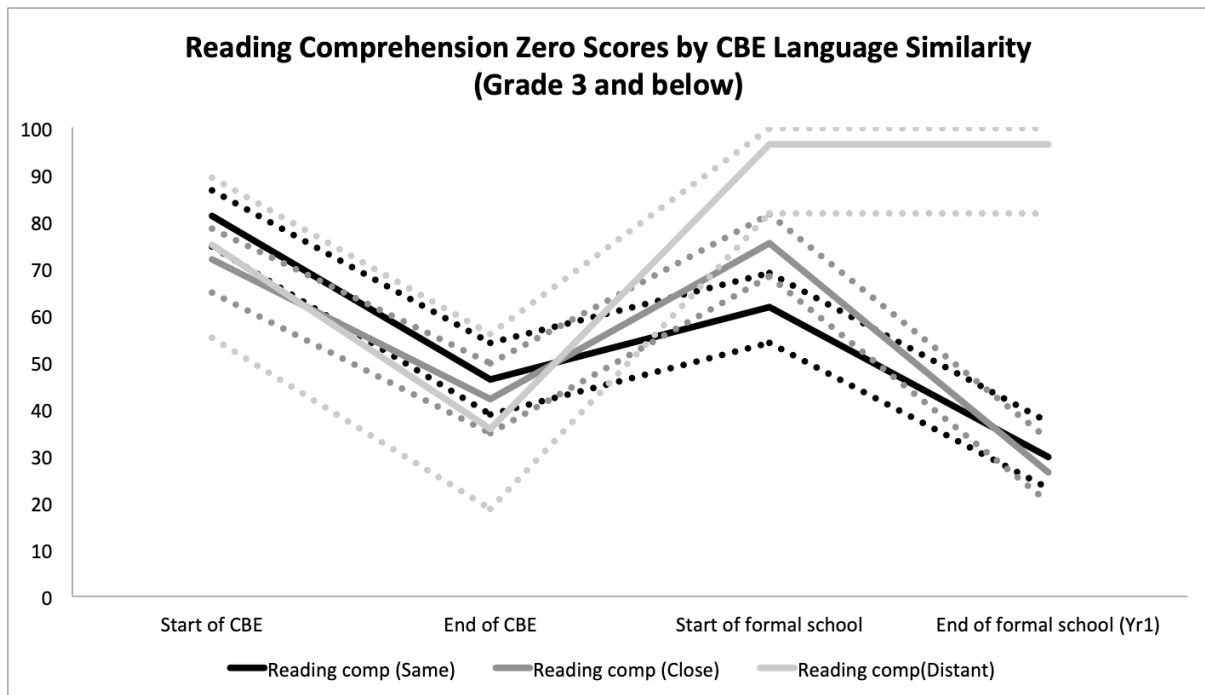
Notes: Authors own classification. Adapted from Adongo (2018) and Ethnologue (2018)

Figure 2: Zero scores for letter-sound identification by language similarity (Grade 3 and below)



*Dotted lines indicate the 95% confidence intervals

Figure 3: Zero scores for reading comprehension by language similarity (Grade 3 and below)



*Dotted lines indicate the 95% confidence intervals