University of St Andrews

School of Economics and Finance Discussion Papers

Mother Tongue Reading Materials as a Bridge to Literacy

Margaret Leighton

School of Economics and Finance Discussion Paper No. 2101
29 Jun 2021
JEL Classification: I21, I28, I25, I24
Keywords: primary education, language of instruction, literacy, intervention

School of Economics and Finance
Online Discussion Paper Series
issn 2055-303X
http://ideas.repec.org/s/san/wpecon.html
info: econ@st-andrews.ac.uk

Mother Tongue Reading Materials as a Bridge to Literacy*

Margaret Leighton†

June 29, 2021

Abstract

Children whose mother tongue is different from the language of instruction at school face significant challenges in developing literacy skills. One educational approach to address this favours immersion in the language of instruction, while another favours a transitional period of bilingual education. This paper evaluates the impact of an early primary school literacy intervention. In addition to multi-faceted literacy support, the programme included a component of transitional bilingual education for students whose mother tongue is not the usual language of instruction. Over two years, the intervention raised literacy scores in the language of instruction by +0.44 sd, and literacy in mother tongue for minority language speakers by +0.68 sd. Both native speakers and language learners saw equivalent gains in literacy in the language of instruction, despite language learners also seeing large gains in mother tongue literacy. Our findings suggests that a light-touch transitional bilingual component does not interfere with literacy acquisition in the language of instruction; however, neither is it likely to close gaps between native speakers and language learners.

*This paper would not have been possible without the excellent team from Save the Children who carried out the intervention and data collection: Deba Chakma, Akter Hossain, Holly-Jane Howell, Akidul Islam, Patrick Mroz-Dawes, Kirsten Mucyo, Dalim Tripura, as well as Aliza Nahar, Azmal Hossain, Zulfiqar Bushra, Islam Shahin, Mo Mo Shey and Shishir Mittro. The project on which the paper is based was made possible through the participation of the Government of Bangladesh, participating schools and staff, parents and children. I am indebted to Jakob Henning, Angus Holford, Himangshu Kumar, Lorenzo Neri and Jamin Speer, as well as seminar participants at the University College London Institute of Education, the Royal Economic Society, the Scottish Economic Society and the European Society for Population Economics for very helpful comments and discussions. The data used in this paper are drawn from a pilot study of the Children's Book Initiative in Bangladesh. The intervention and the data collection were carried out by Save the Children UK and Save the Children in Bangladesh. The collaboration between the author and Save the Children was facilitated through funding from the Save the Children Research and Technical Fund and the University of St Andrews' Impact and Knowledge Exchange fund. This study has been approved by St Andrews' University Teaching and Research Ethics Committee, approval code: EC14847. Conflict of interest statement: the author worked as a paid consultant to Save the Children during the baseline study for the intervention, and has also worked as a paid or unpaid consultant on other Save the Children monitoring and evaluation reports.

†University of St Andrews: mal22@st-andrews.ac.uk

1 Introduction

It is estimated that nearly 40% of the world's population cannot access education in their mother tongue (Walter and Benson (2012)). Primary school children whose mother tongue is not the language of instruction face a double hurdle on the road to academic achievement: learning the curriculum, and learning the language in which the curriculum is taught. Approaches to integrating language learners into unilingual education systems broadly fall into two categories: immersion, in which they are taught immediately in the new language, and transitional bilingual education, in which they are taught in both their mother tongue and the new language for some period of time. Proponents of immersion claim that mother tongue and dominant language instruction are substitutes, while those who favour bilingual education argue they are complements.

This paper evaluates the impact of a literacy intervention piloted by Save the Children in a multi-lingual region with a unilingual school system. The Bangladesh Children's Book Initiative (CBI) sought to promote literacy through storybooks, and worked throughout the book chain to achieve this: developing and publishing high quality, age-appropriate storybooks, training teachers and school leaders in storybook use, engaging with communities, and seeking to influence policy at a national scale. The programme, which targeted primary school grades 1-3 with two years of intervention, was randomly assigned to 51 schools in the Khagrachari District of Bangladesh.

A particular focus of the intervention was to ensure that books were produced in bilingual editions combining the language of instruction with local languages. Teachers were also trained to support speakers of non-dominant languages to read in their mother tongue. For minority language speakers this was a big departure from business-as-usual: speakers of these languages do not normally encounter their mother tongue in the classroom. Because of this, speakers of minority languages effectively received an enhanced intervention, with respect to their dominant language-speaking peers: along with the core components of the intervention described above, these children were also exposed to transitional bilingual education for the first time.

Did this mother tongue component compete with, or complement, literacy development in the language of instruction? The multilingual setting (which includes both native speakers of the language of instruction as well as minority language speakers), combined with the randomised design of the pilot, allows us to investigate this. First, we estimate the average

treatment effect of the programme on literacy development, both in the language of instruction (for all students) and in mother tongue (for minority language speakers). Second, we test whether minority language students in treatment schools, who received transitional bilingual education in addition to other intervention components, gained or lost ground in dominant language literacy with respect to their peers in control schools.

Looking across all students, the intervention led to large and significant increases in literacy in the language of instruction (+0.44 sd). Minority language speakers also saw substantial gains in literacy in their mother tongue (+0.68 sd). To test whether these gains in mother tongue literacy came at a cost to minority language students' progress in the language of instruction, we estimate a difference-in-differences equation comparing minority language and dominant language speakers, across treatment and control. We find no difference in the dominant language treatment effect by language group: despite making substantial gains in literacy in their mother tongue, minority language speakers saw equivalent growth in their dominant language literacy skills as did native speakers of the dominant language.

This paper makes two contributions. First, we report the effectiveness of a multi-faceted literacy intervention that underwent rigorous experimental evaluation. Compared to many studies in the literature, the effect sizes estimated here are large: in a recent systematic review of randomised experiments in primary schools in developing countries, McEwan (2015) calculates a mean effect size of 0.123 for teacher training programmes, and 0.078 for provision of instructional materials. The costs of the intervention were not trivial, averaging 50.41 GBP per child. As a small pilot study, fixed costs represent a relatively large share of these: back-of-the-envelope calculations suggest the marginal cost is closer to 17.21-34.50 GBP per child.

Second, our results show that a modest element of transitional bilingual education does not hamper minority language students' acquisition of literacy skills in the dominant language. This is an important finding in light of the debate on mother tongue education. While our results do not suggest that bilingual education *accelerated* literacy in the dominant language (as would be the case if mother tongue education was a strong complement to literacy in another language), we find no evidence that bilingual education delayed such literacy either (as it would have if the two were substitutes). Fears that mother tongue education necessarily competes with the standard curriculum are, in our context, unfounded.

Previous work has shown that estimated treatment effects can vary substantially with

the details of programme implementation (Kerwin and Thornton (2021)): it would therefore be incorrect to conclude that our findings on mother tongue education would hold for transitional bilingual education programmes in general. Nevertheless, this study makes a novel contribution to the field by presenting evidence on a mother tongue curriculum supplement that is modest with respect to the curriculum as a whole: while the Children's Book Initiative itself is relatively intensive, the provision of mother tongue teaching materials is a minor component: one that, in theory, could be added or removed from this or similar interventions at little cost. This sets our study apart from previous work which has experimentally evaluated literacy interventions with a mother tongue or transitional bilingual component in developing country contexts. The few studies which have done so to date are based on programmes fully or substantially delivered in the mother tongue (see, for example, Piper et al. (2018), Kerwin and Thornton (2021) and Laitin et al. (2019), discussed in more detail in below). Our study is therefore a useful benchmark for policy-makers considering light-touch mother tongue or local language curriculum supplements.

The remainder of the paper proceeds as follows. Section 2 introduces multilingual education, with a focus on mother tongue or transitional bilingual education, and reviews recent work in this area. Section 3 describes the intervention and data collection, and gives an overview of the data. Section 4 sets out the empirical approach. Section 5 presents the main results, while Section 6 presents extensions and robustness checks. Section 7 provides a short discussion, including cost effectiveness estimates. Section 8 concludes.

2 Multilingual education

Historically, there has been considerable debate in the education literature on whether immersion or bilingual education provides a more effective early educational experience for children whose native language is not the language of instruction at school. Transitional bilingual education[1] offers children the opportunity to first learn to read in their native language, a skill which, theoretically, can then be more easily transferred to the new language. Immersion, on the other hand, requires students to learn the language of instruction quickly in order to access the curriculum. Bilingual education is therefore seen by proponents as a stepping stone to literacy, while its detractors argue that it sends students on an unnecessary

[1] Bilingual education used in this way is often referred to as transitional, to distinguish it from other forms of bilingual education, e.g. second or third language acquisition in secondary school.

academic detour, delaying progress in other areas.

Much of the literature on bilingual versus immersion education comes from the US, where the focus has been on integrating Spanish-speaking children into mainstream English education. While the superiority of either approach has been the source of considerable debate (see Slavin and Cheung (2005) for an overview), recent evidence from methodologically rigorous studies finds little enduring difference between the two (Slavin et al. (2011), Chin et al. (2013)). Based on this evidence, policy advice from the US emphasises the importance of education quality, rather than the immersion/bilingual dichotomy, as a primary driver of English language learner success (Chin (2015)).

This conclusion leaves many open questions regarding language of instruction in developing countries. In many developing countries, the goal of providing quality education is hampered by scarce resources, from classroom time and teacher skills, to the availability of quality teaching and learning materials. Furthermore, while the US literature has focussed on the integration of Spanish speakers into English schools, the linguistic landscape in many other countries is much more complex. In areas with multiple local languages, mother-tongue bilingual education is considerably more challenging, as it requires class groups and teachers for each language. Furthermore, many countries juggle multiple official languages, with students moving through two or more languages of instruction during their education. Contexts such as this resist one-size-fits-all recommendations on bilingual versus immersion education; however, limited resources mean that a better understanding of the tradeoffs between the two could have significant impacts on learning.

The evidence on mother tongue - or transitional bilingual - education in developing countries is mixed. Some studies suggest that mother tongue education is in fact a complement to skills acquisition in other areas. Using observational data from the Young Lives Survey, Hynsjö and Damon (2016) find that Indigenous children in Peru who attend Quechua-medium primary schools have substantially higher mathematics scores, but find no effect on Spanish language scores. This finding refutes local perceptions that bilingual education slows down the acquisition of Spanish.

A series of studies has exploited quasi-experimental policy variation arising from legal changes in language of instruction in South Africa and Ethiopia. Taylor and von Fintel (2016) use within-school changes in language policy to estimate the returns to mother tongue education in early primary school on English language skills in South Africa. They find that

students whose first three years of primary school were taught in an African language, as opposed to English, performed better at English language tests taken after the students transitioned to English medium instruction. Eriksson (2014) reaches a similar conclusion, but this time in the longer term, by considering the effect of South Africa's 1953 Bantu Education Act, which mandated 8 years of mother tongue education (from previously 4 or 6, depending on the state). She finds that the longer period of mother tongue education is associated with higher literacy levels, more years of education, and higher earnings, measured on 28-48 years olds in the 1980 census. She also finds greater English proficiency for individuals affected by the reform in English-dominated parts of the country (but not everywhere).

Other studies investigate a 1994 law which increased the offering of mother tongue primary education in Ethiopia. Ramachandran (2017) finds the change substantially increased literacy in one of the five major languages, defined as being able to read a simple sentence. He also shows that this increase in literacy is accompanied by a greater use of written materials and better health knowledge. Seid (2018) investigates the effect of this same reform on test scores in grade 5, after students have transitioned to English-language education. He finds improved grade 5 mathematics scores for students who were first taught in their mother tongue, as opposed to a non-English second language. In contrast, Chicoine (2019) finds that the shift towards mother tongue instruction led to a decrease in years of schooling, particularly in areas which simultaneously adopted a new script. Chicoine's more recent study invokes data indicating the policy was adopted at an earlier date in many communities, and separately estimates the effect of the removal of school fees that occurred around the same time: two factors which can explain the contrasting conclusions from earlier work.

Evidence from an experimental programme in Cameroon, which provided local language instruction in the first three years of primary school in Cameroon as opposed to business-as-usual English, finds in favour of transitional bilingual education, although with a caveat. While Laitin et al. (2019) find that the programme increased test scores in mathematics and English substantially in 1st and 3rd grade, with smaller differences persisting once treated students had reverted to English-medium education in grade 5, the learning outcomes in the area remained very low. The authors conclude that the small difference in test scores in grade 5, from such a low baseline, are unlikely to lead to measurable differences in human capital formation.

In contrast, evidence from a medium-scale randomised control trial in Kenya suggests that mother tongue education may delay skills development in other academic subjects. Piper et al. (2018) report on a learning intervention in Kenya rolled out in experimental conditions with two treatment arms: one taught in the usual language of instruction, while the other was taught in mother tongue. The intervention was conducted at considerable scale, with 242 schools receiving some version of the programme. Compared to the usual language arm, the mother tongue arm did show gains in mother tongue language skills; however, this group performed no better in English or Kiswahili, and worse in mathematics. As it was carried out within the government school system, this intervention was trialed in 'real world' conditions: as a consequence, teachers often had low levels of familiarity with local languages. While this is likely to affect the impact of mother tongue instruction, it also reflects how such a programme could be implemented at scale in Kenya.

While they do not seek to isolate the impact of mother tongue education specifically, a recent study by Kerwin and Thornton (2021) helps make sense of contrasting findings from this literature. Their paper reports on a randomised control trial of a literacy intervention in Northern Uganda. Among other inputs, the intervention delivered the first three years of primary school in the local language: an official policy in Uganda, but one which is not strictly adhered to. In an attempt to understand how the details of programme delivery can influence treatment effects, the study had two treatment arms: a full-cost intervention designed to address a wide range of challenges in the local education setting, and a reduced-cost version which followed similar principles but cut back on the most costly elements, reducing the per-student costs by about 60%. Kerwin and Thornton's results are striking: although the full-cost intervention resulted in substantial gains in reading and writing (0.64 sd and 0.45 sd, respectively), the reduced-cost version showed insignificant reading gains, with evidence of negative effects on writing. While the local language component does not appear to be responsible for this difference (both treatment arms increased the use of local language to a similar extent, compared with control), their results demonstrate how modest adjustments to an intervention can have dramatic implications for effectiveness.

Of the papers reviewed here, the Piper et al. (2018) study comes closest to our own in a number of respects. Both studies are based on the random assignment of a scalable intervention implemented in government-run schools.[2] In addition to being carried out

[2] A difference being that Piper et al. (2018) directly randomised the mother tongue component of the intervention, implementing a 3 arm trial (PRIMR, PRIMR with mother tongue, control group) while in our case the

in a different context, however, the intervention studied here is lighter-touch than Piper et al. (2018)'s Primary Math and Reading (PRIMR) Initiative, both in scope (PRIMR covered English, Kiswahili and mathematics, while the CBI's focus was language only) and in intensity (the CBI teacher training was 2 vs 10 days, and the only materials provided were storybooks as compared with the more comprehensive PRIMR materials). Despite these differences, our results on language skills beyond the mother tongue align closely with theirs.

There are many reasons for which multilingual education might be desirable that are unrelated to educational efficiency (for an excellent overview, see Ginsburgh and Weber (2020)). While necessarily important in policy making, these considerations are beyond the scope of this paper, which focusses on a technical question: does a period of bilingual education help non-dominant language speakers transition to learning in the language of instruction?

3 Context and Data

3.1 The Intervention

The data used in this paper come from a pilot literacy intervention carried out by Save the Children in Bangladesh. The intervention, called the Bangladesh Children's Book Initiative (CBI), aimed to improve poor literacy outcomes in early primary school through the development, provision, and pedagogical use of high quality storybooks. The CBI was designed for multi-lingual environments where school is taught in a language different from the mother tongue of some or all children. By design it included an element of bilingual education, with storybooks provided in both the language of instruction and the local language(s). Notwithstanding this bilingual element, the bulk of the intervention (e.g. teacher training, provision of materials) would benefit children regardless of their language background.

While the intervention included a number of elements common to many literacy programmes, it had a unique focus on storybooks. The CBI espoused a 'whole book chain' approach to the provision of storybooks, working with local writers, illustrators and publishers to develop and produce bilingual, gender sensitive storybooks with interesting and

intervention had only 2 arms (CBI and control), with the estimate of the mother-tongue component coming from a difference-in-difference estimate comparing Bangla native and non-native speakers.

accessible text. Schools and communities were then engaged through training programmes which focussed on getting books into the hands of children (head teachers and librarians) and using storybooks for language lessons (teachers). In some villages, this included working with community based organisations to hold classes on ethnic scripts. Parents were engaged through community sensitisation sessions to support book use at home. Child-friendly book shelves were installed in participating school classrooms, and populated with books. Classroom teachers were expected to send books home with children, and were taught to keep a book lending register.

The CBI emphasised quality over quantity, and invested considerable resources in ensuring the storybooks were carefully edited, age-appropriate and attractive. It was important for the programme that these were written and illustrated locally, and ownership of the books remained with the authors and publishing houses. Given the emphasis on book production, the intervention targeted areas where there were minimal educational resources in the non-dominant local languages. The Bangladesh pilot was carried out in the Chittagong Hill Tracts, a mostly rural, remote hill country in the southeastern part of the country. The area is also ethnically diverse, with a number of indigenous groups as well as a substantial Bengali population. Books created as part of the intervention included text in one of three minority languages (Chakma, Marma and Tripura) printed alongside text in Bangla.

The schools in which the intervention was piloted enrolled students from a mix of ethnic and linguistic backgrounds. The largest group of students were native Bangla speakers, but a substantial number were from local indigenous groups and therefore arrived at school with little or no Bangla language skills. Native Bangla-speakers experienced the educational enrichment provided by the storybooks, teacher training, and school and community engagement. Non-native Bangla speakers, specifically those children from Chakma, Marma and Tripura-speaking families, were also privy to these features of the programme, but in addition they received an additional educational intervention they had not experienced before: educational materials in their mother-tongue.

Since the different experience of native Bangla speakers and minority language speakers is central to the analysis in this paper, it is worth setting out the structure of language education clearly before we move on. The language of instruction in the schools in the pilot study is Bangla. Children whose first language is not Bangla are by default enrolled in Bangla-immersion when they begin formal schooling. Teachers are not allocated to schools

based on language familiarity, and so while some local teachers may speak one of the minority languages, most minority language students would not expect to have a teacher who speaks their language. Minority languages are not taught in schools, and there is no formal opportunity for students to learn to read in other languages. Situational analysis done in this area has highlighted that this was a source of alienation for non-Bangla ethnic groups, and a possible contribution to the relatively poor performance of children at school (Save the Children (2015)).

The provision of bilingual reading materials in early primary school introduces an element of transitional bilingual education into what is otherwise a pure immersion setting. Non-native Bangla speakers, therefore, received the usual package of CBI interventions, but in contrast to native Bangla speakers, they additionally experienced education in their mother tongue for the first time.

3.2 Study design and data collection

The CBI was implemented for two years, between 2016 and 2018, in randomly selected village schools in the Chittagong Hill Tracts (CHT). Based on a mapping exercise carried out prior to baseline data collection, 103 schools in 99 villages were identified as meeting criteria for inclusion in the pilot.[3] Prior to randomisation, the schools were stratified according to two criteria: first, the languages spoken by children at the school and second, the language(s) spoken by teachers. This ensured that the treatment and control groups included a balance of children speaking different languages, and furthermore that minority language-speaking teachers were also balanced across groups. 51 schools were assigned to treatment, and 52 schools to control.

Data collection was carried out in two waves, approximately 23 months apart, at the start and end of the intervention. The data collection plan called for sampling of up to 14 first grade students in each school, with oversampling of minority-language speakers. In practice, almost half of schools had fewer than 14 first grade students present on the survey day: in these cases, the full class was sampled. Larger samples were drawn at bigger schools to maintain the target sample size. The baseline sample included 1313 children,

[3]Inclusion criteria were determined by the implementation team in Bangladesh, based primarily on location of the villages in one of the participating upazila, and the schools not currently being served by similar literacy interventions. The selected villages are not necessarily representative of the CHT; indeed, it has been suggested that the focus on villages not already enrolled in literacy programmes resulted in a set of villages that are particularly remote and difficult to access.

with per-school sample sizes ranging from 5 to 29. At follow-up, 918 of the original cohort were re-surveyed (70%), and an additional 116 new children were recruited to meet sample targets.[4] When possible this replacement sample was matched on sex, expected grade (in this case, grade 3), and language. Both surveys were organised and managed by Save the Children Bangladesh staff familiar with CHT.

Data used in this paper was collected using the CLASS child literacy assessment. This detailed survey tool covers a range of pre-literacy to literacy skills, starting with letter recognition and introducing more complex literacy skills up to, and including, story reading and comprehension. Bilingual enumerators administered the CLASS assessment to children whose mother tongue was not Bangla. These children completed two versions of the assessment: the Bangla version, and an equivalent translation in their mother tongue.

3.3 Data

Table 1 presents basic descriptive statistics for the first wave sample. The sample includes almost equal numbers of girls and boys; the average age is 6.9 years. The sample is composed of almost one third each Bangla and Marma students, one quarter Chakma students, while Tripura students make up only 14%. Note that these figures are not representative of the population of study schools themselves: according to mapping data, the pre-primary to grade three enrolment at the schools eligible for the study was 48% Bangla, 28% Marma, 15% Chakma, 9% Tripura, and less than 1% other. Purposeful over-sampling of minority-language students led to a much smaller share of Bangla students in our sample. Regression results presented throughout the paper are re-weighted to be representative of the population language shares estimated from the mapping data.[5]

3.3.1 Outcome variables

There are two outcomes of interest: literacy in Bangla, and literacy in mother tongue (for minority language speakers only). A summary statistic for each of these is generated from the rich information collected by the CLASS tool as follows. First, the score on each section of the class tool is converted to a percentage grade. These grades are then standardised with

[4] One control school was not reachable at follow-up.

[5] Treatment effects are are estimated from a balanced panel, re-weighted to match population-level language group representation. For estimates using the wave 1 sample (e.g. the attrition study in Section 3.4 below), a separate set of weights are used to re-weight the wave 1 observations, in order to achieve the same population-wide language proportions. Summary statistics, which are intended to describe the data itself, are not weighted.

Table 1: **Characteristics of the sample at wave 1**

	Wave 1 Sample	Wave 1 Treat	Wave 1 Control
Age	**6.90**	**6.89**	**6.92**
	1.19	*1.12*	*1.27*
Girls	**0.50**	**0.51**	**0.49**
	0.50	*0.50*	*0.50*
Chakma native	**0.24**	**0.25**	**0.24**
	0.43	*0.43*	*0.43*
Marma native	**0.31**	**0.29**	**0.32**
	0.46	*0.46*	*0.47*
Tripura native	**0.14**	**0.14**	**0.14**
	0.35	*0.35*	*0.35*
Bangla native	**0.31**	**0.32**	**0.30**
	0.46	*0.47*	*0.46*
Valid responses	1293	654	639

Average values from wave 1 in bold. Standard deviations in italics. Valid responses are complete records: these are slightly below total sample because some age observations are missing.

respect to the mean and standard deviation of the control group for that wave. Finally, an overall score is created by averaging the percentage scores of each section. These in turn are standardised with respect to the control group. Given the very limited literacy in mother tongue for minority language speakers, along with concerns about the reliability of the final two sections in the wave 1 survey, the measure of literacy in minority languages is calculated using only the first four sections of the literacy tool (letter recognition, letter and word spelling, word reading, word recognition).[6] Further details on the data can be found in Appendix A.1.

Table 2 summarises the main outcome variables both for Bangla and for minority language (averaged across Chakma, Marma and Tripura). The data presented here are the overall percentage scores, prior to standardisation. Averaging treatment and control, Bangla literacy scores increased from 39% at wave 1 to 73% at wave 2, while minority language literacy scores rose from 1% to 24%.[7]

[6] This inflates the raw minority language scores with respect to the Bangla literacy scores, as the sections of the tool which have not been included received in almost all cases a score of zero: effectively, the denominator for the minority language scores is 60, while for the Bangla scores it is 100. When standardised, the omission of the final two sections will have little impact on the ranking of students, as these sections were almost never completed.

[7] It is striking here that the minority language speakers have similar Bangla literacy scores to Bangla natives. It is likely that this is somewhat unique to the study area, where the relationship between ethnicity and socioeconomic status is quite different from elsewhere in Bangladesh. Previous work from elsewhere in Bangladesh has found that children who are not ethnic Bengali have substantially lower academic competencies, compared with Bengali peers (Nath (2013)).

Table 2: **Literacy scores (in %) in waves 1 & 2**

	Panel	Treat	Control	Bangla	Min Lang
Bangla scores wave 1	**39.19**	**39.68**	**38.70**	**39.49**	**39.06**
	20.55	*20.22*	*20.87*	*20.12*	*20.75*
Bangla scores wave 2	**72.67**	**76.55**	**69.00**	**72.21**	**72.86**
	16.39	*14.67*	*17.08*	*16.05*	*16.53*
Minority language scores wave 1	**1.05**	**0.86**	**1.24**	-	**1.05**
	4.13	*3.68*	*4.53*	-	*4.13*
Minority language scores wave 2	**23.71**	**31.28**	**16.64**	-	**23.71**
	26.18	*27.60*	*22.63*	-	*26.18*

Average values from wave 1 (all observations) and wave 2 (resampled observations) in bold. Standard deviations in italics.

3.3.2 Balance

Balance checks on baseline characteristics are shown in Table 3. Stratification ensured that treatment and control schools were well balanced on language groups; randomisation successfully achieved a good balance on age and sex as well. Treatment and control groups were also well balanced in terms of baseline literacy scores: both the overall Bangla and overall minority language scores show no differences at baseline. The last two rows of Table 3 compare the Bangla literacy scores of native Bangla speakers and Bangla learners (students whose mother tongue is one of the minority languages). Here we see some difference between treatment and control, with Bangla learners in the treatment group having 2.9 points higher scores than their peers at control schools. Bangla natives in control schools meanwhile have 3.4 points higher scores than Bangla natives in treatment schools. These differences correspond to 0.14 and 0.17 of a standard deviation, respectively. These are not trivial differences, and motivate the inclusion of baseline literacy scores as a control variable in all estimations.[8]

3.4 Attrition

Of the 1313 children surveyed in the first wave, 918 were followed up in the second wave. This implies an attrition rate of 30%. Our estimation strategy requires that the control group

[8]Tables A.3 and A.4 in the Appendix show balance checks for the different elements of the literacy assessments individually. Table A.3 shows there is good overall balance across treatment and control for each element of the assessments. Table A.4 repeats this for the Bangla assessment, breaking the comparison down by learners and natives. It suggests that the differences in overall scores are coming from certain elements of the the assessments and not all: specifically, native speakers in the control group outperform those in the treatment group at letter recognition and story comprehension. Bangla learners in the treatment group outperform those in the control group at word recognition.

Table 3: **Balance checks: independent variables and main outcome at wave 1**

Variable	(1) Control N	(1) Control Mean/SE	(2) Treat N	(2) Treat Mean/SE	T-test P-value (1)-(2)
Age at baseline	639	6.917	654	6.893	0.717
		(0.050)		(0.044)	
Girls	653	0.485	660	0.514	0.307
		(0.020)		(0.019)	
Chakma speaking	653	0.239	660	0.245	0.782
		(0.017)		(0.017)	
Marma speaking	653	0.325	660	0.294	0.229
		(0.018)		(0.018)	
Tripura speaking	653	0.141	660	0.139	0.938
		(0.014)		(0.013)	
Bangla speaking	653	0.296	660	0.320	0.344
		(0.018)		(0.018)	
Overall Bangla score	653	38.704	660	39.677	0.391
		(0.817)		(0.787)	
Overall Min Lang score	460	1.237	449	0.861	0.170
		(0.211)		(0.174)	
Bangla score: learners	460	37.639	449	40.517	0.036**
		(0.977)		(0.966)	
Bangla score: natives	193	41.242	211	37.889	0.094*
		(1.478)		(1.352)	

Each row presents a t-test for difference in means across treatment and control groups.

is an appropriate counterfactual for the treatment group: if the two groups experienced different levels or kinds of attrition this could bias the estimated treatment effects.

To test whether there is a difference in the level of attrition between treatment and control, we compare the number of students who attrit from the sample in treatment and control schools. In the treatment schools, 30.6% of the sample attrited, while in control schools this same figure is 29.6%. This shows good balance in the *level* of attrition. It also suggests that the intervention itself did not substantially reduce drop out in the treatment schools - at least as measured by attendance on a given day two years later.

To test whether the students who attrit in each group are similar to each other, we explore the correlates of attrition by regressing baseline characteristics on a dummy variable equal to one for children who attrited. Table 4 presents these results, both overall and for treatment and control groups separately. As the relative rates indicated, attrition itself is not related to treatment status (coefficient on Treat in Column 1 is small and statistically insignificant); however, the pattern of attrition is different across treatment and control. Specifically, boys and weaker performing students were more likely to attrit in the control group, while in the treatment group neither wave 1 scores, sex, nor age is strongly predictive of attrition. Minority language speakers are less likely to attrit in both groups (compared with Bangla natives), possibly more so in the treatment group.[9]

This suggests that treatment is affecting the characteristics of students who attrit from the sample, which could bias estimated treatment effects. Two steps are taken to account for this. First, the main analysis will focus on the balanced sample of students surveyed at both waves, using both individual characteristics and wave 1 literacy scores as control variables. The availability of a pre-treatment measure of literacy is a strength here, as it is a proxy for many of the unobservable characteristics whose imbalance we should be concerned about: ability, motivation and family support for education, among other things. If this measure, combined with our control variables on age, sex and language, is sufficiently informative, then the assumption that attrition is independent of potential outcomes *conditional on the control variables* is reasonable. If this assumption holds, estimated treatment effects will not be biased due to attrition patterns. Second, the robustness of the main results to different patterns of attrition is explored. The results of several simulated scenarios are presented in

[9] We also run a balance test comparing the first wave characteristics of students who attrited in treatment vs control (see Appendix Table B.6). As in Table 4 comparison shows that students who attrited in the treatment group were more likely to be girls (50.5% vs 38.3%) and have higher overall Bangla literacy scores.

Table 4: **Correlates of attrition**

	(1) Attrition: full sample	(2) Attrition: control	(3) Attrition: treat
Girls	-0.0687**	-0.132***	-0.0170
	(0.0338)	(0.0463)	(0.0501)
Age at wave 2	0.0315*	0.0241	0.0372
	(0.0167)	(0.0236)	(0.0228)
Overall Bangla score	-0.0533***	-0.0890***	-0.0192
	(0.0163)	(0.0208)	(0.0248)
Minority language	-0.102***	-0.0919*	-0.125***
	(0.0333)	(0.0480)	(0.0432)
Treat	0.0233		
	(0.0353)		
Constant	0.114	0.205	0.0718
	(0.149)	(0.206)	(0.210)
Observations	1305	648	657
R^2	0.039	0.069	0.030

Each column reports estimates from a separate OLS regression with a dummy variable for attrition as the dependant variable. Column 1 includes the full wave 1 sample, Column 2 is wave 1 control group only, while Column 3 is wave 1 treatment group only. Standard errors (in parentheses) are clustered at the school level.

Section 6.2, after the main results.

4 Empirical Approach

4.1 Overview

The estimation of the effect of the programme proceeds in two steps. First, the overall impact of the intervention is established by comparing the literacy scores of students in the treatment and the control groups. Second, the additional impact of the mother tongue component is estimated using a difference-in-differences design, where the first difference is treatment vs control, and the second difference is minority language vs dominant language speakers. The intuition and the identifying assumptions for these two steps are outlined below.

The first estimation relies on the randomisation of treatment. The intervention itself was randomly assigned at the school level. Schools were stratified based on the languages spoken by students at the school prior to randomisation, ensuring that the treatment and control groups were balanced with respect to language use. With a large enough sample, randomisation also makes treatment itself exogenous to other individual characteristics. Balance checks at wave 1 suggest that randomisation was largely successful (see Table 3).

The second estimation relies on the following intuition. While all students in the school received the Children's Book Initiative intervention, Bangla speakers have always had learning materials in their mother tongue, while language learners received these through the intervention for the first time. This means that, compared with Bangla speakers, minority language speakers in treated schools received an additional element of the intervention: for the first time, these students had access to learning materials in their mother tongue. Minority language students received a transitional bilingual intervention, along with the other elements of the Children's Book Initiative.

The effect of this transitional bilingual element of the intervention is therefore estimated using a difference-in-differences strategy. The first difference compares minority language speakers at treatment schools to minority language speakers at control schools, while the second difference compares minority language speakers at treatment schools to Bangla natives at treatment schools.

The validity of this estimation strategy relies on the parallel trend assumption. This

requires that the difference in literacy scores between Bangla natives and minority language speakers in the control schools is a valid counterfactual for the difference in literacy scores between these two groups in the treatment schools. This could be violated if there are substantial non-linearities in learning gains that make gaps between different groups incomparable. For example, large percentage differences in letter recognition could exist within a group of children with similar pre-literacy skills, while large percentage differences in word recognition might indicate more substantially different literacy skills.

While this assumption cannot be tested directly, it is reassuring to see that Bangla-natives and minority language speakers have broadly similar achievement levels at wave 1 (as shown in Table 2). Indeed, while some language groups have lower average literacy scores, one has higher average scores (see Appendix B.1, Table B.1).

4.2 Estimating equations

The estimation proceeds in two steps: the first establishes the overall effect of the intervention, while the second estimates the effect of the transitional bilingual element alone. The first estimating equation takes the following form:

$$y_{i,t} = \alpha_0 + \alpha_1 y_{i,t-1} + \alpha_2 \text{treat}_i + \Gamma X_i + \epsilon_{i,t}, \tag{1}$$

where y_{it} is a student's literacy measure at time $t = 1, 2$, $treat = 1$ for treated schools, and X_i is a short list of predetermined individual-level controls (sex and age in first grade, along with a language dummy that captures broad socioeconomic differences between ethnic groups). The coefficient of interest when estimating Equation 1 is $\hat{\alpha_2}$, the estimated effect of treatment on literacy measures in wave 2.

To estimate the effect of the transitional bilingual learning materials alone, a second estimation is carried out:

$$y_{i,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 \text{treat}_i + \beta_3 \text{mlang}_i + \beta_4 \text{mlang}_i * \text{treat}_i + \Delta X_i + \eta_{i,t}, \tag{2}$$

where $mlang = 1$ for children who are minority language speakers and $mlang = 0$ for Bangla natives. This difference-in-differences equation estimates the effect of the intervention on

minority language speakers relative to Bangla natives, given by the estimated coefficient $\hat{\beta}_4$.

In estimates of both Equations 1 and 2, the sample is re-weighted to correct for oversampling of minority language speakers. This re-weighting adjusts the panel frequencies of each language group to match the population levels, which display a higher proportion of Bangla natives. Results from a selection of unweighted regressions, which are nearly identical, are summarised in Appendix Table B.10.

5 Main Results

5.1 Impact of the intervention: treatment vs control

To understand the overall effect of the intervention on literacy scores, Equation 1 is estimated using the balanced panel of observations. While the primary indicator of interest is the overall literacy score, the analysis is repeated for each of the components of the literacy assessment: these sum to the overall score, and so are not independent of the primary indicator. Results of these estimations are presented in Table 5: note that all outcome variables are standardised with respect to the control group for that wave. Column 1 of this table presents the main results: with the overall Bangla literacy score as a measure of literacy, the intervention raised literacy levels by 0.44 of a standard deviation. Relative age and sex are not significant predictors of literacy when controlling for other factors, however language groups show some different trends. Notably, Chakma speakers (language group 1) perform significantly better than Bangla natives (omitted category language group 4). The coefficient on Tripura (language group 3) is negative and substantial in size, but imprecisely estimated, while Marma speakers appear to be quite similar in performance to Bangla natives.

The component literacy indicators, shown in Columns 2-7 of Table 5, generally tell a similar story. The individual effect sizes vary, with the intervention having a small and insignificant effect on Bangla letter recognition (which most students had already mastered in wave 1; for details see Appendix Table A.1); all other component indicators show a significant improvement due to the intervention, with increases between 0.22 and 0.45 standard deviations.

Table 5 also highlights the persistence in literacy attainment across waves. Using the overall measure, a one standard deviation higher literacy score in wave 1 is associated with a

0.45 standard deviation higher score two years later. The component scores also demonstrate considerable persistence, although the coefficients are slightly smaller. The exception here is comprehension: wave 1 comprehension scores were very low (6% on average, with most students scoring 0), and show no association with wave 2 comprehension scores.

Table 5: **Treatment effects: literacy in Bangla**

	(1) Overall	(2) Letter recognition	(3) Spelling	(4) Read words	(5) Point words	(6) Read story	(7) Comprehension
Treat	0.441***	0.122	0.335***	0.209***	0.340***	0.228***	0.462***
	(0.0825)	(0.0979)	(0.0824)	(0.0791)	(0.0733)	(0.0819)	(0.105)
Girls	0.0671	-0.0168	0.00565	0.0778	0.0564	0.197***	0.0428
	(0.0646)	(0.0687)	(0.0703)	(0.0733)	(0.0690)	(0.0708)	(0.0722)
Age at wave 2	-0.0313	-0.0535	-0.0115	-0.0409	-0.0359	-0.0409	0.0272
	(0.0316)	(0.0396)	(0.0360)	(0.0368)	(0.0347)	(0.0340)	(0.0370)
Language=1	0.285***	0.233**	0.337***	0.173	0.0740	0.242**	0.257**
	(0.0891)	(0.115)	(0.0941)	(0.114)	(0.0858)	(0.105)	(0.105)
Language=2	0.0705	0.352***	0.214**	0.0595	0.0197	0.0591	-0.148
	(0.0953)	(0.109)	(0.0865)	(0.0986)	(0.0885)	(0.108)	(0.120)
Language=3	-0.129	0.231**	0.0129	-0.00841	0.0622	-0.148	-0.629***
	(0.104)	(0.114)	(0.117)	(0.109)	(0.102)	(0.105)	(0.123)
Overall Bangla score	0.453***						
	(0.0420)						
Bangla letter recognition		0.246***					
		(0.0697)					
Bangla letter and word spelling			0.370***				
			(0.0410)				
Bangla word reading				0.385***			
				(0.0324)			
Bangla word recognition					0.247***		
					(0.0321)		
Bangla story reading (1 min)						0.410***	
						(0.0462)	
Bangla story comprehension							0.0443
							(0.0373)
Constant	0.0982	0.216	-0.110	0.229	0.252	0.184	-0.221
	(0.290)	(0.319)	(0.322)	(0.359)	(0.317)	(0.343)	(0.356)
Observations	918	918	918	918	918	918	918
R^2	0.281	0.089	0.175	0.199	0.116	0.213	0.104

Each column reports estimates from a separate OLS regression with the dependent variable a normalised literacy outcome. The dependent variable in Column 1 is the overall Bangla literacy score, while the dependent variables in Columns 2-7 are the components which contribute to this overall score. All control variables are shown. Standard errors (in parentheses) are clustered at the school level.

Table 6 repeats the analysis in Table 5, but now with minority language literacy as the outcome variable. This analysis is restricted to minority language speakers, hence the total sample size is smaller. Column 1 of Table 6 shows estimates from the primary specification, where the overall measure of literacy is the dependant variable, while Columns 2-5 replicate this for each component indicator. The intervention increased minority language literacy by

0.68 of a standard deviation. In contrast to the Bangla literacy scores, there is a sex effect in minority language literacy gains, with girls making larger gains than boys between waves. This effect is not trivial, with girls having 0.16 of a standard deviation greater increase. The lagged literacy measure is statistically insignificant in this case: as in the case of Bangla comprehension scores, few students had any literacy skills in minority languages in wave 1, giving this measure little variation.

With respect to Chakma students (the omitted category), Marma students (language group 2) have similar growth in literacy skills, while Tripura students perform considerably better. Tripura students had much more substantial growth in mother tongue literacy skills between wave 1 and wave 2, with wave 2 scores on average 1.7 sd higher than Chakma students (see Appendix Table B.1 for summary statistics on literacy scores by language group). This contrasts with Bangla literacy, where Chakma students were the highest performers. One reason for the strong growth in Tripura literacy could be the fact that Tripura is transcribed using the latin alphabet, and therefore would be commonly seen in the local written environment. Chakma and Marma, with unique scripts, would be less common.

5.2 Impact of transitional bilingual materials: difference-in-differences

Tables 5 & 6 show large and significant treatment effects of the intervention on both Bangla and minority language literacy skills. The question remains: did the mother tongue aspect of the intervention help or hinder minority language speakers acquisition of literacy skills in Bangla? To answer this question we estimate Equation 2, which includes an interaction between treatment and minority language status. Since the comparison here is between Bangla natives and minority language speakers, only Bangla literacy outcomes are analysed. Table 7 presents estimates of Equation 2 for the overall literacy measure (Column 1), as well as its component parts (Columns 2-7).

Estimates presented in Column 1 show that the intervention had a similar effect on minority language speakers' acquisition of literacy skills in Bangla as it did on Bangla natives. While the point estimate for the interaction is negative, it is small in magnitude and statistically insignificant. The overall treatment effect estimated in this equation remains large at 0.49 standard deviations. Sex and age remain statistically insignificant predictors of achievement, while the persistence of scores across waves is similar to that estimated in

Table 6: **Treatment effects: literacy in mother tongue**

	(1) Overall	(2) Letter recognition	(3) Spelling	(4) Read words	(5) Point words
Treat	0.678***	0.529***	0.755***	0.242***	0.657***
	(0.118)	(0.100)	(0.167)	(0.0716)	(0.153)
Girls	0.158**	0.123*	0.152	0.0515	0.200*
	(0.0741)	(0.0681)	(0.0940)	(0.0617)	(0.104)
Age at wave 2	-0.0206	-0.0303	-0.0670	0.0358	0.00312
	(0.0413)	(0.0390)	(0.0485)	(0.0346)	(0.0552)
Language=2	0.0195	0.0654	-0.0228	0.0724	-0.0755
	(0.128)	(0.130)	(0.181)	(0.0707)	(0.140)
Language=3	1.737***	1.258***	1.561***	1.562***	1.155***
	(0.143)	(0.137)	(0.162)	(0.141)	(0.210)
Overall Min Lang score	0.0884				
	(0.0582)				
Min Lang letter recognition		0.0772			
		(0.0517)			
Min Lang letter and word spelling			0.0116		
			(0.0506)		
Min Lang word reading				0.0428	
				(0.0844)	
Min Lang word recognition					0.0791*
					(0.0461)
Constant	-0.249	-0.119	0.234	-0.650**	-0.299
	(0.369)	(0.349)	(0.407)	(0.302)	(0.486)
Observations	661	661	661	661	661
R^2	0.419	0.276	0.294	0.345	0.204

Each column reports estimates from a separate OLS regression with the dependent variable a normalised literacy outcome. The dependent variable in Column 1 is the overall minority language literacy score, while the dependent variables in Columns 2-5 are the components which contribute to this overall score. All control variables are shown. Standard errors (in parentheses) are clustered at the school level.

Column 1 of Table 5. Note that, due to the inclusion of language dummies, the coefficient on minority language is picking up the Tripura-speaker effect (language group 3), while the omitted language group is Bangla.

The component literacy indicators, shown in Columns 2-7, show a less uniform story than the overall treatment effects in Table 5. The coefficients on the interaction of treatment and minority language status have a range of point estimates, both positive and negative, but are all statistically insignificant with the exception of story comprehension (Column 7). For that one component, the interaction term is substantial in size, negative and significant: suggesting that minority language speakers compared with Bangla natives in treated schools had relatively smaller gains in Bangla comprehension, versus minority language speakers compared with Bangla natives in control schools. This finding is suggestive at best: the study is not designed to estimate treatment effects on many different indicators, only a single treatment effect.

The general picture from Table 7 is that of a statistically insignificant interaction effect between treatment and minority language status: minority language students had equivalent treatment gains in Bangla literacy as did their Bangla-native peers. While the standard errors for the coefficients on the interaction terms are relatively large, the point estimates are modest (although not trivial for some components): they also are not systematically positive or negative across components.

6 Extended results and robustness checks

6.1 Extended results

6.1.1 Heterogeneous effects

While Section 5 assesses the impact of the intervention on the population as a whole, and on minority language speakers in particular, the analysis suggests that the evolution of literacy scores may be associated with other student characteristics, such as sex and prior achievement. In this section, we investigate further whether the treatment had heterogeneous impacts on students, based on the predetermined characteristics collected in wave 1. To do this, we estimate a version of Equation 2 where, instead of interacting treatment with minority language status, we iteratively interact it with prior literacy scores, child sex and

Table 7: **Treatment effects: literacy in Bangla**

	(1) Overall	(2) Letter recognition	(3) Spelling	(4) Read words	(5) Point words	(6) Read story	(7) Comprehension
Treat=1	0.490***	0.199	0.254*	0.109	0.376***	0.176	0.624***
	(0.116)	(0.174)	(0.139)	(0.121)	(0.109)	(0.128)	(0.147)
Minority language=1	-0.0818	0.304*	-0.0641	-0.104	0.0963	-0.197	-0.473***
	(0.127)	(0.159)	(0.155)	(0.142)	(0.117)	(0.130)	(0.153)
Treat=1 × Minority language=1	-0.0949	-0.147	0.155	0.192	-0.0687	0.101	-0.312*
	(0.150)	(0.196)	(0.154)	(0.155)	(0.137)	(0.167)	(0.175)
Girls	0.0675	-0.0162	0.00498	0.0768	0.0567	0.196***	0.0439
	(0.0651)	(0.0693)	(0.0695)	(0.0723)	(0.0692)	(0.0708)	(0.0715)
Age at wave 2	-0.0314	-0.0537	-0.0111	-0.0407	-0.0361	-0.0407	0.0267
	(0.0316)	(0.0403)	(0.0358)	(0.0366)	(0.0345)	(0.0340)	(0.0366)
Language=1	0.415***	0.00355	0.322***	0.179	0.0126	0.388***	0.891***
	(0.107)	(0.116)	(0.113)	(0.132)	(0.105)	(0.107)	(0.137)
Language=2	0.199*	0.121	0.202*	0.0678	-0.0426	0.207*	0.480***
	(0.113)	(0.100)	(0.113)	(0.113)	(0.112)	(0.113)	(0.144)
Overall Bangla score	0.455***						
	(0.0418)						
Bangla letter recognition		0.250***					
		(0.0699)					
Bangla letter and word spelling			0.367***				
			(0.0405)				
Bangla word reading				0.383***			
				(0.0318)			
Bangla word recognition					0.248***		
					(0.0317)		
Bangla story reading (1 min)						0.409***	
						(0.0458)	
Bangla story comprehension							0.0506
							(0.0359)
Constant	0.0740	0.179	-0.0715	0.278	0.235	0.209	-0.300
	(0.300)	(0.326)	(0.316)	(0.358)	(0.317)	(0.353)	(0.373)
Observations	918	918	918	918	918	918	918
R^2	0.281	0.091	0.177	0.202	0.116	0.214	0.109

Each column reports estimates from a separate OLS regression with the dependent variable a normalised literacy outcome. The dependent variable in Column 1 is the overall Bangla literacy score, while the dependent variables in Columns 2-7 are the components which contribute to this overall score. All control variables are shown. Standard errors (in parentheses) are clustered at the school level.

child age.

Table 8 show the coefficients of interest from six separate regressions. Columns 1 & 2 interact treatment with wave 1 literacy scores in Bangla and minority language (respectively), and estimate the treatment effect on those same scores in wave 2. While we have seen in Tables 5 & 6 that wave 1 Bangla scores are strongly predictive of wave 2 scores, we do not find any difference in treatment effects for either language based on prior achievement: the coefficients on the interaction terms are modest (-0.12 and -0.11, for Bangla and minority language respectively) and statistically insignificant.

In contrast, the interaction between treatment and sex (in this case girls) is relatively large, and statistically significant, for both Bangla and minority language literacy. With respect to boys, girls show 0.23 sd (Bangla) and 0.29 sd (minority language) larger treatment effects. This is substantial: it is equivalent to approximately half the average treatment effect for Bangla, and only slightly less than that for minority language literacy.

Finally, Columns 5 & 6 of Table 8 show results from regressions interacting treatment with age. Column 5 shows that older children have a larger treatment effect in Bangla literacy: this effect is very large, equivalent to 0.22 sd larger treatment effect per year of age (recall from Table 1 that the average age at wave 1 is 6.9 years, with a standard deviation of 1.2 years). There is no equivalent effect of age on minority language literacy (Column 6): the coefficient on the interaction term is close to 0 and statistically insignificant.

To summarise Table 8, the intervention had a particularly large effect on Bangla literacy for girls and for older students; girls also benefited more from the intervention in terms of minority language literacy. There is no strong evidence of heterogeneous treatment effects by initial literacy level in either language, although it is interesting to note the negative point estimates for both languages: this does not appear to be an intervention that boosted literacy only for those students who were already at a literacy advantage.

6.1.2 Minority language interaction: subsample analysis

The analysis in Table 7, estimating the interaction between treatment and minority language status, suggests that the provision of bilingual teaching to minority language students did not detract from their acquisition of literacy skills in Bangla, on average. It could be the case, however, that some groups of students responded differently to this joint intervention. While the modest sample size available for analysis prevents a rigorous analysis of heterogeneity

Table 8: **Main treatment effects: interaction with student characteristics**

	(1) Bangla	(2) MinLang	(3) Bangla	(4) MinLang	(5) Bangla	(6) MinLang
Treat	0.454***	0.674***	0.320***	0.524***	-1.473***	0.966
	(0.0872)	(0.118)	(0.104)	(0.122)	(0.478)	(0.778)
Treat x wave 1 Bangla	-0.123					
	(0.0831)					
Treat x wave 1 MinLang		-0.109				
		(0.0995)				
Treat x Girls			0.229*	0.293*		
			(0.129)	(0.148)		
Treat x Age					0.219***	-0.0335
					(0.0553)	(0.0907)
Observations	918	661	918	661	918	661
R^2	0.285	0.421	0.284	0.423	0.295	0.419

Each column reports estimates from a separate OLS regression with either wave 2 Bangla literacy (Columns 1, 3, 5) or minority language literacy (Columns 2, 4, 6) as the dependant variable. All regressions include controls for child sex, age, wave 1 literacy scores, and dummy variables for the four languages. Standard errors (in parentheses) are clustered at the school level.

in the interaction effect, we shed light on this by splitting the sample into subgroups and replicating the estimate of Equation 2.

Table 9 presents the coefficients of interest from these estimations. The sample is split at the midpoint of two variables: wave 1 Bangla scores (Columns 1 & 2) and age (Columns 5 & 6), while the binary sex variable is used to split the sample into girls and boys (Columns 3 & 4). Compared with the full-sample analysis, the main treatment effect remains of similar magnitude across all subsamples except age: as found in Table 8, the treatment is much stronger for older children: in fact, the treatment effect is insignificant when considering only the younger half of the sample.

Our main interest in Table 9 is the interaction term: do we find any evidence of heterogeneity? In all cases, the interaction term between treatment and minority language remains statistically insignificant, although the standard errors are fairly large. This suggests that any heterogeneity that may be present is not large enough to overcome the small sample size. The point estimates offer some suggestions for future investigation. Those students with higher wave 1 literacy scores have a larger negative point estimate, compared to those with lower scores; similarly, boys have a larger negative point estimate than girls. Younger

students, who in the split sample do not show a statistically significant treatment effect, have a larger negative point estimate than do older students - whose point estimate for the interaction term is actually positive.

To summarise, Table 9 does not overturn the main finding of an insignificant interaction effect: the provision of bilingual learning materials has not hindered the progress minority language speakers would have made in Bangla absent this bilingual component. There is also no support for the bilingual component providing a substantial boost to Bangla literacy skills for any subgroup (the positive coefficient on the older student being very small and statistically insignificant). While speculative, there is some suggestion that children with higher early literacy scores, boys and younger children saw slower Bangla literacy growth as a result of the bilingual component.

Table 9: **Mother tongue treatment effects: sample split by student characteristics**

	(1)	(2)	(3)	(4)	(5)	(6)
	High score	Low score	Girls	Boys	Young	Old
Treat=1	0.504***	0.531***	0.577***	0.414**	0.192	0.663***
	(0.133)	(0.186)	(0.163)	(0.165)	(0.142)	(0.144)
Minority language=1	-0.0376	-0.0992	-0.214	0.0600	-0.183	-0.121
	(0.162)	(0.168)	(0.145)	(0.169)	(0.137)	(0.183)
Treat=1 × Minority language=1	-0.241	-0.0111	-0.0577	-0.140	-0.102	0.0564
	(0.153)	(0.235)	(0.194)	(0.198)	(0.167)	(0.197)
Observations	416	502	480	438	453	465
R^2	0.206	0.157	0.300	0.282	0.302	0.313

Each column reports estimates from a separate OLS regression with wave 2 Bangla literacy as the dependant variable. Regressions are run using half the sample. This is either split at the midpoint of a given variable (wave 1 Bangla literacy scores in Columns 1 & 2 and child age in Columns 5 & 6) or by child sex (Columns 3 & 4). All regressions include controls for child sex, age, wave 1 literacy scores, and dummy variables for the four languages. Standard errors (in parentheses) are clustered at the school level.

6.2 Addressing Attrition

Section 3.4 highlighted that differential attrition across treatment and control groups presents a threat to the estimation of unbiased treatment effects. In this section we carry out a number of counterfactual experiments to explore the severity of this threat. We first consider various attrition scenarios applied to both treatment and control group equally. We then consider differential attrition across treatment and control. In this case the focus is on those scenarios most likely to threaten our results: when attrition alone would lead to a posi-

tive performance difference between treatment and control. The plausibility of the different scenario is also discussed.

6.2.1 Selection into the panel

To be included in the panel, students had to be present at school on the day enumerators visited. Students who have dropped out of school, either permanently or temporarily, or who have a high rate of absenteeism, are therefore more likely to attrit that regular attendees. Because of this, it is reasonable to expect that attriters have, on average, worse academic performance than panel members. The data support this: a comparison of wave 1 literacy scores between those who attrit and those who remain shows that attriters have 0.26 sd lower test scores than panel members.

Our first approach at accounting for attrition is to use the information we have from wave 1 to predict wave 2 test scores. We do this by first estimating Equation 1 to generate predicted literacy scores, then filling in missing values for attrited individuals using the predicted literacy scores. We then re-estimate our main results using the resulting panel, composed of actual data for those respondents who remained in the sample, and wave 1 plus predicted wave 2 data for attriters. Full results can be found in Appendix Table B.7; the point estimates for treatment effects are almost identical to those estimated in the main specification.

This approach would fully correct for attrition, under the assumption that attrition is independent of potential outcomes, conditional on our control variables. It would not, however, if the attriters had unobservable characteristics associated with worse (or better) literacy development, or in the presence of heterogeneous treatment effects. These are not possibilities we can confidently reject, both in light of the heterogeneity in treatment effects suggested by Table 8 above, and due to the many inputs into the formation of literacy skills.

Before considering differential selection into attrition across treatment and control, we consider two counterfactuals that assume that attriters in the treatment group did not enjoy the treatment effect estimated from Equation 1. First, we impute missing wave 2 literacy scores using wave 1 scores. Since both outcome variables have been standardised using data from the control group, this effectively places attrited students in the same place in the (untreated) distribution at wave 2, as they were at wave 1. Second, we impute missing wave 2 test scores with the average test score value from the control group in wave 2.

This approach is more generous than the previous one, given that attrited students had below-average scores at wave 1.

The full results from these experiments can be found in Appendix B.2, Tables B.8 and B.9. Compared with our baseline finding of an overall treatment effect of 0.44 sd for Bangla literacy, and 0.68 sd for minority language literacy, both counterfactuals give smaller, although still strongly significant, treatment effects. For the counterfactual where wave 1 relative scores are imputed for wave 2 attriters, estimated treatment effects are 0.33 sd and 0.45 sd (for Bangla and minority language, respectively). In the second counterfactual, where attriters are assigned the average score of the control group, estimated treatment effects are 0.27 sd and 0.48 sd, respectively. In neither counterfactual does the interaction of treatment and minority language status become statistically significant.

6.2.2 Differential selection into the panel

The results shown above suggest that the main findings are qualitatively robust to a range of attrition scenarios, applied equally to treatment and control groups. The evidence presented in Table 4 suggests, however, that attrition patterns differed across treatment and control. This is of particular concern if attrition alone is leading to better outcomes in the treatment group: to the extent that these differences are not fully controlled for, it would lead to an upwards bias in the estimated treatment effects, increasing the risk that estimated effects are spurious. If, on the contrary, attrition is raising the average performance of the control group, this would still bias the estimated treatment effects - however, the bias would be towards zero, making our estimates a lower-bound on the true effect.

The correlates of attrition shown in Table 4 indicate that lower performing students are more likely to attrit in the control group compared with the treatment group. This is corroborated in Appendix Table B.6, which shows that attriters in the control group have 0.27 sd lower wave 1 Bangla literacy scores compared with attriters in the treatment group. This suggests that differential attrition is not driving wave 2 differences in literacy scores across treatment and control, as the relatively high attrition of students with poor test scores in the control group should raise the average scores.

While this is reassuring, there remains a real concern that selection into attrition across treatment and control - which shows measurable differences on the control variables we have - could lead to imbalance in variables we do not observe, yet are correlated with literacy

achievement. While we cannot control for this beyond the data available, we carry out some counterfactuals to understand how severe differential selection would need to be in order to overturn, or reverse, the estimated treatment effects. Keeping in mind that the threat to the qualitative nature of our results comes from the weakest students attriting from the treatment group, raising the average of those who remain in the panel, we proceed as follows. First, we impute missing values wave 2 literacy scores for all attriters, using Equation 1 (just as described in Section 6.2.1), and estimate our treatment effects of interest. We then make progressively more pessimistic assumptions about the attriters *in the treatment group*. We do this by lowering the predicted scores of this group by an increasingly large margin: starting with 0.25 sd, and increasing to 2.0 sd.

The results from this series of counterfactuals are presented in Table 10. The first row shows our baseline counterfactual, with missing values imputed from our primary regression specification (recall that these are nearly identical to our results with attriters dropped from the sample). The second row replicates this, with a 0.25 sd penalty applied to attriters in the treatment group: our coefficients of interest have decreased slightly (from 0.44 sd to 0.36 sd in Bangla literacy, and from 0.68 sd to 0.61 sd in minority language literacy), but remain positive and highly significant. The qualitative nature of the results is unchanged up to a penalty of 1.0 sd, at which point the treatment effect on Bangla literacy drops to 0.11 sd (just significant at conventional levels), while the minority language treatment effect remains quite large, at 0.41 sd. Beyond this, the Bangla literacy treatment effect becomes statistically insignificant. In none of the counterfactuals does the interaction term between treatment and minority language status become significant, although it changes sign and becomes positive as the attriter penalty in the treatment group grows (recall from Table 4 that minority language students are less likely to attrit, in both treatment and control, so penalising attriters in the treatment group will lead to relatively higher performance for minority language students).

These counterfactual simulations strongly suggest that our results are not driven by differential attrition across treatment and control. Although we cannot perfectly control for attrition, there is evidence that differential attrition should be biasing estimated treatment effects towards zero. Attriters in the treatment group would need to have substantially lower literacy scores than attriters in the control group to reverse the results: as it is, attriters in the treatment group had substantially *higher* scores at wave 1.

Table 10: **Main results estimated with correction for increasingly severe negative selection into attrition from treatment group**

	Bangla	Minority Language	Bangla: interaction
Predicted values	**0.44**	**0.68**	**-0.06**
	0.06	*0.09*	*0.10*
Attrited T -0.25 sd	**0.36**	**0.61**	**-0.03**
	0.06	*0.09*	*0.10*
Attrited T -0.5 sd	**0.28**	**0.54**	**0.00**
	0.06	*0.09*	*0.10*
Attrited T -1.0 sd	**0.11**	**0.41**	**0.07**
	0.06	*0.09*	*0.11*
Attrited T -1.5 sd	**-0.05**	**0.27**	**0.14**
	0.07	*0.10*	*0.12*
Attrited T -2.0 sd	**-0.21**	**0.14**	**0.20**
	0.08	*0.11*	*0.13*

Each cell reports the treatment effect estimate of interest from a separate OLS regression, with either Bangla literacy (Columns 1 & 3) or minority language literacy (Column 2) at wave 2 as the dependant variable. Values missing due to attrition are filled in with predicted values (Row 1), or predicted values subject to adjustment (Rows 2-6). Adjustments are made only to the attrited in the treatment group, demonstrating increasingly severe positive selection into the panel: Row 2 reduces the predicted wave 2 scores by 0.25 standard deviations, Row 3 by 0.5, and so forth. All regressions include controls for child sex, age, wave 1 literacy scores, and dummy variables for the four languages. Standard errors (in italics, listed below the point estimate) are clustered at the school level.

Why we see this particular pattern of attrition remains an interesting question. Given that poor academic performance can lead students to drop out or disengage with school, in some ways it is surprising that we find such similar rates of attrition across treatment and control: with such large improvements in literacy scores among the treatment group, it would not be unusual to see lower rates of attrition among treated students. Attrition in the treatment group also appears more unpredictable (with respect to observables) than in the control group. If attrition from the sample is a proxy for attendance, this would be an interesting aspect of the intervention to investigate in future programming.

7 Discussion

7.1 The intervention

Our main treatment effects estimates (Tables 5 and 6) show that the intervention had a large positive effect on literacy development of children in the treatment schools. The treatment effect of a 0.44 sd increase in Bangla literacy skills is substantial; the 0.68 sd improvement in minority language literacy is very large.

How might we expect these improvements to effect labour market outcomes, particularly for minority language speakers? Although rigorous identification in this field is challenging, previous studies have found that fluency in the nationally dominant language can have substantial labour market returns. Berman et al. (2003) study Hebrew (dominant language) fluency among immigrants from the Soviet Union to Israel. They find that Hebrew fluency had no effect on wages in low-skilled occupations, but a substantial effect in high-skilled occupations. As part of a wide-ranging paper on language policy, Laitin and Ramachandran (2016) estimate the returns to dominant language fluency in India. The authors use linguistic distance as an instrument for the difficulty of acquiring fluency in the dominant language for otherwise-similar minority language speakers. They find that fluency in the dominant language is associated with greater educational attainment and better occupational outcomes.

In our context in of Bangladesh, the language of instruction is also the dominant language at a national level. Improvement in Bangla fluency are therefore of particular interest in our context: these skills have the potential to translate into improved labour market outcomes. Bangla fluency is likely to be especially critical for minority language speakers with the

prospect of entering higher skilled occupations, as well as those who migrate away from their home communities.

Nevertheless, primary school learning remains modest in the sample, even in the treatment group. While the data show that students are indeed acquiring literacy skills, a look at the mean outcome values in wave 2 is sobering: mean scores across both treatment and control for intermediate literacy skills such as simple word reading and word recognition are only 78%-86%, despite the fact these children have been in school for at least 3 years (see Appendix Table A.2). This is in keeping with estimates of literacy among school children in Bangladesh. Data from UNICEF's Multiple Indicators Cluster 6 Survey (MICS6) indicate that, by the end of 2nd and 3rd grade, only 25% of children in Bangladesh could read 90% of words in a story and answer inferential questions. For comparison, 57% of children at a similar grade level in Thailand could perform the same reading comprehension task.[10] While interventions such as this one can help children make significant progress on literacy, there remains much to be done.

7.2 Transitional bilingual materials

Are transitional bilingual materials a substitute or a complement to the production of Bangla literacy skills? Estimates from Table 7 demonstrates that the additional mother tongue component of the intervention did not hinder, and nor did it accelerate, non-native Bangla speaker's development of Bangla literacy. These results are similar to those found by Piper et al. (2018) in Kenya, although the interventions themselves are quite different.

One of the primary arguments in favour of immersion over bilingual education is that it will bring students up to speed in the language of instruction more quickly, without taking a costly detour through bilingual literacy. Our results suggest that, at least for a very light touch bilingual programme such as this, this concern is unfounded. Despite substantial gains in literacy in their mother tongue, non-Bangla children in the treated schools saw equivalent growth in Bangla literacy skills as their Bangla-native peers.

Does this mean that mother tongue literacy and literacy in the dominant language are neither substitutes nor complements? Our analysis does not answer this definitively. Specifically, we do not know how teachers used these bilingual materials, and to what extent they did or did not divert class time from instruction in Bangla towards instruction in minority

[10] Data from the 'UNICEF Global database on foundational learning skills' (April 2021 version), accessed at https://mics.unicef.org/surveys.

languages. It could be that Bangla instruction time was decreased for minority language children; however complementarities between the two meant that these children easily kept up.

It could also be that teachers did not adjust teaching to make use of these bilingual materials, with no resulting loss in Bangla-focussed instruction. The bilingual learning materials could instead have brought new learning resources for minority language children, either through their parents or through an increased motivation to engage with their more-inclusive school.

Ultimately, our results are (mildly) good new for bilingual education: minority language learners benefited by acquiring literacy in their mother tongue, and did not suffer any negative consequences with respect to their literacy skills in the language of instruction. Given the potential benefits of mother tongue education beyond its role as a bridge to mastery of the dominant language, this is a success in itself. We do not find strong evidence, however, for a light touch bilingual programme such as this acting as a vital stepping stone that provides language learners a fast track to literacy.

7.3 Cost effectiveness

The Bangladesh Children's Book Initiative was carried out as a pilot study over two years. The ultimate goal of the intervention was 'more children reading more and better books'. To achieve this, the programme sought to improve reading outcomes for children in grades 1-3 by addressing the challenges of access to and effective use of quality, appropriate children's storybooks. Adopting a 'whole book chain' approach, the intervention included the development of storybooks by local authors, the publication and distribution of books, training of teachers and school leaders in the use of storybooks, community sensitisation and advocacy at various levels of government. Over two years, the project is estimated to have directly reached 5893 children, 5513 parents, 1031 teachers and school leaders. As part of this, 20,123 bilingual storybooks were distributed.

Calculating the cost effectiveness of complex interventions such as this is challenging: with multiple goals measured through different sorts of outcomes, and different inputs helping achieve more than one goal, linking costs to effects is difficult. It is further complicated in a pilot setting, where fixed costs tend to be a larger share of the total budget than during implementation at scale. It is nevertheless helpful to put the effects estimated here into

context with respect to the cost of the programme. The most straightforward approach is to calculate the total cost of the intervention per child served. Reaching 5893 children with a total budget of 297,047 GBP gives a per-child total cost of 50.41 GBP (as the project ran for two years, 25.21 GBP per child-year).

Estimating the cost per unit of outcome is more challenging, as it requires making assumptions about the relative importance of different outcomes. Our focus here has been estimating the literacy gains for children. If the sole purpose of the intervention was to raise Bangla literacy scores, the cost per child's 1 standard deviation increase in literacy would be: 50.41 GBP/ 0.44 sd = 114.57 GBP/sd improvement in Bangla literacy (0.0087 sd/GBP). Assuming that the gains in mother tongue literacy were equally important, and assuming (as found in the mapping exercise) that 52% of treated children were minority language speakers, would give a different calculation. Suppose 5893 children gained 0.44 sd in Bangla, and (0.52*5893 = 3064) children gained 0.68 sd in minority language literacy. Then the cost per sd in literacy improvement is closer to 297,047 GPB / (0.44*5893 sd + 0.68*3064 sd)= 297,047 GPB / (2417 sd + 2084 sd) =297,047 GPB / (4501 sd) = 66.00 GBP/sd improvement in literacy in any language (0.015 sd/GBP).

As a pilot study, the costs of programme development, technical assistance, and monitoring and evaluation are likely to be a high proportion of total costs. What would be the cost of extending the programme to an additional student, or school? To answer this question, we consider the original budget for the project, and make assumptions about which costs are fixed and which are variable. For the purposes here, we are interested in a marginal student or school, and so costs such as programme manager salary and monitoring and evaluation will be considered fixed: if the project were rolled out at a large scale, these costs would scale up as well, but that is beyond what we consider here. Using broad categories of costs (e.g. output level costs, non-thematic costs) gives an estimate of fixed costs as 32% of the total budget. A line-by-line assessment (e.g. development of books, production of books, teacher training, manager salary) gives a higher estimate of fixed costs, at 66% of total.

Using these two figure as upper and lower bound estimates, reaching an additional child would cost the programme between 17.21-34.50 GBP. If the sole outcome of interest is Bangla literacy gains, the marginal cost per sd is therefore between 39.11-78.41 GBP (0.026-0.013 sd/GBP). Assuming both minority language and Bangla literacy are equally important, the

marginal cost per sd in literacy improvement is 22.53-45.17 GBP (0.044-0.022 sd/GBP).[11]

8 Conclusion

This paper presents evidence from literacy intervention piloted in a multi-lingual area of rural Bangladesh. A complex literacy intervention with a light-touch transitional bilingual component, it achieved substantial gains in literacy: after two years of implementation, treated students had 0.44 sd higher literacy scores than students in control schools. Minority language students, who benefited from the provision of storybooks in both the dominant language and their mother tongue, saw considerable gains in early literacy skill development in their mother tongue, scoring 0.68 sd higher than control group students.

We find no evidence that the bilingual element of the programme either accelerated or decelerated non-native Bangla speaker's acquisition of Bangla literacy skills: native and non-native Bangla speakers enjoyed similar Bangla literacy gains from the intervention. On the one hand, this is positive news for bilingual education: minority language speakers in the study gained skills in their native language, at no cost to their literacy in the language of instruction. On the other hand, it suggests that hopes that bilingual education could be a highly-efficient bridge to literacy in other languages may be unfounded.

This study has a number of limitations. First, the intervention ran for only two years, with data collection taking place immediately afterwards. It is possible that complementarities between mother tongue literacy and the acquisition of other skills would grow with time - or indeed that the costs of investing in bilingualism might emerge at higher grade levels. Due to this timeline, we also have no information on how long the measured literacy gains persist, in either language. Second, learning outcome for other subjects have not been collected. In Kenya, Piper et al. (2018) found that mother tongue education reduced achievement in mathematics: a channel we cannot test with our data. Finally, the study was designed to evaluate the intervention as a whole, and as such further assumptions are required for our analysis of the mother tongue component separately.

[11] For comparison, Kerwin and Thornton (2021)'s pilot study in Northern Uganda achieved an improvement in reading of 0.032 SD per USD (full-cost version of programme) and 0.018 SD per USD (reduced-cost version). The intervention for that study took place in 2013, while the CBI started in 2016: in that year the USD:GPB conversion ranged from approximately 0.7 to 0.8. Assuming a conversion rate of 0.75, the Uganda intervention had a GBP cost effectiveness of 0.043 (full-cost) to 0.024 (reduced-cost). These measures are for reading in mother tongue only, as the programme relied on a mother-tongue-first approach, with the dominant language of instruction introduced in later years.

Furthermore, there are a number of limitations in applying the lessons learned from a pilot study such as this to transitional bilingual education programmes more widely. As Kerwin and Thornton (2021) starkly demonstrate, the context and the design of the intervention matter. The bilingual education introduced through this intervention was very minimal: students were provided with some learning materials, but may or may not have had any literate speakers of their language at home or at school in order to make these materials relevant. While our data show that these materials were nevertheless effective, it is not straightforward to extrapolate these findings to other contexts - or even to the same context with a modified intervention design.

This study adds to the growing body of work suggesting that bilingual education is not a panacea for the challenges faced by low-attainment, multi-lingual schools. Language of instruction is, however, an issue with implications far beyond the scores on literacy assessments. It would be beyond the scope of this research to make recommendations on whether mother tongue education is in the best interest of students from minority language groups. With respect to literacy outcomes in the dominant language of instruction, we find no benefits from a small bilingual addition to the curriculum: on the other hand, we find no costs either. The benefits in other domains, from mother tongue literacy to community buy-in for primary education, could be large.

The results presented above raise two important questions for further research. First, the gains in mother tongue literacy are large, given the nature of the intervention: how did they come about? Bilingual storybooks are a small input into the production of literacy skills: through what mechanism did they help children gain literacy in their mother tongue? In particular, it would be useful to estimate the relative contributions of school and household factors in this process. A better understanding of this would be useful when applying lessons from this study in other contexts.

Second, the study here was limited to short-run outcomes on two related competencies: literacy in Bangla and in the mother tongue. The long term impacts of transitional bilingual education are critical to language policy decisions. The design of this study makes further follow-ups impossible; future research should prioritise establishing middle to long-term effects on a range of outcomes.

References

Berman, E., K. Lang, and E. Siniver (2003): "Language-skill complementarity: returns to immigrant language acquisition," *Labour Economics*, 10, 265–290.

Chicoine, L. (2019): "Schooling with Learning: The Effect of Free Primary Education and Mother Tongue Instruction Reforms in Ethiopia," *Economics of Education Review*, 69, 94–107.

Chin, A. (2015): "Impact of bilingual education on student achievement," *IZA World of Labor*, 1–10.

Chin, A., N. M. Daysal, and S. A. Imberman (2013): "Impact of bilingual education programs on limited English proficient students and their peers: Regression discontinuity evidence from Texas," *Journal of Public Economics*, 107, 63–78.

Eriksson, K. (2014): "Does the Language of Instruction in Primary School Affect Later Labour Market Outcomes? Evidence From South Africa," *Economic History of Developing Regions*, 29, 311–335.

Ginsburgh, V. and S. Weber (2020): "The Economics of Language," *Journal of Economic Literature*, 58, 348–404.

Hynsjö, D. and A. Damon (2016): "Bilingual education in Peru: Evidence on how Quechua-medium education affects indigenous children's academic achievement," *Economics of Education Review*, 53, 116–132.

Kerwin, J. T. and R. L. Thornton (2021): "Making the Grade: the Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures," *Review of Economics and Statistics*, 103, 251–264.

Laitin, D. D. and R. Ramachandran (2016): "Language Policy and Human Development," *American Political Science Review*, 110, 457–480.

Laitin, D. D., R. Ramachandran, and S. L. Walter (2019): "The Legacy of Colonial Language Policies and Their Impact on Student Learning: Evidence from an Experimental Program in Cameroon," *Economic Development and Cultural Change*, 68, 239–272.

McEwan, P. J. (2015): "Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomized Experiments," *Review of Educational Research*, 85, 353–394.

Nath, S. R. (2013): "Factors influencing primary students' learning achievement in Bangladesh," *Research in Education*, 50–63.

Piper, B., S. S. Zuilkowski, D. Kwayumba, and A. Oyanga (2018): "Examining the secondary effects of mother-tongue literacy instruction in Kenya: Impacts on student learning in English, Kiswahili, and mathematics," *International Journal of Educational Development*, 59, 110–127.

Ramachandran, R. (2017): "Language use in education and human capital formation: Evidence from the Ethiopian educational reform," *World Development*, 98, 195–213.

Save the Children (2015): "Language Use and Literacy Skills: READ Baseline Assessment and Situational Analysis ," Tech. rep.

Seid, Y. (2018): "The impact of learning first in mother tongue: evidence from a natural experiment in Ethiopia," *Applied Economics*, 51, 577–593.

Slavin, R. E. and A. Cheung (2005): "A Synthesis of Research on Language of Reading Instruction for English Language Learners," *Review of Educational Research*, 75, 247–284.

Slavin, R. E., N. Madden, M. Calderón, A. Chamberlain, and M. Hennessy (2011): "Reading and Language Outcomes of a Multiyear Randomized Evaluation of Transitional Bilingual Education," *Educational Evaluation and Policy Analysis*, 33, 47–58.

Taylor, S. and M. von Fintel (2016): "Estimating the impact of language of instruction in South African primary schools: A fixed effects approach," *Economics of Education Review*, 50, 1–15.

Walter, S. L. and C. Benson (2012): "Language policy and medium of instruction in formal education," in *The Cambridge Handbook of Language Policy*, ed. by B. Spolsky, The University of St Andrews, 278–300.

A Data Appendix

A.1 Literacy data

The CLASS tool is designed to provide a rich assessment of a young student's literacy level, and therefore measures literacy in a number of ways. The survey opens with a series of questions about access to books and motivation to read. The formal literacy measures begin with letter recognition: children are presented with each letter in the appropriate alphabet, ordered from most to least commonly used. They are asked to sound out the letter if they are able to, and if not, to say so and move on. A percentage score is generated from this test based on the share of correctly pronounced letters.[12]

The letter recognition task is followed by a simple writing exercise. First, students are asked to write three different letters; next to spell their name; and finally to spell out a long word in three steps. Each of these seven tasks receives a 0/1 grade, which are aggregated to form a percent score for the letter and word spelling section.

The next assessment tests word recognition. A cloud of words is shown to the child. These words are drawn from the grades 1-3 curriculum, so that a child reading at grade 1 level should be able to recognise some but not all, while a child reading at grade 3 level should be able to recognise all the words. The child is asked to read out each word in turn, pronouncing as carefully as they can, while the enumerator notes the success of each one. Just like for letter recognition, a percentage score is generated based on the share of correctly pronounced words.

Word recognition is further explored with two other exercises. In the first, the enumerator reads a word, and the student must identify it from a list of five words with some similarity. This is done three times, with words commonly used in grades 1, 2, and 3. Finally, a similar exercise is carried out where the student is shown an image, and must identify the corresponding word from a list of five. These two exercises are combined to give a percentage score for printed word recognition.

The literacy assessment finishes with a fluency and comprehension task. The student is given a short story written in several paragraphs (approximately 15 sentences), and asked to read it aloud to the enumerator. The enumerator marks how far the student has read

[12] The number of letters used in this exercised varied by language, as each language uses a different alphabet. The Bangla exercise included 50 letters, Chakma 33, Marma 48, and Tripura (which is written using the Roman alphabet) 24.

after one minute, then allows the student to continue. When the student has finished or doesn't wish to continue, they are then asked four simple comprehension questions about the story. Both of these are converted to percentage scores.

While these different indicators are interesting in themselves, as they capture different aspects of early literacy, it is useful to generate a single score. We do this by averaging the percentage scores from the six exercises described above. Another way of thinking about this is an overall score that weights the six exercises equally. For analysis, all outcome variables are standardised against the data from the control group in the same wave.

For the purposes of this study, the CLASS tool was also translated into the three common minority languages. Each task was first presented in Bangla; minority ethnic students were then asked similar questions in their mother tongue. Due to problems with translation, the fluency and comprehension scores from the first wave are not valid;[13] however very few students succeeded at recognising even a few letters in their mother tongue, so these advanced questions would have been inaccessible to even the most advanced students in wave 1. Nevertheless, to maintain comparability of scores between waves, we will only use data from the first four tasks (letter recognition, letter and word spelling, word reading, and word recognition). The overall measure of mother tongue literacy averages scores of these four, making it not directly comparable in level to the Bangla literacy score.

Wave 1 values for these variables are shown in Table A.1. Students in the sample had reasonably good Bangla letter recognition and some ability to spell letters and simple words. Data collection took place in November, near the end of the academic year: these were first grade students, so those who were not repeating the year should have now had one full year of primary school. Students could read an average 33% of the words they were presented in Bangla; however they only managed to read an average of 4% of the story, and for the most part could not answer any comprehension questions. Minority language speakers in general had no knowledge of their mother tongue alphabet: on average they could recognise 0.5%-2.6% of the letters presented. These averages are driven up by a few individuals with some familiarity with their mother tongue alphabet: 94% of minority language students scored a 0 on this task. Table A.2 gives outcome variables at wave 2, for the panel observations. A further breakdown of the overall scores by language group is presented in Section B.1.

[13] At wave 1, an error in translation led to the Bangla story being translated into the minority languages. This led to the comprehension questions in non-Bangla languages being rendered useless, as children were already familiar with the story from the previous section. In part for this reason, analysis of the story reading and story comprehension sections of the minority language assessment have been dropped.

Table A.1: **Outcome variables in wave 1: full wave 1 sample**

Variable	Mean	Std. Dev.	Min.	Max.	N
Overall Bangla score	39.19	20.55	0	86.32	1313
Bangla letter recognition	84.8	26.5	0	100	1313
Bangla letter and word spelling	61.96	27.34	0	100	1313
Bangla word reading	37.4	41.18	0	100	1313
Bangla word recognition	41.3	37.05	0	100	1313
Bangla story reading (1 min)	3.49	6.05	0	56.34	1313
Bangla story comprehension	6.21	18.25	0	100	1313
Overall Min Lang score	1.05	4.13	0	48.74	910
Min Lang letter recognition	1.16	8.18	0	93.75	910
Min Lang letter and word spelling	1.16	7.25	0	100	910
Min Lang word reading	0.01	0.13	0	2.33	910
Min Lang word recognition	1.87	8.35	0	100	910

Average values from wave 1.

Table A.2: **Selected outcome variables at wave 2 (% scores)**

Variable	Mean	Std. Dev.	Min.	Max.	N
Overall Bangla score	72.51	16.34	9.76	94.73	918
Bangla letter recognition	95.23	9.55	12	100	918
Bangla letter and word spelling	85.05	19.19	0	100	918
Bangla word reading	78.25	29.94	0	100	918
Bangla word recognition	85.88	24.81	0	100	918
Bangla story reading (1 min)	24.33	15.85	0	69.72	918
Bangla story comprehension	66.34	33.35	0	100	918
Overall Min Lang score	24.27	26.56	0	98.90	662
Min Lang letter recognition	41.96	40.25	0	100	662
Min Lang letter and word spelling	21.46	30.64	0	100	662
Min Lang word reading	14.63	29.97	0	100	662
Min Lang word recognition	19.03	31.96	0	100	662

Average values from wave 2 (panel).

A.2 Data Quality

Data collection was carried out in-house by Save the Children, not by a research-focussed enumeration team. The complexity of the CLASS tool, the multi-lingual assessment, as well as the panel design, made this a challenging data collection effort, and there are necessarily some caveats to be made about the quality of the data. A few comments are noted here; further information is available on request.

Data collection in wave 1 was collected on paper, which needed to be entered and, in some cases, translated. In wave 2 data was collected electronically. A number of inconsistencies in pre-determined characteristics were present following the linkage of the two waves of data. The majority of these were resolved by the local monitoring and evaluation coordinator, however it is likely that some number of matches are erroneous.

A.3 Further balance checks

A number of balance checks have been carried out in order to compare treatment and control groups. An overview of these is presented in the main text. Tables A.3 and A.4 look at balance within subsections of the wave 1 literacy assessments. Table A.3 compares average scores of each section of the Bangla and minority language literacy assessments, while Table A.4 repeats this for Bangla literacy assessments, but looking separately at native speakers and Bangla-learners.

Table A.3: **Balance checks II: detailed dependent variables at first wave**

Variable	(1) Control N	(1) Control Mean/SE	(2) Treat N	(2) Treat Mean/SE	T-test P-value (1)-(2)
Overall Bangla score	653	38.704	660	39.677	0.391
		(0.817)		(0.787)	
Bangla letter recognition	653	84.634	660	84.964	0.822
		(1.061)		(1.008)	
Bangla letter and word spelling	653	61.453	660	62.457	0.506
		(1.088)		(1.047)	
Bangla word reading	653	36.499	660	38.294	0.430
		(1.617)		(1.598)	
Bangla word recognition	653	39.816	660	42.778	0.148
		(1.469)		(1.422)	
Bangla story reading (1 min)	653	3.506	660	3.470	0.914
		(0.230)		(0.242)	
Bangla story comprehension	653	6.317	660	6.098	0.828
		(0.697)		(0.727)	
Overall Min Lang score	460	1.237	449	0.861	0.170
		(0.211)		(0.174)	
Min Lang letter recognition	460	1.544	449	0.767	0.153
		(0.445)		(0.307)	
Min Lang letter and word spelling	460	1.180	449	1.145	0.943
		(0.300)		(0.378)	
Min Lang word reading	460	0.014	449	0.008	0.483
		(0.006)		(0.006)	
Min Lang word recognition	460	2.210	449	1.522	0.215
		(0.409)		(0.374)	

Each row presents a t-test for difference in means across treatment and control groups.

Table A.4: **Balance checks III: detailed dependent variables at first wave by Bangla native or learners**

Variable	(1) Control N	(1) Control Mean/SE	(2) Treat N	(2) Treat Mean/SE	T-test P-value (1)-(2)
Bangla score: learners	460	37.639	449	40.517	0.036**
		(0.977)		(0.966)	
Bangla score: natives	193	41.242	211	37.889	0.094*
		(1.478)		(1.352)	
Bangla letter recognition: learners	460	83.470	449	86.116	0.146
		(1.354)		(1.207)	
Bangla letter recognition: natives	193	87.409	211	82.512	0.044**
		(1.563)		(1.819)	
Bangla letter and word spelling: learners	460	60.533	449	63.156	0.150
		(1.312)		(1.260)	
Bangla letter and word spelling: natives	193	63.644	211	60.968	0.322
		(1.939)		(1.878)	
Bangla word reading: learners	460	35.043	449	39.343	0.118
		(1.932)		(1.952)	
Bangla word reading: natives	193	39.968	211	36.060	0.335
		(2.943)		(2.783)	
Bangla word recognition: learners	460	38.152	449	44.358	0.012**
		(1.727)		(1.746)	
Bangla word recognition: natives	193	43.782	211	39.415	0.235
		(2.772)		(2.434)	
Bangla story reading (1 min): learners	460	3.582	449	3.723	0.731
		(0.277)		(0.302)	
Bangla story reading (1 min): natives	193	3.324	211	2.930	0.491
		(0.410)		(0.398)	
Bangla story comprehension: learners	460	5.054	449	6.403	0.274
		(0.807)		(0.932)	
Bangla story comprehension: natives	193	9.326	211	5.450	0.026**
		(1.344)		(1.114)	

Each row presents a t-test for difference in means across treatment and control groups.

B Extended Results

B.1 Extensions by language group

B.1.1 Summary statistics by language group

Tables B.1 & B.2 show the literacy scores for children in each wave, by language group, for control and treatment groups respectively. Looking at control group children only, Table B.1 demonstrates differences between language groups, absent the treatment. While Table 2 showed that minority language students as a group had almost identical wave 1 Bangla literacy scores as Bangla natives, Table B.1 shows that some groups (e.g. Chakma natives) perform as well or better in Bangla over the two waves as Bangla natives, while others (Marma and Tripura natives) lag behind. This is consistent with anecdotal evidence on the relative socioeconomic position of the four ethnic groups, as well as the attitudes towards education. Similar trends are present in wave 1 data for the treatment group in Table B.2; however the ordering of the different groups is slightly different.

A further notable statistic from Table B.1 is the large increase in Tripura literacy in the control group (similarly outstanding in the treatment group). This can be at least partly attributed to the fact that the Tripura language is transcribed using the latin alphabet. Given the prevalence of latin letters in the environment, through English language materials and adverts, Tripura students would have a much greater exposure to their mother tongue alphabet than would student of the other minority language groups.

Table B.1: **Literacy scores by language group in wave 1 and wave 2: control only**

	Chakma native	Marma native	Tripura native	Bangla native
Wave 1 Bangla	**41.91**	**37.19**	**31.44**	**41.24**
	20.85	*21.32*	*18.63*	*20.53*
Wave 2 Bangla	**74.36**	**67.62**	**61.76**	**69.26**
	15.78	*17.68*	*16.32*	*16.45*
Wave 1 Minority Lang	**0.46**	**1.16**	**2.72**	
	3.51	*4.86*	*4.94*	
Wave 2 Minority Lang	**10.01**	**9.18**	**48.59**	
	13.69	*17.13*	*21.33*	
N: wave 1	156	212	92	193
N: wave 2	132	178	69	152

Average values from wave 1 (full sample) and wave 2 (panel): control group only.

Table B.2: **Literacy scores by language group in wave 1 and wave 2: treatment only**

	Chakma native	Marma native	Tripura native	Bangla native
Wave 1 Bangla	**43.12**	**40.54**	**35.50**	**37.89**
	20.61	*20.70*	*18.69*	*19.64*
Wave 2 Bangla	**79.67**	**76.73**	**72.18**	**75.30**
	13.82	*13.90*	*16.22*	*15.08*
Wave 1 Minority Lang	**0.14**	**0.95**	**1.94**	
	0.85	*4.76*	*3.91*	
Wave 2 Minority Lang	**24.62**	**23.99**	**64.23**	
	25.70	*23.01*	*16.17*	
N: wave 1	162	194	92	211
N: wave 2	137	155	62	145

Average values from wave 1 (full sample) and wave 2 (panel): treatment group only.

B.1.2 Results by language group

Throughout the main analysis, the minority languages are grouped together as 'non-Bangla natives.' Although all regressions include dummy variables for each language group, it is possible that this is not sufficient to capture differences between them. Of particular concern is the risk that different language groups have heterogeneous response to treatment which may be going in opposite directions. To rule this out, we replicate the main analysis separately for each language group (for overall treatment effects) and for each minority language versus Bangla for the interaction effects.

The results from this language-specific regressions are presented in Tables B.3-B.5. Despite the reduced sample size in each regression, the literacy results remain strongly significant and qualitatively similar to the headline results. None of the interaction effects becomes significant when estimated using only pairs of languages, although this is perhaps not surprising: the standard errors were large when the full sample was used, and are larger if anything in these small-sample estimations. These regressions do reassure us, however that, despite differences between them, aggregating the minority languages is not summing effects that are going in opposite directions.

It is worth noting that Chakma natives show the smallest point estimate for Bangla literacy treatment effects; the Chakma interaction effect is also the largest in absolute value (-0.17), and the closest to being statistically significant. Chakma students, as we see in Tables B.1 & B.2 also have the highest Bangla literacy scores in both wave 1 and wave 2. Without seeking to over-interpret a statistically insignificant result, this combines to form suggestive evidence that Chakma students benefited slightly less than others from the

programme, perhaps due to their literacy level being above average to begin with.

Table B.3: **Treatment effects: Bangla literacy by language group**

	(1) Chakma natives	(2) Marma natives	(3) Tripura natives	(4) Bangla natives
Treat	0.330***	0.412**	0.465**	0.505***
	(0.122)	(0.156)	(0.190)	(0.116)
Observations	233	303	129	253
R^2	0.259	0.211	0.318	0.305

Each column reports the estimated coefficients of interest from a separate OLS regression with wave 2 Bangla literacy as the dependant variable. Regressions are run separately by language group, as indicated by the column headings. All regressions include controls for child sex, age, wave 1 Bangla literacy scores. Standard errors (in parentheses) are clustered at the school level.

Table B.4: **Treatment effects: minority language literacy by language group**

	(1) Chakma natives	(2) Marma natives	(3) Tripura natives
Treat	0.756***	0.646***	0.684***
	(0.177)	(0.180)	(0.204)
Observations	233	300	128
R^2	0.198	0.130	0.147

Each column reports the estimated coefficients of interest from a separate OLS regression with wave 2 minority language literacy as the dependant variable. Regressions are run separately by language group, as indicated by the column headings. All regressions include controls for child sex, age, wave 1 minority language literacy scores. Standard errors (in parentheses) are clustered at the school level.

B.2 Extended attrition checks

The balance of baseline characteristics of students who attrited from the panel is given in Table B.6. Reinforcing the regression results from Table 4, these balance checks show that attrited students in control were more likely to be boys, and to have lower Bangla literacy scores. We can see in Table B.6 that the difference in literacy scores is coming from the most basic elements of the assessment (letter recognition, letter and word spelling, word reading and word recognition). Although few students were able to complete questions beyond this at baseline, this suggests that average differences in test scores between attrited in treatment and control are not being driven by differences at the top end of the ability distribution.

Table B.5: **Treatment effects: Bangla literacy treatment interaction by language group**

	(1) Chakma & Bangla	(2) Marma & Bangla	(3) Tripura & Bangla
Treat=1	0.496***	0.482***	0.506***
	(0.117)	(0.118)	(0.117)
Treat=1 × Minority language=1	-0.168	-0.0794	-0.0257
	(0.170)	(0.189)	(0.214)
Observations	486	556	382
R^2	0.306	0.244	0.317

Each column reports the estimated coefficients of interest from a separate OLS regression with wave 2 Bangla literacy as the dependant variable. Regressions are run separately for pairs of languages, as indicated by the column headings. All regressions include controls for child sex, age, wave 1 Bangla literacy scores and language. Standard errors (in parentheses) are clustered at the school level.

B.2.1 Selection into panel

Tables B.7-B.9 present the results from some further checks exploring the robustness of the main results to different assumptions on attrition. Table B.7 presents results from the three headline estimations, with missing values from wave 2 imputed through prediction. This is achieved as follows. First, the relation between wave 2 literacy scores observables is estimated using panel observations, as described in Equation 1. Second, wave 2 test scores that are missing due to attrition are predicted using the estimated values for coefficients from step 1. Finally, Equations 1 and 2 are re-estimated, using the full wave 1 sample, with missing wave 2 test scores imputed. Effectively, this attributes the estimated treatment effect from the panel to the attrited sample members, while using information from wave 1 (sex, age and initial literacy scores) to predict wave 2 outcomes. It is perhaps not surprising, then, that the estimated treatment effects are almost identical to those found using the panel alone.

A more conservative approach would be to assume that the attrited sample would have benefitted less from treatment than panel members. Tables B.8 and B.9 present results from two experiments along these lines, each relying on a different approach to imputation. Table B.8 replicates the main results, imputing missing wave 2 literacy scores using wave 1 scores. Given that scores are standardised against the control group in each wave, this effectively assigns the attrited individual to the same *relative position* in the ranking of literacy scores compared with the control group. If all students had an identical growth in literacy scores, and if test scores are a perfect measure of this, this imputation strategy would perfectly

Table B.6: **Attrition: characteristics of attrited sample by treatment status**

Variable	(1) Control N	(1) Control Mean/SE	(2) Treat N	(2) Treat Mean/SE	T-test P-value (1)-(2)
Age at wave 2	188	8.968	199	8.910	0.643
		(0.098)		(0.081)	
Girls	193	0.383	202	0.505	0.015**
		(0.035)		(0.035)	
Chakma	193	0.228	202	0.203	0.547
		(0.030)		(0.028)	
Marma	193	0.290	202	0.233	0.194
		(0.033)		(0.030)	
Tripura	193	0.130	202	0.149	0.587
		(0.024)		(0.025)	
Bangla	193	0.352	202	0.411	0.232
		(0.034)		(0.035)	
Overall Bangla score	193	-0.296	202	-0.030	0.008***
		(0.075)		(0.066)	
Bangla letter recognition	193	77.057	202	83.594	0.026**
		(2.328)		(1.806)	
Bangla letter and word spelling	193	51.974	202	60.974	0.001***
		(2.127)		(1.835)	
Bangla word reading	193	22.332	202	29.792	0.020**
		(2.276)		(2.252)	
Bangla word recognition	193	31.606	202	40.017	0.023**
		(2.685)		(2.520)	
Bangla story reading (1 min)	193	2.733	202	2.894	0.782
		(0.435)		(0.388)	
Bangla story comprehension	193	4.275	202	4.084	0.900
		(1.097)		(1.050)	
Chakma letter recognition	44	0.000	41	0.074	0.303
		(0.000)		(0.074)	
Marma letter recognition	56	4.613	47	0.310	0.096*
		(2.332)		(0.310)	
Tripura letter recognition	25	0.000	30	0.417	0.220
		(0.000)		(0.306)	

Each row presents a t-test for difference in means across treatment and control groups.

Table B.7: **Attrition: replication of main results, attrited wave 2 scores replaced with predicted values**

	(1) Bangla	(2) MinLang	(3) Bangla interaction
Treat	0.438***	0.678***	0.471***
	(0.0570)	(0.0865)	(0.0733)
Minority language=1			-0.0971
			(0.0876)
Treat=1 × Minority language=1			-0.0644
			(0.103)
Observations	1305	899	1305
R^2	0.373	0.498	0.374

Each column reports the estimated coefficients of interest from a separate OLS regression with wave 2 Bangla literacy (Columns 1 & 3) or minority language literacy (Column 2) as the dependant variables. Missing values for the outcome variable are imputed using predicted values, as described the text. All regressions include controls for child sex, age, language and wave 1 values of the dependent variable. Standard errors (in parentheses) are clustered at the school level.

predict the growth in literacy of attrited students. Estimates using this imputation strategy give lower values for treatment effects (+0.33 sd for Bangla, +0.45 sd for minority language). Given that this approach assumes that attrited students from the treatment group would have had a treatment effect of 0, and that the sample had 30% attrition, this reduction of about 1/3 the magnitude of the treatment effect is predictable. However, the estimated treatment effects remains strongly significant, and quite large; the interaction effect remains small and statistically insignificant.

Finally, we assign the control group average wave 2 score to attrited panel members. This approach again assumes a treatment effect of 0 for attrited members of the treatment group. Given that students attrited from control had lower wave 1 test scores than students attrited from treatment, this approach is even more conservative than the pervious one, as it assigns all these students the same score. Results from a replication of the three main regressions using a panel with missing values imputed in this way are given in Table B.9. With respect to the previous experiment, the estimated treatment effects on literacy are lower still for Bangla (+0.27 sd), but slightly higher for minority languages (+0.48 sd); in both cases qualitatively very similar, and strongly significant. As above, the interaction effect remains small and statistically insignificant.

Table B.8: **Attrition: replication of main results, with attrited sample assigned wave 1 standardised score in wave 2**

	(1) Bangla	(2) MinLang	(3) Bangla interaction
Treat	0.332***	0.447***	0.346***
	(0.0627)	(0.0941)	(0.0759)
Minority language=1			-0.0910
			(0.101)
Treat=1 × Minority language=1			-0.0261
			(0.111)
Observations	1305	903	1305
R^2	0.443	0.353	0.443

Each column reports the estimated coefficients of interest from a separate OLS regression with wave 2 Bangla literacy (Columns 1 & 3) or minority language literacy (Column 2) as the dependant variables. Missing values for the outcome variable are imputed using wave 1 values, as described the text. All regressions include controls for child sex, age, language and wave 1 values of the dependent variable. Standard errors (in parentheses) are clustered at the school level.

Table B.9: **Attrition: replication of main results, with attrited sample assigned control group average wave 2 score**

	(1) Bangla	(2) MinLang	(3) Bangla interaction
Treat	0.272***	0.475***	0.266***
	(0.0574)	(0.0882)	(0.0753)
Minority language=1			-0.0755
			(0.0807)
Treat=1 × Minority language=1			0.0108
			(0.106)
Observations	1305	903	1305
R^2	0.191	0.287	0.191

Each column reports the estimated coefficients of interest from a separate OLS regression with wave 2 Bangla literacy (Columns 1 & 3) or minority language literacy (Column 2) as the dependant variables. Missing values for the outcome variable are imputed using wave 2 average values from the control group, as described the text. All regressions include controls for child sex, age, language and wave 1 values of the dependent variable. Standard errors (in parentheses) are clustered at the school level.

B.3 Results: unweighted sample

Regression results in the paper are all weighted to correct for purposeful oversampling of minority language students. This reweighting effectively gives more weight to the Bangla observations, and less weight to the Chakma and Tripura observations. Marma observations (with 28% frequency in the population and 31% frequency in the sample) are given almost unitary weight. These weights allow us to interpret our estimates as an average treatment effect; however, it is useful to know how influential the weights are in and of themselves.

Table B.10 replicates the three main regressions without using weights, in other words, keeping the sample frequencies of each language group. The point estimates for the treatment effects on literacy are almost unchanged, and the interaction effect remains small and statistically insignificantly different from zero.

Table B.10: **Main findings: unweighted regressions**

	(1) Bangla	(2) MinLang	(3) Bangla interaction
Treat	0.420***	0.685***	0.487***
	(0.0828)	(0.113)	(0.118)
Minority language=1			-0.0862
			(0.126)
Treat=1 × Minority language=1			-0.0923
			(0.148)
Girls	0.0826	0.159**	0.0828
	(0.0548)	(0.0738)	(0.0551)
Age at wave 2	-0.0235	-0.0207	-0.0235
	(0.0308)	(0.0401)	(0.0309)
Language=1	0.292***	-1.731***	0.425***
	(0.0896)	(0.143)	(0.107)
Language=2	0.0708	-1.714***	0.203*
	(0.0949)	(0.131)	(0.112)
Language=3	-0.133		
	(0.103)		
Overall Bangla score	0.438***		0.440***
	(0.0419)		(0.0422)
Overall Min Lang score		0.102	
		(0.0643)	
Constant	0.0333	1.483***	-0.00100
	(0.282)	(0.376)	(0.298)
Observations	918	661	918
R^2	0.282	0.439	0.282

Each column reports the estimated coefficients from a separate OLS regression with wave 2 Bangla literacy (Columns 1 & 3) or minority language literacy (Column 2) as the dependant variables. Regressions are done without weights to adjust for oversampling of minority language speakers. All control variables are shown. Standard errors (in parentheses) are clustered at the school level.