

Promoting Mental Health and Educational Success among Young People in the Global South



Jennifer Opare-Kumi

Exeter College

University of Oxford

An article-based thesis submitted for the degree
of Doctor of Philosophy in Public Policy

Hilary 2024

Abstract

This article-based thesis aims to study a range of mental health and education reforms and interventions addressing the overall mental health and educational success of young people in Global South countries.

Article 1 studies the relationship between foundational learning and mental health. Using a Difference in Difference (DD) research design, this article leverages an ongoing pedagogical intervention - Teaching at the Right Level (TaRL) - to explore the effects of targeted instruction on the mental health and educational success among 1297 primary school students in Botswana. Relative to the learners yet to be exposed to the targeted instruction pedagogy, I find that the foundational learning intervention significantly improves the mental health outcomes of students. To my knowledge this is the first study to explore the impact of TaRL on mental health. This research bridges an important gap at the intersection of education and public health literature and underscores the importance of context-relevant policy reforms that prioritise pedagogic approaches fostering holistic learning and development. **Article 2** studies the impact of a mindfulness-based programme on the mental health and educational success of secondary school students in low-cost private schools in Pakistan. A sample of 214 students are randomly assigned to a Mindfulness-based Stress Reduction (MBSR) group and a neutral class activity control group. I find that MBSR significantly improves student prosociality and behaviour change indicators such as improved mood and reduced stress at home. Additionally, students exposed to the MBSR intervention demonstrated higher levels of grit. These findings are particularly meaningful for Global South countries where fewer resources are typically allocated towards the mental health of young people. **Article 3** studies the impact of English Medium Instruction (EMI) policies on the education outcomes of upper primary school students in Ethiopia. This article studies EMI transition policies using a longitudinal dataset from Ethiopia, leveraging regional variation in education language policy reforms for causal identification. Employing a dynamic value-added model, I find that students in schools using EMI as a medium of instruction have significantly lower mathematics test scores compared to students in mother tongue education schools. Furthermore, English medium learners do not outperform their mother tongue education counterparts in their English test scores. Interestingly, treatment heterogeneity suggests some positive long-term effects of an early grade transition from

mother tongue to EMI, relative to mother tongue-only instruction. However, the overall results support the prolonged utilisation of mother tongue instruction in primary school education. These findings add some nuance to the debate on the "right" grade to transition to English medium instruction and the results are particularly relevant for policy makers in linguistically diverse contexts.

Together, these three articles improve our understanding on policies and interventions supporting the education and mental health outcomes for the whole child. Particularly in low-middle income countries where this data and research is sparse, this research bridges a significant gap.

0.1 Declaration

The work described in my thesis was carried out between October 2019 and September 2023 under the supervision of Professor Stefan Dercon and Associate Professor Kate Orkin. I hereby declare that this thesis entitled *Promoting Mental Health and Educational Success among Youth in the Global South* is the result of my own work.

I can confirm that no part of my thesis has been accepted or is currently submitted for any degree, diploma or certificate, or other qualification at the University of Oxford, or any other university or similar institute of learning.

Article 2 of this thesis is co-authored work with Dr. Zahra Mansoor. It was conceptualised in equal collaboration with Dr. Mansoor who was a former DPhil and Post-doctoral research fellow at the Blavatnik School of Government (BSG). As collaborators, I was the lead researcher, responsible for the literature review, identifying and constructing the outcomes measures and for the main regression analysis. Dr. Mansoor provided feedback on all the data analysis and written work and led on the in-country (Pakistan) logistics and coordination.

Article 3 uses data from the Young Lives Longitudinal Study. This data is publicly archived and available from the UK Data Archive. This article has been submitted to a peer-review journal and I have received comments to revise and resubmit the article.

This article-based thesis does not exceed the prescribed word limit of 100 000 words.

I have no conflicts of interest to declare.

0.2 Acknowledgements

This DPhil journey has been an arduous but rewarding journey and I am sincerely grateful to many people for their thoughtful feedback, words of encouragement and support.

My DPhil was supported by funding from the Blavatnik School of Government with support from the Exeter College John Kufuor Fellowship. I am grateful for the generous funding that made this research possible. My first article was made possible through funding from the RISE Programme and the Centre for the Study of African Economies (special thanks to Rose and Sue). My second article was funded by the John Fell fund.

Stefan Dercon and Kate Orkin have been extraordinary supervisors. They have kindly provided years of constructive and detailed feedback that has shaped this thesis. Thank you for always expecting more and pushing me in ways that has made this work so much better.

In Botswana, Pakistan and Ethiopia, I am sincerely grateful and indebted to every student, teacher, school and policymaker who made this research possible. I hope this work contributes towards improving the lives of young people. Thank you for your openness and enthusiastic participation.

In Botswana, my sincere appreciation goes to Director Labane Mokgosi, David Ananius and Rapelang Bogatsu for supporting the research and granting the permissions that made the Botswana chapter possible. I worked with an amazing team of collaborators and friends who went above and beyond to help me execute this research. I thank Katlego Sengadi, Otsile Mphusu, Boamogeleng Motlhobogwa, Thato Letsomo Israel, Lorato Gaolebe, and Efua Bortsie.

In Pakistan, I am deeply grateful to my co-author, Dr Zahra Mansoor who has been such a joy to work with and learn from. Thank you for trusting me to work on this project with you. I also extend my thanks to Be-Khudi Meditation and Sirajia Institutions for their valuable insights during the research design phase and careful implementation during programme roll-out. Special thanks to Tehmina Mansoor, Zohaib Zubay, Sarfraz Bhatti and their teams for their belief in the importance of this work.

My work was also significantly improved through conversations with academics, colleagues

and friends. I am grateful for the helpful feedback and support received from David Johnson, Martin Williams, Julien Labonne, Lucie Cluver, Marta Favara, Caine Rolleston, Dapo Akande, Gabby Areng, Zahra Mansoor, Julia Ruiz Pozuelo, John Walker, Gabatshwane Tsayang, Noam Angrist, Rosemond Opare-Kumi, and from participants at RDS, DRIPS, CSAE, OPM, AGW India, ADE Pakistan, ICERDA Ghana, RISE Programme seminars, and many conferences.

To my Oxford community of friends, and other my friends all around the globe, I am so grateful for all your love, prayers, coding fix conversations, the allotment dates, coffees, and general banter. This has sustained me and kept me whole. Thank you. Thank you Gabby. We sat in a coffee shop in 2017 dreaming about doing our PhDs one day. It is miraculous that the dream became a reality for both of us. I am so blessed to have gone through this doctoral journey with you. Thank you for cheering me on every step of the way, for our guiding orikis, meadow walks with our sweet pupp Monet, daydreaming and every practice and ritual that reminded me I was capable and deserving, and allowed me to stay near to myself. I couldn't have done any of this without you- THANK YOU.

Finally, my love to the entire Opare-Kumi and Owusu-Bamfo family who have all at some point lent me their faith and belief, carrying me through this DPhil. Thank you for believing that this would be possible for me. My paternal grandmother passed away in 2020 during the COVID pandemic. From a young age, she nicknamed me *doctor* and only called me *doctor* for years. At some point I promised her that I would one day get my doctorate to finally become deserving of this title she had bestowed on me. Sadly she passed away during the COVID pandemic in 2020, but I know what I have accomplished through this DPhil has made her proud.

Table of Contents

0.1	Declaration	1
0.2	Acknowledgements	2
	List of Figures	7
	List of Tables	9
1	Introduction	11
2	Foundational Learning and Mental Health	18
2.1	Introduction	19
2.2	Background	23
2.2.1	Mental Health	23
2.2.2	Botswana context	24
2.2.3	Teaching at the Right Level - a foundational learning intervention	26
2.3	Outcome measurement and Psychometric validation	29
2.3.1	Measurement	29
2.3.2	Psychometric validation	31
2.3.3	Ethics	33
2.4	Research Design and Econometric Specification	33
2.4.1	Treatment Assignment	33
2.4.2	Research Design	35
2.4.3	Econometric Specification and Hypothesis	37
2.4.3.1	Hypotheses	38
2.5	Data and Descriptive Statistics	39
2.5.1	Data	39
2.5.2	Descriptive Statistics	40
2.6	Results	43
2.6.1	Treatment Heterogeneity	45
2.7	Discussion	47

2.8	Conclusion	49
2.9	References	50
3	Improving the mental health and educational success of young people: experimental evidence from a mindfulness-based programme in Pak- istan	58
3.1	Introduction	59
3.2	Context	63
3.2.1	Literature on mindfulness and mindfulness-based stress reduction interventions	63
3.2.2	Mental health and in Pakistan	65
3.3	Theory and Research Questions	66
3.3.1	Research Questions	66
3.4	Experimental Design	67
3.4.1	Intervention	67
3.5	Randomization	68
3.6	Data and Balance Checks	69
3.6.1	Measurement	69
3.6.2	Summary Statistics at Baseline	73
3.6.3	Data Validity	74
3.6.4	Balance Checks	76
3.7	Results	77
3.7.1	Treatment Effects on Mental Health Outcomes	78
3.7.2	Treatment Effects on Behaviour Change	78
3.7.3	Treatment Effects on Capacity to Learn and Educational Success Outcomes	80
3.8	Treatment Heterogeneity, by Sex	81
3.9	Discussion and Concluding Remarks	84
3.10	References	86
4	English Medium Instruction in Multilingual Contexts: Empirical Evidence from Ethiopia	91

4.1	Introduction	92
4.2	Policy Context and Review of Quantitative Evidence	96
4.2.1	Education policy context in Africa	96
4.2.2	Review of the quantitative evidence on the effect of English Medium Instruction (EMI)	97
4.2.3	Summary of regional motivations for Language of Instruction poli- cies in Ethiopia	99
4.3	Data, Sampling Design and Assessment Design	101
4.3.1	Data and Sampling Design	101
4.3.2	Assessment Design	104
4.4	Empirical Framework	105
4.4.1	Empirical Strategy	107
4.5	Descriptive Statistics	108
4.6	Results	112
4.6.1	Robustness Checks	116
4.6.2	Treatment Heterogeneity	118
4.6.2.1	Heterogeneity by EMI Grade of Transition	121
4.7	Discussion and Conclusion	124
4.8	References	127
5	Conclusion	132
	Appendix A Foundational Learning and Mental Health: Appendices	135
	Appendix B Mindfulness: Appendices	151
	Appendix C English Medium Instruction: Appendices	152

List of Figures

2.1	Map of Botswana	25
2.2	TaRL assessment and intervention calendar	28
2.3	Timeline of assessments	31
2.4	Parallel pre-trends	36
2.5	Kdensity plots on Total Difficulties (standardised)	41
2.6	Kdensity plots on Anxiety and Depression (standardised RCADS score)	42
3.1	Hypothesized Categories of Outcomes	66
4.1	Young Lives study sites in Ethiopia (Young Lives, 2018)	102
4.2	Kdensity plots on the mathematics outcome	113
4.3	Kdensity plots on the English outcome	114
A.1	Mental health related terms	135
A.2	Number of Primary schools by ownership and Region	136
A.3	Kdensity plots - Total Difficulties (SDQ), raw scores	136
A.4	Kdensity plots - Anxiety and Depression score (RCADS), raw scores	137
A.5	Impact of TaRL on Mental Health Outcomes, raw scores	137
A.6	Teacher reported student difficulties	138
A.7	Impact of TaRL on Educational success outcomes (raw)	138
A.8	Student tool	139
A.9	Teacher Tool on observed student behaviour	141
A.10	Sample of student exam	142
B.1	Combined index score utilising a means score of each outcome	151
C.1	The 1994 Language Policy in relation to the Constitution	152
C.2	Language of Instruction used in Primary Schooling and in Teacher Training	153
C.3	Distribution of EMI by Grade in Somali Region	154
C.4	Impact of EMI on Math Outcomes (winsorized 5% in each tail)	154
C.5	Impact of EMI on English Outcomes (winsorized 5% in each tail)	155
C.6	EMI Treatment Heterogeneity by the English Language Improvement Programme (ELIP)	155

C.7	Heterogeneity by EMI Grade of Transition	156
C.8	Teacher statistics on years of teaching experience	156
C.9	Dependent and independent variables used to in study sample	157

List of Tables

2.1	Outcome measures	30
2.2	Reliability of study instruments	33
2.3	Sample size by data collection round	35
2.4	Data description for all variables included for analysis	40
2.5	Baseline comparison between TaRL students (treated) and non-TaRL students (comparison)	41
2.6	Impact of TaRL on Mental Health Outcomes	43
2.7	Impact of TaRL on Education Success Outcomes	44
2.8	Heterogeneous effects on students with high baseline mental health outcomes	46
2.9	Heterogeneous Effects on Mental Health, by sex	47
3.1	MBSR Curriculum	68
3.2	Sample Size	69
3.3	Summary of variables and measurement methodology	73
3.4	Sample Characteristics at Baseline	74
3.5	Data Validity	76
3.6	Balance Checks	77
3.7	MBSR Treatment Effects on Mental Health Outcomes	78
3.8	MBSR Treatment Effects on Behaviour Change	79
3.9	MBSR Treatment Effects on Time-use and Impulse Control	80
3.10	MBSR Treatment Effects on Capacity to Learn	81
3.11	MBSR Treatment Effects on Academic Outcomes	81
3.12	Heterogenous Treatment Effects on Mental Health, by sex	82
3.13	Heterogenous Treatment Effects on Behaviour Change Outcomes, by sex	83
3.14	Heterogenous Treatment Effects on Capacity to Learn Outcomes, by sex	83
4.1	Distribution of schools and students by region	102
4.2	Distribution of schools using English medium instruction by Grade . . .	104
4.3	Baseline Comparison between EMI (Treatment) and Mother-Tongue Instruction (Control) Schools	110
4.4	Impact of English Medium Instruction (EMI) on Math Outcomes	115

4.5	Impact of English Medium Instruction (EMI) on English Outcomes . . .	116
4.6	Impact of English medium instruction on student outcomes (including Somali region)	117
4.7	Impact of English medium instruction on student outcomes (restricted sample)	118
4.8	Summary of EMI Treatment Heterogeneity by the English Language Improvement Programme	119
4.9	EMI Treatment Heterogeneity by Teachers' Years of Experience	120
4.10	EMI Treatment Heterogeneity by Sex	121
4.11	Heterogeneity by EMI Grade of Transition	122

1 | Introduction

Ninety percent of the global population of children and adolescents live in Low-income and Middle-income countries (Kieling et al., 2011). For young people, while there are several factors that may impact their social and economic development, health and education have been identified as two of the main key drivers for human capital formation and sustainable development (Victoria et al., 2022; World Bank, 2022). Health and education are intrinsically valuable, and are also fundamental to the sustainable development agenda in ensuring that no one is left behind (UNDP, 2018). Another valuable framework for development is Sen's Capability approach which centers human autonomy in choosing a life one values, by expanding one's capabilities and choices (Sen, 1992; Sen, 1999).

Mental health as defined by the World Health Organization (WHO, 2001a) is "a state of well-being in which the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to his or her community." In alignment with this comprehensive definition, the studies within this thesis conceptualise mental health as inclusive of all psychological and socio-emotional constructs, including facets such as anxiety, depression, and stress. Specifically addressing the mental health of children and adolescents, this pertains to the "capacity to achieve and maintain optimal psychological functioning and wellbeing. It is directly related to the level reached and the competence achieved in psychological and social functioning" (WHO, 2015). It is worth noting that the evidence base for child and adolescent mental health in low-income and middle-income countries remains limited. Research in this domain constitutes less than 7% of all published global mental health research, with studies focusing on child and adolescent mental health representing a notably small fraction of this body of research (Saxena et al., 2006; Patel et al., 2007).

Throughout the various phases of a child's life there are potential risks and opportunities that may significantly influence their mental and emotional well-being. In particular, during the early years, caregivers play a pivotal role in establishing a safe and secure home environment. Subsequently, during the school-going years, the

school environment, peer relationships emerge as crucial contexts that provide valuable opportunities for the development of essential life skills (Patel, 2018; Kieling, 2011; Yu et al., 2023). Furthermore, Childhood and adolescence present vulnerable periods during which adverse experiences can detrimentally affect one’s cognitive, emotional, and behavioural development. In low-resource settings, the co-occurrence of childhood adversities such as violence, neglect, abuse, parental stress are constantly associated with poor mental health outcomes (Jokinen, 2021; Massarwi et al., 2022; Yu et al., 2023).

As such, an overarching ethical and methodological principle guiding this thesis was to undertake research on existing policies and interventions affecting the lives of young people. Among the various evidence-based mental health and psychosocial support interventions implemented in low-income and middle-income countries over the past decade, delivery is categorised into individual and family-based interventions, school-based interventions, community-based interventions, and digital interventions. These interventions are further classified as preventative, promotive, or treatment-focused (Yu et al., 2023). The mental health-focused studies within this thesis primarily center around school-based initiatives, aligning them under the promotion umbrella as they aim to promote mental health, peer support and education.

Using a combination of experimental and quasi-experimental evaluation techniques, this thesis carefully explores the mental health and education factors directly affecting young people’s capabilities. Thus, the aim of this thesis is to explore the range of health and education reforms and interventions addressing the mental health and educational success of young people in Botswana, Ethiopia and Pakistan. I was motivated to investigate the wider implications of ongoing interventions in three Global South countries, employing a combination of primary and secondary data sources.

In Botswana, after gaining approvals from the Ministry of Basic Education, parental consent, and student assent, I led the data collection of four rounds of data from 1,297 students between January and July 2022. Additionally, teachers responded to open-ended survey questions on their experiences and observations with common themes identified from this data presented in the Botswana study. Similarly in Pakistan, with

approvals from implementation schools, parental consent, and student assent, I co-led the collection of two rounds of data from 214 students between October and November 2022. In Ethiopia, there was no primary data collection, however a secondary data source was leveraged to examine a popular education reform policy.

Chapter 2 explores the impact of an on-going government-led foundational learning programme on the mental health and educational success outcomes of primary school learners in Botswana. A Difference in Difference (DD) method is used to causally identify the impact of a targeted instruction pedagogical intervention -Teaching at the Right Level (TaRL)- on mental health and learning. The main results show that TaRL learners experience a 6% percent decrease in total emotional and behavioural difficulties (-0.15 standard deviations), compared to non-TaRL learners. Additionally these TaRL learners are also exhibit an 8% increase (0.24 standard deviations) in prosocial behaviour in comparison to their non-TaRL counterparts.

Chapter 3 studies the impact of an adapted mindfulness-based intervention on the mental health and educational success outcomes of secondary school learners in Pakistan. As there is limited evidence on the impact of mindfulness-based programmes on student mental health, well-being, behaviour change and educational success, this research study collaborated with the first ever Urdu-based mindfulness platform in Pakistan -Be-Khudi- to study the impact of a Mindfulness-Based Stress Reduction (MBSR) programme on young people. The MBSR programme was adapted through an 8-month iteration and piloting process, and evaluated through a randomised control trial in a low-cost school setting in Pakistan. The main results show that MBSR significantly improved student prosociality by 0.30 standard deviations in comparison to the control group. Second, MBSR students report an improvement in mood (by 0.52 standard deviations) and less stress at home (by 0.34 standard deviations). Additionally, the study showed that MBSR increased the grit of students by 0.34 standard deviations, relative to their counterparts.

Finally, chapter 4 examines the effects of a language of instruction policy reform on the educational success outcomes of upper primary school learners in Ethiopia.

This article makes use of a rich source of individual level secondary data collected through the Young Lives longitudinal study to measure the impact of English Medium Instruction (EMI) on the mathematics and English test scores of learners. The main results show that students under the EMI policy score 0.24 standard deviations below mother tongue education students in Mathematics. Furthermore, these EMI students do not outperform their mother-tongue education counterparts on English test scores. However, treatment heterogeneity suggestively indicates that an early grade EMI transition has positive effects on long-term English test outcomes with no negative effects on mathematics test outcomes; all relative to sustained mother tongue instruction.

Collectively, all three articles provide important insights for policymakers and practitioners whilst producing valuable empirical research contributing broadly to the education and global health literatures. To the best of my knowledge, the first article, produces the first empirical study linking the targeted instruction pedagogy, TaRL, to mental health outcomes for young people. Additionally, this research study focuses on a Global South context where mental health data and resources are either scarce or insufficient.

The second article on MBSR demonstrates that it is possible to adapt mindfulness-based programming to a low-resource setting. This article contributes specifically to the existing literature on mindfulness, which typically focuses on high-income contexts or adult populations. This research provides causal evidence from a lower income, lower resource context. Furthermore, similar to the first study, these results highlight the role of school-based interventions in addressing student mental health and education outcomes.

The third article complements a small quantitative literature identifying the causal effects of changes in language of instruction policies in Africa. Specifically, this study is show the effects of English language policies on upper primary grades (7 and 8) whereas the bulk of these studies typically focus on immediate learning outcomes in early primary grades (1-4) or on adult literacy level and years of schooling. Additionally, this study is able to add some nuance to the on-going language of instruction debates on the "right

grade" to transition.

References

- Jokinen, T., Alexander, E. C., Manikam, L., Huq, T., Patil, P., Benjumea, D., Das, I., & Davidson, L. L. (2021). A systematic review of household and family alcohol use and adolescent behavioural outcomes in low and middle income countries. *Child Psychiatry & Human Development*, 52(4), 554-570.
- Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., Rahman, A. (2011). Child and adolescent mental health worldwide: Evidence for action. *The Lancet*, 378(9801), 1515-1525.
- Massarwi, A., Cluver, L., Meinck, F., Doubt, J., & Green, O. (2022). Pathways to parenting stress reduction among parents in South Africa. *Child & Family Social Work*. 10.1111/cfs.12952
- Patel, V., Flisher, A. J., Nikapota, A., & Malhotra, S. (2008). Promoting child and adolescent mental health in low and middle income countries. In *Journal of Child Psychology and Psychiatry and Allied Disciplines* (Vol. 49, Issue 3, pp. 313-334). <https://doi.org/10.1111/j.1469-7610.2007.01824>.
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P. Y., Cooper, J. L., Eaton, J., Herrman, H., Herzallah, M. H., Huang, Y., Jordans, M. J. D., Kleinman, A., Medinaâ Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., UnÃ¼tzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598
- Saxena, S., Paraje, G., Sharan, P., Karam, G., & Sadana, R. (2006). The 10/90 divide in mental health research: Trends over a 10-year period. *British Journal of Psychiatry*, 188, 81â82.
- Sen A (1992). *Inequality Reexamined*. Oxford: Oxford University Press.
- Sen A (1999). *Development as Freedom*. Oxford, UK: Oxford University Press.
- United Nations Development Program (2018). What does it mean to leave no one behind? A UNDP discussion paper and framework for implementation.
- Victora, C.G., Hartwig, F. P., Vidaletti, L.P., Martorell, R., Osmond, C., Richter, L.M., Stein, A.D., Barros, A.J.D., Adair, L.S., Barros, F.C., Bhargava, S.K., Horta, B.L.,

Kroker-Lobos,M.F., Lee,N.R., Menezes,A.M. B., Murray,J., Norris,S.A., Sachdev,H.S., Stein,A., Black,R.E. (2022). Effects of early-life poverty on health and human capital in children and adolescents:analyses of national surveys and birth cohort studies in LMICs. *The Lancet* (Vol. 399, Issue 10336, pp. 1741-1752). Elsevier B.V. [https://doi.org/10.1016/S0140-6736\(21\)02716-1](https://doi.org/10.1016/S0140-6736(21)02716-1)

World Bank (2022).Harnessing Catalytic Initiatives to Accelerate Investments in Human Capital. Annual report, 2022.

Yu, R., Perera, C., Sharma, M., Ipince, A., Bakrania, S., ShokraneH, F., Sepulveda, J. S. M., & Anthony, D. (2023). Child and adolescent mental health and psychosocial support interventions: An evidence and gap map of lowâ and middleâincome countries. *Campbell Systematic Reviews*, 19, e1349. <https://doi.org/10.1002/cl2.1349>

Abstract

A significant proportion of mental health problems surface in early childhood and adolescent years, with early onset mental health problems having the potential to affect the long-term development of young people. Currently, we know very little about whether learner-centred, positive school climates drive better mental health outcomes for young people. As such, this research paper leverages an ongoing pedagogical intervention - Teaching at the Right Level (TaRL)- to understand the impact of foundational learning on student mental health and educational success outcomes in Botswana's primary schools. TaRL is an evidence-backed pedagogical intervention designed to improve foundational numeracy and literacy capabilities through targeted, fun, and engaging activities. The underlying hypothesis is that improving the learning environment results in improved student mental health. Using a Difference in Difference (DD) research design, this study of 1,297 students finds that TaRL students have significantly fewer behavioural and emotional total difficulties (0.15 standard deviations) and exhibit 0.24 standard deviations more prosocial behaviour, all relative to similar students yet to be exposed to the pedagogy. This study bridges an important research gap at the intersection of education and public health literature and underscores the importance of context-relevant policy reforms that prioritise pedagogic approaches fostering holistic learning and development.

2.1 Introduction

Approximately 20% of all children and adolescents have mental health challenges (WHO, 2001; Kieling et al., 2011; O'Reilly et al., 2018). In sub-Saharan Africa (SSA), a systematic review across six SSA countries (Ethiopia, Nigeria, Kenya, South Africa, Uganda, and the Democratic Republic of Congo) showed that 1 in 7 children and adolescents up to 16 years of age have significant psychological difficulties. These included emotional problems, depression, anxiety disorders, conduct, disruptive and reactive behaviour disorders, and post-traumatic stress disorder (Cortina et al., 2012). Additionally, we know that early onset mental health problems have the potential to affect the long-term development and wellbeing (Kessler et al., 2007; Gleason et al., 2016) and with 90% of the global population of children and adolescents living in low-income and middle-income countries (LMICs), this presents a considerable global health challenge, particularly in lower resource settings (Kieling et al., 2011).

Furthermore, young people spend considerable amounts of time in schools. Schools are the ideal settings for universally rolling out early preventative interventions, and identifying at risk youth (Zenner et al., 2014). Schools can provide an enabling learning environment, and as such, can be leveraged to support the mental health of children (Becker et al., 2014; Garcia-Escalera et al., 2020). Studies mostly conducted in high-income, Global North countries have found positive associations between school climate ¹ and mental health (Aldridge & McChesney, 2018). Where the psychosocial environment is poor, young people struggle. In LMICs in particular, where high levels of bullying, poor disciplinary strategies, problems with student behaviour, and lower test performance are often prevalent, evidence shows that a positive school climate is associated with positive socio-emotional, behavioural, and academic outcomes (Larson et al., 2020). Positive relationships between teachers and student have been linked to a decrease in mental health issues with mediating factors including, teacher support, a sense of belonging and school connectedness and students' experiences of bullying

¹School climate in this study is conceptualised as the norms, expectations and beliefs that create the psychosocial environment determining the level at which people feel physically, emotionally, and socially safe. The quality of teaching and learning also contribute to this environment (Cohen et al., 2009; Aldridge & McChesney, 2018). As such, school climate is used synonymously with school environment.

behaviour (Aldridge & McChesney, 2018). There is very little data and research from Global South countries.

This research study investigates the causal impact of a foundational learning intervention -Teaching at the Right Level- on the mental health of students. In light of the extensive research demonstrating the association between positive school climates with positive socio-emotional, behavioural, and academic outcomes, I investigate the impact of Teaching at the Right level -a pedagogical approach to teaching and learning designed to cultivate an enabling school climate for learning. However, existing empirical studies have primarily focused on TaRL's immediate effects on educational outcomes. This research endeavours to bridge this gap by examining the causal impact of the foundational learning intervention -Teaching at the Right Level- on the mental health outcomes of students. To my knowledge this is the first empirical study to examine holistic outcomes of targeted instruction interventions.

I collaborate with the Botswana Ministry of Basic Education to collect a unique data set of mental health measures during the roll-out of the TaRL programme in Botswana schools. The mental health measures are adapted and psychometrically validated (assessing internal reliability) for the Botswana context. A Difference in Difference (DD) specification is used to causally identify the impact of the foundational learning intervention TaRL on learning and mental health. Students receiving the TaRL programme are in the treated group and students in the same grades (or standards) and schools wait-listed to receive the programme are in the comparison group. Students in the comparison group are from the same schools as treated schools, indicating that both groups are exposed to the same school quality variables and similar socioeconomic variables.

The main results are as follows. First, students engaged in the TaRL programme experience a 6% decrease in total emotional and behavioural difficulties (0.15 standard deviations) when compared with non-TaRL students. Second, TaRL students exhibit an 8% increase in prosocial behaviour (0.24 standard deviations) in comparison to non-TaRL students. Third, greater effects are experienced by students predisposed to

poorer mental health, with larger effects for those with higher baseline levels of total difficulties and anxiety and depressive symptoms. These results are in line with the literature that shows how student-centered pedagogies that are activity based, include peer learning, and have no corporal punishment have also been shown to improve student psychological outcomes (Erwin et al., 2012; Hanson et al., 2016; Larson et al., 2020). Although this study was not explicitly designed to test for specific mechanisms in the TaRL programme, the ease with which students are able to openly ask questions, the inclusion of structured play activities such as energizers or icebreakers, the use of small group activities to reinforce concepts, the enthusiastic vocal support of instructors, and the behavioural management techniques that are utilised without corporal punishment, may all play a role in improving the learning environment for students.

Finally, there are no significant effects found on the educational success outcomes (test scores and attendance). However, as there is a robust literature linking TaRL to improved foundational skills (Banerjee et al., 2007; Duflo et al., 2011; Banerjee, Banerji, Duflo et al., 2016), these null effects may be explained through implementation heterogeneity, specifically through the delivery model (this study examined a government-led delivery model).

This paper contributes to the literature in education and economics of education on educational interventions by providing one of the first empirical studies linking a targeted instruction pedagogy to the mental health outcomes of young people. This research study will contribute to the large body of work connecting improved educational outcomes and mental health (Valdez et al., 2011; Suldo et al., 2014; Becker et al., 2014). However, most of the causal studies are typically focused on high income countries, and others are correlational or largely qualitative (Valdez et al., 2011; Aldridge & McChesney, 2018; McAllister et al., 2018). This research focuses on a middle-income country in the Global South where less evidence has been generated (Larson et al., 2020). Additionally, in Botswana, the context for this study, there is insufficient mental health data to inform policy-making (Opondo et al., 2020). As such, this study will fill an important research gap in LMICs adding to the evidence base on mental health and education.

The rest of the paper is organised as follows. Section 2.2 provides a background for the study context, and research questions. Section 2.3 provides details on the measurement outcomes and psychometric validation of the research instruments included in this study. Section 2.4 outlines the research design and econometric specification used. Section 2.5 describes and shares descriptive statistics for the study. Section 2.6 presents the results, section 2.7 provides a discussion, and section 2.8 concludes.

2.2 Background

2.2.1 Mental Health

Mental health² is fundamental to good health and wellbeing (WHO, 2014). Sen's capability approach spotlighted the importance of the freedom of people to be or to do the things they have reason to value (Sen, 1992). Having good mental health can allow one to participate in the things that they value (i.e., education) and do things one values (i.e., learn). This allows one to be in good health. Renowned philosopher and educational theorist Dewey wrote about the opportunity educators have in utilising their physical and social environments in supporting the "whole child" (Dewey, 1938). This learner-centered pedagogical approach which is often marked by experiential activities, participation and peer learning has also been shown to improve student psychological outcomes (Erwin et al., 2012; Hanson et al., 2016; Larson et al., 2020).

Furthermore, externalising mental health problems such as disruptive and behavioural difficulties affect academic outcomes, and academic outcomes affect mental health, particularly internalising problems such as anxiety and depression (Suldo et al., 2014). Poor academic performance has been shown to predict depression and overall poorer mental health (Hilsman & Garber 1995; Chen & Kaplan, 2003). It is against this backdrop that this research study investigates impact of a targeted instruction pedagogical intervention on the mental health and educational success outcomes of young people in Botswana. Teaching at the Right Level (TaRL) is a basic numeracy and literacy intervention, that disrupts the traditional rote learning, teacher-centered pedagogy for a student-centered, targeted activity and play-based pedagogy. This research explores the extent to which TaRL impacts the emotional and behavioural mental health of children exposed to the learner-centered pedagogy. The underlying hypothesis is that positively influencing the school climate or learning environment that children are in will also improve the extent to which children feel emotionally and socially safe, subsequently reflecting in their improved mental health.

²"a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to his or her community" (WHO, 2001a)

2.2.2 Botswana context

Botswana is a landlocked democratic country located in Southern Africa with a population of approximately 2.3 million (Statistics Botswana, 2022). Over the years, the proportion of the population living on less than \$1.90 has steadily decreased from 29.8% to 18.2% (World Bank, 2020). Botswana's GDP per capita rose from \$80 in 1966 (at independence) to \$6 924 in 2016 with services as the fastest growing sector contributing to the national GDP (Honde, 2018). Regarding education spending, the government allocates approximately 8.7% of the national budget on public education, which is double the global average government expenditure on education (UIS, 2022). However, despite this investment in education, the Human Capital Index (HCI) reveals that a child born in Botswana has only 5.1 years of learning-adjusted years of schooling and is expected to have 8.1 years of schooling by their 18th birthday, assuming a child enters school at age 4 (World Bank, 2020).

Basic education (the first ten years of school) is available to all children in the country, and students enter primary education at age 6 (Government Paper No.2, 1994). Parents are charged a co-payment for education, but families from low-income groups receive free education and school meals are provided to all learners daily (ETSSP 2015-2020). The net enrolment ratio for primary school learners aged 6-12 years was at 93% in 2012, an increase from 86% in 2009 (ETSSP 2015-2020). Furthermore, at the primary education level, there are two standardised national assessments; the first is the Standard 4 attainment test, and the second is the Primary School Leaving Examination (PSLE) taken at Standard 7 (SACMEQ, 2017).

The Government of Botswana introduced an Education and Training Sector Strategic Plan (ETSSP) from 2015-2020. The goal of this reform strategy was to "improve access, quality, inclusion and equity, accountability and governance in the education system" (ETSSP 2015-2020). This reform plan includes goals for improving foundational numeracy and literacy at the basic education level. In addition, the ETSSP also references making schools "safe environments to inspire students to learn and teachers to teach." (ETSSP 2015-2020). In line with these reform goals, a qualitative study on

improving academic performance in Botswana found that cultivating a positive learning climate and actively listening to the concerns of learners, teachers and stakeholders improved student academic performance (Bakokonyane, 2022).

In the health sector, HIV/AIDS remains the leading cause of morbidity and mortality in Sub-Saharan Africa (James, Abate, et al., 2017). The latest Botswana AIDS Impact Survey (BAIS V) showed Botswana has one of the highest national HIV prevalence rates, with 20.8% of adults aged 15-64 years living with HIV (Statistics Botswana: BAIS V, 2022). In terms of mental health policies, the first mental health policy was developed in 2003 to align mental health care with the general healthcare system (Opondo et al., 2020). It is therefore unsurprising that most mental health research studies in Botswana find links between people living with HIV (PLWHIV) and mental health disorders such as depression and anxiety (Opondo et al., 2020). However, studies have found that there is insufficient mental health data in Botswana to inform policy effectively (Opondo et al., 2020).

Figure 2.1: Map of Botswana



Botswana is divided into 9 districts with 10 education regions: Central, North-West/, Ngamiland, Ghanzi, Kgalagadi, Kweneng, Southern, Kgatleng, North-East, South-East

and Chobe (see figure 3). The education system is organized into four levels: (1) pre-primary, (2) primary education, (3) junior and senior secondary education and (4) tertiary education.

This research study focuses on the Teaching at the Right Level (TaRL) intervention in primary schools in Francistown, located in the North-East education region of Botswana.

2.2.3 Teaching at the Right Level - a foundational learning intervention

Teaching at the Right Level (TaRL³), is a targeted foundational numeracy and literacy intervention, backed by multiple evaluations and proven to improve learning outcomes across multiple contexts and delivery models⁴ (Banerjee et al., 2007; Duflo et al., 2011; Banerjee, Banerji, Duflo et al., 2016). TaRL is a basic numeracy and literacy remedial education intervention designed to ensure all students are fully numerate and literate. The intervention was developed by Pratham Education Foundation approximately 15 years ago in India as a community-based initiative to improve children's learning.

In TaRL classes, instruction is organised by children's level of knowledge rather than by their assigned grade level. As such, children are assessed and grouped by their learning level and receive intensive bursts of engaging, targeted instruction over a period of time. In each session activities are conducted in large groups, small or peer groups and individually (Banerji & Chavan, 2016; GEM: Pratham, 2020) to reinforce learning concepts. TaRL is learner-centered, and the aspects of the pedagogy that may promote student mental health include the removal of the threat of corporal punishment, students eagerly asking questions, structured play activities such as icebreakers or energisers, using learning manipulatives from their surroundings (i.e., sticks and stones) and the positive encouragement of instructors.

TaRL has now been tested through multiple randomised control trials by the Jameel Poverty Action Lab (J-PAL) and affiliated researchers. In India, the programme was

³Within Pratham, the approach is also called CAMaL (Combined Activities for Maximized Learning); which in Hindi also means "magic" or "wonder" (Banerji & Chavan, 2016)

⁴A low-tech variation of the programme was also found to reduce learning loss and encourage parental engagement during the height of the pandemic (Angrist, Bergman, & Matsheng, 2022).

found to improve student learning outcomes by 0.28 standard deviations (Banerjee, Banerji, Berry, Duflo, et al. 2016). Evaluations from Kenya and Ghana also found that grouping children by their learning levels, helped in supporting instructors to target instruction and for children to learn (Duflo, Dupas & Kremer, 2011; IPA, 2018). In addition, the TaRL literature has also found that learning gains persisted even when multiple delivery models are used (Banerjee et al., 2007; Duflo et al., 2011; IPA 2018).

In Botswana, a situational analysis of basic literacy and numeracy report found that only 1 in 10 standard 5 students could do division and only 20% of students could read a paragraph (Pansiri & Tsayang, 2017). In response, the Ministry of Basic Education, supported by the evidence driven NGO - Youth Impact, have committed to adapting and scaling TaRL in all primary schools in Botswana.

The theory of change behind TaRL is that through the (1) diagnostic testing for the assignment of learning groups, (2) targeting instruction tailored to each learning capability group and (3) teaching in a fun and engaging manner, learning outcomes improve. Noteworthy features in a TaRL classroom include energisers, sessions conducted on the floor (especially in areas where students typically sit at desks) and the absence of corporal punishment. Regarding potential implications for holistic outcomes such as mental health outcomes beyond the numeracy and literacy outcomes typically tested, both the education and mental health literature indicate that student-centered pedagogies characterised by activity-based methods, peer learning, and the absence of corporal punishment have demonstrated improvements in student psychological well-being (Erwin et al., 2012; Hanson et al., 2016; Larson et al., 2020). This provides the impetus for examining the mental health outcomes of TaRL in Botswana.

In Botswana the intervention is typically delivered in approximately 30-day bursts, over a 9-week period, with three formative assessments points. The baseline assessment is used to form the initial teaching groups for targeted instruction. After the baseline assessment, students are taught in their learning groups for approximately 1 hour each day. At mid-point of the programme, students are re-assessed and re-grouped as needed. Targeted instruction continues till the endline line assessment, typically in week 9 of

the programme. TaRL numeracy assessments cover number recognition, basic addition, subtraction, multiplication, and long division. TaRL literacy assessments cover letter recognition, reading a word, reading a sentence, and reading a paragraph or short story.

Figure 2.2: TaRL assessment and intervention calendar

Baseline TaRL Assessment Week 1
Week 2
Week 3
Week 4
Midline TaRL Assessment Week 5
Week 6
Week 7
Week 8
Endline TaRL Assessment Week 9

Implementation was led by Tirelo Sechaba Participants - youth volunteers placed in schools by the Ministry of Youth, Gender, Sports, and Culture, to support teachers. The youth volunteers were trained and mentored by Youth Impact⁵. Youth volunteers, are one of many delivery models proven to be effective in delivering the intervention.

The focus of this study is on mental health and wellbeing outcomes. Education related outcomes are included as the foundational learning intervention evaluated directly targets educational success outcomes and is implemented in a school context.

In view of this study context, this research study asks the following research questions.

1. What is the impact of foundational learning (TaRL) on the overall mental health of primary school students?
2. What is the impact of foundational learning (TaRL) on the anxiety and depressive symptoms of primary school?
3. What is the impact of foundational learning (TaRL) on students predisposed to mental health difficulties (including behavioural and emotional difficulties, anxiety, and depressive symptoms)?

⁵The evidence-based NGO supporting the Ministry of Basic Education to scale TaRL in Botswana. <https://www.youth-impact.org/>

-
4. What is the impact of foundational learning (TaRL) on the educational success (i.e., test scores and attendance record) of students?

2.3 Outcome measurement and Psychometric validation

2.3.1 Measurement

Evidence-based psychological assessments can be conducted for several reasons including diagnosis, screening (identifying those at risk), treatment monitoring, and or evaluation (Hunsley & Mash, 2008; Beidas et al., 2015). This research study selected research instruments that had been validated for screening and treatment evaluation purposes in the literature and were suitable for children under the age of 18 years.

To capture overall mental health and wellbeing, the Strengths and Difficulties Questionnaire (Goodman, 1997) is included. The Strengths and Difficulties Questionnaire (SDQ) captures behavioural and emotional difficulties, covering the following 5 subscales: emotional symptoms, conduct problems, peer relationship problems, hyperactivity/inattention, and prosocial behaviour. It is a self-report instrument, with 25 questions, using likert scale responses 0, 1, 2 for not true, somewhat true, and certainly true (Wolpert et al., 2015). The 'total difficulties' score is generated by summing up the responses for 4 of the subscales (excluding the prosocial behaviour subscale). The prosocial sub-scale is included separately. Higher scores indicate higher levels of total difficulties whereas a higher prosociality score indicates higher levels of prosocial behaviour.

The Revised Child Anxiety and Depression Scale (Chorpita et al., 2000) is included in this study to screen for anxiety and depressive symptoms through 6 subscales: separation anxiety disorder, social phobia, generalised anxiety disorder, panic disorder, obsessive compulsive disorder, and major depressive disorder. The Revised Child Anxiety and Depression Scale (RCADS) is a self-report instrument, with 25 questions, using likert scale responses 0, 1, 2, 3 for never, sometimes, often, and always (Wolpert

et al., 2015). To calculate the outcome, raw scores are summed to generate an overall depression score and an overall anxiety score. This raw score is then converted into a standardised "Test score" appropriate for the gender and grade level of the child (Chiporta, 2015). A higher score indicates greater levels of anxiety and depression.

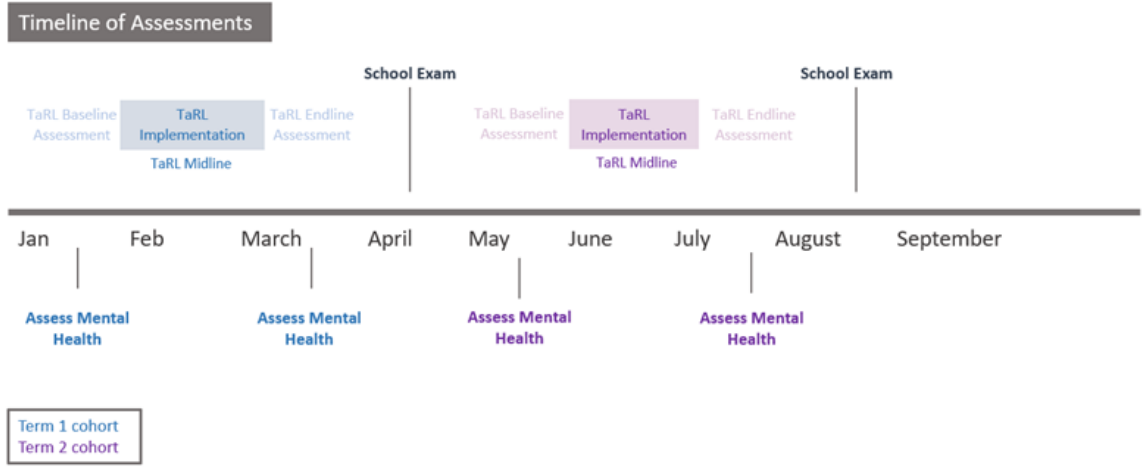
The use of the SDQ and RCADS tools in this research study capture general symptoms in the study sample while when utilised collaboratively with clinicians, the same tools can also be used as diagnostic tools (Cortina et al., 2012). All students in the TaRL programme and the waitlisted comparison group students were screened using the mental health and wellbeing instruments (SDQ & RCADS). In addition, class teachers report their student observations using a shortened SDQ instrument for each student in the study sample.

Table 2.1: Outcome measures

Instrument	Items	Constructs/Details	Timing	Target
Strengths and Difficulties Questionnaire (SDQ)	25 items	Total difficulties (indication of overall mental health and wellbeing)	Baseline & Endline	TaRL & Non-TaRL students
Revised Child Anxiety and Depression Scale (RCADS)	25 items	Anxiety and Depression	Baseline & Endline	TaRL & Non-TaRL students
Strengths and Difficulties Questionnaire (SDQ)- <i>selected questions</i>	3 items	Teacher reported on student emotional and behavioural difficulty	Endline only	TaRL & Non-TaRL students
Administrative Data		School exam results	Term 1 & 2	TaRL & Non-TaRL students

School exam data were selected as the main learning outcome measure as this data are routinely used to inform school level decisions (i.e., student progression) and wider regional and national education policy decisions. This goes beyond other TaRL-related studies that typically only use ASER-based TaRL assessment data. As the TaRL curriculum is very closely aligned to the TaRL assessments, the TaRL assessment data is narrow in scope. As such, the TaRL assessment data are not included in this study. Nonetheless, a limitation of this research study is the exclusion of ASER based TaRL assessment data and future studies should include both school-based test scores as well as TaRL assessment data.

Figure 2.3: Timeline of assessments



At the start of each academic term, a baseline assessment was conducted for all of the mental health outcomes of interest. At the end of each academic term, an endline assessment was conducted for all the mental health outcomes of interest.

2.3.2 Psychometric validation

The use of the Strengths and Difficulties Questionnaire (SDQ) in community and clinical samples has shown that the SDQ performs well on various psychometric properties including internal validity, predictive validity, and test-retest reliability (Goodman, 2001; Muris et al., 2003; Wolpert et al., 2015). Similarly, the Revised Child Anxiety and Depression Scale (RCADS) also performs well on the various validity and reliability tests (Chiporta et al., 2000; Wolpert et al., 2015). The SDQ have also been documented as an evidence-based assessment suitable for screening and evaluation purposes (Beidas et al., 2015).

This research study uses pre-existing psychometrically validated English versions of the SDQ and RCADS. These were then back-translated using principles on cross-cultural back-translation (Brislin, 1970). Two bilingual translators familiar with the TaRL programme and study context individually translated the tools from English into the local vernacular, Setswana. Once the two translators had agreed on the translation, a third translator back-translated the tool into English to ensure consistency in the

translation. All three translators agreed on the final translation.

The translated tools were then assessed for their validity and reliability in the Botswana context. The study specifically tests for the face validity and internal reliability of the measures. Face validity is assessed through piloting the questions with a small sample, ensuring that respondents understood the questions (i.e., comprehension) as anticipated and that it was clear that the questions were designed to gain an understanding of their mental health and well-being. Through multiple pilots, all words and phrases pointed out by respondents as confusing were updated for clarity.

To assess internal reliability, Cronbach's alpha (α) and McDonald's Omega are used. Cronbach's alpha provides an assessment of internal consistency (Cronbach, 1951). The Cronbach alpha statistic is a ratio of variances between 0 and 1. Following the guidance set out by Hunsley & Mash (2005), α values between 0.70-0.79 indicates adequate internal consistency; α values between 0.80-0.89 indicates good internal consistency, and $\alpha > 0.90$ may indicate excellent internal consistency. It should be noted that while $\alpha > 0.90$ may indicate excellent internal validity, it is possible for alpha values to be artificially high which could also be an indication of item redundancy (Streiner, 2003; Hunsley & Mash, 2005). The rule of thumb indicates that a desirable level internal consistency is represented by an alpha value above 0.7 (Nunally and Bernstein, 1978). McDonald's Omega statistic (Ω) is another measure of composite reliability. In some literatures, the Omega coefficient has been shown to be a more precise estimate of reliability as it accounts for measurement error (Dunn et al., 2013; Ravinder & Saraswathi, 2020).

Table 2 shows the assessment of the SDQ and RCADS internal reliability including both alpha and omega coefficients for comparison and discussion. Following the Hunsley & Mash guideline (2005) of adequate, good, and excellent levels of internal consistency, the translated SDQ reliably measures the same targeted constructs in the Botswana context with adequate internal consistency ($\alpha > 0.7$; $\Omega > 0.7$). The translated RCADS performs slightly better with good internal consistency ($\alpha > 0.8$; $\Omega > 0.8$).

Table 2.2: Reliability of study instruments

Instrument	Alpha baseline	Alpha endline	Omega baseline	Omega endline	N (items)
SDQ	0.79	0.76	0.75	0.72	25
RCADS	0.88	0.83	0.89	0.84	25

2.3.3 Ethics

The Botswana Ministry of Basic Education issued a research permit for this study. In addition, the North-East Regional Director assigned a Principal Education Officer (PEO) to the research study who actively participated in multiple training sessions throughout the data collection period. The PEO also provided feedback on the research design and assisted with communications shared with all schools and caregivers. Furthermore, signed parental consent and informed student assent were obtained for all students in the study sample. Parents and students were assured of the confidentiality of their personally identifiable information. The data collection facilitators were also encouraged to build rapport with the students allowing for an open and relaxed atmosphere. A detailed ethics discussion is included in Appendix K.

This research study was approved by the institutional research ethics committee at the University of Oxford (UK, Ref No. SSH/BSG-C1A-22-01, February 7th, 2022).

2.4 Research Design and Econometric Specification

2.4.1 Treatment Assignment

Treatment assignment was not randomised. All schools in the study sample identified TaRL as necessary remedial support for learners to catch up on foundational numeracy and literacy. As such, TaRL was being rolled out in phases with some classes enrolled in the programme, while other classes were waitlisted to receive the programme over the

course of the academic year. School leaders (Principals and Heads of Departments) in each school selected which classes would receive the TaRL intervention each term. This decision was determined based on need, and willingness of class teachers to enroll their students into the programme.

However, treatment and control classes were extremely similar at baseline. In Botswana, students are not tracked into high performing and low performing classes in each grade ⁶. At baseline, test scores show that the students in the treated group and students in the comparison group are similar (see balance Table 5). This means that students are not systematically assigned to classes based on academic grades and or other performance metrics further supporting the study's assumptions that students in the TaRL group and students in the comparison group (non-TaRL) were similar.

The TaRL intervention was implemented at the class level. The students in classes receiving the TaRL programme were in the "treated group" and the classes yet to receive the TaRL programme were in a waitlist "comparison group". Students in the comparison group were engaged in a "business-as-usual" study hour where the students sat quietly at their desks completing homework activities and or copying notes off the blackboard. This is a strength of the design because it means the treatment and control children spent similar time in a classroom environment at school, so children having different exposure to school or having different amounts of time available outside formal class time cannot account for differences between groups.

The comparison group was comprised of students who had not been exposed to TaRL in the data collection rounds. This means that students treated in term 1 do not end up in the term 2 comparison group.

The final study sample had 1297 students with 703 students in treated TaRL classes and 594 students in the comparison, non-TaRL classes. As shown in figure 2.3 above, there is a baseline measurement for each cohort prior to the TaRL programme and

⁶According to the education Policy in Botswana, students are randomly assigned to classes in standard/grade 1. Students who remain in the same schools, progress in the same classes through school. This means a student assigned to standard 1B, moves to standard 2B, 3C etc. This information was confirmed through an interview with North East Regional Chief Education Officer.

an endline measurement post implementation. These two cohorts are pooled together so that there is one sample with a baseline and endline.

Table 2.3: Sample size by data collection round

Round	Treated Group	Comparison Group	Total
Term 1 cohort	244	530	774
Term 2 cohort	459	64	523
Total Students	703	594	1,297

Note: There are 3 academic terms, with this data collected from 2 terms. Students in the comparison group in the term 2 cohort were scheduled to receive the programme in term 3.

2.4.2 Research Design

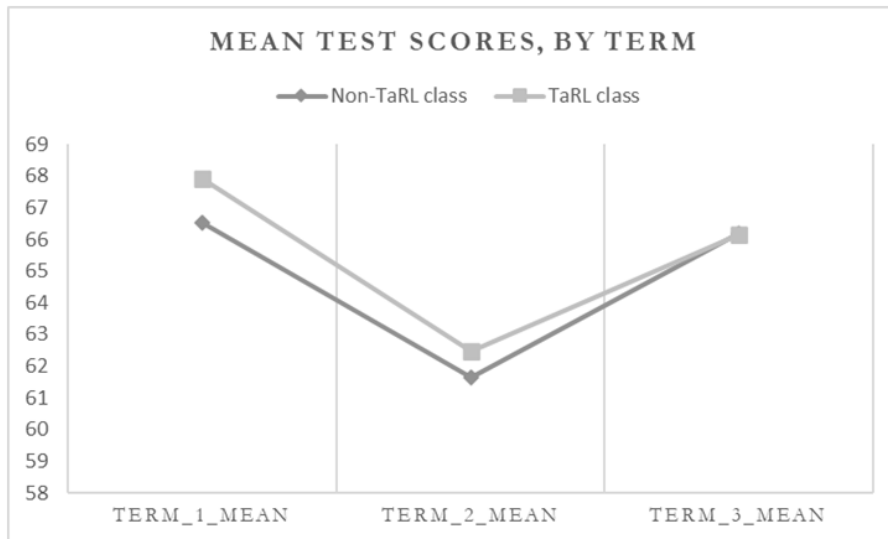
A Difference in Difference (DD) design is used to estimate the effects of TaRL on mental health and wellbeing outcomes, and educational success outcomes. The standard DD framework requires two groups (treatment and control), at least two time periods (before and after) and a change in treatment status such that the treatment group receives a treatment in the second period, while the control group does not receive the treatment in either of the periods. This design can account for biases introduced by time-invariant differences between the groups (Ashenfelter & Card, 1985; Duflo 2001; Angrist & Pischke, 2005).

In this study, the before time period is captured through baseline data collection prior to the TaRL intervention and the after-time period is captured through endline data collection, after the intervention was completed. TaRL receiving classes and the waitlisted comparison classes are balanced on key baseline characteristics (see table 2.4). Students wait-listed to receive the programme and thus in the comparison group, were from the same schools as students in treated classes. This would suggest that both groups are exposed to the same school quality variables and similar socioeconomic variables.

Furthermore, to infer causality, one of the key underlying assumptions of the DD

design is the parallel trend assumption. This identifying assumption states that in the absence of treatment, the outcomes of interest would be similar or moving in a 'parallel' fashion and the introduction of the treatment intervention (i.e., TaRL) is what creates the difference observed between groups (Angrist & Pischke, 2005). Using available school test scores data, the parallel trends assumption is tested by looking at 3 data points prior to the implementation of TaRL in schools. Figure 2.4 shows that students in TaRL classes and students in the comparison group were performing similarly on test scores in the academic year (2021) prior to implementation. There are no statistical differences between the groups.

Figure 2.4: Parallel pre-trends



Overall, I do not find any statistical differences between TaRL students and non-TaRL students in term 1, 2, and 3. In term 1, the non-TaRL students ($M=66.51$, $SD=1.23$) compared to the TaRL students ($M=67.93$, $SD=1.14$) were not statistically different, $t(552)=-0.84$, $p>.05$. In term 2, the non-TaRL students ($M=61.66$, $SD=1.29$) compared to the TaRL students ($M=62.48$, $SD=1.04$) were not statistically different, $t(568)=-0.50$, $p>.05$. In term 3, the non-TaRL students ($M=66.17$, $SD=1.10$) compared to the TaRL students ($M=66.13$, $SD=0.96$) were also not statistically different, $t(575)=0.02$, $p>.05$. I conclude that a year prior to the introduction of TaRL the (treated and comparison) groups were similar.

Baseline measures (refer to table 2.5) also indicate balance between groups on in-

dicators of student mental health. Furthermore, as the TaRL intervention is not designed to directly influence student mental health (the main outcomes of the study), there is no reason to suspect bias as a result from systematic differences in mental health outcomes between the groups.

2.4.3 Econometric Specification and Hypothesis

The unit analysis for each dependent variable is for student i in class c at time t . The following specifications are estimated:

$$Y_{ict} = \beta_0 + \beta_1 A_{ic} + \beta_2 T_t + \beta_3 (A_i \times T_t) + \epsilon_{it} \quad (2.1)$$

The dependent variable Y_{ict} is the dependent variable of interest for student i in class c at time t . β_1 captures pre-treatment differences between treatment and control groups, prior to the TaRL implementation ($A = 1$ for treatment, and 0 for the comparison group). β_2 captures the time trend with T_t as the dummy variable for the second time period (i.e., the period when students receive the TaRL programme). β_0 is the constant term. The treatment effect is estimated the interaction term, captured by the β_3 coefficient. This is the difference between treatment and control classes in the period after the TaRL intervention takes place.

$$Y_{ict} = \beta_0 + \beta_1 A_{ic} + \beta_2 T_t + \beta_3 (A_i \times T_t) + \beta_4 \mathbf{X}_{ic} + \delta_{ic} + \epsilon_{it} \quad (2.2)$$

Specification 2.2 includes student level controls (such as gender, age, standard) and school fixed effects (δ_{ic}), which enhances the precision of the estimation. This is achievable due to the presence of treatment and comparison classes within each school, contributing to increased power by controlling for time-invariant school-specific variation. The main results presented in this study are based on the preferred Specification 2.2.

$$Y_{ict} = \beta_0 + \beta_1 A_{ic} + \beta_2 T_t + \beta_3 (A_i \times T_t) + \beta_4 YH_{ict} + \beta_5 (A_i \times YH_{ict}) + \beta_6 (T_t \times YH_{ict})$$

$$+ \beta_7(A_i \times T_t \times YH_{ict}) + \beta_8\mathbf{X}_{ic} + \delta_{ic} + \epsilon_{it}(2.3)$$

Specification 2.3 explores treatment heterogeneity for students with mental health difficulties at baseline (i.e., high SDQ and RCADS scores). YH_{ict} is for students with SDQ/RCADS scores at "high/very high" thresholds. The coefficient of interest is captured through β_7 .

Robust standard errors are calculated for all specifications, with standard errors clustered at the class level, the level at which the treatment is administered. All outcomes of interest are standardised by the mean and standard deviation of the control group.

2.4.3.1 Hypotheses

Primary Outcomes: Impact on student mental health. The underlying hypotheses is that TaRL will improve the mental health of all students exposed to the intervention.

H1: TaRL reduces the emotional and behavioural difficulties of students participating in the programme, in comparison to non-TaRL students (assessed through the SDQ).

H2: TaRL increases the prosocial behaviour among TaRL students compared to non-TaRL students (assessed through the SDQ)

H3: TaRL reduces the depressive and anxiety symptoms of students participating in the programme, in comparison to non-TaRL students (assessed through the RCADS)

Alternative mechanisms: Impact on Educational Success.

H4: TaRL improves mental health outcomes through improving the test scores of students enrolled in the programme as compared with non-TaRL students.

H5: TaRL improves mental health outcomes through improving the attendance record of students enrolled in the programme, as compared with non-TaRL students.

Other:

H6: TaRL improves the mental health and wellbeing outcomes of students with baseline mental health difficulties compared to with non-TaRL students also with baseline mental health difficulties.

2.5 Data and Descriptive Statistics

Primary schools in the city of Francistown, located in the North-East region, were selected as the focus for this research study. Francistown was selected due to its active implementation of TaRL by the Ministry of Basic Education within the region. There are 20 public schools in Francistown (Statistics Botswana, 2017) and at the time of the study, 15 schools were committed to TaRL implementation. As such, the study sample includes all 15 schools implementing TaRL during the data collection period. A total of 1,297 students from 52 classes in standards 4 and 5 were included in the study.

Moreover, in selecting the North-East region, several factors were also considered. These include: (1) the extent of completion of the 30-day implementation cycle; (2) the successful execution of multiple implementation cycles; and (3) the willingness of regional education officers to actively engage in the research process. Notably, all the schools in the study sample fulfilled these criteria.

2.5.1 Data

Data was collected over two academic terms between January and July 2022. Prior to data collection, I spent 2 months piloting and adapting the research tools with multiple groups of students to ensure face validity and comprehension. I trained 22 facilitators, and 2 coordinators to support the data collection. Baseline data collection began before the 30-day TaRL implementation and endline data collection took place after the completion of the programme. Students in TaRL classes (treated) and non-TaRL classes (comparison group) were assessed using the same mental health and wellbeing instruments as described in table 2.1. Table 2.4 below describes all the variables used in this paper.

Table 2.4: Data description for all variables included for analysis

Variable	Description
Student characteristics	
Age	Student reported age
Sex	Student reported sex (1=girl)
Standard	Student grade level
Outcome variables (standardised)	
Total Difficulties	Total SDQ score assessing total difficulties (4 subscales)
Prosocial behaviour	SDQ subscale
Anxiety	Total anxiety score assessed through RCADS
Depression	Total depression score assessed through RCADS
Total Anxiety and Depression	Total RCADS score (6 subscales)
Test score	Teacher recorded end of term exam results
Attendance	Teacher recorded student attendance
Teacher reported outcome variables (standardised)	
Student Prosocial behaviour	Question on how readily the student shares with other students
Student Conduct problems	Question on how often the student fights with other students
Student Hyperactivity	Question on how easily the student is distracted or has wandering concentration
Student emotional and behavioural difficulty indicator	Composite score of the 3 questions on student behaviour
Treatment Status	
TaRL Class	Students grouped by TaRL Implementation classes (1=TaRL class)

2.5.2 Descriptive Statistics

This study sampled 1297 students, from 52 classes, in standards 4 and 5. At baseline, there were no observed significant differences between the treated and comparison groups on most of the outcomes. Students in TaRL classes are comparable to students in non-TaRL classes on key background characteristics, on the primary mental health and wellbeing outcomes, and on the alternative educational success outcomes. One would expect at least one significant difference between treatment and control groups just by chance, and indeed on the prosocial scale, there is a significant difference (at the 5% level) between groups. However, there is no concern about estimates being biased towards treatment due to the non-TaRL comparison having higher prosocial scores than TaRL students at baseline.

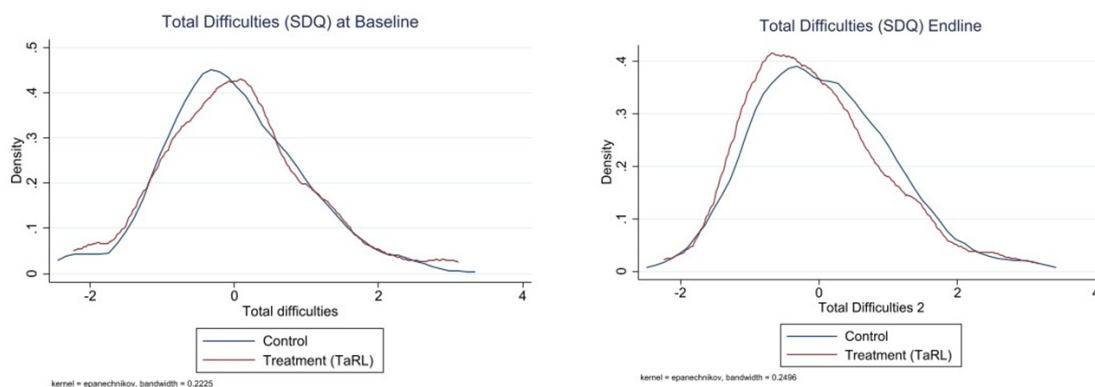
Table 2.5: Baseline comparison between TaRL students (treated) and non-TaRL students (comparison)

Panel A: Student characteristics	Baseline mean/proportion in the comparison group	T-C difference at baseline	N
Girl (1=Girl)	0.49	-0.012 (0.031)	1,296
Age	9.44	0.078 (0.136)	1,287
Standard (grade)	4	-0.110 (0.095)	1,297
Panel B: Mental health and wellbeing Outcomes (standardised)			
Total difficulties (SDQ)	-0.021	0.039 (0.083)	1,282
Prosociality (SDQ)	0.116	-0.215** (0.097)	1,285
Anxiety (RCADS)	-0.037	0.072 (0.106)	1,296
Depression (RCADS)	-0.009	0.015 (0.105)	1,296
Total anxiety and depression (RCADS)	-0.026	0.049 (0.110)	1,296
Panel C: Educational success outcomes (standardised)			
Test score (term 1)	0.074	-0.143 (0.134)	1,134
Attendance (term 1)	0.038	-0.074 (0.109)	1,020

Note: Robust standard errors in parentheses are clustered at the class level. The T-C difference is the baseline “treatment effect” using a minimal specification, including a treatment dummy. Mental health and educational success outcomes are standardised with mean 0, standard deviation 1; * $p < 0.01$ ** $p < 0.05$ *** $p < 0.01$.

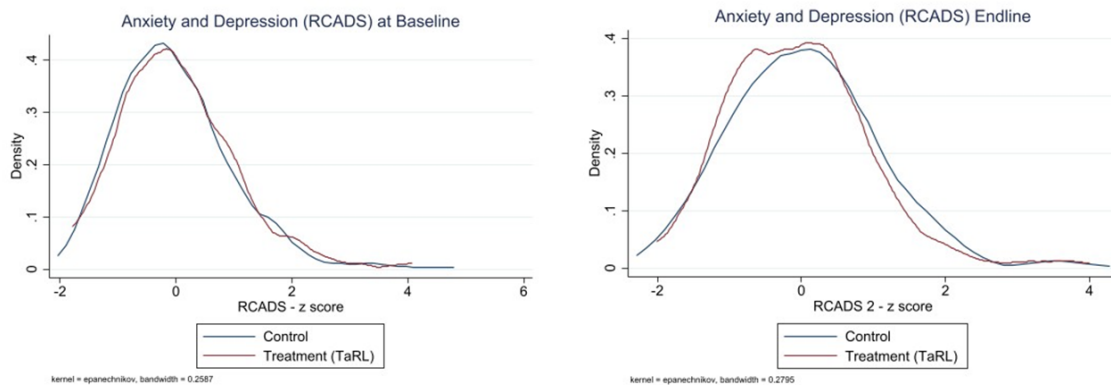
The kernel density plots in Figure 2.5 and Figure 2.6 show the distribution of students’ mental health outcome scores at baseline (prior to the TaRL implementation) and at endline.

Figure 2.5: Kdensity plots on Total Difficulties (standardised)



At the endline assessment, treated TaRL students seem to exhibit fewer total difficulties (indicated by a lower SDQ score) compared to the non-TaRL group. This observation is from the treated students' distribution being visibly skewed to the left of students in the comparison group. A test on equality of the distributions showed that at baseline, the combined Kolmogorov-Smirnov test statistic (D) across both groups was calculated to be 0.041, with a corresponding p-value of 0.632. At endline, the combined Kolmogorov-Smirnov test statistic (D) across both groups was calculated to be 0.072, with a corresponding p-value of 0.075. At baseline there are no significant differences between the groups, but at endline, while the p-value (0.075) shows that there are some differences between the distributions, observed at the 10% level of significance.

Figure 2.6: Kdensity plots on Anxiety and Depression (standardised RCADS score)



In the case of the total anxiety and depression outcome (measured by RCADS), a similar pattern emerges, although differences are non-significant on this measure. TaRL students at the endline assessment seem to experience lower levels of anxiety and depression. At baseline, a test on equality of the distributions showed that the combined Kolmogorov-Smirnov test statistic (D) across both groups was calculated to be 0.037, with a corresponding p-value of 0.766. At endline, the combined Kolmogorov-Smirnov test statistic (D) across both groups was calculated to be 0.056, with a corresponding p-value of 0.269. This indicates that there are no significant differences between the distributions at baseline and endline.

However, these figures show only raw data. The next section employs a more precise estimation strategy to fully assess the impact of TaRL on the outcomes of interest, using the

difference in difference (DD) strategy to account for time trends, student characteristics and time-invariant school fixed effects.

2.6 Results

Mental health outcomes. Regarding the mental health outcomes, a higher score signifies a greater level of overall mental health difficulties, anxiety, or depression. Using specification 2, the coefficients of interest (TaRL x Period) under column (1) show that TaRL students exhibit 0.15 standard deviations fewer total difficulties (significant at the 10% level) compared to non-TaRL students. Furthermore, looking at column (2), the results show that TaRL students demonstrate 0.24 standard deviations higher prosocial behaviour (significant at the 5% level) compared to their non-TaRL counterparts. On the anxiety, depression and total anxiety and depression outcomes, there are no significant differences observed. However, as the coefficients are negative, this suggests small improvements in mental health (i.e., less anxiety and depressive symptoms), but this is imprecisely estimated. Regression results on raw mental health outcomes are shown in appendix E.

Table 2.6: Impact of TaRL on Mental Health Outcomes

	(1) Total Difficulties	(2) Prosocial	(3) Anxiety	(4) Depression	(5) Anxiety & Depression
TaRL	0.022 (0.106)	-0.176** (0.083)	0.052 (0.099)	0.049 (0.092)	0.053 (0.100)
Period	0.086 (0.070)	-0.132* (0.075)	0.045 (0.068)	0.040 (0.066)	0.047 (0.068)
TaRL x Period	-0.153* (0.089)	0.239** (0.101)	-0.097 (0.092)	-0.084 (0.099)	-0.100 (0.096)
Control mean	0	0	0	0	0
Controls	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes
Students	1272	1273	1286	1286	1286

Notes: regression estimates on standardised mental health and wellbeing outcomes. Coefficients of interest are reported with robust standard errors in parentheses, clustered at the class level; *, **, and *** denote significance at the 10; 5; and 1 percent levels respectively. Controls include student age, sex, and standard variables

Additionally, during the endline assessment, teachers were asked three questions from the SDQ pertaining to observed emotional and behavioural difficulties among students in their classes. Questions included: (1) the extent to which students shared with other students; (2) how often the student fought with other students and (3) how easily the student was distracted or had wandering concentration. Their collective responses were combined into a composite score with a minimum score of 0 and maximum score of 6. Appendix F shows the impact of TaRL on teacher observations of student behaviour. There are no significant differences between TaRL students and non-TaRL students.

Educational success outcomes. This study also explores the impact of TaRL on educational success, assessing the impact of TaRL on educational success outcomes as an alternative mechanism enroute to improved mental health. This is used to test for an alternative mechanism: did mental health outcomes improve as a result of educational outcomes improving? The mental health outcomes included are test scores and recorded attendance data. This data was administratively collected by teachers each academic term.

Table 2.7: Impact of TaRL on Education Success Outcomes

	Math Test Score	Attendance
TaRL	-0.219 (0.173)	-0.007 (0.095)
Period	-0.087 (0.087)	0.051 (0.060)
TaRL x Period	0.166 (0.119)	-0.132 (0.159)
Control mean	0	0
Controls	Yes	Yes
School FE	Yes	Yes
Observations	1126	982

Notes: missing observations are a result of administrative data not readily available at the time of data collection. Robust standard errors in parentheses, clustered at class level; *, **, and *** denote significance at the 10; 5; and 1 percent levels respectively.

Results shown in Table 2.7 indicate that there are no meaningful effects of TaRL on student test scores or attendance records, between groups. On the Mathematics test

score we observe a positive (but insignificant) effect of 0.17 standard deviations. On the attendance outcome, a small negative insignificant effect is observed. However, due to the missing observations, it is difficult to precisely estimate the treatment effect on this outcome.

Although this study finds null education effects, there is a robust TaRL literature on the impact of the intervention on improved foundational skills (Banerjee et al., 2007; Duflo et al., 2011; Banerjee, Banerji, Duflo et al., 2016) and a growing literature linking student mental health to their academic attainment and success (Chen & Kaplan, 2003; Moilanen et al., 2010; Valdez et al., 2011; Becker et al., 2014; Daniele et al., 2022). Thus, the possible explanations for the observed null effects could point to implementation heterogeneity as explained by a recent education study by Angrist and Meager (2022). The authors found that implementation fidelity and the variation of delivery models found in education interventions explained much of the heterogeneity in treatment effects. It is then possible that these factors affected this research study.

In addition, another plausible explanation for the null education outcomes could stem from the choice of the outcome measure itself: school-based test scores. The impact of TaRL is typically assessed through a tailored TaRL assessment, designed to test the very same skills targeted in the intervention. However, the rationale for selecting administratively collected test scores was because this remains the primary decision-making mechanism for educators. As such, it was an appropriate indicator of educational success as much of the school-based tests (see Appendix H for example) examine knowledge of the basic mathematical operations such as addition and simple multiplication, and then go beyond the basics to test its application (e.g., volume, time). All concepts tested are covered in the usual class time, outside of the intervention. With regards to this study, there is a clear opportunity for future studies to include targeted assessment tools as well as broader assessments examining a wider range of applied skills.

2.6.1 Treatment Heterogeneity

Next, the study examines the heterogeneous impact of TaRL on children with baseline mental health difficulties. Regarding the comprehensive assessment of mental health

and wellbeing, the SDQ identifies children in the top 5%-10% with high/very high total difficulties (Goodman et al., 1998). On anxiety and depressive disorder symptoms, the cutoff indicating a high or severe anxiety or depression is typically the top 2% of the population or sample (Chorpita et al., 2000). Following the same principles, this research study adopts a similar approach to identify children predisposed to mental health difficulties.

Table 2.8: Heterogeneous effects on students with high baseline mental health outcomes

	(1) SDQ	(2) RCADS
TaRL	0.044 (0.076)	0.022 (0.131)
Period	0.014 (0.045)	0.036 (0.056)
highSDQ	1.389*** (0.035)	
TaRL*Period	-0.078 (0.057)	-0.127 (0.122)
TaRL*highSDQ	-0.050 (0.054)	
Period*highSDQ	-0.037 (0.037)	
TaRL*Period*highSDQ	0.072 (0.048)	
highRCADS		0.998*** (0.053)
TaRL*highRCADS		0.014 (0.059)
Period*highRCADS		0.005 (0.022)
TaRL*Period*highRCADS		0.050 (0.051)
Control mean	0	0
Controls	Yes	Yes
School FE	Yes	Yes
Students	755	758

Notes: Robust standard errors in parentheses. Controls include student age, sex, and standard; *p<0.10 **p<0.05 ***p<0.01

Using specification 2.3, the results show no heterogeneous treatment effects by baseline mental health or by sex. Student's predisposed to mental health difficulties are not

differentially affected by TaRL. These treatment effects are not significantly different from 0. Like-wise, no significant differences are observed between girls and boys in the study sample. The coefficients of interest in Table 2.9 (TaRL*Period*Girl) are negative, indicating a further reduction in total difficulties and anxiety and depressive symptoms relative to boys, but these effects are imprecisely estimated.

Table 2.9: Heterogeneous Effects on Mental Health, by sex

	(1) Total Difficulties (SDQ)	(2) Anxiety & Depression (RCADS)
TaRL	0.035 (0.122)	0.065 (0.133)
Period	0.067 (0.083)	-0.020 (0.087)
Girl	0.039 (0.094)	0.001 (0.105)
TaRL * Girl	-0.030 (0.133)	-0.029 (0.143)
TaRL * Period	-0.106 (0.116)	-0.069 (0.128)
Period * Girl	0.038 (0.103)	0.137 (0.095)
TaRL * Period * Girl	-0.097 (0.155)	-0.061 (0.150)
Control mean	0	0
Controls	Yes	Yes
School FE	Yes	Yes
Students	1272	1286

Notes: Robust standard errors in parentheses. Controls include student age, sex, and standard;

*p<0.10 **p<0.05 ***p<0.01

2.7 Discussion

The primary outcome of this research study was to evaluate the impact of a targeted pedagogy intervention on the mental health outcomes of students. The Strengths and Difficulties Questionnaire (SDQ) was selected as a measure of overall mental health (Beidas et al., 2015) with the Revised Child and Anxiety Scale (RCADS) specifically included to capture any anxiety and depressive symptoms. As discussed, the SDQ and RCADS are evidence-based assessment tools, with robust peer reviewed literature

attesting to their strong psychometric properties. Both instruments have wide-ranging scope, serving as effective tools for screening individuals at risk and for monitoring shifts in symptoms over time (Hunsley & Mash, 2008; Beidas et al., 2015; Wolpert et al., 2015). Regarding measures of internal consistency, in line with the guidance set out by Hunsley & Mash (2005) and the broader principles set out by Nunally & Bernstein (1978), this study determines that the internal consistency measures (alpha and omega coefficients) adequately measured the intended mental health constructs.

In response to the research questions set out in this study, the results confirm the hypotheses that TaRL has a positive impact on the overall mental health and wellbeing of students. The results reveal a significant 0.15 standard deviation reduction in the emotional and behavioural total difficulties experienced by TaRL students compared to non-TaRL students. The study also found a 0.24 standard deviation increase (significant at the 5% level) in prosocial behaviour among TaRL students compared to the non-TaRL students. Although the study does not find any significant average treatment effects on anxiety and depressive symptoms, this is unsurprising as there was a very small sample in this study exhibiting any anxiety and depressive symptoms above typical thresholds.

The main results are in line with the growing literature demonstrating how educational interventions can also support the mental health of young people (Erwin et al., 2012; Hanson et al., 2016; Larson et al., 2020; Daniele et al., 2022). Additionally, data from teachers and regional education officers interviewed during the research study also supports this study's main findings.

"TaRL reinforces key concepts and builds self-esteem and confidence. When they work in small groups, it helps them open up...you can even hear their voices, without that you don't even know if they understand."

– Chief Education Officer, North-East Region

"TaRL pupils seem to be open and free, they are confident and free spirited."

– Teacher ,North-East Region

"Even slow learners are freely participating in class, some used to display bad behaviour

but these days there's positive change."

– Teacher, North-East Region

2.8 Conclusion

Educational environment matters for mental health promotion. The main results reveal targeted instruction interventions such as Teaching at the Right Level have a positive impact on reducing some externalising and internalising mental health difficulties experienced by learners. This paper is able to connect the education and mental health literatures, contributing to the evidence base on improving mental health and educational outcomes particularly in low-income and middle-income settings in Africa. With the dearth of policy relevant mental health data in Africa and this study fills an important research gap. Moreover, this research presents one of the first studies linking Teaching at the Right Level with improved mental health outcomes.

2.9 References

- Aldridge, J.M., & McChesney, K. (2018). The relationships between school climate and adolescent mental health and wellbeing: A systematic literature review. *International Journal of Educational Research*, 88, 121-145. <https://doi.org/10.1016/j.ijer.2018.01.012>
- Angrist, J.D., & Pischke, J.S. (2005). *Mostly harmless econometrics: An empiricist's companion*. Chapter 5: Parallel Worlds: Fixed effects, differences-in-differences, and panel data, pp221-247. <http://dx.doi.org/10.1007/s00362-009-0284-y>
- Angrist, N., Bergman, P. & Matsheng, M. Experimental evidence on learning using low-tech when school is out. *Nature Human Behaviour* 6, 941-950 (2022). <https://doi.org/10.1038/s41562-022-01381-z>
- Angrist, N. and Meager, R. (2022). The role of implementation in generalisability: A synthesis of evidence on targeted educational instruction and a new randomised trial, CEDIL Syntheses Working Paper 4, CEDIL, Oxford. Available at <https://doi.org/10.51744/CSWP4>
- Ashenfelter, O., & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effects of training programs. *Review of Economics and Statistics*, 67, 648-660. <http://dx.doi.org/10.2307/1924810>
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., & Walton, M., 2016. Mainstreaming an effective intervention: Evidence from randomized evaluations of 'Teaching at the Right Level' in India. National Bureau of Economic Research, 22746.
- Banerjee, A., Cole, S., & Duflo, E. (2007). Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics* 122.3 (2007):1235-1264.
- Banerji, R., & Chavan, M. (2016). Improving literacy and math instruction at scale in India's primary schools: The case of Pratham's Read India program. *Journal of Educational Change* 17(4), 453-475.
- Becker, K.D., Brandt, N.E., Stephan, S.H., & Chorpita, B.F. (2014). A review of edu-

-
- cational outcomes in the children's mental health treatment literature. *Advances in School Mental Health Promotion*, 7(1), 5-23.
- Beidas,R.S., Stewart,R.E., Walsh,L., Lucas,S., Downey,M.M., Jackson,K., Fernandez,T., & Mandell, D. S. (2015). Free, brief, and validated: Standardized instruments for low-resource mental health settings. *Cognitive and behavioral practice*, 22(1), 5-19. <https://doi.org/10.1016/j.cbpra.2014.02.002>
- Botswana Government (1994). Revised National Policy on Education (Government Paper No.2).
- Brislin,R.W. (1970). Back-Translation for Cross-Cultural Research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216. <https://doi.org/10.1177/135910457000100301>.
- Bronfenbrenner, U. (1979).The ecology of human development: Experiments by nature and design. Cambridge: Harvard University Press.
- Carneiro P, Heckman JJ. Human capital policy. In: Heckman JJ, Krueger A, eds. *Inequality in America: what role for human capital policy?* Cambridge: MIT Press, 2003.
- Chen,Z., & Kaplan,H.B. (2003). School failure in early adolescence and status attainment in middle adulthood: A longitudinal study. *Sociology of Education*, 76, 110-127.
- Chorpita, B.F., Yim, L.M., Moffitt, C.E., Umemoto L. A., & Francis,S.E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy*, 38, 835-855.
- Cohen, J., McCabe, E. M., Michelli, N.M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record*, 111(1),180-213.
- Cortina, M.A., Sodha,A., Fazel,M., & Ramchandani, P.G. (2012). Prevalence of child mental health problems in sub-Saharan Africa: a systematic review. *Archives of pediatrics & adolescent medicine*, 166(3), 276-281. <https://doi.org/10.1001/archpediatrics.2011.592>.
- Council on Early Childhood, Committee on Psychosocial Aspects of Child and Family
-

-
- Health, & Section on Developmental and Behavioral Pediatrics (2016). Addressing Early Childhood Emotional and Behavioral Problems. *Pediatrics*, 138(6), e20163023. <https://doi.org/10.1542/peds.2016-3023>.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.
- Daniel, B.C., and A.A. Powell. 2003. "Altruism and Prosocial Behavior." In *Handbook of Psychology*, edited by T. Millon and M.J. Lerner, 463-479. Vol. 5. Hoboken, NJ: Wiley.
- Danon, A., Das, J., de Barros, A., Filmer, D., Cognitive and socioemotional skills in low-income countries: Measurement and associations with schooling and earnings 1, *Journal of Development Economics* (2023), doi: <https://doi.org/10.1016/j.jdeveco.2023.103132>.
- DeVon, H.A., Block, M.E., Moyle-Wright, P., Ernst, D. M., Hayden, S.J., Lazzara, D.J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing*, 39(2), 155-164. <https://doi.org/10.1111/j.1547-5069.2007.00161.x>
- Dewey J (1938) *Experience and Education*. New York: Kappa Delta Pi.
- Diener, E., Oishi, S., & Lucas, R.E. (2003). Personality, culture, and subjective well-being: emotional and cognitive evaluations of life. *Annual review of psychology*, 54, 403-425. <https://doi.org/10.1146/annurev.psych.54.101601.145056>.
- Duflo, E. (2001). Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment. *American Economic Review*, 91(4), 795-813. <https://doi.org/10.1073/pnas.0706537105>
- Duflo, E., Dupas, P., Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review* 101(5): 1739-74.
- Eriksson, M., Ghazinour, M., & Hammarstrom, A. (2018). Different uses of Bronfenbrenner's ecological theory in public mental health research: what is their
-

-
- value for guiding public mental health policy and practice? 16, 414-433.
<https://doi.org/10.1057/s41285>
- Erwin,H., Fedewa,A., Beighle,A., & Ahn S. (2012) A Quantitative Review of Physical Activity, Health, and Learning Outcomes Associated With Classroom-Based Physical Activity Interventions, *Journal of Applied School Psychology*, 28:1, 14-36, DOI: 10.1080/15377903.2012.643755
- Garcia-Escalera,J., Valiente,R., Sandin,B., Enhrenrieck-May,J., Chorot, P. (2020). Educational and wellbeing outcomes of an anxiety and depression prevention program for adolescents. *Revista de Psicodidactica (English ed)*, 25(2), 143-149.
<https://doi.org/10.1016/j.psicoe.2020.05.003>.
- Gleason, M.M., Goldson,E., Yogman, M. (2016). Addressing Early Childhood Emotional and Behavioral Problems. *Pediatrics*. 138(6): e20163025-e20163025. DOI:10.1542/peds.2016-3025.
- Global Education Monitoring report. (2020). Teaching at the Right Level: from concern with exclusion to challenges of implementation. "Paper commissioned for the 2020 Global Education Monitoring Report, Inclusion and education".
- Goodman,R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40(11), 1337-1345. <https://doi.org/10.1097/00004583-200111000-00015>
- Hilsman,R., & Garber,J. (1995). A test of the cognitive diathesis-stress model of depression in children: Academic stressors, attributional style, perceived competence, and control. *Journal of Personality and Social Psychology*, 69, 370-380.
- Honde,G.J. (2018). Botswana 2018 African Economic Outlook.
- Hunsley, J., & Mash, E. J. (Eds.). (2008). A guide to assessments that work. Oxford University Press.
- Innovations for Poverty Action (2018). Evaluating the Teacher Community Assistant Initiative. Accessed May 19, 2022. <https://www.poverty-action.org/study/evaluating-teacher-community-assistant-initiative-ghana>
- James,S.L., Abate,D., Abate,K.H., Abay,S.M., Abbafati,C., Abbasi,N., Briggs,A.M.
-

-
- (2018). Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet*, 392(10159), 1789-1858.
- Jana M. Hanson, Teniell L. Trolan, Michael B. Paulsen & Ernest T. Pascarella (2016). Evaluating the influence of peer learning on psychological well-being, *Teaching in Higher Education*, 21:2, 191-206, DOI: 10.1080/13562517.2015.1136274
- Kessler, R.C., Amminger, G.P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., Ustun, T.B. (2007). Age of onset of mental disorders: a review of recent literature. *Current Opinion in Psychiatry*, volume 20, Issue 4, pg. 359-364. DOI: 10.1097/YCO.0b013e32816ebc8c
- Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigbodun, O., Rahman, A. (2011). Child and adolescent mental health worldwide: Evidence for action. *The Lancet*, 378(9801), 1515-1525.
- Larson, K.E., Nguyen, A.J., Orozco Solis, M.G., Humphreys, A., Bradshaw, C.P., & Lindstrom Johnson, S. (2020). A systematic literature review of school climate in low- and middle-income countries. *International Journal of Educational Research*, 102. <https://doi.org/10.1016/j.ijer.2020.101606>
- Massarwi, A., Cluver, L., Meinck, F., Doubt, J. & Green, O. (2022). Pathways to parenting stress reduction among parents in South Africa. *Child & Family Social Work*. 10.1111/cfs.12952
- McAllister M, Knight BA, Hasking P, Withyman C, Dawkins J. Building resilience in regional youth: Impacts of a universal mental health promotion programme. *Int J Ment Health Nurs*. 2018 Jun;27(3):1044-1054. doi: 10.1111/inm.12412. Epub 2017 Nov 24. PMID: 29171720.
- Mokibelo, E. (2015). The Outcomes of Learning a Foreign Language: Cases of Rural Primary Schools in Botswana. *US-China Education Review A*. 5. 10.17265/2161-623X/2015.09A.001.
- Morgan, P.J., & Hansen, V. (2008). Physical education in primary schools: Classroom teachers' benefits and outcomes. *Health Education Journal*, 67, 196-207.
-

-
- Opondo, P.R., Olashore, A.A., Molebatsi, K., Othieno, C.J., & Ayugi, J.O. (2020). Mental health research in Botswana: a semi-systematic scoping review. In *Journal of International Medical Research* (Vol. 48, Issue 10). SAGE Publications Ltd. <https://doi.org/10.1177/0300060520966458>
- O'Reilly, M., Svirydzenka, N., Adams, S., & Dogra, N. (2018). Review of mental health promotion interventions in schools. *Social psychiatry and psychiatric epidemiology*, 53(7), 647-662. <https://doi.org/10.1007/s00127-018-1530-1>
- Pansiri, O.N and Tsayang, G.T. (2017). A Situational Analysis of Basic Literacy and Numeracy levels at early grade levels in Botswana. Background Report.
- Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., Chisholm, D., Collins, P.Y., Cooper, J.L., Eaton, J., Herrman, H., Herzallah, M.M., Huang, Y., Jordans, M.J.D., Kleinman, A., Medina-Mora, M. E., Morgan, E., Niaz, U., Omigbodun, O., Unutzer, J. (2018). The Lancet Commission on global mental health and sustainable development. *The Lancet*, 392(10157), 1553-1598.
- Population and Housing Census: Preliminary Results. Statistics Botswana (2022). <https://www.statsbots.org.bw/sites/default/files/2022%20Population%20and%20Housing%20Census%20Preliminary%20Results.pdf>
- Republic of Botswana (2015). Education and Training Sector Strategic Plan, (ETSSP 2015-2020), Gaborone: Government Printing and Publishing Services.
- Sandin, B., & Chorot, P., Valiente, R., & Chorpita, B. (2010). Development of a 30-item version of the Revised Child Anxiety and Depression Scale. *Revista de Psicopatologia y Psicologia Clinica*. 15. 10.5944/rppc.vol.15.num.3.2010.4095.
- Sen A. Inequality Reexamined. Oxford: Oxford University Press; (1992).
- Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) (2018). SACMEQ IV Project in Botswana: A study of the conditions of schooling and quality of education. Department of Educational Planning and Research Services Research Unit.
- Statistics Botswana (2022). Botswana AIDS Impact Survey V (BAIS V) Summary sheet. <https://www.statsbots.org.bw/sites/default/files/>
-

BAIS%20V%20Preliminary%20Report.pdf

Statistics Botswana (2017). Primary School Stats Brief 2017.

<https://www.statsbots.org.bw/sites/default/files/publications/Primary%20School%20Stats%20Brief%20%202017.pdf>

Suldo, S.M., Thalji, A., & Ferron, J. (2011). Longitudinal academic outcomes predicted by early adolescents' subjective well-being, psychopathology, and mental health status yielded from a dual-factor model. *Journal of Positive Psychology*, 6, 17-30. DOI: 10.1080/17439760.2010.536774.

Suldo, S., Gormley, M., DuPaul, G., & Anderson-Butcher, D. (2014). The Impact of School Mental Health on Student and School-Level Academic Outcomes: Current Status of the Research and Future Directions. *School Mental Health*, 6, 84-98.

United Nations Educational, Scientific and Cultural Organization (UNESCO). UNESCO science report (2005). Paris: UNESCO Publishing; 2005. Available: <http://www.unesco.org/science/psd/publications/scrp05.shtml> [accessed 29.04.08].

UNESCO Institute of Statistics (2022). SDG 4- Country Profiles: Botswana. <https://uis.unesco.org/sites/default/files/country-profile/Botswana-SDG4-Profile.pdf>

Valdez, C.R., Lambert, S. F., & Ialongo, N. S. (2011). Identifying patterns of early risk for mental health and academic problems in adolescence: a longitudinal study of urban youth. *Child psychiatry and human development*, 42(5), 521-538. <https://doi.org/10.1007/s10578-011-0230-9>.

Wolpert, M., Cheng, H. and Deighton, J. (2015), Measurement Issues: Review of four patient reported outcome measures: SDQ, RCADS, C/ORS and GBO - their strengths and limitations for clinical use and service evaluation. *Child Adolescent Mental Health*, 20: 63-70. <https://doi.org/10.1111/camh.12065>

World Bank, (2020). Human Capital Index: Botswana. Accessed 3 December 2020. <https://databank.worldbank.org/data/download/hci/HCI-2pager-BWA.pdf>.

World Bank. 2020. The Human Capital Index 2020 Update: Human Capital in the Time

of COVID-19. World Bank, Washington, DC. <http://hdl.handle.net/10986/34432> License: CC BY 3.0 IGO.

World Health Organization (2001). World Health Report (2001). Mental Health: New Understanding, New Hope. Geneva, Switzerland.

World Health Organization. Prevention of Mental Disorders: effective interventions and policy options, summary report. Geneva: World Health Organization, 2004.

World Health Organization (WHO) and Calouste Foundation. (2014). Social determinants of mental health. Geneva: WHO.

World Health Organization. (2021). Mental health atlas (2020). Geneva: World Health Organization, 2021.

Zenner, C., Herrnleben-Kurz, S., & Walach, H. (2014). Mindfulness-based interventions in schools: A systematic review and meta-analysis. *Frontiers in Psychology*, 5, 1-20. <https://doi.org/10.3389/fpsyg.2014.00603>

3 | Improving the mental health and educational success of young people: experimental evidence from a mindfulness-based programme in Pakistan

Abstract¹

Mindfulness-based Stress Reduction are a class of group-based interventions that have emerged to help people manage stress, improve their mood and quality of life. However, these interventions are typically implemented in adult populations. As such, there is limited evidence on the impact of such mindfulness-based programmes on student mental health, and educational success. This research evaluates the impact of a mindfulness-based programme on student mental health, behaviour change, capacity to learn, and educational success in two low-cost private schools in Faisalabad, Pakistan. A total of 214 secondary school students were randomly assigned to a mindfulness programme treatment group, and a neutral class activity control group. The study shows three main results. First, it demonstrates that the intervention improves student's prosociality. Second, the mindfulness programme impacts behaviour change, specifically improving mood and stress at home. Lastly, students who received the programme exhibit a greater capacity to learn, particularly demonstrating higher levels of grit. No effects are found on primary mental health outcomes, anxiety and depression. These findings have implications for how to implement and scale-up mindfulness-based programmes for young people, particularly in low-resource settings.

¹This is co-authored work with Dr. Zahra Mansoor. This study was funded by the John Fell Fund. The study was registered in the AEA RCT registry at: AEARCTR-0010019. Ethics approval was granted by the institutional research ethics committee at the University of Oxford (UK, Ref. No. SSH/BSGC1A-21-18, 02 December 2021)

3.1 Introduction

A significant proportion of mental health disorders surface in early childhood and adolescent years (Kileling et al., 2011). Systematic reviews show that interventions focused on student mental health can have promising effects on the mental health and education outcomes of children (Becker et al., 2014). However, less than 50% of young people with such disorders receive help, and there remains a substantial gap between need and resource availability (Ye et al., 2004).

Among psychotherapy-based programmes to address mental health and well-being, Mindfulness- Based Stress Reduction (MBSR) are a class of group-based interventions that help patients cope with stress, improve mood, and quality of life (Kabat-Zinn, 1982). Although originally designed for patients, the programme has been adapted for general populations and has been used to reduce physical pain and psychological distress including stress, anxiety, and depression (Teasdale et al., 2000; Baer, 2003; Virgili, 2015).

Executive control serves as a mechanism linking mindfulness to emotion regulation-a skill pivotal for cultivating stress resilience, effective coping under stress (Creswell, 2017) and mood regulation, subsequently reducing the probability of anxiety and depression (Monk et al., 2008; Holzel et al., 2011). Notably, the existing evidence highlights that MBSR can be effective for regulating emotional responses and reducing stress, both across different types of patients and young people (see meta-analysis by Grossman et al., 2004). Studies focused on young people broadly show that MBSR programmes can have positive impacts on measures such as stress and anxiety (Kuyken et al., 2013; Gouda et al., 2016; Xiang et al., 2020). However, the evidence remains limited with majority of the studies either focusing on adults, observing impacts on a specific set of mental health outcomes, or having methodological limitations². The gap in evidence is particularly stark when it comes to young people in low-middle income countries.

This research addresses this gap. Through partnerships with a local NGO in Punjab, Pakistan - Sirajia Institutions - that manages low-cost private schools in the

²Such as pre-post designs which makes it hard to establish causality, small sample sizes, or questions on program fidelity (Kabat-Zinn et al., 1998; Teasdale et al., 2000; Williams et al., 2000, Baer, 2003).

city of Faisalabad, and the first ever Urdu-based mindfulness-based platform in Pakistan - Be-Khudi, we study the impact of a mindfulness-based stress reduction (MBSR) programme on young people. The MBSR programme is adapted to the local language (Urdu) and to a low-cost school setting. We ask three research questions. First, what is the impact of MBSR on the mental health and well-being outcomes of young people (as captured through measures of anxiety, depression, prosociality, perceived stress, and test anxiety)? Second, what is the impact of MBSR on the behaviour change of young people (as captured through the perceived change in the relationship with parents and siblings, stress at home, mood, impulse control, and time-use)? Third, what is the impact of MBSR on the capacity to learn outcomes and educational success of young people (as captured through measures of growth mindset, general self-efficacy, grit, and test scores)?

Pakistan, the context of this study, is a lower-middle income country with a total population estimated at 240.5 million (UNFPA, 2023). An estimated 10-16% of the population are reported to suffer from some mental illnesses (Ansari, 2015). However, to date, less than 1% of the annual health budget is allocated to adult mental health policy and support with no specific allocations made for adolescents and young people (Khan, Shehzad & Chaudry; 2008). This further highlights the relevance of our research questions in the study context.

To implement the programme, Sirajia Institutions randomised a total of 214 students in grades 8-10 into either the MBSR treatment group or to a status quo control group. Students who received the MBSR treatment were enrolled in an 8-week mindfulness programme between October to November school term in 2022, in which students received an hour-long session on mindfulness each week. The programme was broadly based on the principles of standard MBSR programmes (as originally developed by Kabat-Zinn, 1982), but went through an intensive 8-month adaptation and piloting period to tailor it to the local language and low-cost school setting. To ensure that our control group served as an appropriate counterfactual, two points were necessary: first, students in the control group needed to be in a classroom setting to control for any possible peer effects by simply being in a classroom. Second, we needed to ensure that the students received a neutral experience (i.e., no direct support from teachers that

could impact their mindfulness). To achieve these objectives, control group students were assigned to a silent Quran reading class. A class where students would be asked to read their Quran is a typical activity in the Pakistan context and as such could act as a neutral control.

The study highlights the following results. Among mental health and wellbeing outcomes, we find that MBSR significantly improves prosociality by 0.30 standard deviations in comparison to the control group. We do not observe significant effects of MBSR on total difficulties (SDQ), depression and anxiety (RCADS), perceived stress (PSS), or test anxiety (STABS). Multiple hypothesis test correction following Anderson (2008) shows that the treatment effect of MBSR on prosociality remains significant at the 10% level.

Second, we observe that students who receive the MBSR treatment report an improvement on our combined behaviour change index by 0.30 standard deviations in comparison to the control group. Looking at the individual behaviour change questions, we observe that students who receive the MBSR treatment specifically report an improvement in mood (by 0.52 standard deviations) and less stress at home (by 0.34 standard deviations). Overall, this indicates that students who received the MBSR programme were able to foster a more positive home environment relative to students who were not exposed to MBSR. Finally, on student capacity to learn and educational success outcomes, MBSR increases the grit of treated students by 0.34 standard deviations relative to the students in the control group. There are no significant effects on growth mindset and generalised self-efficacy (GSE). We observe no treatment effects on learning outcomes in Maths, English, Urdu, and Science.

This study makes the following contributions. First, it demonstrates that it is possible to adapt a mindfulness-based curriculum to a low-resource setting. The programme was implemented successfully - it was adapted to the local context and ran smoothly over the 8-week period. While mindfulness-based programmes have typically been implemented in high-income settings (Xiang et al., 2020), this study highlights that such group-based programmes could also potentially be considered as part of the suite of school-based in-

terventions that can address student mental health and wellbeing in low-resource settings.

Second, the study provides causal evidence on the impact of mindfulness-based programmes on student mental health, behaviour change and educational success. It contributes specifically to the existing literature on the impact of mindfulness programmes on student mental health and educational success (Kuyken et al., 2013; Gouda et al., 2016; Xiang et al., 2020), as well as the wider literature on the impact of non-cognitive skills such as grit and self-control on student educational outcomes (Heckman et al., 2011; Alan, Boneva & Ertac, 2019). While existing studies mostly focus on mindfulness interventions for adult populations (Baer, 2003), or on student populations in high income contexts (Gouda et al., 2016; Xiang et al., 2020), this study adapts a mindfulness-based intervention for a student population in low-cost private schools in a low-middle income country, demonstrating promising impacts on student behaviour and capacity to learn (particularly grit). This is non-trivial given the lack of mental health support for young people in this setting.

More broadly, the study generates important lessons for how mindfulness-based programmes may be scaled up. In particular, the null results of the study on some measures of student mental health, well-being and educational success highlight important lessons for future programme iterations. For example, measures such as student anxiety, depression, and test scores may be harder to move in a short time-span of 3 months, particularly in community samples with anxiety and depressive symptoms falling in the low severity threshold. Future iterations could vary programme length or build stronger focus on aspects of behaviour change and grit which appear to be the programme’s strength and could become causal mechanisms towards shifting wider measures of student mental health, behaviour change, and educational success. Put together, this has implications for how mindfulness-based programmes may be scaled up across low-resource school settings.

The rest of the paper is structured as follows. Section 3.2 presents the study context with a summary of the mindfulness literature and Pakistani context. Section 3.3 presents the study research questions. In section 3.4, the experimental design is presented, first by discussing the intervention, and then the randomization strategy. Section

3.5 describes the outcome measures and data used in this study, including summary statistics at baseline, tests for data validity, and balance checks for randomization. Section 3.6 presents the results, and Section 3.7 concludes.

3.2 Context

3.2.1 Literature on mindfulness and mindfulness-based stress reduction interventions

Mental health is "a state of well-being in which the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to his or her community" (WHO, 2001a). Following this definition, this study conceptualises mental health and well-being as encompassing all psychological and socioemotional constructs including anxiety, depression, and stress.

Mental health and well-being interventions take many forms. Behavioural and or psychological interventions (i.e., activities designed to change behaviours, feelings, and emotional states) can lead to improvements in mental health, a reduction in mental illness or both. A recent systematic review and meta-analysis on psychosocial interventions highlighted the role of mindfulness-based interventions in causing significant improvements in the mental health and well-being of populations with mental illness as well as populations with physical illness (van Agteren et al., 2021).

Kabat-Zinn (1994) defines mindfulness as "paying attention in a particular way: on purpose, in the present moment, and nonjudgmentally". Mindfulness as a practice can be operationalised in several ways but the common facets of mindfulness are (1) present moment awareness and (2) non-judgmental acceptance of emotions and thoughts (Cardaciotto et al., 2008; Teper et al., 2013). Present moment awareness can include body sensations, emotional reactions, mental images, mental talk, and perceptual experiences. Cognitive scientists observe that the difficulty in cultivating an awareness of the present moment can engender self-regulation skills (Bjork & Bjork 2011). The second facet of non-judgmental acceptance suggests the importance of embracing an

openness or acceptance towards individual experiences. Developing this attitude of acceptance is not a "passive resignation to one's current circumstances but rather one of inviting in experiences even if they are difficult" (Creswell, 2017). Mindful acceptance thus leads to a non-judgmental attitude (Teper et al., 2013).

One of the psychological benefits of the practice of mindfulness is its impact on emotion regulation (Chambers et al., 2009) and executive control³. Theorists suggest that mindfulness can improve executive control as it promotes present moment awareness and acceptance and nonjudgement openness. Through the mechanism of executive control, mindfulness is linked to emotion regulation; the skill that develops stress resilience, coping under stress (Creswell, 2017) and regulating mood thereby decreasing the probability of anxiety and depression (Monk et al., 2008; Holzel et al., 2011).

Mindfulness-based stress reduction (MBSR) interventions is one of many (including treatment as usual, mindfulness retreats, etc.) interventions most used in the mindfulness scientific literature. The 8-10 week programme first developed Kabat-Zinn (1982, 1990) has been adapted to treat depression, and improve relationship functioning among other applications (Baer, 2003; Creswell, 2017). Researchers and clinicians who have introduced mindfulness activities and practice into mental health interventions and programmes often teach these skills independently of any of its religions and cultural origins (Kabat-Zinn, 1982; Baer, 2003). MBSR programmes designed to reduce suffering and improve wellbeing typically include practice in meditation skills, discussions on stress, coping activities, and homework assignments (Baer, 2003). A study by Goldin & Gross (2010) and Holzel et al (2010) found meditation practice reduced stress and was also correlated with reductions in amygdala activation - evidence of MBSR's positive effects on stress. Finally, a recent systematic review showed confirmed mindfulness-based interventions as cost-effective and scalable (Zhang et al., 2022).

However, it is worth noting that the empirical literature on mindfulness-based training contains some methodological limitations. These include pre-post designs, small sample sizes, and questions on program fidelity (Kabat-Zinn et al., 1998; Teasdale

³Executive control has also proven impact on academic attainment and healthy eating habits (Hofmann et al., 2012; Teper et al., 2013)

et al., 2000; Williams et al., 2000, Baer, 2003). Overall, this literature suggests that mindfulness-based programmes and interventions can lead to a reduction in physical pain and psychological distress including anxiety, stress, and depression (Kabat-Zinn, 1982; Kabat-Zinn et al., 1992; Kristeller & Hallet, 1999; Shapiro et al., 1998; Teasdale et al., 2000; Baer, 2003; Virgili 2015). These studies however are typically conducted in general adult populations. This presents a gap in the literature to explore the implications for adolescents particularly in low-resource settings.

3.2.2 Mental health and in Pakistan

Pakistan is a lower-middle income country with a total population estimated at 240.5 million (UNFPA, 2023). An estimated 10-16% of the population are reported to suffer from some mental illnesses (Ansari, 2015). Moreover, a recent study in Pakistan by Danon et al. (2023) showed that more years of schooling is linked with greater cognitive skills and some socioemotional skills. In addition, long-run labour earnings were also associated with schooling and higher socioemotional skills. The study supports prioritising increased investments in improving both cognitive and socioemotional skills (Danon et al., 2023).

Furthermore, a Pakistani study extrapolating the 2020 economic consequences of mental illness using associated costs from 2006 data found that the economic costs would reach 619.9 billion PKR rising from 250.5 billion (PKR) in 2006 (Alvi et al., 2023). The staggering economic burden highlights the need for a greater focus on mental health policy; particularly that which provides sufficient services as well as identifying effective interventions particularly for young people (Alvi et al., 2023). However, to date, less than 1% of the annual health budget is allocated broadly to adult mental health with no specific allocations made for adolescents and young people (Khan, Shehzad & Chaudry; 2008). On research outputs, WHO's mental health atlas (2020) reports that mental health research makes up 4.53% of the total research outputs in Pakistan.

As such, our study is directly related to this policy context. It is the first of its kind to adapt a mindfulness-based stress reduction intervention in Urdu for low-resources secondary schools in Faisalabad, Pakistan. It fills a critical evidence gap,

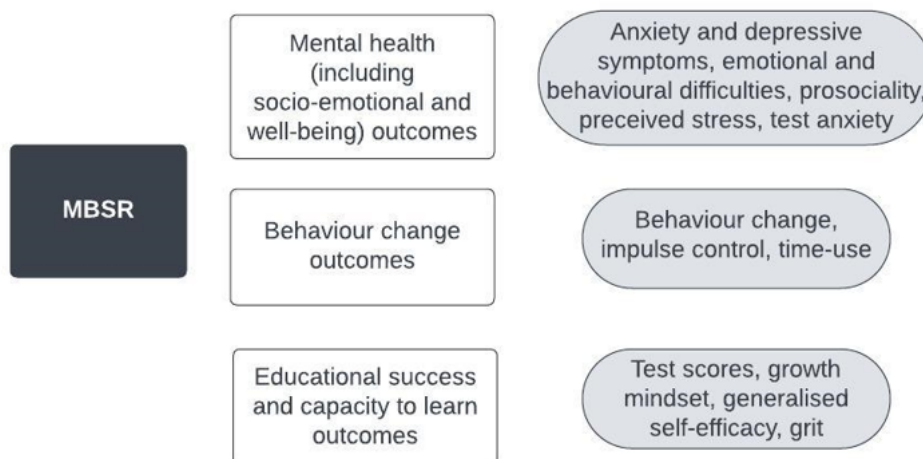
using a combination of self-report and task-based instruments to further understand the impact of a mindfulness-based programme on the mental health, behaviour change and educational success of young people.

3.3 Theory and Research Questions

3.3.1 Research Questions

Section 3.2 showed that MBSR has the potential to impact a range of mental health outcomes such as anxiety and depression, emotional and behavioural difficulties, and stress. However, no single study has examined how such a programme affects the full range of these outcomes, and the possible relationships between these outcomes. Additionally, there is growing literature showing the interconnections between mental health and education outcomes, where changes in one area affects the other (Valdez et al., 2011; Suldo et al., 2014). As such, we take an expanded approach, categorising the potential outcomes as: (1) mental health outcomes; (2) behaviour change outcomes and (3) education and capacity to learn outcomes. The constructs targeted are shown in Figure 3.1 below and are explained in detail in Section 3.5.

Figure 3.1: Hypothesized Categories of Outcomes



We evaluate the following research questions:

1. What is the impact of MBSR on the mental health outcomes such as anxiety,

depression, prosociality, perceived stress, and test anxiety on young people?

2. What is the impact of MBSR on the behaviour change (including perceived change in relationship with parents and siblings, impulse control, and time-use) on young people?
3. What is the impact of MBSR on the educational success and capacity to learn outcomes (test scores, growth mindset, self-efficacy, and grit) of young people?

In addition, for subgroup analysis, we look at the gendered effects of MBSR on our main outcomes. Evidence indicates that among the adult population, women are at a higher risk for mental health problems such as depression and anxiety whereas men are at a higher risk for substance abuse disorders (Patel, Saxena, Lund et al., 2018).

3.4 Experimental Design

3.4.1 Intervention

The MBSR curriculum was developed through multiple iterations and piloting over an eight-month period from January - August, 2022. First, Be-Khudi (our implementation partner) conducted a teacher training with 9 teachers. All the teachers completed 8 sessions (1 hour each week) of the mindfulness and mediation training. At the end of the training, 4 teachers were selected as Master Trainers (MTs) to deliver the programme to the students. Throughout the training sessions, lesson plans for students were constantly updated incorporating suggestions from the MTs.

The MBSR curriculum was then piloted in schools with students prior to implementation ⁴. MTs and students provided practical and content-specific feedback that was used to further improve lesson plans. The iterative curriculum adaptation process allowed for research participants (i.e., students) and MTs expert and contextually-relevant feedback to feed into the process of developing materials. In addition, Be-Khudi also recorded 7-minute introductory videos for each topic in the curriculum. These were included to standardize the introduction of each concept in the curriculum. All the

⁴The curriculum was piloted in a school not included in the study sample.

content and materials were developed in Urdu to ensure maximum student understanding and engagement. The MBSR curriculum topics are shown in Table 3.1.

The final MBSR programme was then rolled out over an 8-week period during the October to November school term in 2022. Each week, before the session was rolled out, Be-Khudi met with the MTs to provide support and to adapt lesson plans as needed. In addition, we were able to monitor implementation quality and fidelity by having a trained observer track attendance, take notes on what was working well each week, challenges, and suggestions for improving each week’s materials.

Table 3.1: MBSR Curriculum

Week 1	Fundamental of meditation & mindfulness
Week 2	Understanding attention span & focus
Week 3	Training attention span
Week 4	Observing thought patterns
Week 5	Handling difficult feelings
Week 6	Understanding self-compassion
Week 7	The power of gratitude
Week 8	Designing your own mindful practice

3.5 Randomization

Two hundred and fourteen students across grades 8, 9, and 10 in two low-cost schools in Faisalabad were randomized into an MBSR treatment and a control group (see Table 3.2)⁵. Randomization was carried out at the student-level, using the classroom as a strata. In addition, we blocked on a set of key baseline mental health outcome measures to ensure a balanced randomization. However, as mindfulness-based interventions can be advantageous for all students, the MBSR treatment was offered to all students irrespective of their possible predisposition to mental health difficulties.

To ensure that our control group served as an appropriate counterfactual, two points were necessary: first, students in the control group needed to be in a classroom setting to control for any possible peer effects by simply being in a classroom. Second, we needed to ensure that the students received a neutral experience (i.e., no direct support

⁵One is a girls-only school, and one is a boys-only school.

from teachers that could impact their mindfulness). To achieve these objectives, control group students were assigned to a silent Quran reading class. We coordinated with the schools to create an additional silent reading Quran class where the students sat at their desks for the hour engaged in their personal reading while students in the treatment group were engaged in the MBSR programme. This activity served as a neutral control for the study, considering that a class where students are instructed to read their Quran is a common practice in the context of Pakistan.

Table 3.2: Sample Size

	(1)	(2)	(3)
	Treatment (MBSR)	Control	Total
Grade 8	39	38	77
Grade 9	46	46	92
Grade 10	23	22	45
Total	108	106	214

3.6 Data and Balance Checks

3.6.1 Measurement

The MBSR programme was implemented between October and November 2022. We collected baseline data before implementation in October 2022 and endline data was collected in early December 2022. In this sub-section, we describe our main measures under our three categories of outcomes: mental health and well-being outcomes, behaviour change outcomes, and capacity to learn and educational success outcomes (see Table 3 for summary of our measurement). All measurement tools were piloted prior to the evaluation.

Mental health and well-being outcomes. We included the Strengths and Difficulties scale (SDQ), the Revised Child Anxiety and Depression Scale (RCADS), the Perceived Stress scale (PSS) and the Test Anxiety scale (STABS). These outcomes were measured at baseline and endline. The mental health and wellbeing tools included were used as screening tools, and are not intended to be interpreted as diagnostic. For some measures we use tools that have been psychometrically validated (SDQ, RCADS and GSE) and for the remaining measures we translate pre-existing psychometrically

validated English tools into Urdu (PSS, STABS). We follow the principles of cross-cultural back-translations (Brislin, 1970). Furthermore, we carefully assess the face validity and internal consistency of all the tools in the Pakistan context (see section 3.5.3).

The SDQ (Goodman, 2001) is a widely used validated self-reported questionnaire measuring emotional and behavioural difficulty. The scale has 25 questions covering 5 subscales: emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, and prosocial behaviour. The sum of responses to these questions is calculated for each subscale and the total score captures the total strength and difficulties score, "total difficulties". Higher scores for the subscales (except for the prosocial behaviour scale) reflects greater difficulties whereas a higher prosocial behaviour score indicates a strength. SDQ scores 14 and above are outside of the 'normal'/average scores for young people. In this study, we use the psychometrically validated Urdu tool approved by SDQ developer, Goodman (2020). This Urdu tool was also cross-culturally validated by Naveed et al., 2019.

The RCADS (Chorpita, Yim, Moffitt, Umemoto & Francis, 2000) is a self-reported 25-item questionnaire including 6 subscales measuring social phobia, panic disorder, major depression, separation anxiety, generalised anxiety, and obsessive-compulsive disorder. The raw scores are added to generate an overall depression score and an overall anxiety score. This raw score is then converted into a standardised "Test score" appropriate for the gender and grade level of the child (Chorpita, 2015). A test score above 65 is the borderline clinical to clinical cut off. A higher score indicates greater levels of anxiety and depression. We use the psychometrically validated Urdu tool approved by RCADS developers Chorpita et al., 2000. This Urdu translation has also been cross-culturally validated by Mehmood & Sultan (2014).

The PSS (Cohen, 1983) is a global measure of perceived stress. The shortened version (4 questions) used in this research study measures the degree to which respondents find situations stressful. The total perceived stress scores is calculated by summing all items (with 2 items reverse coded). A higher score indicates more perceived stress. Finally, STABS, the test anxiety scale (Suinn, 1969) was designed to measure the degree

of anxiety respondents experience before taking a test. Suinn (1969) describes test anxiety as a barrier to higher mental processes. For this study, 3 items were selected from the 20-item scale to capture some indications of test anxiety. The scores are summed, with a higher score indicating higher levels of test anxiety. For our measurement of perceived stress (PSS) and test anxiety (STABS), we translate the tools into Urdu following principles on cross-cultural back-translation (Brislin, 1970).

Behaviour change outcomes. To capture behaviour change outcomes, we included a set of questions to measure student behaviour change, time-use, and impulse control using task-based game (GoNoGo). These questions were only administered at endline. To capture student behaviour change, we asked students a range of questions to understand their behaviour at home. These questions were informed by early pilots and formative qualitative evaluations. The questions ask students to reflect on any observed changes (better/worse/no change) to their appetite, mood, arguments with siblings or parents, stress at home, and ability to focus on school work.

To capture student time-use allocation, we ask students to think about the time they typically allocate to activities such as care for others, domestic tasks, paid work, school, study time, play time and religious activities in one day. We used this to see if there has been a shift in the allocation of time spent studying as a result of the MBSR intervention. Finally, we include a task-based activity - the GoNoGo task (Gorden & Caramazza, 1982; Gorden, 1983) - that aims to measure impulse control. Students were presented with a tablet where squares appear on the screen for a short period of time (a few seconds). If the square is black ("no go"), they must not touch the screen, but if the square is any other colour, they must touch the screen as quickly as possible. A total of 72 trials were completed and the main outcome we calculate is the average response time⁶.

⁶Before playing the game, the facilitator explained the instructions after which each student played a practice round. For each game, students received a total of 47 Go and 25 no Go trials.

Capacity to learn and educational success outcomes. To measure student capacity to learn we measure growth mind set, self-efficacy, and student grit. To measure growth mindset, we include the 3-item growth mindset scale. Having a growth mindset is the belief that one can develop their skills through hard work and support (Dweck et al., 1995; Dweck, 1999). A total score is calculated by taking the average rating of the items (summing the items and dividing by the total number of items). We translate the psychometrically validated tool into Urdu, following principles on cross-cultural back-translation (Brislin, 1970).

To understand student perceived self-efficacy, we used the short form 6-item generalised self-efficacy scale (Schwarzer & Jerusalem, 1995). The generalised self-efficacy scale (GSE) is intended to capture the general sense of self-efficacy respondents feel, with the goal of predicting how one copes and adapts to daily challenges. The total score calculated by adding up all the items with a higher score indicating more perceived self-efficacy. We use the validated Urdu translation, adapted for Urdu-speaking populations (Tabassum et al., 2003).

Lastly, we included a task-based game to measure of grit among students. Grit has been defined as the "combination of passion and perseverance for long term goals" (Duckworth & Quinn, 2009). We use the Alan & Ertac Grit task (Alan, Boneva & Ertac, 2019) task where students are presented with multiple numbers in a grid on a tablet. The goal of the game is to find all the pairs of numbers that add up to 100. Students can choose between an easy game and a difficult game. The outcome of interest is the probability of choosing the difficult game in all five rounds of the game.

To capture educational success, school-based test scores were collected at baseline and endline. These test results were from exams set at the school level. We collect test results for English, Mathematics, Science and Urdu. Urdu test results were only available at endline.

Table 3.3: Summary of variables and measurement methodology

Instrument	Construct	Mode	Timing
Panel A: Mental well-being outcomes			
Strengths and Difficulties Questionnaire (SDQ)	Total emotional and behavioural difficulties and Prosociality	Self-reported	Baseline & Endline
Revised Child Anxiety and Depression Scale (RCADS)	Anxiety & Depression	Self-reported	Baseline & Endline
Perceived Stress Scale (PSS)	Stress	Self-reported	Baseline & Endline
The Suinn Test Anxiety Behaviour Scale (STABS)	Test anxiety	Self-reported	Baseline & Endline
Panel B: Behaviour change outcomes			
Behaviour change questions	Behaviour change	Self-reported	Endline only
Time- use questions	Allocation of time	Self-reported	Endline only
GoNoGo Task	Impulse/self- control	Task-based	Endline only
Panel C: Educational success and capacity to learn outcomes			
Test scores (% scored)	Learning	Administrative data	Baseline & Endline
Growth Mindset Scale	Growth mindset	Self-reported	Baseline & Endline
The Generalised Self-Efficacy Scale	Self-efficacy / self-belief	Self-reported	Baseline & Endline
Alan & Ertac Grit Task "Addition Game"	Grit	Task-based	Endline only

Notes: Behaviour change questions include a set of questions that inquire about arguments with parents and siblings at home, studying at home, mood at home, and stress at home. Test scores includes English, Maths, Science and Urdu administrative data.

3.6.2 Summary Statistics at Baseline

Table 3.4 shows that the study sample is evenly split between male and female students, each comprising 50% of the sample with the mean age of 15 years. At baseline, we observe a total difficulty mean score of 20.47 which in the four-fold classification based off a UK community sample falls within the "very high" range of total difficulties. These difficulties capture a range of emotional and behavioural difficulties described in the aforementioned measurement section. The mean anxiety test score in our sample is 61.29, mean depression score is 51.84, and a total anxiety and depression score at 59.21. These scores fall in the low severity threshold which is recorded a score between 0-64 (Chorpita & Spence, 1998).

For the perceived stress scale (PSS), we record a median score of 7, with a score of 9 at the 75th percentile. On the test anxiety scale (STABS), we record a median score of 10 and a score of 12 at the 75th percentile. For both scales, a higher score indicates higher levels of stress or more test anxiety. On the growth mindset and generalised

self-efficacy scales we record baseline mean scores of 2 and 16 respectively. For both scales, a higher score reflects a growth mindset or more self-efficacy.

At endline, we gathered some fixed characteristics of the caregivers in the sample. Caregivers were approximately 41 years of age, with 76% having a secondary school education and above, and 63% of them identifying as stay at home carers. The average monthly household income in the sample is 42,541 PKR (approximately 150 USD).

Table 3.4: Sample Characteristics at Baseline

	Obs	Mean	sd	25p	50p	75p
Panel A: Student characteristics						
1[Male]	214	0.50				
1[Age]	214	14.47	1.15	14	15	15
Panel B: Mental well-being						
Prosociality	214	3.47	1.56	2	4	5
Total SDQ	214	20.47	5.70	16	20	25
Anxiety (RCADS)	214	61.29	11.63	52.84	59.72	68.90
Depression(RCADS)	214	51.84	9.46	44.02	51.20	56.73
Total Anxiety and Depression	214	59.21	11.07	50.94	57.59	66.18
Perceived Stress (PSS)	214	7.52	2.63	6	7	9
Test anxiety (STABS)	214	10	2.72	8	10	12
Panel C: Capacity to learn and educational success						
Growth mindset	214	2.43	0.73	2	2	3
Generalised self-efficacy	214	15.70	3.29	13	16	18
English (%)	158	74.70	23	74	74	89
Math (%)	158	76.16	17.71	74.34	74.34	90
Science (%)	158	76.05	18.04	76	76	90

Notes: We do not have baseline test scores for the full sample due to gaps in the administrative data.

3.6.3 Data Validity

We assess the validity and reliability of the range of measures used in this research study. To assess validity, we specifically discuss the face validity of our instruments. Face validity is a psychometric property that confirms that respondents correctly perceive what the instrument intends to measure (Danon et al., 2022). To ensure face validity, prior to the evaluation, we piloted the tool with a small sample of students to confirm that students understood questions as intended. Where there were misunderstandings

(i.e., with translation), we updated the language to ensure full comprehension.

We assess internal reliability using Cronbach's alpha and McDonalds Omega. Internal reliability refers to the extent to which all items in the measure contribute in a consistent way to the data obtained (Hunsley & Mash, 2008). Cronbach's alpha is a test of internal consistency, providing an assessment of reliability of the scales. The Cronbach's alpha statistic ranges between 0 and 1. Following the guidance set out by Hunsley & Mash (2005), α values between 0.70-0.79 indicates adequate internal consistency; α between 0.80-0.89 indicates good internal consistency, and $\alpha > 0.90$ may indicate excellent consistency. It should be noted that while $\alpha > 0.90$ may indicate excellent internal validity, it is possible for alpha values to be artificially high which could also be an indication of item redundancy (Streiner, 2003; Hunsley & Mash, 2005).

However, Streiner (2003) cautions against the blind use of the coefficient α in dismissing scales. As such, we include McDonald's Omega statistic -another measure of composite reliability. The Omega coefficient has been shown to be a more precise estimate of reliability as it accounts for measurement error (Ravinder & Saraswathi, 2020; Dunn et al., 2013). We take a measured approach by including both alpha and omega coefficients for comparison and discussion.

Table 3.5 shows most of our instruments have moderate to acceptable levels of internal consistency. Specifically, items in the mental health outcome scales (SDQ and RCADS) and behaviour and capacity to learn outcomes (behaviour change questions and generalised self-efficacy) have adequate levels of internal consistency (Hunsley & Mash, 2005). Other psychological outcomes such as PSS shows weaker internal consistency with α ranging from 0.22 - 0.42; and omega ranging from 0.39 - 0.42. The growth mindset scale similarly shows lower reliability with α and Ω coefficients ranging from 0.40 - 0.52.

We keep all the study instruments for our main results. However, we are cautious not to over interpret results from the perceived stress scale (PSS) as it scores the lowest on both reliability tests.

Table 3.5: Data Validity

Instrument	Alpha baseline	Alpha endline	Omega baseline	Omega endline	# of items	Baseline students	Endline students
SDQ	0.67	0.64	0.62	0.59	25	211 - 214	209
RCADS	0.83	0.86	0.84	0.86	25	214	209
PSS	0.22	0.42	0.39	0.42	4	214	209
STABS	0.57	0.70	0.60	0.70	3	214	209
Behaviour change Qs	0.60		0.60		6		208
Growth mindset	0.40	0.50	0.43	0.52	3	212 - 214	209
Generalised Self-efficacy	0.67	0.77	0.68	0.78	6	214	209

Notes: SDQ (total emotional and behavioural difficulties); RCADS (total anxiety and depression); PSS (perceived stress); STABS (test anxiety); GSE (generalised self-efficacy).

3.6.4 Balance Checks

In this sub-section, we confirm balance between our treatment and control groups on relevant variables and outcomes of interest. Table 6 shows that there are no significant differences across treatment and control groups across our measures of mental wellbeing, capacity to learn, and educational success. We also show a joint F-stat of 0.41 which confirms that there are no significant differences between the treatment and control groups.

Table 3.6: Balance Checks

Variable	(1) Control	(2) Treatment	(3) Mean difference	(4) N
Prosocial subscale (SDQ)	0.106 (0.933)	-0.104 (1.056)	-0.209 (0.126)	214
Total Difficulties Score (SDQ)	0.050 (1.013)	-0.049 (0.989)	-0.100 (0.466)	214
Total Depression (RCADS)	-0.008 (1.020)	0.008 (0.985)	0.015 (0.911)	214
Total Anxiety (RCADS)	0.052 (1.009)	-0.051 (0.993)	-0.103 (0.453)	214
Total Depression and Anxiety (RCADS)	0.031 (1.022)	-0.031 (0.981)	-0.062 (0.651)	214
Perceived Stress Score (PSS)	0.022 (1.112)	-0.021 (0.881)	-0.043 (0.755)	214
Test Anxiety (STABS)	-0.052 (1.063)	0.051 (0.936)	0.103 (0.452)	214
Growth mindset	-0.040 (0.973)	0.039 (1.029)	0.078 (0.567)	214
Generalised self-efficay	-0.025 (1.025)	0.022 (0.978)	0.047 (0.731)	214
English test score	-0.081 (1.105)	0.081 (0.882)	0.161 (0.292)	172
Math test score	0.186 (1.227)	0.221 (1.005)	0.036 (0.835)	172
Science test score	0.000 (1.035)	-0.000 (0.970)	-0.001 (0.995)	172
Joint F	0.41			

Notes: P-value difference is presented in parenthesis in column 3. The joint F is conducted for all variables apart from test scores to be able to include the full sample of 214.

3.7 Results

To estimate treatment effects, we use the following specification across our three main categories of outcomes.

$$y_{is} = \alpha + \gamma \underline{y}_{is} + \beta_1 T_{is} + \nu_s + \epsilon_{is} \quad (3.1)$$

where y_{is} is the dependent variable of interest for student i in school s . $\underline{y}_{is}$ is the baseline measure of y_{is} for the ANCOVA estimation. T_{is} is the treatment indicator that takes the value 1 if student i of school s is assigned to the MBSR treatment, and 0 other-

wise. ν_s are school fixed effects, and standard errors are clustered at the session level at which treatment is administered. β_1 is the coefficient of interest which shows the relationship between our outcome of interest and the treatment. All outcomes of interest are standardised by the mean and standard deviation of the control group⁷.

3.7.1 Treatment Effects on Mental Health Outcomes

Table 3.7 shows treatment effects of MBSR on our mental health outcomes based on specification 1. We find that MBSR significantly improves prosociality by 0.30 standard deviations (see column 1). However, on the remaining mental health outcomes, we observe no significant treatment effects on total difficulties (SDQ), depression and anxiety (RCADS), perceived stress (PSS) or test anxiety (STABS). To address challenges of multiple hypothesis testing, we compute sharpened False Discovery Rate (FDR) adjusted q-values (as suggested by Anderson 2008) across our different measures of mental health outcomes. We find that the treatment effects on prosociality remain significant at the 10% level.

Table 3.7: MBSR Treatment Effects on Mental Health Outcomes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Prosocial	SDQ	Anxiety	Depression	Total RCADs	PSS	STABS
Treatment	0.303** (0.102)	0.163 (0.124)	-0.103 (0.118)	0.028 (0.117)	-0.096 (0.159)	-0.090 (0.165)	0.123 (0.156)
Observations	199	199	199	199	199	199	199
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Baseline Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: All regressions are an ANCOVA estimation. Each column represents a different mental well-being dependent variable as indicated by the column header. The dependent variable in all columns is standardized against the control group. Standard errors are in parentheses, clustered at class level. Controls include baseline mental health and psychological outcomes. Adjusted q-values are as follows: 0.091, 0.695, 0.695, 0.814, 0.695, 0.695, 0.695. Estimates are significant at the *10%, **5%, and ***1% level.

3.7.2 Treatment Effects on Behaviour Change

Next, we observe treatment effects on our behaviour change outcomes, including student self-reported behaviour at home, time-use, and impulse control. We use the same specification as above, but with the behavioural outcomes of interest as the dependent variable.

⁷Note that we do not have baseline measures for our behaviour change outcomes and task-based measures. For these measures, we estimate treatment effects using endline measures only.

Table 3.8 presents results on the set of self-reported behaviour change questions. We show that in comparison to the control group, students who received the MBSR treatment report a 0.52 standard deviations improvement in mood (column 2). In addition, we also find that students who received the MBSR treatment report 0.34 standard deviations less stress at home (column 6). The coefficients on student appetite, arguments with parents and siblings and ability to focus on school work are also positive but insignificant.

Given all these questions were administered on the same scale, we compute a behaviour change index, using the methods as in Anderson (2008). We find that students who received the MBSR programme show an 0.49 standard deviations improvement (significant at the 1% level) in reported behaviour in comparison to the control group (column 7)⁸. Overall, this indicates that students who received the MBSR programme were able to foster a more positive home environment; relative to students not exposed to MBSR.

Table 3.8: MBSR Treatment Effects on Behaviour Change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Improved appetite	Improved mood	Less arguments w/ siblings	Less arguments parents	Focus on school work	Less stress at home	Combined Index
Treatment	0.185 (0.125)	0.515** (0.168)	0.265 (0.161)	0.335 (0.195)	0.156 (0.089)	0.336*** (0.106)	0.492*** (0.132)
Observations	208	208	208	208	208	208	208
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column represents a different behaviour change dependent variable as indicated by the column header. All outcomes are standardised against the mean and standard deviation of the control group. Standard errors are in parentheses, clustered at class level. The combined index is computed based on the individual behaviour indicators using the method as in Anderson(2008). Adjusted q-valued are as follows: 0.168, 0.033, 0.154, 0.154, 0.154, 0.033. Estimates are significant at the *10%, **5%, and ***1% level.

Table 3.9 presents results on our remaining behavioural outcomes: time-use and impulse control (measured through task-based GoNoGo game). We see non-significant differences on our measures of time-use allocation, including the proportion of time spent in care for others and domestic activity, religious activity, study-time, playtime/leisure. Addition-

⁸While observing treatment effects on the combined index should address concerns around multiple hypothesis testing (Anderson, 2008), as an additional check we also compute sharpened FDR adjusted q-values for each of six individual measures. Our treatment effects on improved mood and less stress at home remain significant at the 5% level.

ally, our task-based activity measuring impulse control reveals no significant differences between MBSR students and students in the control group.

Table 3.9: MBSR Treatment Effects on Time-use and Impulse Control

	Time-Use				GoNoGo
	Care for others & domestic activity	Religious activity	Studying	Playtime/Leisure	Response time
MBSR	0.005 (0.010)	0.002 (0.008)	0.005 (0.008)	-0.005 (0.004)	-0.003 (0.108)
Observations	200	200	200	200	191
School FE	Yes	Yes	Yes	Yes	Yes
Control mean	0.097	0.067	0.137	0.084	0

Notes: Each column represents a different dependent variable as indicated by the column header. The response time variable captures impulse control which is measured as the response time in the GoNoGo game. Standard errors are in parentheses, clustered at class level. Estimates are significant at the *10%, **5%, and ***1% level.

3.7.3 Treatment Effects on Capacity to Learn and Educational Success Outcomes

Next, we observe treatment effects on capacity to learn and educational success outcomes. Table 10 presents results on our three different measures that capture student capacity to learn - growth mindset, generalised self-efficacy, and grit. We find that MBSR significantly increases the grit of the treated students by 0.34 standard deviations relative to the students in the control group (column 3). The coefficient on generalised self-efficacy (GSE) is also positive but insignificant, whereas the coefficient on growth mindset is close to 0. In column 4, we combine our three measures of capacity to learn outcomes into a combined index of capacity to learn (following Anderson, 2008). We find that students who received the MBSR treatment show a 0.27 standard deviations higher capacity to learn relative to the control group⁹.

⁹Note that as in the case of Table 3.8, observing effects on a combined index should address concerns around multiple hypothesis testing. However, we also compute FDR adjusted q-values and find that and find the treatment effect on grit is only marginally significant (p-value of 0.13).

Table 3.10: MBSR Treatment Effects on Capacity to Learn

	(1)	(2)	(3)	(4)
	Growth mindset	Self-efficacy	Grit	Combined Index
Treatment	0.005 (0.144)	0.194 (0.145)	0.343** (0.155)	0.274* (0.125)
Observations	199	199	191	191
School FE	Yes	Yes	Yes	Yes

Notes: Each column represents a different dependent variable as indicated by the column header. The first two estimations are an ANCOVA estimation. In the final estimation, we did not have baseline measure of grit. Standard errors are in parentheses, clustered at class level. The combined index combines our different capacity to learn measures using the method as in Anderson(2008). Adjusted q-values are as follows: 0.97,0.31,0.15. Estimates are significant at the *10%, **5%, and ***1% level.

Finally, we look at treatment effects on education outcomes in Table 11. We measure the impact of MBSR on English, mathematics, science, and Urdu school-based test scores. Although we observe positive coefficients on English and Urdu, overall, we do not observe any significant effects on any of the subjects between groups.

Table 3.11: MBSR Treatment Effects on Academic Outcomes

	(1)	(2)	(3)	(4)
	English	Math	Science	Urdu
Treatment	0.232 (0.191)	0.071 (0.164)	-0.017 (0.286)	0.149 (0.200)
Observations	182	182	182	182
School FE	Yes	Yes	Yes	Yes

Notes: Standard errors are in parentheses, clustered at class level. Estimates are significant at the *10%, **5%, and ***1% level..

3.8 Treatment Heterogeneity, by Sex

Finally, we use the following specification to study the heterogenous effects of MBSR on our main mental health, behaviour change outcomes, capacity to learn and educational success outcomes, by sex.

$$y_{is} = \alpha + \gamma \underline{y}_{is} + \beta_1 T_{is} + \beta_2 F_i + \beta_3 (T_i * F_i) + \epsilon_{is} \quad (3.2)$$

where y_{is} is the dependent variable of interest for student i in school s . $\underline{y}_{is}$ is the baseline measure of y_{is} for the ANCOVA estimation. T_i is a dummy variable which is equal to 1 if a student is exposed to MBSR and 0 otherwise, and F_i is a dummy variable which is equal to 1 if the student from the girls school, and 0 otherwise. The interaction coefficient β_3 is our outcome of interest which shows the differential treatment effects between girls and boys. The main outcomes are standardised by the mean and standard deviation of the control group.

Table 3.12: Heterogenous Treatment Effects on Mental Health, by sex

	Prosociality	Total Difficulties	Anxiety & Depression	Perceived Stress	Test Anxiety
MBSR	0.252 (0.191)	0.226 (0.167)	0.216 (0.149)	0.017 (0.147)	0.292** (0.098)
Girl	0.110 (0.070)	0.248 (0.161)	0.321* (0.173)	0.147 (0.231)	0.544** (0.193)
MBSR x Girl	0.100 (0.203)	-0.124 (0.239)	-0.620** (0.269)	-0.213 (0.326)	-0.337 (0.303)
Observations	199	199	199	199	199

Notes: each column represents a different mental health dependant variable. Standard errors are in parentheses, clustered at the class level. Estimates are significant at the *10%, **5%, and ***1% level.

Table 3.12 shows the heterogenous treatment effects on our main mental health outcomes. We observe that in comparison to boys, girls who received the MBSR programme experienced 0.62 standard deviations lower anxiety and depressive symptoms. On the remaining outcomes (prosociality, total difficulties (SDQ), perceived stress and test anxiety), we find no evidence of differential treatment effects between girls and boys.

Table 3.13: Heterogenous Treatment Effects on Behaviour Change Outcomes, by sex

	Appetite	Mood	Less arguments w/siblings	Less arguments w/parents	Focus on school	Stress at home
MBSR	0.073 (0.162)	0.366 (0.234)	0.130 (0.252)	0.388* (0.207)	0.105 (0.133)	0.195 (0.163)
Girl	-0.029 (0.177)	0.163 (0.248)	0.055 (0.238)	-0.103 (0.316)	0.154** (0.066)	0.090 (0.191)
MBSR x Girl	0.221 (0.234)	0.297 (0.332)	0.268 (0.314)	-0.104 (0.387)	0.101 (0.180)	0.278 (0.199)
Observations	208	208	208	208	208	208

Notes: each column represents a behaviour change dependant variable. Standard errors are in parentheses, clustered at the class level Estimates are significant at the *10%, **5%, and ***1% level.

For our set of behaviour change outcomes (see Table 3.13) and capacity to learn outcomes (see Table 3.14), we observe no evidence of differential treatment effects between boys and girls.

Table 3.14: Heterogenous Treatment Effects on Capacity to Learn Outcomes, by sex

	Growth mindset	Self-efficacy	Grit
MBSR	-0.209 (0.162)	0.195 (0.179)	0.469* (0.233)
Girl	-0.254 (0.164)	-0.387 (0.242)	-0.590** (0.265)
MBSR x Girl	0.366 (0.255)	0.095 (0.346)	-0.251 (0.301)
Observations	199	199	192

Notes: each column represents a capacity to learn dependant variable. Standard errors are in parentheses, clustered at the class level Estimates are significant at the *10%, **5%, and ***1% level.

3.9 Discussion and Concluding Remarks

This study combined a battery of research instruments, including self-reported mental health tools, objective task-based measures, and a set of behaviour change questions informed by the context and intervention. We grouped our main outcomes into (1) mental health outcomes, (2) behaviour change outcomes and (3) educational success and capacity to learn outcomes. On mental health outcomes, the most notable result was the impact of MBSR increasing students' prosocial behaviour by 0.30 standard deviations (significant at the 5% level).

Through the curriculum adaptation process, pilots and qualitative interviews, students and teachers repeatedly reported improved mood, less arguments with family, increased appetite and focus on school work and less stress at home. As such, we rigorously test this by including questions on these reported behaviour changes. We observe that MBSR significantly overall student behaviour by 0.49 standard deviations (see combined index coefficient) with specific improvements in mood, and stress at home. On the remaining behaviour change outcomes, time-use, and impulse control, we did not observe any significant treatment effects.

On educational success and capacity to learn outcomes, the most notable result was on the objective task-based grit activity. We detected a significant 0.34 standard deviation increase in the grit of MBSR students, relative to control group students. This is meaningful as grit (combination of passion and perseverance for long term goals) has been shown to be correlated to educational attainment (Danon et al., 2022). Although we do not directly observe any effects on most of our school-based test scores in this study, this may be as a result of the time-span of the study (approximately 3 months), which may not be sufficient to drive learning gains.

Finally, we also explored treatment heterogeneity by sex. The results show that MBSR is more impactful for girls in comparison to boys in reducing anxiety and depressive symptoms. However, we do not observe any differential treatment effects between girls and boys for the other outcomes.

Nonetheless, this study is not without some limitations. Firstly, the study had a small sample size and it is possible that it was not well-powered to detect treatment effects across our full range of outcomes. Second, the time-span of the study was only 3 months. However, mindfulness is a practice with benefits growing over time. As such, it is feasible that over time, stronger effects may be detected. Future iterations of a scaled-up programme could take these limitations into consideration by administering the programme to a larger sample of students and over a longer time-span. In addition, they could also strengthen certain components of the programme that focus on behaviour change and student grit which appear to be the strength of the programme. Future rigorous evaluations of such scaled-up programmes would generate important findings on the impact of MBSR programmes across a range of outcomes. Finally, some of the observed null effects could be explained through positive spillovers. In this case, it is possible that students in the MBSR group shared content and activities with their friends in the control group, and this may have positively biased the effects in the control group.

To conclude, we set out to evaluate an adapted MBSR curriculum in two low-cost private schools, utilising a range of mental well-being, behaviour change, and capacity to learn outcomes scarcely tested and adapted for use in low-resource settings. We successfully adapt and implement an Urdu MBSR programme in schools. However, this was a small study with limited funding. Future studies would utilise a larger sample size to be sufficiently powered for subsequent mediation analysis. The lessons generated can be used to inform mindfulness-based programme design and evaluation in low-resource settings.

3.10 References

- Alan,S., Boneva,T., Ertac,S., 2019. Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics*, 134(3), 1121-1162.
- Alvi,M.H., Ashraf,T., Kiran,T., Iqbal,N., Gumber, A., Patel,A., & Husain,N. (2023). Economic burden of mental illness in Pakistan: an estimation for the year 2020 from existing evidence. *BJPsych International*, 1-3.<https://doi.org/10.1192/bji.2023.4>
- Ansari,I.(2015). Mental health Pakistan: optimizing brains. *Int J Emerg Ment Health*, 17, 288.
- Bajaj,B., Robins,R.W., & Pande,N. (2016). Mediating role of self-esteem on the relationship between mindfulness, anxiety, and depression. *Personality and Individual Differences*, 96, 127-131.
- Bergen-Cico, Dessa, Possemato, Kyle, & Cheon, Sanghyeon.(2013). Examining the Efficacy of a Brief Mindfulness-Based Stress Reduction (Brief MBSR) Program on Psychological Health. *Journal of American College Health*, 61(6), 348-360.
- Baer,R.A., Smith,G. T., Hopkins,J., Krietemeyer,J., & Toney,L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, 13(1), 27-45. doi: /10.1177/1073191105283504
- Baer,R.A., Smith,G.T., Lykins,E., Button,D., Krietemeyer,J., Sauer,S., Walsh,E., Duggan,D. & Williams,J.M.G. (2008). Construct validity of the Five Facet Mindfulness Questionnaire in meditating and nonmeditating samples. *Assessment*, 15, 329-342. doi: 10.1177/1073191107313003
- Bear,H., Krause,A., Karolin R., Wolpert, M., Edbrooke-Childs, J., & Norton,S. (2019). Systematic review and meta-analysis: Outcomes of routine specialist mental health care for young people with depression and/or anxiety. *Journal of the American Academy of Child and Adolescent Psychiatry*, 59, 810-841.
- Biegel,G.M., Brown,W.K., Shapiro,S.L, & Schubert,C. M. (2009). Mindfulness-Based Stress Reduction for the Treatment of Adolescent Psychiatric Outpatients. *Journal*

-
- of Consulting and Clinical Psychology, 77(5), 855-866.
- Bluth, Karen, Campo, Rebecca A, Pruteanu-Malinici, Sarah, Reams, Amanda, Mullarkey, Michael, & Broderick, Patricia C. (2016). A School-Based Mindfulness Pilot Study for Ethnically Diverse At-Risk Adolescents. *Mindfulness*, 7(1), 90-104.
- Calear, A.L., & Christensen, H. (2009). Systematic review of school-based prevention and early intervention programs for depression. *Journal of Adolescence* (London, England.), 33(3), 429-438.
- Chorpita, B.F., Yim, L.M., Moffitt, C.E., Umemoto L. A., & Francis, S.E. (2000). Assessment of symptoms of DSM-IV anxiety and depression in children: A Revised Child Anxiety and Depression Scale. *Behaviour Research and Therapy*, 38, 835-855.
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior*, 24, 385-396. [Link to full-text \(pdf\)](#) (2.)
- Cohen, S., & Williamson, G. (1988).
- Cohen, S., & Williamson, G. (1988). Perceived stress in a probability sample of the U.S. In S. Spacapan & S. Oskamp (Eds.), *The social psychology of health: Claremont Symposium on Applied Social Psychology*. Newbury Park, CA: Sage.
- Danon, A., Das, J., de Barros, A., Filmer, D., Cognitive and socioemotional skills in low-income countries: Measurement and associations with schooling and earnings 1, *Journal of Development Economics* (2023), doi: <https://doi.org/10.1016/j.jdeveco.2023.103132>.
- Diaz-Gonzalez, M.C, Perez Duenas, C., Sanchez-Raya, A., Moriana E., Juan A., & SÃ¡nchez VÃ¡zquez, V. (2018). Mindfulness-based stress reduction in adolescents with mental disorders: A randomised clinical trial. *Psicothema*, 30(2), 165-170.
- Grossman, Paul, Niemann, Ludger, Schmidt, Stefan, & Walach, Harald. (2004). Mindfulness-based stress reduction and health benefits: A meta-analysis. *Journal of Psychosomatic Research*, 57(1), 35-43.
- Gordon B, Caramazza A. Lexical decision for open- and closed-class words: Failure to replicate differential frequency sensitivity. *Brain and Language*. 1982;15:143-160.
- Gordon B. Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Jour-*
-

-
- nal of Verbal Learning and Verbal Behavior. 1983;22:24-44.
- Gouda,S., Luong, M.T., Schmidt, S., & Bauer,J. (2016). Students and Teachers Benefit from Mindfulness-Based Stress Reduction in a School-Embedded Pilot Study. *Frontiers in Psychology*, 7, 590.
- Heckman,J.J., Humphries,J.E. and Mader,N.S. (2011). The GED. In E. Hanushek, S. Machin and L. Woessmann, eds., *Handbook of the Economics of Education*, 3, 423-484.
- Hjeltnes, Aslak, Molde, Helge, Schanche, Elisabeth, Vollestad, Jon, Lillebostad Svendsen, Julie, Moltu, Christian, & Binder, Per-Einar. (2017). An open trial of mindfulness-based stress reduction for young adults with social anxiety disorder. *Scandinavian Journal of Psychology*, 58(1), 80-90.
- Hunsley, J., & Mash, E. J. (Eds.). (2008). *A guide to assessments that work*. Oxford University Press.
- Johnson, Catherine, Burke, Christine, Brinkman, Sally, & Wade, Tracey. (2016). Effectiveness of a school-based mindfulness program for transdiagnostic prevention in young adolescents. *Behaviour Research and Therapy*, 81, 1-11.
- Kabat-Zinn, Jon. (1982). An outpatient program in behavioural medicine for chronic pain patients based on the practice of mindfulness meditation: Theoretical considerations and preliminary results. *General Hospital Psychiatry*, 4(1), 33-47.
- Khan, F., Shehzad, R. K., & Chaudhry, H. R. (2008). Child and adolescent mental health services in Pakistan: current situation, future directions and possible solutions. *International Psychiatry*, 5(4), 86-88. <https://doi.org/10.1192/s1749367600002253>
- Khalid, A, Qadir, F, Chan, W & Schwannauer, M 2018, 'Adolescents' mental health and well-being in developing countries: A cross-sectional survey from Pakistan', *Journal of Mental Health*. <https://doi.org/10.1080/09638237.2018.1521919>
- Kuyken, Willem, Weare, Katherine, Ukoumunne, Obioha C, Vicary, Rachael, Motton, Nicola, Burnett, Richard, Huppert, Felicia. (2013). Effectiveness of the Mindfulness in Schools Programme: Non-randomised controlled feasibility study. *British Journal of Psychiatry*, 203(2), 126-131.
- Mehmood, T., & Sultan, S. (2014). Translation and Adaptation of Revised Children's
-

Anxiety and Depression Scale.

- Naveed, S., A. Waqas, A. R. Memon, M. Jabeen and M. H. Sheikh (2019). "Cross-cultural validation of the Urdu translation of the Patient Health Questionnaire for Adolescents among children and adolescents at a Pakistani school." *Public Health* 168: 59-66.DOI: 10.1016/j.puhe.2018.11.022
- Streiner D.L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of personality assessment*. 80 (3) 217-222. [PubMed: 12763696]
- Tabassum, U., Rehman, G., Schwarzer, R., & Jerusalem, M. (2003). Adaptation of the General Self-Efficacy Scale for Urdu-speaking populations. *Journal of Psychosomatic Research*, 54(2), 131-134.
- United Nations Population Fund, (2023). 8 Billion Lives, Infinite Possibilities: the case for rights and choices. State of World Population report 2023. www.unfpa.org/swp2023/YouGovData
- Virgili, Mario. (2015). Mindfulness-Based Interventions Reduce Psychological Distress in Working Adults: A Meta-Analysis of Intervention Studies. *Mindfulness*, 6(2), 326-337.
- Woodward, Lianne J, & Fergusson, David M. (2001). Life Course Outcomes of Young People With Anxiety Disorders in Adolescence. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40(9), 1086-1093.
- World Health Organisation, (2020). Mental Health Atlas.
<https://cdn.who.int/media/docs/default-source/mental-health/mental-health-atlas-2020-country-profiles/pak.pdf?sfvrsn=62378896-6&download=true>
- Zhang, L., Lopes, S., Lavelle, T., Jones, K. O., Chen, L., Jindal, M., Zinzow, H., & Shi, L. (2022). Economic Evaluations of Mindfulness-Based Interventions: a Systematic Review. In *Mindfulness* (Vol. 13, Issue 10). Springer. <https://doi.org/10.1007/s12671-022-01960-1>
- Zhou, Xiang, Guo, Jieyu, Lu, Guangli, Chen, Chaoran, Xie, Zhenxing, Liu, Jiangmin, & Zhang, Chuning. (2020). Effects of mindfulness-based stress reduction on anxi-

ety symptoms in young people: A systematic review and meta-analysis. *Psychiatry Research*, 289, Psychiatry research. Volume 289

4 | English Medium Instruction in Multilingual Contexts: Empirical Evidence from Ethiopia

Abstract

Language acquisition and learning literatures favour mother-tongue education policies, particularly in the early years of schooling. However, English Medium Instruction (EMI) is a popular language policy position in multilingual contexts. This paper studies EMI transition policies using a longitudinal dataset from Ethiopia, by leveraging regional variation in education language policy reforms for causal identification. Employing a dynamic value-added model, this paper shows that students in schools using English as a medium of instruction have lower mathematics test scores (-0.25 standard deviations) compared to students in mother tongue education schools. Furthermore, English medium learners do not perform any better in their English test scores compared to mother tongue learners. These findings are in line with the international education literature on skill development, learning and second language acquisition. Treatment heterogeneity suggests some positive long-term effects of an early grade transition from mother tongue to EMI compared to a later primary grade transition to EMI, relative to mother tongue-only instruction. However, the overall results support the prolonged utilisation of mother tongue instruction in primary education. These results are particularly relevant for policy makers in linguistically diverse contexts.

4.1 Introduction

In multilingual countries where English is not the first spoken language, English Medium Instruction (EMI) is increasingly prevalent (Macaro, Curle, Pun, An, & Dearden, 2018). However, EMI is a highly contested education language policy position. The body of evidence on language education favours mother-tongue education over English medium instruction in primary education. Mother-tongue instruction has been shown to lead to improved cognitive development (Cummins, 2000; Benson, 2000), positive test scores in education outcomes (Vujcich, 2013; Eriksson, 2014; Seid, 2016; Taylor & von Fintel, 2016; Seid, 2019) and improved downstream human capital formation indicators such as adult literacy, and completed years of schooling (Ramachandran, 2017). However, education language policies in Africa, specifically in countries with English as a national language, trend towards an early transition from mother tongue¹ or proximate language instruction to English medium instruction (Clegg & Simpson, 2016).

One of the common reasons cited for the push to EMI in early primary grades stems from parents and teachers who believe that academic attainment is directly related to the time of exposure to English medium instruction (Brock-Utne & Holmarsdottir, 2004; Piper, Zuilkowski, & Ong'Ele, 2016). These beliefs are often championed by politicians and policy makers (Clegg & Simpson, 2016), but result in education language policies that may not serve the very beneficiaries they were designed to support (Cummins, 2000; Hakuta, Butler, & Witt, 2000; Taylor & von Fintel, 2016). Even in areas with high linguistic diversity, a second language (L2) such as the use of Amharic in Addis Ababa for example, may function as a lingua franca to enable communication between communities. The use of a local bridging language then becomes logistical necessity, with communities advocating for the use of international languages, such as English in schooling despite its limited use in civil society (Heugh, Leader, & Benson, 2007). Against this backdrop, this paper investigates the impact of education language policies on children's learning outcomes in Ethiopian primary schools.

Ethiopia, a multilingual country with 91 registered linguistic groups and at least

¹Mother tongue refers to the first spoken language (L1), usually a child's first acquired language(s)

80 spoken languages, is a prime example of the challenges surrounding language of instruction (Heugh et al, 2007; Rossiter, Azubuike, & Rolleston, 2017). Using data collected through the Young Lives² longitudinal study, this paper employs a dynamic value-added model of student attainment to estimate the impact of English medium instruction among Grade 7 and 8 students. A 1994 Education and Training Policy³ gave regions the autonomy to decide when (i.e., grade level) to transition to using English as a medium of instruction with a mandatory transition at secondary school in Grade 9 (Woldemariam, 2007; Vujcich, 2013). The reform policy created a natural experiment with plausibly exogenous variation between states; as some regions maintained their mother tongue instruction policy while other regions transitioned to EMI at various grade levels in primary school. It is through this variation that the causal effect is identified. Using data collected through the Young Lives³ longitudinal study, this paper employs a dynamic value-added model of student attainment to estimate the impact of English medium instruction among Grade 7 and 8 students.

Findings reveal language of instruction as a key input in education outcomes. The preferred lagged value-added specification in the main analysis compares Grade 7 and 8 mother tongue instruction students to students who transition to English Medium Instruction (EMI) either in Grade 1, Grade 5, or Grade 7. These results show that schools exiting from mother tongue instruction exit policies in primary school perform worse in mathematics when compared with to schools maintaining mother tongue instruction policies. There are no significant overall effects on English test outcomes. The results in this study are robust to other specifications. The main results, particularly for mathematics outcomes, are in line with UNESCO's Global Education Monitoring report (2016) recommending at least six years of mother tongue education or, in the case of Ethiopia, at least 8 years of mother tongue teaching as recommended by Heugh et al., (2007).

In addition, there is suggestive evidence that the grade of EMI transition affects learning outcomes, with an early transition having positive effects on English outcomes

²Young Lives is a longitudinal study of childhood poverty conducted in Ethiopia, India, Peru, and Vietnam.

³See Education and Training Policy in the Appendix A.

and no significantly negative effects on mathematics outcomes relative to sustained mother tongue instruction. On the other hand, a later EMI transition compared with sustained mother tongue instruction, has a negative effect on mathematics outcomes with no meaningful effects on English outcomes. These heterogeneous treatment effects are interpreted conservatively as there are not many regions in the study sample implementing the early transition to EMI, so effects may be caused by regional differences rather than the language policy. However, this heterogeneity highlights an area for future research. Treatment heterogeneity results are in line with research studies that suggest that multilingual exposure can take place in the early developmental years as children are able to acquire both language skills simultaneously (Alidou, Boly, Brock-Utne, & Diallo, 2006).

Using the rich individual level dataset, the paper is also able to investigate other differential treatment effects as a result of the language policy. The findings show no significant differences in treatment effects between students with higher or lower baseline scores, suggesting that even stronger students struggle with the transition in language of instruction. Furthermore, effects are similar for boys and girls and there are no significant differential treatment effects for students with better resourced or supported teachers. This suggests that existing policy measures to mitigate potentially negative effects of EMI are not particularly effective.

This paper complements a small literature identifying the causal effects of changes in language policy in Africa. Other studies focus on immediate learning outcomes in early primary grades or, alternatively, on adult literacy levels and total years of education. In Cameroon, Laitin, Ramachandran and Walter (2019) find significant positive effects of local language instruction among Grade 1- 3 students on Mathematics (ranging from 0.17 standard deviations - 0.71 standard deviations.) and English outcomes (ranging from 0.68 standard deviations - 1.53 standard deviations). Similarly, Seid (2019) finds that learning first in mother-tongue in the early years (primary 1 to 4) relative to learning in a second language improves mathematics test scores by 0.16 standard deviations in Grade 5. Using the 2012/2013 Young Lives dataset, this evaluation shows that students who are first taught in their mother-tongue increase their Mathematics test scores by 0.16 standard deviations. Focusing on adults, Ramachandran (2017) uses the Demographic

and Health Survey (DHS) data to show that mother tongue instruction leads to a 40 per cent increase in adult literacy and more than half a year increase in completed years of schooling. Additionally, using the Ethiopian Labour Force Survey, Seid (2022) also finds that exposure to mother-tongue instruction in primary school years increases labour market outcomes.

Thus, this research contributes to the growing experimental and quasi-experimental literature on language of instruction and student outcomes in the economics of education, and general education literature. This paper builds on the previous language of instruction work in three ways. First, it leverages the rare opportunity to evaluate the language of instruction effect on upper primary Grade 7 and 8 students with a control group that have had 7-8 years of mother tongue instruction. To my knowledge, no studies have examined the effects of EMI on children's outcomes in the later years of primary school relative to teaching in mother tongue. This is particularly important for policy as most African countries teach in mother tongue in Grade 1-4 but very few teach in mother tongue up to Grade 6 or later as recommended by UNESCO (2016). Second, this paper uses a rich source of individual level data with robust and comprehensive outcome measures (test scores as opposed to grades attained or basic measures of literacy) and an array of background characteristics relevant for accounting for potential differences between the groups of interest. Finally, this paper studies both mathematics and English outcomes: other work except for Laitin et al. (2019) focuses on either adult literacy or mathematics outcomes.

The paper is structured as follows. Section 4.2 summarises the quantitative EMI literature, and Section 4.3 describes the data sampling, and assessment design used to capture the outcomes in the paper. Section 4.4 discusses the empirical framework and strategy with details on the preferred value-added specification. Sections 4.5 and 4.6 cover descriptive statistics, the main results, robustness checks, and treatment heterogeneity. Section 4.7 provides a discussion and conclusion.

4.2 Policy Context and Review of Quantitative Evidence

4.2.1 Education policy context in Africa

Currently, UNESCO recommends at least six years of mother tongue instruction for ethnically diverse communities (UNESCO, 2016). This is based on an extensive body of work in education theory which has long argued that learning in mother tongue or proximate local languages aids with student comprehension, particularly in new concepts (Cummins 1981, Alidou et al., 2006). The idea has a long history, dating back to an influential UNESCO report on "the use of vernacular languages in education" (UNESCO, 1953) which advocated for starting formal education with mother-tongue instruction. Students are found to participate more in their learning when familiar languages are used in the classroom (Brock-Utne, 2005c). More specifically, there is also evidence of mother-tongue instruction leading to more effective teaching of mathematics and the sciences (Prophet & Dow, 1994; Mwinsheikhe, 2003; Alidou et al., 2006). Additionally, for parents not fluent in English, mother-tongue instruction is shown to encourage parental involvement in the education of their children (Gfeller & Robinson, 1998; Eriksson, 2014).

Despite this UNESCO recommendation, countries have varying education language policies. In the sub-Saharan Africa context, most 'Francophone' and 'Lusophone' countries use 1-3 years of mother-tongue instruction followed by French or Portuguese language instruction. In 'Anglophone' countries mother-tongue language policies last approximately 4-6 years before a transition to English medium instruction. However, most mother-tongue education policies in Anglophone countries now tend to be 4 years or less (Heugh et al., 2007).

One of the reasons for the early transition to a foreign or international language for education may be an artefact of colonialism. This left languages such as English, French, Portuguese, and Spanish as the national and education language policy (Roy-Campbell, 2006; Clegg & Simpson, 2016). Moreover, even in countries with no real colonial history such as Ethiopia and in countries with smaller English histories such as

Tanzania, Namibia, and Rwanda; policy makers still widely adopt EMI policies. Beyond learning English as a subject, it is widely believed that learning in English leads to future success. The use of the English language is often regarded as the language of progress, development, and economic opportunities (Barkhuizen, 2002; Chitera, 2011; Clegg & Simpson, 2016). The growth in the number of low-cost private schools usually advertised as English-medium schools, competes with the public sector for students, and contributes to the overall push for EMI in schooling. Moreover, parents often demand EMI in schools and politicians are eager to capitalize on this (Jamil & Saeed, 2015; Clegg & Simpson, 2016).

In linguistically diverse areas, people often learn a second language that functions as a lingua franca to enable communication between communities. These are typically local languages such as with Amharic in Ethiopia and are widely spoken and understood (Heugh et al., 2007). However, parents and teachers value foreign language acquisition and believe that English proficiency is linked to the length of exposure to the language (Piper et al., 2016) and that full immersion leads to desirable learning outcomes and socioeconomic opportunities (Bunyi, 2005; Clegg & Simpson, 2016). Proponents of English language instruction also believe local languages to be insufficiently developed to convey complex concepts in the school curriculum (Benson, 2000). Nonetheless, pedagogic, linguistic researchers and some educators continue to assert that sufficient mother tongue instruction is a necessary condition for learners' education (Brock-Utne, Birgit., 2005c.; Alidou et al., 2006).

4.2.2 Review of the quantitative evidence on the effect of English Medium Instruction (EMI)

The empirical studies over the past two decades largely find positive effects of mother-tongue instruction on student's reading abilities, numeracy, increased participation in class, improved attendance, an overall improvement in student performance, and years of schooling (Garcia, 2000; Alidou et al., 2006; Desai and Qorro, 2006; August & Shanahan, 2008; Ramachandran, 2017; Seid, 2019; Laitin et al., 2019)⁴.

⁴Another compelling example is observed from South Africa during the Apartheid years of Bantu education (1955-1976)- the educational policy that divided the primary school curriculum into the different

In Ethiopia, Seid (2016) finds that mother tongue instruction increased the probability of enrollment by 2.6 percentage points (between regional states) and increased the probability of children being in the "right" grade for their age by 2.4 percentage points in primary schools. The same author (Seid, 2019) later uses the 2012/2013 Young Lives study dataset to investigate the impact of mother tongue instruction in Grades 1 to 4 on the performance of Grade 5 students. By exploiting variation in one region in Ethiopia, the paper finds that mother tongue instruction in early grades compared to second language education improves learning in mathematics by .16 standard deviations. Finally, Seid (2022) uses the 2013 Ethiopian Labour Force Survey to show that mother-tongue instruction improves labour market outcomes. However, the effect decreases as the years of exposure to mother-tongue instruction increases.

Ramachandran's (2017) research using the Demographic Health Survey in Ethiopia also finds that mother tongue instruction leads to gains in adult literacy by 40 percent, an increase in newspaper readership by approximately 25 per cent and more than half a year increase in completed years of schooling. In addition, Laitin, Ramachandran and Walter (2019) investigating language policies in Cameroon find that mother tongue instruction in early grades improves mathematics by 1.17 - 1.71 of a standard deviation and English outcomes by 0.68 - 1.53 of a standard deviation quite significantly.

Additionally, Taylor and von Fintel (2016) find in over 9000 schools in South Africa that children exposed to English medium instruction in early grades tested 0.17 standard deviations below mother tongue instruction children in Grades 4, 5 and 6. Other empirical studies also find mother tongue education to improve literacy rates (Kerwin & Thornton, 2015; Brunette et al., 2019).

On the other hand, a study by Piper, Zuilkowski, Kwayumba and Oyanga (2018) in Kenya found among early primary school students that mother tongue instruction did not significantly improve English learning and mother tongue students even had

native languages. The language policy intended to create further divisions between the various language groups had unintended effects as Black African learners achieved greater academic success compared to non-Black learners. This resulted in the former Apartheid government reducing the mother-tongue instruction policy from 8 years down to 4 years (Heugh, 2008; UNESCO, 2010).

lower mathematics outcomes compared to non-mother tongue students. Another study from the United States (Slavin, Madden, Calderon, Chamberlain, & Hennessy, 2011) found that the early impact of mother-tongue instruction for first graders diminished by year three at which point there were no significant differences between mother-tongue instruction students and English-only instruction students on English reading measures. The implications of these studies caution policy makers to also focus on the quality of instruction when enacting language of instruction policies (Slavin et al., 2011).

While most of the literature leans towards mother tongue instruction policies, it is important to mention that even in contexts where there is a willingness to implement mother-tongue instruction policies, the instructional environment and implementation challenges encountered may be substantial (Piper et. al, 2018). These can include poor teacher deployment policies, under-resourced teacher training for mother-tongue instruction and inadequate instructional materials (Iyamu & Ogiegbaen, 2007; Nyaga & Anthonissen, 2012; Begi, 2014; Ssentanda et al, 2019). This was also illustrated in a multilingual Uganda where a study by Ssentanda, Southwood & Huddleston (2019) found that one of the reasons for students' difficulties in learning English was because of insufficient learning and teaching support materials teachers. In addition, Ssentanda and Wenske (2021) disrupt the either/or debates around mother tongue and English instruction. The authors suggest that there is no strict binary perspective, proposing that mother tongue education can be seen as a form of bilingual education cultivating literacy in two languages, thereby challenging the notion that mother tongue education only serves as a transitional phase toward English literacy (Ssentanda and Wenske, 2021).

4.2.3 Summary of regional motivations for Language of Instruction policies in Ethiopia

In line with federal policy, English is introduced nationally as a subject in Grade 1 while regional authorities have autonomy over language of instruction policy within the first 8 years of primary school (Vujcich, 2013). Prior to the 1994 education reform, all regions used Amharic as the medium of instruction. However, over a 20-year period from 1990-2010, 32 additional languages were developed allowing regions discretionary use

of mother tongue instruction once the education policy reform came into effect. Even so, Ethiopia's linguistic diversity meant that for many regions proximate and or second languages were developed for instructional purposes. (Skutnabb-Kangas & Heugh, 2013; Seid, 2019).

The English Medium Instruction (EMI) variation between regions presents an interesting and unique opportunity to explore the extent to which, if any, using EMI in primary school affects mathematics and English test outcomes. In the regions captured in the study dataset, schools in Amhara, the Southern Nations, Nationalities and Peoples' Region (SNNPR), and Afar region transition to EMI in Grade 1, Grade 5, or Grade 7 while schools in Oromia and Tigray maintain mother tongue instruction through all 8 years of primary education. A report commissioned by the Ethiopian Ministry of Education (Heugh et al, 2007) revealed multi-faceted reasons for the regional language of instruction policy decisions.

In regions transitioning from mother tongue to EMI policies (Amhara, SNNPR and Afar), there seemed to be public debate and pressure from parents, students, and some of the teachers to introduce English medium instruction in the primary school years. In Amhara, while educators advocated for mother-tongue instruction, middle class parents in urban areas were eager for English language instruction to begin as soon as possible. In SNNPR, parents expressed that mother tongue instruction policies weakened the value of education while according to a Teaching College official, the region's policy was decided by politicians with no research backing. In Afar, the general attitude towards language of instruction was for EMI to be introduced in upper primary grades.

In the mother tongue education regions such as Oromia, the general attitude was that mother-tongue instruction was critical to ensuring quality education. Similarly in Tigray, teachers and education officers interviewed expressed that mother-tongue instruction was appropriate for schooling (Heugh et al., 2007).

4.3 Data, Sampling Design and Assessment Design

4.3.1 Data and Sampling Design

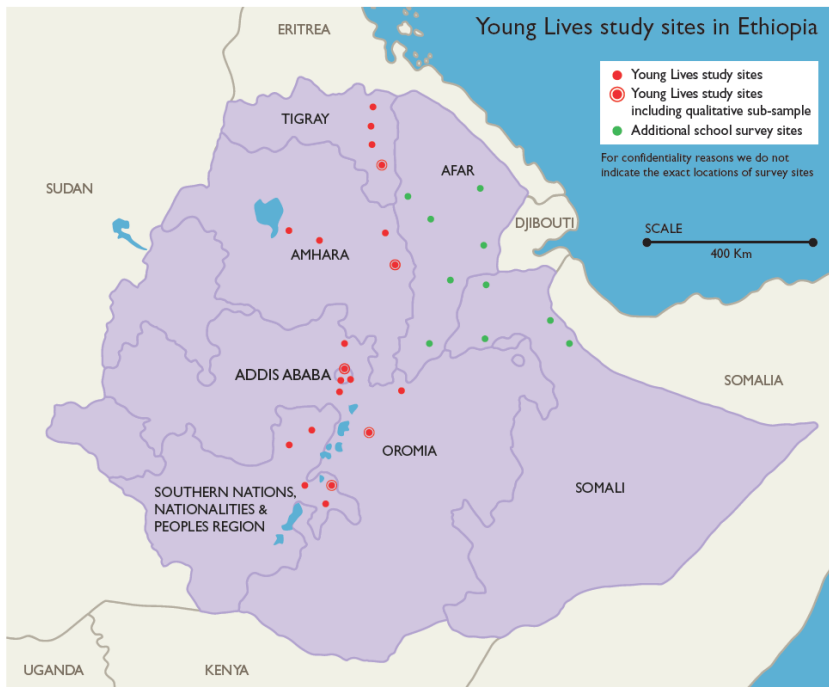
This paper uses data from the Young Lives longitudinal study on childhood poverty. The Young Lives study⁵ followed the lives of children located in five regions: Addis Ababa, Amhara, Oromia, Southern Nations, Nationalities and Peoples' Region (SNNPR) and Tigray; collecting data on indicators of health, nutrition, education, and livelihoods.

The Young Lives study included surveys with parents and children as well as a school-based component, conducting observations in schools, surveying students, interviewing teachers, and school heads to fully capture children's experiences with education. School surveys were conducted in three rounds: 2010/2011 (round 1), 2012/2013 (round 2) and 2016/2017 (round 3). Three to five study sites in each region (see figure 4.1) were purposively selected to reflect the cultural and geographic diversity in Ethiopia. Altogether, 30 study sites were selected with all the schools within the site's geographic boundary included in the survey (Rossiter et al, 2017). Study sites were selected with a bias on including poor areas with food insecurity (Outes-Leon and Sanchez, 2008). This paper utilizes round 3 of the Young Lives school data collected in the 2016/2017 academic year from Amhara, SNNPR, Oromia and Tigray and Afar.

The data used in this paper captures Grade 7 and 8 learning in two assessment waves. Wave 1 was conducted at the beginning of the school academic year in October/November 2016 and wave 2 at the end of the school year in May/June 2017 (Rossiter et al., 2017). The survey targeted all Grade 7 and 8 students physically present (87% of enrolled students) on the first day of the assessment at each school, with wave 2 following up the same set of students without replacements (Rossiter et al., 2017). This means that if a student was absent during wave 1 or absent during wave 2 of the survey round, their data is excluded from the sample.

⁵<https://www.younglives.org.uk/content/about-us>

Figure 4.1: Young Lives study sites in Ethiopia (Young Lives, 2018)



In addition, the study includes data from Grade 7 & 8 mathematics and English teachers and school heads. Background instruments include a school head questionnaire, a teacher questionnaire, a student questionnaire, and a school facilities tool. Figure 4.1 shows the study sites selected in the study. The final sample (table 4.1) includes 38 schools and 4,420 students from the Amhara, Oromia, SNNPR, Tigray and Afar regions.

Table 4.1: Distribution of schools and students by region

Region	Schools	Students
Amhara	7	717
Oromia	5	578
SNNPR	11	1,233
Tigray	7	1,204
Afar	8	688
Total	38	4,420

This paper restricts the sample to the aforementioned regions with Addis Ababa and Somali dropped from the main study sample. Addis Ababa, as the urban administrative capital, has been dropped due to the ethnic diversity found in the region (Vujcich, 2013). The linguistic diversity in the capital means that for some students, Amharic instruction

is in their mother tongue while for others, this may be their second or even third language. This heterogeneity makes Addis Ababa different and difficult to compare to the other regions in the sample where more language homogeneity within regions is found.

Somali region is also dropped from the main study sample as there are some schools with an EMI policy and other schools maintaining a mother tongue education only policy (see appendix C for Somali variation). In Somali, students/parents may be motivated to move to schools with their preferred education language policy, and this selection into schools may be what drives differences between EMI and mother-tongue instruction children. As such, Somali has been excluded from the main study sample. The study sample only includes schools in regions with one education language policy (i.e., a mother tongue instruction policy or a mother tongue to EMI policy).

Table 4.2 shows the distribution of the English Medium Instruction (EMI) treatment variable by grade of EMI transition in each region. In schools using EMI, the table shows the grade at which schools transition from mother-tongue instruction to English medium instruction. Most schools transition to EMI in Grades 5 or 7, with a few schools starting with English medium instruction as early as Grade 1 (in SNNPR). Schools that utilise mother tongue-instruction, use only mother-tongue teaching (i.e., these schools do not teach mathematics or any other academic subject using English at any grade in primary school). This applies to all students within the study sample who at the time of the assessments were in Grades 7 and 8.

The treatment variable is captured through a question posed to the school heads in the study sample. The question describes the language of instruction policy within the school.

"In which primary grade does this school start using English as a medium of instruction for Mathematics?"

Table 4.2: Distribution of schools using English medium instruction by Grade

Region	EMI Transition Grade	Schools in sample
SNNPR	Grade 1	3
	Grade 5	8
Amhara	Grade 7	7
Afar	Grade 7	8
Oromia	Mother-tongue instruction only	5
Tigray	Mother-tongue instruction only	7
Total		38

*All schools in this study are within the Young Lives study sample. Treatment regions have been highlighted.

In SNPPR, there is variation in EMI grade of transition (Grade 1 or 5), however all schools within the region have a primary school EMI transition policy. SNPPR has 14 regional languages in use in the first cycle of primary education from Grade 1 to 4, and these languages are all used as a medium of instruction and also taught as a subject in schools (Heugh et al, 2007).

Schools in Amhara, SNNPR and Afar teaching mathematics using EMI are captured as treatment schools and schools in Oromia and Tigray teaching mathematics using mother tongue instruction only are captured as control schools.

4.3.2 Assessment Design

The mathematics assessment covered three cognitive domains: (1) knowing facts, concepts, and procedures; (2) applying knowledge and conceptual understanding to solve problems and (3) reasoning (Gronmo et al, 2015). The Young Lives research team received input from the Ethiopian Ministry of Education’s Mathematics and Science Improvement Centre (MISC) to select the assessment content. The content domains for the mathematics assessment included basic number competency, algebra, volume, surface area to problem solving and daily life applications. After a series of pilots, 40 multiple-choice questions were included in the final assessment.

Similarly, through a series of consultations, the English assessment covered four skill domains: (1) word identification through simple vocabulary; (2) word meaning and contextual vocabulary through identifying the meaning of words in context; (3) sentence construction and comprehension through the correct completion of sentences

with appropriate grammar; and (4) reading and comprehension through reading and comprehending stories, posters, and factual passages. After a series of pilots, 40 multiple-choice questions were included in the final assessment.

The content of the Mathematics and English tests administered at the beginning and end of the year (wave 1 and wave 2) were identical (Rossiter et al, 2017), maintaining the difficulty level across the assessment waves. The same assessment is administered to all Grade 7 and 8 students in each round of data collection. Test results in Mathematics and English are the outcomes of interest.

4.4 Empirical Framework

A value-added model of achievement production is used in the empirical analysis of this study. Value-Added (VA) models are now commonly used in the education production function literature (Todd & Wolpin, 2003; Andrabi et al, 2011). Todd and Wolpin (2003) provide an economic model for cognitive achievement where knowledge acquisition is created as a "production process in which current and past inputs are combined with an individual's genetic endowment of mental capacity to produce a cognitive outcome". The inputs described capture choices made by caregivers and schools. The knowledge acquisition model is given as:

$$Y_{*ist} = F[C_i(t), S_i(t), \mu_{is0}, \epsilon_{ist}] \quad (4.1)$$

where Y_{*is} is the measure of achievement for child i in school s at time t , which proxies for cognitive outcomes. Inputs into the production of achievement are captured through vectors $\mathbf{C}_i(t)$ for caregiver and home-based inputs and $\mathbf{S}_i(t)$ for school-based inputs. In addition, unobserved child ability (i.e., endowed mental capacity) is denoted as μ_{is0} and a time-varying error term denoted as ϵ_{ist} . However, such a model is impossible to accurately estimate as data on inputs (caregiver, home, and school based) are incomplete with inevitably missing histories, and child ability endowed at birth is not observed (Todd & Wolpin, 2003, 2007).

The missing data problem on input histories and ability have led to the adoption

of VA models in estimating knowledge production functions. The intuition behind including a baseline achievement measure in the model is that this lagged achievement measure can be used as an adequate proxy for the unobserved inputs and endowed mental capacity⁶ (Todd & Wolpin, 2003).

The value-added model is given as⁷:

$$Y^*_{ist} = \alpha' \mathbf{X}_{ist} + \lambda \mathbf{P}_{ist} + \beta Y_{is,t-1} + \mu_{ist} \quad (4.2)$$

where all inputs (student, family, and school-based) are captured in the vector $\mathbf{X}_{ist}$ with $\mathbf{P}_{ist}$ as the treatment effect, capturing a policy intervention isolated for student i in school s at time t . Y^*_{ist} is the unobserved achievement for student i in school s at time t and $Y_{is,t-1}$ is the lagged test score, with μ_{ist} capturing productivity shocks. The α is the input coefficient and β is the parameter linking achievement across rounds as the persistence coefficient (i.e., if how learning gains persist stay over time). The model thus accounts for past inputs and any past unobservable endowments (Todd & Wolpin, 2003; Andrabi et al, 2011; Singh, 2015).

While this model is better than linking current inputs to current knowledge achievement (i.e., as with contemporaneous specifications), the value-added (VA) model is still susceptible to threats in identifying the causal effect of a policy. The first potential threat in estimation is that there is unobserved heterogeneity between children in how much they learn between time point $t-1$ and point t which is correlated with both the treatment and the outcome. The lagged term only accounts for learning heterogeneity entered through a one-time process or mental capacity endowment at time $t-1$. However, suppose academically gifted children learn faster between $t-1$ and t . This is not fully captured in the β so will be captured by the error term. If the error term is correlated with treatment (e.g., academically gifted children are more likely to be in treatment schools), then estimates of λ will be biased. However, in using this model for the study, an unbiased estimate of the treatment effect can still be estimated as it is unlikely that

⁶Mental capacity can be interpreted as a quasi-fixed input, which may differ from mental capacity at birth; or that mental endowment is fixed at birth with varying impact on achievement dependent on age (Todd & Wolpin, 2003).

⁷Notation used is consistent with education production and dynamic panel literature.

child-level heterogeneity in learning is correlated with being in a treatment (EMI) school. All schools in the main study sample apply the same policy within region (mother tongue education only or a mother tongue to EMI policy). Thus, it is highly unlikely that children to self-select into schools based on language of instruction policy.

The second potential threat is that achievement test scores are a noisy measure of latent achievement, potentially introducing attenuation bias on the right-hand side of the regression which bias diminish the estimate of β (the persistence parameter) towards zero while also making the estimates of the input parameters less precise (Andrabi et al, 2011; Singh 2015). With these potential threats, the inputs captured in the input vector $\mathbf{X}$ becomes even more important. Todd and Wolpin (2003) argue that with good data on relevant inputs, outcome measures, and including a baseline measure, if one relies on the (plausible) assumption of coefficients' geometric decay, one can identify the casual relationship of interest. Thus, in this study, these threats are unlikely to threaten identification in this particular context.

4.4.1 Empirical Strategy

The following specifications are estimated in this paper:

$$Y_{ist} = \alpha_1 + \beta_1 EMI_{ist} + \epsilon_{ist} \quad (4.3)$$

$$Y_{ist} = \alpha_2 + \beta_1 EMI_{ist} + \beta_2 \mathbf{X}_{ist} + \epsilon_{ist} \quad (4.4)$$

$$Y_{ist} = \alpha_3 + \beta_1 EMI_{ist} + \beta_2 \mathbf{X}_{ist} + \beta_3 Y_{is,t-1} + \epsilon_{ist} \quad (4.5)$$

EMI is an indicator variable that equals 1 for students in schools where English is used as a medium of instruction (EMI) for Mathematics and equals to 0 for students in schools with mother-tongue instruction. $\mathbf{X}_{ist}$ is a vector of background characteristics related to students, their families, teachers, and school-quality variables. These characteristics are all captured in table 3, and include student age, gender, parents' literacy levels, teachers'

age, gender, and highest qualifications as well as indicators of school quality. Finally, Y_{ist} and $Y_{is,t-1}$ captures achievement and the lagged achievement score respectively. In this paper, the mathematics assessment scores from the wave 1 assessment are used as the lagged measure for the mathematics outcome and the English assessment scores from the wave 1 assessment are used as the lagged measure for the English outcome measure.

The parameter of interest β_1 captures the treatment effect. Each specification gradually adds controls, increasing the reliability of the treatment effect estimate. The VA model as specified in equation 4.5 is the preferred specification for this study as it includes student relevant inputs as controls as well as a lagged assessment measure capturing past inputs and unobservable endowments. Lastly, standard errors are clustered at the school level⁸. All input variables indicated in table 3 are included as controls in all regression outputs specifying controls.

4.5 Descriptive Statistics

Descriptive statistics on student characteristics, family, teacher, and school related variables are shown in table 3. Each panel in table 3 shows baseline differences by listing the control group mean or proportion and the treatment minus control (T-C) baseline differences on all input variables. Treatment schools are schools in regions with English medium instruction while control schools are schools in regions with mother-tongue instruction. Student and family related variables are detailed in Panel A, teacher-related variables differentiated by subject (mathematics or English) in Panel B, and school quality-related variables are captured in Panel C.

Overall, the groups appear balanced and there are no systematic patterns of students in either treatment or control schools performing "better" or having advantageous access to resources. Students are approximately 14 years of age with a fairly even proportion of boys and girls in the sample. In some input characteristics, we observe EMI schools as slightly better (e.g., more urban, higher teacher training qualifications, etc.) while in other characteristics they appear to be below mother-tongue instruction

⁸Most regions share the same treatment status. But this is not the case in SNNPR, where there is grade variation in EMI transition by school, making school-level clustering appropriate.

schools (e.g., in family literacy levels, teachers' specialisation in mathematics, etc.).

Descriptively, there are only four significant differences found between treatment and control groups and this small number of statistically significant differences can be expected by chance. Nonetheless, the value-added specification proposed includes all student, family, teacher, and school quality- related variables captured in the table below as control variables.

Table 4.3: Baseline Comparison between EMI (Treatment) and Mother-Tongue Instruction (Control) Schools

Panel A Student and Family Characteristics	(1) Baseline mean/proportion controls	(2) T-C difference at baseline (wave 1)	(3) N
Girl	0.546	-0.012 (0.017)	4,420
Age in years	13.929	0.279 (0.178)	4,420
Age at Grade 1	6.902	-0.230*** (0.078)	4,420
School Grade	7.512	-0.059 (0.039)	4,420
Help with homework at home (1=Yes)	0.714	-0.003 (0.034)	4,420
Read books out of school? (1=Yes)	0.898	-0.009 (0.028)	4,420
Attended pre-school (1=Yes)	0.447	0.087 (0.059)	4,420
Mother can read and write in any language (1=Yes)	0.545	-0.020 (0.056)	4,420
Father can read and write in any language (1=Yes)	0.725	-0.040 (0.038)	4,420
Sibling can read and write in any language (1=Yes)	0.838	-0.075** (0.036)	4,420
Mathematics test score (wave 1)	-0.000	0.001 (0.127)	4,420
English test score (wave 1)	0.000	0.431** (0.205)	4,420

Notes: Standard errors in parentheses are clustered at the school level. The T-C difference is the baseline (wave 1) “treatment effect” using a minimal specification, including treatment dummy, and excluding wave 1 conditioning variables. Asterisks in column 2 indicates a significant difference; * $p < 0.01$ ** $p < 0.05$ *** $p < 0.01$.

Panel B Teacher Characteristics	(1) Baseline mean/proportion controls	(2) T-C difference at baseline (wave 1)	(3) N
Mathematics Teachers			
Math Teacher Age (Years)	33.348	-0.552 (2.806)	4,420
Female Math Teacher	0.167	-0.094 (0.097)	4,420
Math Teacher training qualification (1=University Degree)	0.291	0.105 (0.178)	4,420
Specialisation in Math (1=Yes)	1.000	-0.236*** (0.078)	4,420
Math Teacher Experience (Years)	13.474	-1.037 (2.815)	4,420
English Language Improvement Programme (1=Started/Completed)	0.278	-0.167 (0.137)	4,420
English Teachers			
English Teacher Age (Years)	34.704	1.919 (2.396)	4,420
Female English Teacher	0.356	-0.049 (0.129)	4,420
English Teacher training qualification (1=University Degree)	0.161	0.135 (0.131)	4,420
English Language Improvement Programme (1=Started/Completed)	0.655	-0.056 (0.144)	4,420
Teacher Experience (Years)	14.897	0.846 (2.558)	4,420

Notes: Standard errors in parentheses are clustered at the school level. The T-C difference is the baseline (wave 1) “treatment effect” using a minimal specification, including treatment dummy, and excluding wave 1 conditioning variables. Asterisks in column 2 indicates a significant difference; *p<0.01 **p<0.05 ***p<0.01.

Panel C School Quality Characteristics	(1) Baseline mean/proportion controls	(2) T-C difference at baseline (wave 1)	(3) N
Shift/Full Day (1=Shift)	1.000	-0.085* (0.047)	4,420
Number of working latrines/toilets for student use	3.504	0.482 (1.833)	4,420
Functional pedagogical centre (1=Yes)	0.984	-0.239*** (0.116)	4,420
Functional library (1=Yes)	0.941	-0.095 (0.087)	4,420
Sports/Play area (1=Yes)	0.857	0.083 (0.124)	4,420
Electricity normally connected (1=Yes)	0.920	-0.028 (0.073)	4,420
Locality (1=Urban)	0.464	0.265 (0.205)	4,420

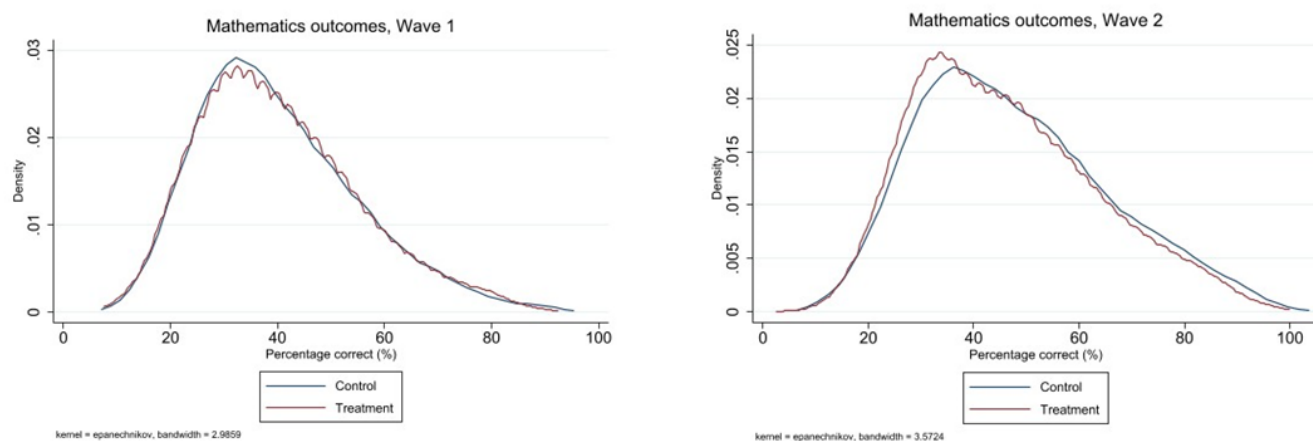
Notes: Standard errors in parentheses are clustered at the school level. The T-C difference is the baseline (wave 1) “treatment effect” using a minimal specification, including treatment dummy, and excluding wave 1 conditioning variables. Asterisks in column 2 indicates a significant difference; * $p < 0.01$ ** $p < 0.05$ *** $p < 0.01$.

4.6 Results

Overall, the results from the preferred value-added specification (5) show that teaching using English as a medium of instruction, when compared to mother tongue instruction, reduces Grade 7 and 8 student test scores by 0.25 standard deviations. This is a large effect. In evaluations of school-based interventions on learning, effect sizes from 0.1 to 0.2 standard deviations are considered large and noteworthy (McEwan, 2014). Also, international education studies in low-and middle-income countries find median treatment effects of 0.1 standard deviations (Evans & Yuan, 2020). Moreover, effects are measured over a year, whereas evaluations typically show aggregated level effects on learning. This paper shows the policy effect on student’s results over the academic year; a policy that could affect some students as early as Grade 1.

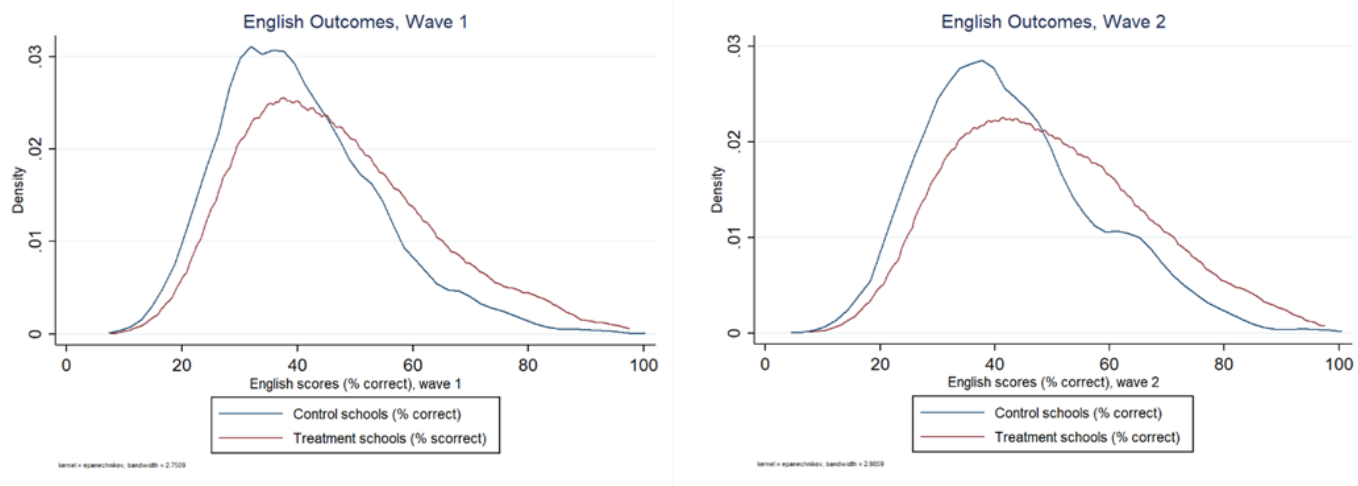
The following section details the results, first by examining the patterns in the descriptive data, which is largely supported by the main regressions specified. The kernel density plots in figures 4.2 and 4.3 show the distribution of students' mathematics and English test scores at wave 1 and wave 2 of the assessments. For mathematics wave 1 scores (baseline), there are few discernable differences. At this point, most students in the sample have been taught in mother tongue but are beginning a year where academic subjects will be taught in English. However, at wave 2 (endline), EMI schools (treatment) appear to have marginally worse outcomes after a year of being taught in English, compared to mother-tongue (control) schools.

Figure 4.2: Kdensity plots on the mathematics outcome



Looking at the distribution of the English outcome, it is less clear which group performs better as there are discernable differences between groups at baseline (wave 1). It is possible that differences in proficiency at baseline are linked to more English exposure in urban schools, higher baseline proficiency or even other region-specific differences. As such, extensive control variables are included to account for students' prior test scores, parent-level, school-level, and regional inputs in causally identifying the impact of the education language policy on student mathematics and English outcomes.

Figure 4.3: Kdensity plots on the English outcome



The main results (Tables 4.4 and 4.5) show the results of the specifications outlined in section 4. The effects found in the raw descriptive data above largely persist even when student, family and school related controls are included. For mathematics, there is a significant negative effect of EMI on test scores. For English, there is no significant effect: while there are some differences between treatment and control schools in the raw descriptive data, these are largely accounted for by differences in child, parent and school characteristics that are present before the implementation of the policy rather than the policy itself. This highlights the importance of not only examining the descriptive data, but also running the regression specification.

The first column (Tables 4.4 and 4.5) shows the unconditional cross-sectional difference in mathematics and English test outcomes between English medium instruction (EMI) schools and mother-tongue instruction schools. The second column includes

controls for relevant student, family, teacher, and school quality covariates (captured in table 3). This specification may also be interpreted as effects on cumulative learning for Grade 7 & 8 students in the study sample. The value-added specification in Column 3 (in Tables 4.4 and 4.5) is the most robust estimate. It introduces the lagged student achievement measure, region fixed effects, and specified controls into the model, thereby capturing all previous inputs as well as any unobservable endowments or shocks. This specification may also be interpreted as capturing the policy effect over one academic year.

Table 4.4: Impact of English Medium Instruction (EMI) on Math Outcomes

Variables	(1)	(2)	(3)	(4)
EMI	-0.088 (0.136)	-0.277*** (0.079)	-0.248*** (0.064)	-0.255*** (0.064)
Lagged math score			0.612*** (0.022)	0.654*** (0.029)
Interaction (EMI*Lagged score)				-0.072* (0.041)
Control mean	0	0	0	0
Controls	No	Yes	Yes	Yes
Region FE	No	No	Yes	Yes
Schools	38	38	38	38
Students	4,420	4,413	4,413	4,413

The outcome has been standardised using the control group mean and standard deviation. Columns 2 includes controls, as in Table 3. Column 3 specifies the value-added model by including a lagged achievement score and region fixed effects. Column 4 includes interactions. Robust standard errors in parentheses are clustered at the school level;

*** p<0.01, ** p<0.05, * p<0.1.

After controlling for individual, school, and regional level differences, the results show that English as a Medium of Instruction (EMI) for Mathematics reduces test scores by 0.25 standard deviations among Grades 7 and 8 students (see column 3). On English test outcomes, there are no observable meaningful differences between students in English medium instruction schools and students in mother-tongue instruction schools (see table 4.5, column 3). Once controls are introduced a seemingly significant positive outcome flips into a null result.

Column 4 in both results tables shows the interaction term between the treatment dummy and lagged test score. On mathematics outcomes, the results show baseline test scores had a significant effect on mathematics outcomes. On English test outcomes, there were no significant differences in treatment effects between students with

higher or lower baseline scores in both mathematics and English outcomes. Furthermore, outcomes winsorized at 5% in each tail (shown in Appendix D and E) show similar results.

Table 4.5: Impact of English Medium Instruction (EMI) on English Outcomes

Variables	(1)	(2)	(3)	(4)
EMI	0.428** (0.196)	0.159 (0.108)	-0.025 (0.061) 0.510*** (0.023)	-0.022 (0.061) 0.493*** (0.04)
Lagged English score Interaction (EMI*Lagged score)				0.027 (0.044)
Control mean	0	0	0	0
Controls	No	Yes	Yes	Yes
Region FE	No	No	Yes	Yes
Schools	38	38	38	38
Students	4,420	4,420	4,420	4,420

The outcome has been standardised using the control group mean and standard deviation. Columns 2 includes controls, as in Table 3. Column 3 specifies the value-added model by including a lagged achievement score and region fixed effects. Column 4 includes interactions. Robust standard errors in parentheses are clustered at the school level; *** p<0.01, ** p<0.05, * p<0.1

4.6.1 Robustness Checks

The main effects are also robust to other specifications. Table 6.6 runs the value-added specification, re-introducing schools in the Somali region back into the sample. Somali was initially dropped from the main analysis sample for having varying language of instruction policies within the region (i.e., some schools using mother-tongue instruction with other schools choosing to transition to EMI at varying grades). The new sample with Somali is then restricted to students who remained in the same schools from Grade 1 through to Grade 7 or 8. This restriction limits potential selection bias from students/parents who may have self-selected into schools using their preferred language of instruction policy. Column 1 shows value-added estimates on mathematics and column 2 reports the English outcome estimates.

Although the results reveal slightly smaller effect sizes, these results remain robust with a significant negative effect of 0.12 standard deviations on mathematics outcomes and a positive non-significant effect 0.07 standard deviations on English

outcomes. These are consistent with the main results shown in Table 4.4 and 4.5.

Table 4.6: Impact of English medium instruction on student outcomes (including Somali region)

Variables	Mathematics	English
EMI	-0.122** (0.053)	0.066 (0.057)
Lagged math score	0.581*** (0.029)	
Lagged English score		0.464*** (0.028)
Control mean	0	0
Controls	Yes	Yes
Schools	41	41
Students	3,570	3,570

Outcome has been standardised using the control group mean and standard deviation. Robust standard errors in parentheses are clustered at the school level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

To further assess the robustness of the estimated effects, table 4.7 runs the value-added specification on a restricted sample dropping schools from the Somali and SNNP regions. These regions vary EMI policy, with some EMI transition variation at grade level (i.e., SNNPR and Somali) or a variation in education language policy (i.e., Somali). Results on mathematics outcomes remain robust in direction and magnitude of effect (-0.21 standard deviations). On English outcomes, the direction of the effect of EMI changes. However, this result remains non-significant as reported in the main analysis.

Table 4.7: Impact of English medium instruction on student outcomes (restricted sample)

Variables	Mathematics	English
EMI	-0.205*** (0.053)	-0.038 (0.056)
Lagged math score	0.637*** (0.027)	
Lagged English score		0.462*** (0.036)
Control mean	0	0
Controls	Yes	Yes
Schools	24	24
Students	2,208	2,208

Outcome has been standardised using the control group mean and standard deviation. Robust standard errors in parentheses are clustered at the school level; *** p<0.01, ** p<0.05, * p<0.1.

4.6.2 Treatment Heterogeneity

This section explores treatment heterogeneity. In particular, this paper examines the effects of a range of school quality characteristics that might mitigate the negative effects of EMI, such as more teacher experience or specialized teacher training for teaching in English. No significant differences are found for students with teachers enrolled in an English language improvement programme (ELIP), or with teachers with more years of teaching experience. In addition, there are no significant gender differences among the students. Finally, heterogenous treatment effects by the grade of EMI transition suggests that an early primary grade transition has positive effects on English test results and no significantly negative effects on mathematics outcomes for Grade 7 & 8 students; relative to mother tongue education only.

Table 4.8: Summary of EMI Treatment Heterogeneity by the English Language Improvement Programme

	Mathematics	English
Mother Tongue & No ELIP	0	0
Mother Tongue & ELIP	-0.118	0.066
EMI & No ELIP	-0.223***	0.077
EMI & ELIP	-0.341	0.207
Control mean	0	0
Controls	Yes	Yes
Schools	38	38
Observations	4,413	4,413

Note: only treatment estimates shown for mathematics and English outcomes, treatment coefficients with standard errors are shown captured in main regression output in Appendix H*

When the Ethiopian government mandated an EMI policy in all secondary and higher education institutions, the Ministry of Education strategized to improve the English language proficiency and teaching skills of its English teachers through the ELIP established in 2003 (Ahmad, 2013). This study finds no heterogeneous effects on ELIP such that having an ELIP teacher in an EMI school compared to an ELIP teacher in a mother-tongue school makes no real difference. On student mathematics outcomes, a joint significance test $F(2,36) = 1.36$, $p > 0.05$ confirms that there are no significant differences between teachers in mother-tongue schools with ELIP and teachers in EMI schools with ELIP. Similarly, on English outcomes there are no significant differences between ELIP teachers in the two groups, with $F(2, 37) = 1.01$, $p > 0.05$. However, these results are interpreted cautiously as estimates may be somewhat under-powered due to the small number of schools in each subgroup.

In addition, using the median years of teaching experience⁹, this paper does not find any meaningful differences between teachers above or below the median 12 years (of mathematics) or 16 years (of English) teaching experience.

⁹Teacher years of experience statistics reported in appendix J.

Table 4.9: EMI Treatment Heterogeneity by Teachers' Years of Experience

Variables	Mathematics	English
EMI	-0.187*** (0.047)	0.093 (0.058)
EMI * Years of Math Teacher Experience (above P50)	-0.022 (-0.066)	
Years of Math Teacher Experience (1= above P50)	-0.007 (0.010)	
Lagged Mathematics Score	0.611*** (0.022)	
EMI * Years of English Teacher Experience (above P50)		0.040 (0.117)
Years of English Teacher Experience (above P50)		0.001 (0.008)
Lagged English Score		0.519*** (0.024)
Control mean	0	0
Controls	Yes	Yes
Schools	37	38
Students	4,413	4,420

Outcome has been standardised using the control group mean and standard deviation. Columns 2 and 3 use the value-added specification including the full set of covariates and a lagged achievement measure. Robust standard errors in parentheses are clustered at the school level ; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Differentiating the effect of English medium instruction (EMI) by sex is important in ensuring equity in access to schooling and learning opportunities. Table 4.10 shows that on English test outcomes, girls and boys perform similarly but on mathematics outcomes, girls slightly outperform boys in EMI schools, relative to mother tongue girls and boys. However, as the coefficient on the sex and treatment variable is found to be non-significant, the effect of EMI between girls and boys is indistinguishably the same on both mathematics and English outcomes. Thus, there are no meaningful differential effects between girls and boys in EMI and girls and boys mother tongue instruction schools.

Table 4.10: EMI Treatment Heterogeneity by Sex

Variables	Mathematics	English
EMI	-0.217*** (0.055)	0.099 (0.068)
Girl	-0.126*** (0.045)	-0.057 (0.048)
EMI * Girl	0.062 (0.052)	-0.020 (0.064)
Lagged Math Score	0.611*** (0.022)	
Lagged English Score		0.518*** (0.025)
Control Mean	0	0
Controls	Yes	Yes
Schools	37	38
Students	4,413	4,420

Outcome has been standardised using the control group mean and standard deviation. Columns 2 and 3 show the VA estimates including the full set of covariates and a lagged achievement measure. Robust standard errors in parentheses are clustered at the school level ; *** p<0.01, ** p<0.05, * p<0.1.

4.6.2.1 Heterogeneity by EMI Grade of Transition

Finally, the study investigates the effect, if any, the grade of EMI transition has on the mathematics and English test outcomes of Grade 7 & 8 students. As previously referenced most Anglophone countries in sub-Sahara Africa transition to English medium instruction after providing approximately 4 years of mother tongue instruction (Heugh et al., 2007). The question of EMI grade of transition is then a pertinent policy question as the contention around education language policy is further complicated by the debate on the "right" grade to transition to EMI.

Specification 4.6 introduces dummy variables for students in schools transitioning to English medium instruction in Grade 1 or Grade 5, and Grade 7 EMI transition¹⁰.

$$Y_{ist} = \alpha_4 + \beta_{1.1}Grd1_{ist} + \beta_{1.2}Grd5_{ist} + \beta_{1.3}Grd7_{ist} + \beta_{\mathbf{x}}\mathbf{x}_{ist} + \beta_3y_{s,t-1} + \epsilon_{ist} \quad (4.6)$$

where $Grd1_{ist}$, $Grd5_{ist}$, $Grd7_{ist}$ are dummy variables for students in schools transitioning

¹⁰In this sample, Grade 1 and Grade 5 EMI transitions can only be found in SNNPR with Grade 7 EMI transitions in Amhara and Afar.

to EMI in Grade 1, Grade 5, and Grade 7 respectively. X_{ist} is a vector for all student, family, and school-based inputs, and $y_{is,t-1}$ the lagged achievement score.

Heterogeneity by the grade of EMI transition reveal interesting results. This paper does not focus much on students who transition to EMI in Grade 1 as there are only 3 schools applying this policy (see table 4.2). Rather, the focus is on the difference between students who transition in Grade 5 and Grade 7, as there are sufficient schools applying both policies to make meaningful comparisons.

On mathematics outcomes, both Grade 5 and 7 students experiencing English medium instruction perform somewhat below mother tongue students, although the Grade 5 effect is not statistically significant. The negative effects of EMI relative to mother tongue instruction are larger the later the transition occurs, with the negative effects becoming significant (possibly driven) by the Grade 7 EMI transition (0.2 standard deviations.). Additionally, I find that the difference between a Grade 5 and a Grade 7 student to be significantly different, $F(2,36)=9.76$, $p<0.00$. This supports the notion that the grade in which learners transition from mother-tongue instruction to EMI influences their mathematics outcomes. Estimates in column 1 captures cumulative learning while column 2 captures learning over the academic year.

Table 4.11: Heterogeneity by EMI Grade of Transition

Variables	Mathematics		English	
	(1)	(2)	(3)	(4)
Grade 1 EMI Transition	0.771*** (0.160)	0.085 (0.160)	1.801*** (0.221)	0.646** (0.254)
Grade 5 EMI Transition	-0.242 (0.144)	-0.133 (0.079)	0.540*** (0.161)	0.304*** (0.086)
Grade 7 EMI Transition	-0.271*** (0.059)	-0.220*** (0.052)	-0.026 (0.096)	-0.010 (0.059)
Lagged Mathematics Score		0.609*** (0.022)		
Lagged English Score				0.504*** (0.023)
Students	4,413	4,413	4,420	4,420
Schools	37	37	37	37

Outcome has been standardised using the control group mean and standard deviation. Column (1) includes full set of covariates and column (2) uses the VA specification. Robust standard errors in parentheses are clustered at the school level; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

On English outcomes, differences in effects by grade of transition are even starker. It appears that effects of EMI on English outcomes are actually positive with a Grade 1 or Grade 5 transition to EMI. This effect turns to zero for learners who transition in Grade 7. This is true for cumulative learning (column 3) and learning over the course of the academic year (column 4). A joint significant test between Grade 5 and Grade 7 EMI transitions reveals a significant difference with $F(2, 37) = 6.34$, $p < 0.00$. This suggests that students' English scores actually benefit if students transition early, relative to mother tongue instruction. But if students transition to EMI in higher grades, it becomes more challenging for them to perform in English tests.

This indicates that the grade of transition from mother-tongue education to English medium instruction may considerably impact student learning. These results add some nuance to the main results (EMI on average reduces student learning outcomes). This paper is able to show that an early (i.e., Grade 5 or below) could help EMI students outperform mother tongue students in English while having no real effects on mathematics outcomes. However, a later primary school EMI transition (as shown in the main results) leads to negative mathematics outcomes with no real effects on English test outcomes. One plausible reason for this may be that the later primary EMI transitions lead to some cognitive overload¹¹. Learners may be burdened by the increasing complexity of the curriculum in addition to learning in a less familiar language.

However, it is important to caveat these results by pointing out that early grade EMI transitions (Grade 1 and 5) are found only in the SNNP region, therefore it is possible that the effects observed are driven by other characteristics of the region rather than the EMI policy. Nonetheless, this nuanced finding points to the overall effect of the grade of EMI transition as an area for future research.

¹¹"Cognitive load theory assumes a limited working memory connected to an unlimited long-term memory (Braddeley, 1986). As a result of this limitation, instruction should be designed such that working memory is capable of processing the instruction" (Kirschner, 2002).

4.7 Discussion and Conclusion

Pedagogic theories indicate that an early transition to international languages such as English Medium Instruction (EMI) leads to underachievement (Alidou et al., 2006; UNESCO, 2016). Beyond pedagogical gains, mother-tongue instruction is lauded for how it centers ethnolinguistic identity (James, 2018). The main results in this paper support leading pedagogic theories and UNESCO’s recommendation on prolonging the years of mother-tongue education in schools with EMI learners performing significantly below mother-tongue learners.

The results shown in this paper are also in line with the broader economics and education literature on language of instruction and similar experimental papers from the last decade on mother-tongue instruction (Seid 2019, 2018; Ramachandran, 2017; Laitin et. al., 2019), which also find that mother-tongue instruction is causally linked to higher student achievement in mathematics outcomes. In addition, Seid (2022) finds that exposure to mother-tongue instruction in primary school also improves later labour market outcomes compared to those not exposed. However, the treatment effect decreased with years of exposure, with 4 years of mother-tongue instruction yielding the most returns (Seid, 2022). The direction of this result is similar to the suggestive evidence provided in this study where students with 5 years of mother-tongue instruction had significantly positive gains in their English test outcomes at the end of primary school.

Furthermore, a study on literacy trajectories in Ghana showed that students benefited from continuous mother tongue education. They found that learners in schools where the language of instruction was different from their mother tongue did not perform as well in literacy (Carter, Sabates, Rose & Akyeampong, 2020). Another study conducted in Malawi and Tanzania discovered that in classrooms where the language was foreign to the students and teachers, students were likely to be quieter and more scared of giving wrong answers (Alidou & Brock-Utne, 2005). The observational study by Alidou and Brock-Utne (2006) conducted in Benin, Burkina Faso, Guinea-Bissau, Mali, Mozambique, Niger, South Africa, Togo, Tanzania, Ethiopia, Ghana, and Botswana revealed that teaching using languages foreign to teachers (i.e., teaching in English),

forced them to resort to teacher-centric teaching techniques, thereby undermining teaching efforts and student learning. These mechanisms could be studied via classroom observation of the learning process in classrooms with different language policies.¹²

Another layer to this discussion is to consider the interaction of language policy and available resources. There are many determinants driving language policy choices, chief of these include teaching and learning materials and support for teachers, quality of implementation and teacher deployment policies. Similarly, James (2018) on language and learning in Ethiopia found that the variation in the effects language use was linked to the development and standardization of the language for teaching and learning purposes. Another key challenge in the execution of mother tongue instruction policies is the imperative need for the language policy to facilitate the necessary access to and proficiency in the languages of wider communication (i.e., Amharic and English) deemed as crucial within the country (James, 2021).

This study suggests a number of directions for future research. First, this paper points to the need to further research the impact EMI grade of transition has on student learning outcomes. A limitation found with this study is that the early primary grade EMI transition sample is only found in the SNNP region, this means that it is possible that some of the effects may have been driven by regional differences unrelated to the EMI policy. Thus, a larger country case study with varying grades of EMI transition would aid future research efforts. Second, the paper points to the need to study the mechanisms such as teaching techniques (through classroom observations and discussions with practitioners) as it may help to uncover why results are mixed across different country, language, and resource contexts.

This paper makes a valuable contribution to the experimental literature linking mother-tongue instruction to student attainment. Using a rich individual level database, this paper is able to evaluate 8 years of mother tongue instruction against English medium instruction on mathematics and English test outcomes. In view of the results in

¹²An Ethiopian study using classroom observation data in Ethiopia (Frost & Little, 2014) found that education policies focused on upgrading teacher qualifications, reducing class sizes, maximizing limited class time as well as increasing the allocation of qualified female teachers were important. However, the study did not consider language of instruction.

this study, this paper allows policy makers intent on an English medium education policy at the primary school level to consider an early language policy transition. Seid (2022) also showed that labour market outcomes were not affected by prolonged mother-tongue instruction. This paper's results suggest that an early EMI transition may be beneficial relative to a later transition from mother-tongue instruction. This assertion is however made conservatively within the limitations of this study.

Overall, these results make the dominant narrative around mother-tongue education and positive learning outcomes less authoritative leaving room for further research. Given the socio-political implications of language policies, there is not one optimal language policy for all multilingual contexts. As such, it is prudent for policy makers to consider context relevant evidence with regards to education language policy.

4.8 References

- Ahmad, S. (2013). Teacher education in Ethiopia: Growth and development. *African Journal of Teacher Education*.
- Alidou, H., Boly, A., Brock-Utne, B., Diallo, Y. (2006). Optimizing learning and education in Africa-The language factor. Paris: ADEA.
- Andrabi, T., Das, J., Khwaja, A.I., Zajonc, T. (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics*, Vol. 3 (3), 29-54.
- August, D., & Shanahan, T. (2008). Developing reading and writing in second-language learners: Lessons from the report of the national literacy panel on language-minority children and youth. Taylor & Francis.
- Barkhuizen, G. P. (2002). Language-in-education policy: students' perceptions of the status and role of Xhosa and English. *Science direct* 30(4), 499-515.
- Begi, N. (2014). Use of Mother Tongue As A Language Of Instruction In Early Years Of School To Preserve The Kenyan Culture. *Journal of Education and Practice*, 5, 37-49.
- Benson, C. (2000). The primary bilingual education experiment in Mozambique: 1993 to 1997. *International Journal of Bilingual Education and Bilingualism* 3, 149-166.
- Benson, C. (2004). The Importance of Mother Tongue-Based Schooling for Educational Quality. Background paper for the 2005 Education for All Global Monitoring Report, UNESCO, Paris.
- Brock-Utne, B. (2005c). Learning through a Familiar Language versus Learning through a Foreign Language - a Look into some Secondary School Classrooms in Tanzania. Paper presented at the CIES 2005 Conference at Stanford University 22- 25 March.
- Brock-Utne, B., & Holmarsdottir, H.B. (2004). Language policies and practices in Tanzania and South Africa: problems and challenges. *International Journal of Educational Development*, 24, 67-83.
- Bunyi, G. (2005). "Language Practices in Kenya." In *Decolonisation, Globalisation:*

-
- Language in Education Policy and Practice, edited by A. Lin and P. Martin, 131-152. Clevedon: Multilingual Matters.
- Carter, E., Sabates, R., Rose, P., & Akyeampong, K. (2020). Sustaining literacy from mother tongue instruction in complementary education into official language of instruction in government schools in Ghana. *International Journal of Educational Development*, 76. <https://doi.org/10.1016/j.ijedudev.2020.102195>
- Chitera, N. (2011). Language of learning and teaching in schools: An issue for research in mathematics teacher education? *Journal of Mathematics Teacher Education*, 14(3), 231-246. <https://doi.org/10.1007/s10857-010-9167-3>
- Clegg, J., & Simpson, J. (2016). Improving the effectiveness of English as a medium of instruction in sub-Saharan Africa. *Comparative Education*, 52(3), 359-374. <https://doi.org/10.1080/03050068.2016.1185268>
- Cummins, J. (2000). *Language, Power, and Pedagogy: Bilingual Children in the Cross-fire*. Vol. 23. Bristol, UK: Multilingual Matters.
- Cummins, J. (1984). *Bilingualism and Special Education: Issues in Assessment and Pedagogy*. Clevedon: Multilingual Matters.
- Damon, A., Glewwe, P., Wisniewski, S., & Sun, B. (2019). What education policies and programmes affect learning and time in school in developing countries? A review of evaluations from 1990 to 2014. *Review of Education*, 7(2), 295-387.
- Desai, Z., & Qorro, M. A. (2006). Focus on fresh data on the language of instruction debate in Tanzania and South Africa: *African Minds*.
- 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* 29:2, pages 311-335.
- Frost, M., Little, A.W. (2014). Children's learning practices in Ethiopia: observations from primary school classes, *Oxford Review of Education*, 40:1, 91-111.
- Garcia, G. (2000). Bilingual children's reading. *Handbook of Reading Research*, 38, 13-834.
-

-
- Gfeller, E. , & Robinson, C. (1998). Which language for teaching? The cultural messages transmitted by the languages used in education. *Language and Education*, 12 (1), 18-32, DOI: 10.1080/09500789808666737.
- Gronmo, L.S., Lindquist, M., Arora, A. and Mullis, I.V.S. (2015) TIMSS 2015 Mathematics Framework. Boston, MA: TIMSS and PIRLS International Study Centre.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). How Long Does It Take English Learners To Attain Proficiency?
- Heugh, K. A. (2008). *Encyclopedia of Language and Education*. Encyclopedia of Language and Education.
- Heugh, K., Benson, C., Bogale, B., & Yohannes, M. A. G. (2007). Final report: Study on medium of instruction in primary schools in Ethiopia.
- Ede O. S. Iyamu, & Sam. E. Aduwa Ogiegbaen (2007) Parents and Teachers' Perceptions of Mother-Tongue Medium of Instruction Policy in Nigerian Primary Schools, *Language, Culture and Curriculum*, 20:2, 97-108, DOI: 10.2167/lcc328.0.
- Jamil, B., and S. Saeed. (2015). "Reflections on Access and Learning within the Low-Cost Private English Medium Sector in Pakistan." Unpublished paper presented at the 11th International Language and Development Conference, Delhi, November 14-16, 2015.
- Laitin, D., Ramachandran, R., & Walter, S. L. (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(1), 239-272.
- Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2018). A systematic review of English medium instruction in higher education. *Language Teaching*, 51(1), 36-76. DOI: 10.1017/S0261444817000350
- Macaro, E. (2018). *English medium instruction: Language and content in policy and practice*. Oxford: Oxford University Press.
- McEwan, Patrick J. (2014). "Improving Learning in Primary Schools of Developing Countries A Meta-Analysis of Randomized Experiments." *Review of Educational Research*, (in press).
-

-
- Mwinsheikhe, H.M. (2003). Science and the Language Barrier: Using Kiswahili as a Medium of Instruction in Tanzania Secondary Schools as a Strategy of Improving Student Participation and Performance in Science. In: Language of Instruction in Tanzania and South Africa(LOITASA), ed. by Birgit Brock-Utne, Zubeida Desai and Martha Qorro, 129-149. Dar es Salaam: E & D Publishers.
- Nyaga, S., & Anthonissen, C. (2012) Teaching in linguistically diverse classrooms: difficulties in the implementation of the language-in-education policy in multilingual Kenyan primary school classrooms, *Compare: A Journal of Comparative and International Education*, 42:6, 863-879, DOI: 10.1080/03057925.2012.707457.
- Piper, Benjamin & Miksic, E. (2011). Mother Tongue and Reading: Using Early Grade Reading Assessments to investigate Language-of-instruction Policy in East Africa. *The Early Grade Reading Assessment: Applications and Interventions to improve Basic Literacy*. Research Triangle Park, NC: RTI Press, 139-182.
- Piper, B., Zuilkowski, S. S., & Ong'Ele, S. (2016). Implementing mother tongue instruction in the real world: Results from a medium-scale randomized controlled trial in Kenya. *Comparative Education Review*, 60(4), 776-807. <https://doi.org/10.1086/688493>
- Prophet, R., & Dow, J. (1994). Mother Tongue Language and Concept Development in Science. A Botswana Case Study. *Language, Culture and Curriculum* 7(3): 205-217.
- Ramachandran, R. (2017). "Language Use in Education and Human Capital Formation: Evidence from the Ethiopian Educational Reform." *World Development* 98: 195-213.
- Rossiter, J., Azubuike, O. B., & Rolleston, C. (2017). Young Lives School Survey, 2016-17: Evidence from Vietnam. *Young Lives*.
- Seid, Y. (2016). Does learning in mother tongue matter? Evidence from a natural experiment in Ethiopia. *Economics of Education Review*, 55,21-38.
- Seid, Y. (2019). The impact of learning first in mother tongue: evidence from a natural experiment in Ethiopia. *Applied Economics*, 51(6), 577-593.
- Seid, Y (2022) Mother-tongue instruction and later labour market outcomes: evidence from a natural experiment in Ethiopia, *Applied Economics*, 54:54, 6254-6271, DOI:
-

- Singh, A. (2015). Private school effects in urban and rural India: Panel estimates at primary and secondary school ages. *Journal of Development Economics*, 113, 16-32. <https://doi.org/10.1016/j.jdeveco.2014.10.004>
- Skutnabb-Kangas, T., and K. Heugh. 2013. *Multilingual Education and Sustainable Diversity Work: From Periphery to Center*. New York, USA: Routledge.
- Slavin, R. E., Madden, N., Calderon, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, 33(1), 47-58. <https://doi.org/10.3102/0162373711398127>
- Taylor, S., & von Fintel, M. (2016). Estimating the impact of language of instruction in South African primary schools: A fixed effects approach. *Economics of Education Review*, 50, 75-89.
- Todd, P.E., Wolpin, K.I. (2003). On the specification and estimation of the production function for cognitive achievement. *Economic Journal*, 113 (485), F3-F33.
- UNESCO. (2010). *Enhancing learning of children from diverse language backgrounds: mother tongue-based bilingual or multilingual education in early years*.
- UNESCO. (2016). *If you don't understand, how can you learn? Policy Paper 24*.
- Vujcich, D. (2013). *Policy and Practice on Language of Instruction in Ethiopian Schools. Young Lives*.
- Woldemariam, H. (2007). The challenges of mother-tongue education in Ethiopia: The case of North Omo area, *Language Matters*, 38:2, 210-235.
- James, Z. C. (2018). *Language and learning in Ethiopia* (Doctoral dissertation, UCL (University College London)).
- James, Z. (2021). Stakeholder Perspectives on Medium of Instruction Policy in Ethiopia. In *Language Issues in Comparative Education II* (pp. 146-162). Brill.

5 | Conclusion

Quality education and mental health have intrinsic value and are fundamental for human capital accumulation and sustainable development. Through this thesis, I explore questions pertaining to the mental health and educational success outcomes of young people in the Global South. The first article studies the impact of an education intervention on the mental health and education outcomes of primary school learners in Botswana. The second article studies the impact of a mental health intervention on the mental health and education outcomes of secondary school learners in Pakistan. Finally, the third article focuses on the impact of a popular education language policy on the education outcomes of upper primary school learners in Ethiopia.

Using a combination of experimental and quasi-experimental methods, my main findings are summarised as follows. First, I find empirical evidence linking the foundational learning intervention, Teaching at the Right Level (TaRL) to improved mental health outcomes. My underlying hypothesis is that improving the school climate (through TaRL), improves student mental health. My main results support this hypothesis. I find that TaRL students have significantly less emotional and behavioural total difficulties and exhibit more prosocial behaviour compared to their non-TaRL counterparts. Although, my study was not designed to directly test for mechanisms, these results are in line with the literature on student-centred pedagogies. This literature finds that teacher support, a sense of belonging and school connectedness, and students' experiences of bullying behaviour as mediating factors. Consequently, these findings are also in line with the descriptive data shared by teachers in the study sample. According to teachers, TaRL students were more confident, open, exhibited improved self-esteem and behaviour.

Second, I find that it is possible to adapt a mindfulness-based curriculum to a low-resource setting. Mindfulness-based Stress Reduction (MBSR) has a strong evidence base in high resource or high-income settings. Together with my co-author (Dr Zahra Mansoor), we are able to work collaboratively with an Urdu-based mindfulness NGO to develop and test an MBSR programme in Pakistan. We find that MBSR significantly improves student prosociality and increases grit. We also find that MBSR students are

able to foster a more positive home environment, all relative to other students not in the programme.

Third, through the language policy study, my main findings show that English Medium policies do not necessarily lead to academic attainment, particularly in mathematics. My results show that English medium students fare worse than their mother tongue education counterparts in mathematics and do not perform any better in English test scores. However, when exploring the grade at which students transition into English medium instruction, the results tell a different story. Early exit from mother tongue education may lead to improved English test scores and does no harm to math test scores. Overall, these results add some nuance to the on-going language policy debates, specifically on the priorities within the education system.

With respect to the scalability, cost-effectiveness, and the generalisability of the findings, TaRL is currently being scaled up in multiple African countries where similar conditions of poor learning outcomes persist. Given that the underlying assumptions in Botswana hold, and possibly even stronger in the other African countries, it can be argued that effects would be similar if not greater. Future research would focus on a larger scale evaluation in multiple countries testing potential mechanisms such structured play, small group activities, and behaviour management techniques. These mechanisms, as found in the child and adolescent school climate literature improve both educational and psycho-social outcomes of young people. Additionally, targeted instruction interventions like TaRL have been recommended by the Global Education Evidence Advisory Panel (GEEAP) ‘as a great buy’ in comparison to good buys and bad buys in the 2023 Cost-Effective Approaches to Improve Global Learning report. This underscores the importance of generating more context-relevant research on these foundational learning interventions.

With regards to questions around scalability and cost-effectiveness in the Pakistan study, mindfulness-based interventions in adult populations have been confirmed as cost-effective and or cost saving in comparison to usual standard care including cognitive-behavioural therapy, or non-mindfulness-based treatments. However, given

that the findings presented in this thesis pertain to young adolescents in a small-scale study designed for learning and adapting the suitable model for broader implementation, future studies will strive to assess the cost-effectiveness of the model intended for scaling. Finally, in the Ethiopia study, the transition from mother tongue instruction to English medium instruction in the early years of primary school has become a prevailing policy practice across many African countries. As such, the insights derived from this thesis are applicable beyond the Ethiopia case. The findings show that given the socio-political implications of language and its usage, there exists no universal optimal language policy suitable for all multilingual contexts. Consequently, it is advisable for policymakers to attentively consider context-specific evidence when formulating education language policy.

Collectively, these three articles improve our understanding on the range of mental health and education interventions or reforms addressing the mental health and educational success of young people in the Global South. The results are policy relevant and the findings offer rigorous causal evidence for under-served and under-researched populations.

A | Foundational Learning and Mental Health: Appendices

1. Appendix A

Figure A.1: Mental health related terms

	Working Definitions
Mental health	“a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively, and is able to make a contribution to his or her community” (WHO, 2001a) “an asset or resource that enables positive states of wellbeing and provides the capability for people to achieve their full potential” (Patel et al., 2018)
Mental disorder	Disturbances of thought, emotion, behaviour, and relationships with others that lead to substantial suffering and functional impairment in one or more major life activities (WHO, 2004); identified by the WHO International Classification of Diseases and the Diagnostic and Statistical Manual of Mental Disorders (Patel et al., 2018)
Externalising mental health problems	Disruptive behaviours, inattention/hyperactivity, other behavioural and emotional difficulties (Becker et al., 2014; Garcia-Escalera et al., 2020)
Internalising mental health problems	Anxiety, depression, trauma (Becker et al., 2014; Garcia-Escalera et al., 2020)

2. Appendix B

Figure A.2: Number of Primary schools by ownership and Region

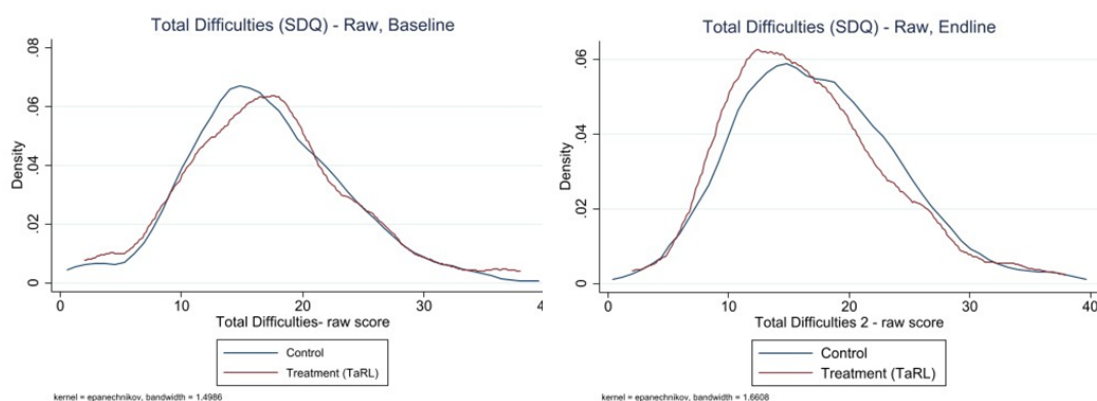
Table 2. 1: Number of Primary Schools by Ownership and Region – 2017

Region	Public	%	Government Aided	%	Private	%	Total	%
South East	44	6	4	40	24	37	72	9
North East	61	8	2	20	5	8	68	8
South	122	16	1	10	6	9	129	16
Kweneng	92	12	-	-	9	14	101	12
Kgatleng	37	5	-	-	1	2	38	5
North West	68	9	-	-	5	8	73	9
Chobe	10	1	-	-	1	2	11	1
Gantsi	22	3	-	-	1	2	23	3
Kgalagadi	42	6	-	-	-	-	42	5
Central	248	33	3	30	13	20	264	32
Total	746	100	10	100	65	100	821	100

Source: Statistics Botswana, May 2021. *Primary School Stats Brief, 2017*.

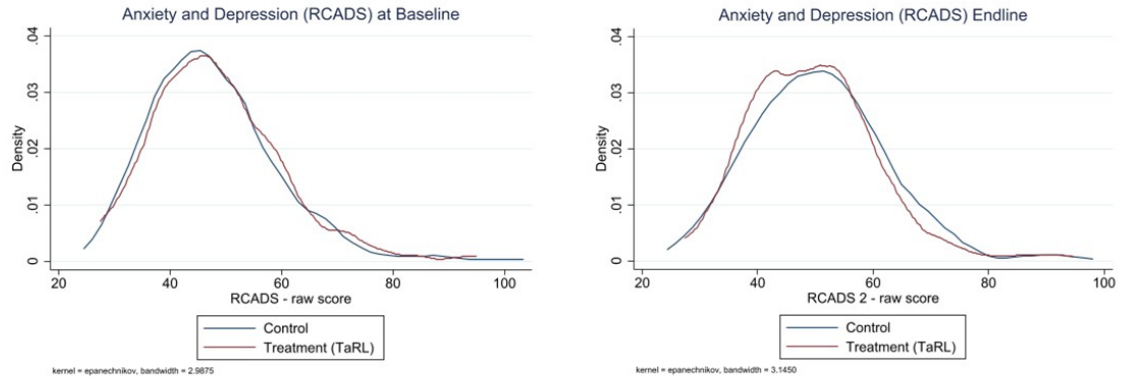
3. Appendix C

Figure A.3: Kdensity plots - Total Difficulties (SDQ), raw scores



4. Appendix D

Figure A.4: Kdensity plots - Anxiety and Depression score (RCADS), raw scores



5. Appendix E

Figure A.5: Impact of TaRL on Mental Health Outcomes, raw scores

	(1) SDQ	(2) Prosocial	(3) Anxiety	(4) Depression	(5) RCADS
TaRL	0.147 (0.713)	-0.445** (0.210)	0.586 (1.109)	0.610 (1.115)	0.615 (1.168)
Period	0.468 (0.466)	-0.195 (0.183)	1.626** (0.758)	2.369*** (0.793)	2.075** (0.794)
TaRL x Period	-1.021* (0.594)	0.603** (0.247)	-1.085 (1.019)	-1.046 (1.203)	-1.158 (1.110)
Control mean	17.098	7.847	50.225	47.480	49.114
Controls	Yes	Yes	Yes	Yes	Yes
School FE	Yes	Yes	Yes	Yes	Yes
Students	1272	1273	1286	1286	1286

Notes: regression estimates on standardised mental health and wellbeing outcomes. Coefficients of interest Are reported with robust standard errors in parentheses, clustered at the class level; *, **, and *** denote significance at the 10; 5; and 1 percent levels respectively. Controls include student age, sex, and standard variables

6. Appendix F

Figure A.6: Teacher reported student difficulties

	Composite SDQ (raw)	Composite SDQ (standardised)
TaRL	-0.096 (0.155)	-0.088 (0.142)
Control mean	2.680	0
Controls	Yes	Yes
Students	1,262	1,262

Robust standard errors in parentheses. Controls include student age, sex, and standard variables; *** p<0.01, ** p<0.05, * p<0.1

7. Appendix G

Figure A.7: Impact of TaRL on Educational success outcomes (raw)

	Math Test Score	Attendance
TaRL	-4.093 (3.258)	-0.141 (0.300)
Period	6.850*** (1.651)	0.297 (0.190)
TaRL x Period	3.104 (2.244)	-0.270 (0.478)
Control mean	54.125	99.403
Controls	Yes	Yes
School FE	Yes	Yes
Students	1126	982

Robust standard errors in parentheses. Controls include student age, sex, and standard variables; *** p<0.01, ** p<0.05, * p<0.1

8. Appendix H

Figure A.8: Student tool

STUDENT QUESTIONNAIRE				
North East Francistown School Name:		Are you a girl or a boy		Girl
Full name:				Boy
Are you a TaRL student? ☐ YES ☐ NO		Age:	Baseline	Endline
Instruction: Please select the answers that are true for you. It would help us if you answered all the items as best as you can even if you are not absolutely certain, or the question seems silly. Please answer based on how you have been feeling over the past three months (over the last school term). If you do not want to answer any of the questions, no one will be upset with you. Thank you for your time. Are you ready to start?				
1	I try to be nice to other people. I care about their feelings	☐ Not True	☐ Somewhat True	☐ Certainly True
2	I am restless, I cannot stay still for long	☐ Not True	☐ Somewhat True	☐ Certainly True
3	I get a lot of headaches, stomach-aches, or sickness	☐ Not True	☐ Somewhat True	☐ Certainly True
4	I usually share with others (food, games, pens/pencils, etc.)	☐ Not True	☐ Somewhat True	☐ Certainly True
5	I get very angry and often lose my temper	☐ Not True	☐ Somewhat True	☐ Certainly True
6	I am usually on my own. I generally play alone or keep to myself	☐ Not True	☐ Somewhat True	☐ Certainly True
7	I usually do as I am told	☐ Not True	☐ Somewhat True	☐ Certainly True
8	I worry a lot	☐ Not True	☐ Somewhat True	☐ Certainly True
9	I am helpful if someone is hurt, upset, or feeling ill	☐ Not True	☐ Somewhat True	☐ Certainly True
10	I am constantly fidgeting or squirming	☐ Not True	☐ Somewhat True	☐ Certainly True
11	I have one good friend or more	☐ Not True	☐ Somewhat True	☐ Certainly True
12	I fight a lot. I can make other people do what I want	☐ Not True	☐ Somewhat True	☐ Certainly True
13	I am often unhappy, down-hearted or tearful	☐ Not True	☐ Somewhat True	☐ Certainly True
14	Other people my age generally like me	☐ Not True	☐ Somewhat True	☐ Certainly True
15	I am easily distracted; I find it difficult to concentrate	☐ Not True	☐ Somewhat True	☐ Certainly True
16	I am nervous in new situations. I easily lose confidence	☐ Not True	☐ Somewhat True	☐ Certainly True
17	I am kind to younger children	☐ Not True	☐ Somewhat True	☐ Certainly True
18	I am often accused of lying or cheating	☐ Not True	☐ Somewhat True	☐ Certainly True
19	Other children or young people pick on me or bully me	☐ Not True	☐ Somewhat True	☐ Certainly True

20	I often volunteer to help others (parents, teachers, children)	☐ Not True	☐ Somewhat True	☐ Certainly True	
21	I think before I do things	☐ Not True	☐ Somewhat True	☐ Certainly True	
22	I take things that are not mine from home, school or elsewhere	☐ Not True	☐ Somewhat True	☐ Certainly True	
23	I get on better with adults than with people my own age	☐ Not True	☐ Somewhat True	☐ Certainly True	
24	I have many fears, I am easily scared	☐ Not True	☐ Somewhat True	☐ Certainly True	
25	I finish the work I'm doing. My attention is good	☐ Not True	☐ Somewhat True	☐ Certainly True	
26	I feel sad or empty	☐ Never	☐ Sometimes	☐ Often	☐ Always
27	I worry when I think I have done poorly at something	☐ Never	☐ Sometimes	☐ Often	☐ Always
28	I would feel afraid of being on my own at home	☐ Never	☐ Sometimes	☐ Often	☐ Always
29	Nothing is much fun anymore	☐ Never	☐ Sometimes	☐ Often	☐ Always
30	I worry that something awful will happen to someone in my family	☐ Never	☐ Sometimes	☐ Often	☐ Always
31	I am afraid of being in crowded places (like shopping centers, the movies, buses, busy playgrounds)	☐ Never	☐ Sometimes	☐ Often	☐ Always
32	I worry what other people think of me	☐ Never	☐ Sometimes	☐ Often	☐ Always
33	I have trouble sleeping	☐ Never	☐ Sometimes	☐ Often	☐ Always
34	I feel scared if I have to sleep on my own	☐ Never	☐ Sometimes	☐ Often	☐ Always
35	I have problems with my appetite	☐ Never	☐ Sometimes	☐ Often	☐ Always
36	I suddenly become dizzy or faint when there is no reason for this	☐ Never	☐ Sometimes	☐ Often	☐ Always
37	I have to do some things over and over again (like washing my hands, cleaning or putting things in a certain order)	☐ Never	☐ Sometimes	☐ Often	☐ Always
38	I have no energy for things	☐ Never	☐ Sometimes	☐ Often	☐ Always
39	I suddenly start to tremble or shake when there is no reason for this	☐ Never	☐ Sometimes	☐ Often	☐ Always
40	I cannot think clearly	☐ Never	☐ Sometimes	☐ Often	☐ Always
41	I feel worthless	☐ Never	☐ Sometimes	☐ Often	☐ Always
42	I have to think of special thoughts (like numbers or words) to stop bad things from happening	☐ Never	☐ Sometimes	☐ Often	☐ Always
43	I think about death	☐ Never	☐ Sometimes	☐ Often	☐ Always
44	I feel like I don't want to move	☐ Never	☐ Sometimes	☐ Often	☐ Always
45	I worry that I will suddenly get a scared feeling when there is nothing to be afraid of	☐ Never	☐ Sometimes	☐ Often	☐ Always
46	I am tired a lot	☐ Never	☐ Sometimes	☐ Often	☐ Always

47	I feel afraid that I will make a fool of myself in front of people	☐ Never	☐ Sometimes	☐ Often	☐ Always
48	I have to do some things in just the right way to stop bad things from happening	☐ Never	☐ Sometimes	☐ Often	☐ Always
49	I feel restless	☐ Never	☐ Sometimes	☐ Often	☐ Always
50	I worry that something bad will happen to me	☐ Never	☐ Sometimes	☐ Often	☐ Always
Facilitator Name:			Date of completion:		

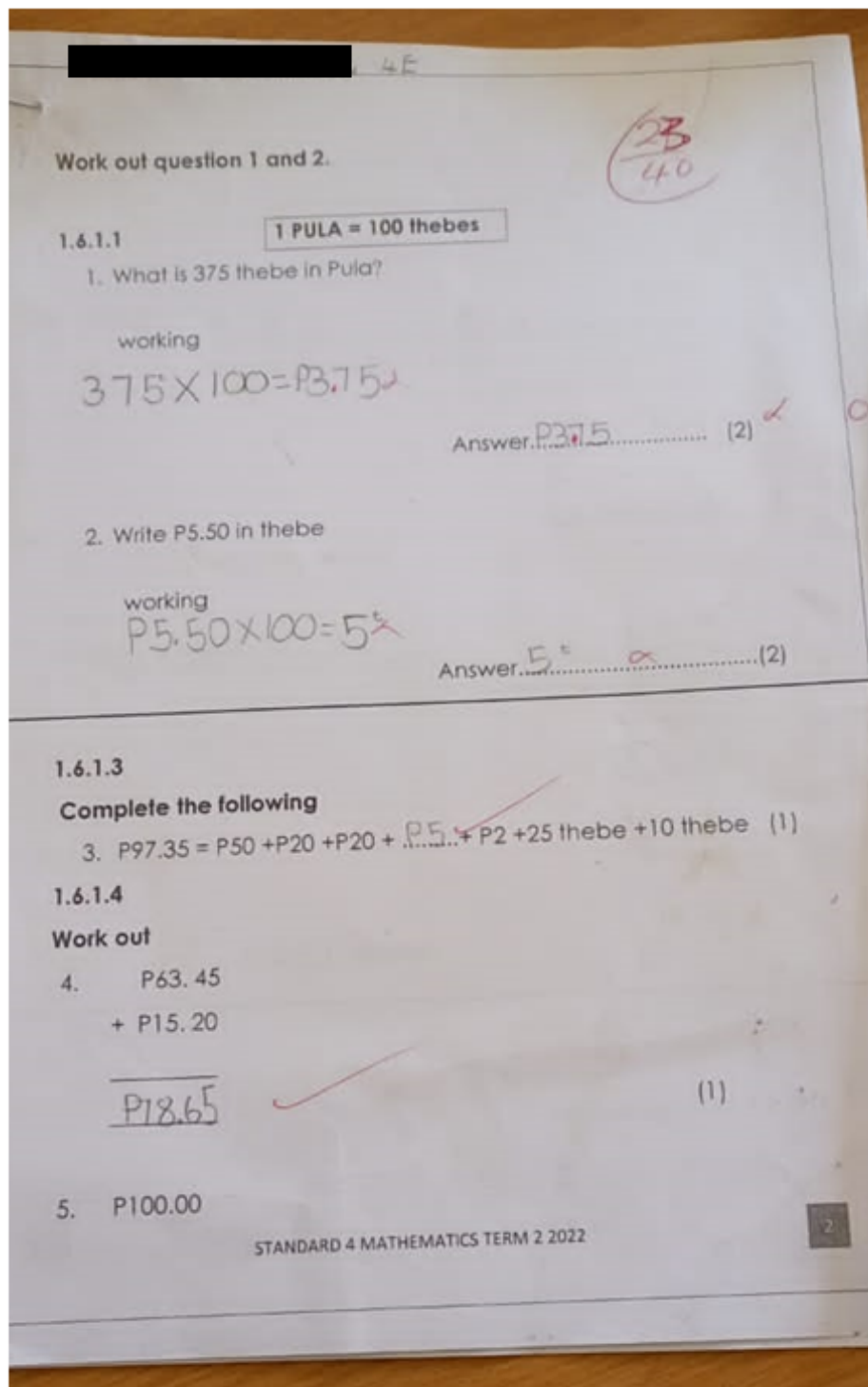
9. Appendix I

Figure A.9: Teacher Tool on observed student behaviour

TEACHER QUESTIONNAIRE				
Student Name:		Teacher Name:		Date:
Std. and class:	Baseline / Endline	Is your student a TaRL student?		Yes / No
For each question, select Not True, Somewhat True or Certainly True. Please give your answers on the basis of how things have been this year (2022) for the student. The student...				
Questions				
1	Shares readily with other children (treats, toys, pencils etc.) O kgaogana sentle le ba bangwe(dinekere, diToy, dipensele etc)	☐ _A_] Not True	☐ _B_] Somewhat True	☐ _C_] Certainly True
2	Often fights with other children or bullies them Gantsi o lwa le bana ba bangwe kana o a ba tshwenya	☐ _A_] Not True	☐ _B_] Somewhat True	☐ _C_] Certainly True
3	Is easily distracted, concentration wanders O iteega tsebe motlhofo, o eta mogopolo	☐ _A_] Not True	☐ _B_] Somewhat True	☐ _C_] Certainly True

10. Appendix J

Figure A.10: Sample of student exam



8. Ms Neo gave $\frac{5}{8}$ of a cake to Shawn, she gave $\frac{3}{8}$ of the cake to her son, what fraction of the cake did Ms Neo give to her children altogether?

Working

$$\frac{5}{8} + \frac{3}{8} = \frac{8}{8}$$

Answer... $\frac{8}{8}$ (2)

1.4.1.5

Work out the following

9. $350 \div 10 = \boxed{35}$

(1)
$$\begin{array}{r} 35 \\ 10 \overline{) 350} \\ \underline{-30} \\ 50 \\ \underline{-50} \\ 0 \end{array}$$

1.4.14

10. $428 \div 5 =$

Working

$$\begin{array}{r} 85 \\ 5 \overline{) 428} \\ \underline{-40} \\ 28 \\ \underline{-25} \\ 30 \\ \underline{-30} \\ 0 \end{array}$$

Answer... 85 (2)

$$11. \frac{3}{6} + \frac{2}{6} = \frac{5}{6}$$

(1) ✓

The table below shows a bill. Use it to answer question 12

1.6.1.6

ITEM	COST PER ITEM	NUMBER OF ITEMS	TOTAL COST
Shirt	P21.50	2	P21.50
Pair of trousers	P36.75	1	P36.75
		TOTAL	P58.25

12. What is the total cost of all the items?

working

$$\begin{array}{r} P21.50 \\ + P36.75 \\ \hline P58.25 \end{array}$$

Answer: P58.25 (2)

Use the information given below to answer question 13.

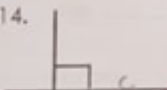
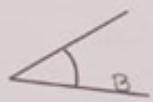
$$10\text{mm} = 1\text{cm}$$

3.1.1.6

13. Convert **140 mm** to cm

$$140\text{ mm} = \dots \text{1.40 cm} \dots \text{cm (1)}$$

Match the angles with their correct names

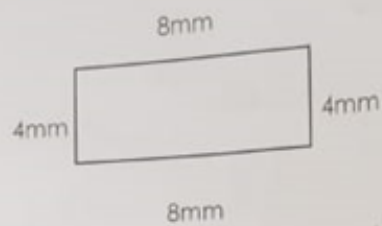
ANGLES	NAMES
14. 	A. Obtuse angle
15. 	B. Acute angle
16. 	C. Right angle

3.1.1.4

STANDARD 4 MATHEMATICS TERM 2 2022



17. Find the perimeter of the following shape



Answer 24mm ✓ (1) |

1.2.1.6

Use the list of numbers below to answer question 18 and 19.

8, 9, 10, 11, 12

18. Which of the numbers are odd numbers?

9, 11 ✓ (2) 2

19. Which of the numbers are multiples of 4?

8, 12 ✓ (2) 2

1.2.1.3

20. Write **704** in words

Answer Seven hundred and four ✓ (1) |

21. Write **one hundred and forty five** in numerals.

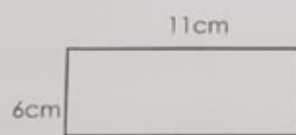
STANDARD 4 MATHEMATICS TERM 2 2022

7 (4)

Answer...145.....✓ (1) /

3.2.1.2

22. Calculate the area of the rectangle below



Working

L x W ✓

$$11\text{cm} \times 6\text{cm} = 2\text{km}$$

Answer...2km...✗ (2) /

23. Find the area of the square below

25. From the given containers, which one has more capacity?



Answer. B ~~A~~ C (1) 0

3.4.1.3

26. Neo bought 9 kg of sugar and gave her neighbour 3 kg. What is the mass of sugar left for Neo?

Working

$$9\text{ kg} - 3\text{ kg} = 6\text{ kg}$$

Answer. 6 kg (2) 2

27. Gaba has 10 kg of potatoes, Reba has 5 kg of potatoes and Chedu has 7 kg of potatoes. What is the total mass of all the bags of potatoes?

Working

$$\begin{array}{r} 10\text{ kg} \\ 5\text{ kg} \\ + 7\text{ kg} \\ \hline 22\text{ kg} \end{array}$$

Answer. 22 kg (2) 2

MARKING KEY MATHEMATICS STD4

STANDARD 4 MATHEMATICS TERM 2 2022

10

4

11. Appendix K

Ethics Discussion- Brief Explanations

As recommended by Asiedu et al. (2022) and seen in Crawford et al. (2023), the study reflects on the following ethical questions when collecting of primary data.

(1) Policy equipoise. Policy equipoise refers to uncertainty regarding the net benefits of the policy/intervention. In the case of this research study, this is the first study that empirically investigates the impact of the foundational learning intervention (TaRL) on the mental health of young people. As such this evidence generation was necessary.

(2) The role of researchers with respect to implementation. As the solo author of this study, I did not have any direct decision-making power over how or whether to implement the intervention. This study was fully independent from the programme implementers.

(3) Potential harms to research participants. The intervention under study does not pose any potential harm to participants or non-participants.

(4) Potential harms from data collection. The research was approved by the Botswana Ministry of Basic Education. My data collection team did not observe any distress and the research was respectful of community norms. The research study obtained written consent from parents/caretakers of all the children in the study, and assent was obtained by the children prior to data collection. The data collection team did not observe any distress from the children during the data collection periods but there was a risk assessment protocol in place following the guidelines recommended by Principal Education Officer. In addition, the was approved by the institutional research ethics committee at the University of Oxford (UK, Ref No SSH/BSG-C1A-22-01, February 7th, 2022).

(5) Financial and reputational conflicts of interest. There were no financial conflicts of interest with respect to results of the research. Also, there was no potential reputational conflicts of interest with respect to the results of the research. That said, I did previously work as the research and programme manager for the

implementing organisation (Youth-Impact). However, at the time of this research study, I had not been employed by the organisation for 3 years.

(6) Intellectual freedom. There are no contractual limitations regarding the reporting of this research. Research funds were raised independently, and the implementing organisation was not involved in the analysis or interpretation of the data.

(7) Feedback to participants or communities. A research report has been shared with the Regional Education office that approved the study. Pending additional funds, a dissemination event will be held in the region to further disseminate the findings in the community.

(8) Foreseeable misuse of research results. There is no foreseeable risk that the results of this research will be misused or deliberately misinterpreted by interested parties to the detriment of other interested parties.

B | Mindfulness: Appendices

1. Appendix A

Figure B.1: Combined index score utilising a means score of each outcome

Table 8: MBSR treatment effects on behaviour change

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Appetite	Improved mood	Less arguments w/ siblings)	Less arguments parents	Focus on school work	Less stress at home	Combined Index
Treatment	0.185 (0.125)	0.515** (0.168)	0.265 (0.161)	0.335 (0.195)	0.156 (0.089)	0.336*** (0.106)	0.299*** (0.077)
Observations	208	208	208	208	208	208	208
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Each column represents a different behaviour change dependent variable as indicated by the column header. All outcomes are standardised against the mean and standard deviation of the control group. Standard errors are in parentheses, clustered at class level. Adjust q-values are as follows: .178, .104, .051, .018, .142, .086. Estimates are significant at the *10%, **5%, and ***1% level..

The combined index score equally weights each behaviour change dependent variable.

1. Appendix A

Figure C.1: The 1994 Language Policy in relation to the Constitution

Constitution of the Federal Democratic Republic of Ethiopia (1994)	Education and Training Policy (1994)
Article 5: All Ethiopian languages shall enjoy equal state recognition. ▪ Amharic shall be the working language in the federal government. ▪ Members of the federation may by law determine their respective work languages. Article 39: Rights of Nations, Nationalities and Peoples: ▪ Every Nation, Nationality and People in Ethiopia has an unconditional right to self-determination, including the right to secession. ▪ Every Nation, Nationality and People in Ethiopia has the right to speak, to write and to develop its own language; to express, to develop and to promote its culture and to preserve its history.	3.5 LANGUAGES AND EDUCATION 3.5.1 Cognizant of the pedagogical advantage of the child in learning in mother tongue and the rights of nationalities to promote the use of their languages, primary education will be given in nationality languages. 3.5.2 Making the necessary preparation, nations and nationalities can either learn in their own language or can choose from among those selected on the basis of national and countrywide distribution. 3.5.3. The language of teacher training for kindergarten and primary education will be the nationality language used in the area. 3.5.4 Amharic shall be taught as a language of countrywide communication. 3.5.5 English will be the medium of instruction for secondary and higher education. 3.5.6 Students can choose and learn at least one nationality language and one foreign language for cultural and international relations. 3.5.7 English will be taught as a subject starting from Grade one. 3.5.8 The necessary steps will be taken to strengthen language teaching at all levels.

source: Heugh et al., 2007

2. Appendix B

Figure C.2: Language of Instruction used in Primary Schooling and in Teacher Training

Region	Primary Grade			Teacher Training	
	Grade 1 - 4	Grade 5 - 6	Grade 7 - 8	Grade 1 - 4	Grade 5 - 8
SNNPR	Amharic, Dawro, Gamo, Gedeo, Gofa, Hadiyya, Kembata, Kafinono, Kontigna, Korete, Sidama, Silti, Wolatya Hadiya Sidama Wolayta	English	English (all subjects)	Amharic, Dawro, Gamo, Gedeo, Gofa, Hadiyya, Kembata, Kafinono, Kontigna, Korete, Sidama, Silti, Wolatya Hadiya Sidama Wolayta	English
Amhara	Amharic, Awingi, Hammittena, Oromiffa	Amharic, Awingi, Hammittena, Oromiffa	English for mathematics and science Amharic, Awingi, Hammittena, Oromiffa for the other subjects	Amharic, Awingi	English
Afar	Amharic Afar (ABE*)	Amharic	English (all subjects)		
Oromia	Oromiffa, Amharic				
Tigray	Tigringna				
Somali	Somali, Amharic	Somali, Amharic	English for mathematics and science Somali for the other subjects		

*ABE= Alternate Basic Education ; Source: adapted from Heugh, 2007 & Vujcich, 2013.

3. Appendix C

Figure C.3: Distribution of EMI by Grade in Somali Region

In which primary grade does this school start to use English as the language of instruction?				
Grade	Schools	Students	Percent	Cum.
Grade 1	3	335	13.03	13.03
Grade 3	1	139	5.41	18.44
Grade 5	3	339	13.19	31.62
Grade 7	3	883	34.34	65.97
Mother tongue only	4	875	34.03	100
Total		2,571	100	

4. Appendix D

Figure C.4: Impact of EMI on Math Outcomes (winsorized 5% in each tail)

	(1)	(2)	(3)
EMI	-0.081 (0.127)	-0.255*** (0.075)	-0.170*** (0.044)
Lagged math score			0.621*** (0.022)
Control mean	0	0	0
Schools	38	38	38
Observations	4,413	4,413	4,413

Outcome has been standardised using the control group mean and standard deviation. Columns 2 and 3 regress test scores on full set of covariates. Column 3 specifies the value-added model including a lagged achievement score. Robust standard errors in parentheses are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1

5. Appendix E

Figure C.5: Impact of EMI on English Outcomes (winsorized 5% in each tail)

	(1)	(2)	(3)
EMI	0.394** (0.185)	0.147 (0.101)	0.079 (0.057)
Lagged English score			0.518*** (0.025)
Control mean	0	0	0
Schools	38	38	38
Observations	4,420	4,420	4,420

Outcome has been standardised using the control group mean and standard deviation. Columns 2 and 3 regress test scores on full set of covariates. Column 3 specifies the value-added model including a lagged achievement score. Robust standard errors in parentheses are clustered at the school level . *** p<0.01, ** p<0.05, * p<0.1

6. Appendix F

Figure C.6: EMI Treatment Heterogeneity by the English Language Improvement Programme (ELIP)

	Mathematics	English
EMI	-0.223*** (0.050)	0.077 (0.072)
EMI * ELIP (Mathematics teachers)	0.159 (0.103)	0.051 (0.091)
ELIP (Mathematics teachers)	-0.118 (0.075)	
Lagged Mathematics Score	0.612*** (0.022)	
EMI * ELIP (English teachers)	0.017 (0.125)	
ELIP (English teachers)		0.066 (0.064)
Lagged English Score		0.518*** (0.025)
Control mean	0	0
Controls	Yes	Yes
Schools	38	38
Observations	4,413	4,413

Outcome has been standardised using the control group mean and standard deviation. Columns 2 and 3 show the VA estimates including the full set of covariates and a lagged achievement measure. Robust standard errors in parentheses are clustered at the school level ; *** p<0.01, ** p<0.05, * p<0.1.

7. Appendix G

Figure C.7: Heterogeneity by EMI Grade of Transition

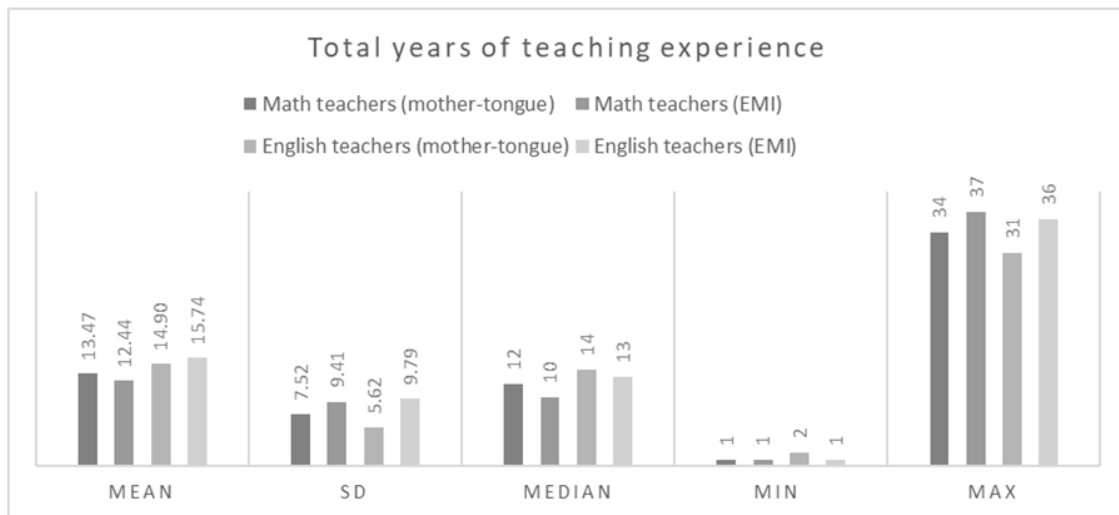
	Mathematics	English
EMI Grade 1 Transition	0.085 (0.160)	0.619*** (0.198)
EMI Grade 5 Transition	-0.133 (0.079)	0.274*** (0.090)
EMI Grade 7 Transition	-0.220*** (0.115)	-0.087 (0.142)
Lagged mathematics score	0.609*** (0.022)	
Lagged English score		0.505*** (0.024)
Control mean	0	0
Controls	Yes	Yes
Schools	37	37
Students	4,413	4,413

Outcome has been standardised using the control group mean and standard deviation.

Robust standard errors in parentheses are clustered at the school level; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

8. Appendix H

Figure C.8: Teacher statistics on years of teaching experience



9. Appendix I

Figure C.9: Dependent and independent variables used to in study sample

Panel A: Amhara Region (Treatment)					
	Obs.	Mean/ proportion	Std. Dev.	Min	Max
Math score wave 1	717	-0.15	0.93	-2.20	2.85
Math score wave 2	717	-0.13	0.93	-2.54	2.96
English score wave 1	717	-0.07	0.90	-1.87	3.26
English score wave 2	717	0.09	0.92	-1.87	3.64
Sex (1= Girl)	717	0.56	0.50	0	1
Age in years	717	14.05	1.43	11	22
Age at Grade 1	717	6.64	1.15	4	12
School Grade	717	7.47	0.50	7	8
Help with homework at home (1=Yes)	717	0.60	0.49	0	1
Read books out of school (1=Yes)	717	0.81	0.39	0	1
Attended pre-school (1=Yes)	717	0.21	0.40	0	1
Mother can read and write in any language (1=Yes)	717	0.34	0.47	0	1
Father can read and write in any language (1=Yes)	717	0.63	0.48	0	1
Sibling can read and write in any language (1=Yes)	717	0.90	0.31	0	1
Math Teacher Age in Years	717	28.26	3.98	24	40
Female Math Teacher	717	0.13	0.33	0	1
Math Teacher training qualification (1=University Degree)	717	0	0	0	0
Specialisation in Math (1=Yes)	717	0.27	0.44	0	1
Math Teacher Experience (Years)	717	9.18	4.78	3	21
Math Teacher ELIP (1=started/completed)	717	0.06	0.24	0	1
English Teacher Age in Years	717	36.74	6.88	21	50
Female English Teacher	717	0.43	0.50	0	1
English Teacher training qualification (1=University Degree)	717	0	0	0	0
English Teacher ELIP (1=started/completed)	717	0.35	0.48	0	1
English Teacher Experience (Years)	717	14.12	6.01	2	28
Shift/Full Day (1=Shift)	717	1	0	1	1
Number of working latrines/toilets for student use	717	2.18	1.40	0	4
Functional pedagogical center (1=Yes)	717	0.79	0.41	0	1
Functional library (1=Yes)	717	0.79	0.41	0	1
Sports/Play area	717	1	0	1	1
Electricity normally connected (1=Yes)	717	0.74	0.44	0	1
Locality (1=Urban)	717	0.34	0.47	0	1

Panel B: SNNP Region (Treatment)					
	Obs.	Mean/ proportion	Std. Dev.	Min	Max
Math score wave 1	1,233	0.18	1.08	-2.03	3.52
Math score wave 2	1,233	0.01	1.04	-2.40	2.96
English score wave 1	1,233	0.92	1.22	-2.05	4.17
English score wave 2	1,233	0.82	1.16	-2.04	3.64
Sex (1= Girl)	1,233	0.54	0.50	0	1
Age in years	1,233	14.21	1.48	11	27
Age at Grade 1	1,233	6.63	1.38	4	20
School Grade	1,233	7.43	0.49	7	8
Help with homework at home (1=Yes)	1,233	0.74	0.44	0	1
Read books out of school (1=Yes)	1,233	0.93	0.26	0	1
Attended pre-school (1=Yes)	1,233	0.74	0.44	0	1
Mother can read and write in any language (1=Yes)	1,233	0.68	0.47	0	1
Father can read and write in any language (1=Yes)	1,233	0.77	0.42	0	1
Sibling can read and write in any language (1=Yes)	1,233	0.77	0.42	0	1
Math Teacher Age in Years	1,233	36.54	11.37	20	59
Female Math Teacher	1,233	0.08	0.27	0	1
Math Teacher training qualification (1=University Degree)	1,233	0.35	0.48	0	1
Specialisation in Math (1=Yes)	1,233	0.92	0.27	0	1
Math Teacher Experience (Years)	1,233	15.75	11.66	1	37
Math Teacher ELIP (1=started/completed)	1,233	0.20	0.40	0	1
English Teacher Age in Years	1,233	38.24	11.22	23	54
Female English Teacher	1,233	0.28	0.45	0	1
English Teacher training qualification (1=University Degree)	1,233	0.43	0.50	0	1
English Teacher ELIP (1=started/completed)	1,233	0.67	0.47	0	1
English Teacher Experience (Years)	1,233	18.00	11.40	3	36
Shift/Full Day (1=Shift)	1,233	0.82	0.39	0	1
Number of working latrines/toilets for student use	1,233	1.72	0.53	1	3
Functional pedagogical center (1=Yes)	1,233	0.73	0.44	0	1
Functional library (1=Yes)	1,233	0.95	0.22	0	1
Sports/Play area	1,233	1	0	1	1
Electricity normally connected (1=Yes)	1,233	0.96	0.20	0	1
Locality (1=Urban)	1,233	0.84	0.37	0	1

Panel C: Afar Region (Treatment)					
	Obs.	Mean/ proportion	Std. Dev.	Min	Max
Math score wave 1	688	-0.157	0.918	-2.032	3.015
Math score wave 2	688	-0.221	0.869	-2.118	2.392
English score wave 1	688	0.078	1.011	-2.416	3.075
English score wave 2	688	0.079	1.036	-2.375	3.307
Sex (1= Girl)	688	0.503	0.500	0	1
Age in years	688	14.378	1.803	11	29
Age at Grade 1	688	6.781	1.443	4	19
School Grade	688	7.487	0.500	7	8
Help with homework at home (1=Yes)	688	0.776	0.417	0	1
Read books out of school (1=Yes)	688	0.906	0.293	0	1
Attended pre-school (1=Yes)	688	0.510	0.500	0	1
Mother can read and write in any language (1=Yes)	688	0.443	0.497	0	1
Father can read and write in any language (1=Yes)	688	0.577	0.494	0	1
Sibling can read and write in any language (1=Yes)	688	0.622	0.485	0	1
Math Teacher Age in Years	688	30.817	5.680	23	45
Female Math Teacher	688	0	0	0	0
Math Teacher training qualification (1=University Degree)	688	0.894	0.308	0	1
Specialisation in Math (1=Yes)	688	1	0	1	1
Math Teacher Experience (Years)	688	9.901	5.914	2	25
Math Teacher ELIP (1=started/completed)	681	0	0	0	0
English Teacher Age in Years	688	33.596	6.682	21	42
Female English Teacher	688	0.234	0.424	0	1
English Teacher training qualification (1=University Degree)	688	0.363	0.481	0	1
English Teacher ELIP (1=started/completed)	688	0.733	0.443	0	1
English Teacher Experience (Years)	688	13.394	8.937	1	29
Shift/Full Day (1=Shift)	688	1	0	1	1
Number of working latrines/toilets for student use	688	9.927	5.779	2	16
Functional pedagogical center (1=Yes)	688	0.722	0.448	0	1
Functional library (1=Yes)	688	0.722	0.448	0	1
Sports/Play area	688	0.770	0.421	0	1
Electricity normally connected (1=Yes)	688	0.936	0.245	0	1
Locality (1=Urban)	688	0.936	0.245	0	1

Panel D: Oromia Region (Control)					
	Obs.	Mean/ proportion	Std. Dev.	Min	Max
Math score wave 1	578	0.23	1.12	-2.03	3.52
Math score wave 2	578	0.26	1.07	-1.98	2.96
English score wave 1	578	0.33	1.14	-1.87	3.81
English score wave 2	578	0.26	1.03	-2.37	3.64
Sex (1= Girl)	578	0.55	0.50	0	1
Age in years	578	14.46	1.24	12	19
Age at Grade 1	578	7.05	1.02	4	12
School Grade	578	7.56	0.50	7	8
Help with homework at home (1=Yes)	578	0.69	0.46	0	1
Read books out of school (1=Yes)	578	0.93	0.26	0	1
Attended pre-school (1=Yes)	578	0.47	0.50	0	1
Mother can read and write in any language (1=Yes)	578	0.65	0.48	0	1
Father can read and write in any language (1=Yes)	578	0.70	0.46	0	1
Sibling can read and write in any language (1=Yes)	578	0.77	0.42	0	1
Math Teacher Age in Years	578	37.43	10.65	23	54
Female Math Teacher	578	0.13	0.34	0	1
Math Teacher training qualification (1=University Degree)	578	0.71	0.46	0	1
Specialisation in Math (1=Yes)	578	1	0	1	1
Math Teacher Experience (Years)	578	16.90	11.15	1	34
Math Teacher ELIP (1=started/completed)	578	0.63	0.48	0	1
English Teacher Age in Years	578	39.86	5.89	24	49
Female English Teacher	578	0.39	0.49	0	1
English Teacher training qualification (1=University Degree)	578	0.50	0.50	0	1
English Teacher ELIP (1=started/completed)	578	0.85	0.35	0	1
English Teacher Experience (Years)	578	19.84	7.30	2	31
Shift/Full Day (1=Shift)	578	1	0	1	1
Number of working latrines/toilets for student use	578	2.89	1.72	2	8
Functional pedagogical center (1=Yes)	578	0.95	0.22	0	1
Functional library (1=Yes)	578	0.82	0.39	0	1
Sports/Play area	578	1	0	1	1
Electricity normally connected (1=Yes)	578	0.85	0.35	0	1
Locality (1=Urban)	578	0.45	0.50	0	1

Panel E: Tigray Region (Control)					
	Obs.	Mean/ proportion	Std. Dev.	Min	Max
Math score wave 1	1,204	-0.11	0.92	-2.03	3.35
Math score wave 2	1,204	-0.12	0.94	-2.12	2.67
English score wave 1	1,204	-0.16	0.88	-2.23	4.17
English score wave 2	1,204	-0.12	0.96	-2.04	3.64
Sex (1= Girl)	1,204	0.55	0.50	0	1
Age in years	1,204	13.68	1.01	12	21
Age at Grade 1	1,204	6.83	0.70	4	15
School Grade	1,204	7.49	0.50	7	8
Help with homework at home (1=Yes)	1,204	0.73	0.45	0	1
Read books out of school (1=Yes)	1,204	0.88	0.32	0	1
Attended pre-school (1=Yes)	1,204	0.44	0.50	0	1
Mother can read and write in any language (1=Yes)	1,204	0.49	0.50	0	1
Father can read and write in any language (1=Yes)	1,204	0.74	0.44	0	1
Sibling can read and write in any language (1=Yes)	1,204	0.87	0.34	0	1
Math Teacher Age in Years	1,204	31.39	3.60	24	35
Female Math Teacher	1,204	0.18	0.39	0	1
Math Teacher training qualification (1=University Degree)	1,204	0.09	0.29	0	1
Specialisation in Math (1=Yes)	1,204	1	0	1	1
Math Teacher Experience (Years)	1,204	11.83	3.97	1	15
Math Teacher ELIP (1=started/completed)	1,204	0.11	0.31	0	1
English Teacher Age in Years	1,204	32.23	2.65	28	37
Female English Teacher	1,204	0.34	0.47	0	1
English Teacher training qualification (1=University Degree)	1,204	0	0	0	0
English Teacher ELIP (1=started/completed)	1,204	0.56	0.50	0	1
English Teacher Experience (Years)	1,204	12.52	1.97	8	15
Shift/Full Day (1=Shift)	1,204	1	0	1	1
Number of working latrines/toilets for student use	1,204	3.80	2.81	1	8
Functional pedagogical center (1=Yes)	1,204	1	0	1	1
Functional library (1=Yes)	1,204	1	0	1	1
Sports/Play area	1,204	0.79	0.41	0	1
Electricity normally connected (1=Yes)	1,204	0.95	0.22	0	1
Locality (1=Urban)	1,204	0.47	0.50	0	1

article xcolor