The Long Shadow of Early Education: Evidence from a Natural Experiment in the Philippines*

Thomas Lloyd Dean Yang

September 11, 2024

Abstract

How does early educational quality affect longer-term academic outcomes? We shed light on this question via a natural experiment in the Philippines—the flawed implementation of a mother tongue education policy in public schools in kindergarten to Grade 3. This policy led to an unexpected decline in educational quality, but differentially in a subset of schools strongly predicted by pre-policy student language composition. We use language composition variables as instrumental variables for treatment. Leveraging panel data and confirming robustness to pre-trends, we find that the policy led to declines in standardized test scores, and reduced student enrollment and teacher retention in public primary schools. Employing a triple-difference strategy with Philippine Census data (across cohorts, localities, and decadal censuses), we show that by 2020, cohorts fully exposed to the policy completed 0.3 fewer years of schooling, most prominently due to reduced completion of late-primary and early-secondary grades (Grades 6-8). By revealing how a policy-induced reduction in early education quality reduces educational attainment in later years, our results underscore the importance of investing in the quality of education in the first years of schooling.

Keywords: education quality, mother tongue education, Philippines
JEL Codes: I21, I28, O15, C26, H75

*Thomas Lloyd: University of Michigan (tmlloyd@umich.edu); Dean Yang: University of Michigan, NBER and BREAD (deanyang@umich.edu). Yu Shi and Jiaxin Guo provided exemplary research assistance. We appreciate thoughtful feedback from Kevin Carney, Taryn Dinkelman, Susan Godlonton, Brian A. Jacob, and Yusuf Neggers, as well as seminar participants at Michigan's Causal Inference in Education Research Seminar (CIERS), the Michigan Development Workshop, and the 2024 MDev Conference. We thank DepEd officials, Undersecretary for Curriculum and Teaching (CT), Gina O. Gonong, Undersecretary for School Infrastructure and Facilities, Epimaco V. Densing III, Undersecretary for Administration, Nolasco A. Mempin, Assistant Secretary for Administration, Noel T. Baluyan, and former Director of Planning Services, Roger B. Masapol; the staff at the Planning Service - Policy Research and Development Division (PS-PRDD) namely, Karla Sio, Marie Christ Apit-Pangilinan, Maria Katrina Gregorio, John Lawrence Carandang, Lia Angela Daus; Camille Angela Magpayo; the Education Management Information System Division (EMISD), Marieta Atienza, Marion Gapultos, Deogracias Genito; the Information and Communication Technology Service, Sara Borja; the Bureau of Education Assessment (BEA), Director of BEA, Nelia Benito, Gretchen Cordero, Enrico Totanes, Sheryll Ann Legarte, Alexander Sucalit; the Bureau of Learning Delivery (BLD), Director of BLD, Leila Areola, Gaudencio Serrano; the Bureau of Learning Resources, for sharing data and coordinating this study. We also thank the staff at Innovations for Poverty Action (IPA) Philippines for field support and notably Karisha Anne Cruz for stellar project management.

1 Introduction

Human capital investments in early life are widely believed to play a crucial role in shaping long-run academic and economic outcomes. Prior research has demonstrated positive relationships between early educational experiences and later life success (e.g., Heckman (2006); Chetty et al. (2011)). However, our understanding of the impact of early human capital investments remains incomplete. While some studies have shown short-term benefits of high-quality early education programs (Schweinhart et al., 2005), evidence on the longevity of these effects is mixed, with some research suggesting fade-out of initial gains (Bailey et al., 2020). Much of the existing literature focuses on specific interventions, leaving open questions about the impact of broader changes in educational quality. In particular, there is very little causal evidence on how *mainstream early education quality* affects longer-term outcomes. Alleviating this knowledge gap can help shape educational policies by revealing the gains from investing in mainstream early-years education.

Empirically isolating the causal impact of early educational quality on later-life outcomes is a significant challenge. Omitted variable bias is a major concern: family background, socioeconomic status, peer quality, local public goods, and myriad other factors are associated with early education quality as well as later life outcomes, making it difficult to disentangle the causal effect of early education quality. Experimental studies, while providing strong internal validity, are typically limited in scale and duration. The gradual nature of many educational policy changes also makes it challenging to identify exogenous shocks facilitating causal inference. Consequently, our understanding of the causal relationship between early education quality and long-term outcomes remains incomplete, particularly for broad, system-wide changes in mainstream public education.

To address these challenges, our study leverages a unique natural experiment. In 2012, the Philippine government implemented a mother tongue education policy in public schools for kindergarten through Grade 3. While well-intentioned, the policy's implementation was seriously flawed, and led to an unexpected decline in educational quality. The switch in the medium of instruction – the "treatment" – occurred only in a subset of schools, which can reliably be predicted by pre-policy student native language composition. This variation creates a quasi-experimental setting. We use pre-policy student language composition measures as instrumental variables for "treatment" (exposure to lower-quality early education) at the school level. We conduct empirical analyses with panels of cross-sectional data from two sources: 1) internal Philippine Department of Education (DepEd) datasets on standardized test scores, enrollment, and teacher counts, and 2) the Philippine Census of Population (2010 and 2020 rounds).

Our instrumental variable approach combined with quasi-panel data provides plausible causal identification. First, the set of school-level student native language composition variables provides plausible instruments for treatment. In panel data settings such as ours, tests for parallel trends in the pre-treatment period typically provide the main test for the plausibility of causal identification. We do find violations of the parallel-trends assumption for some of the outcomes derived from DepEd administrative data, so we estimate Rambachan and Roth (2023) "honest" confidence intervals accounting for these violations. Our estimated impacts on test scores, enrollment, and teacher counts are robust to this approach.

Analyses of outcomes from the DepEd administrative data are subject to concerns about self-selection of students into and out of the public school system. We do find that treatment reduces student enrollment in public schools, which raises concerns about selection bias in the estimated effects on test scores. It is therefore important that we also analyze impacts in Philippine Census data. We conduct a triple-difference analysis exploiting variation across student cohorts (younger students are differentially more-treated), across municipalities (which vary in the share of students treated), and across time (the 2010 pre-policy Census vs. the 2020 post-policy Census). This analysis addresses selection concerns. The Census includes the full population, regardless of selection into public or private schools. Moreover, we rely on a respondent's municipality of *birth* (for respondents in the 2020 census) or their municipality of residence (for respondents in the 2010 census) to assign treatment intensity, both of which predate policy implementation and thus are immune to concerns of migration into and out of treated municipalities.

Our empirical analyses yield several key findings. First, we find that exposure to the flawed mother tongue education policy had negative effects on standardized test scores of public primary school students. Such effects do not emerge yet by Grade 3 (the final year of mother tongue education), but are prominent by Grade 6: students in treated schools score 0.67 standard deviations lower (on average across all subjects) than students in control schools. Second, we observe notable declines in student enrollment in later primary school grades (Grades 3-6) and lower teacher retention in treated public schools. Declines in student enrollment may reveal parental or student dissatisfaction with the quality of education in public schools. Teacher losses are likely another indication of declines in educational quality, with may have reduced teacher morale.

Third, our analysis of Census data reveals longer-term effects on educational attainment. Our triple-difference estimates indicate that by 2020 (eight years after the start of the policy), younger cohorts in fully-treated municipalities completed 0.3 fewer years of schooling compared to students in untreated municipalities. This result is both statistically significant and economically meaningful. We estimate negative effects on grade completion for Grades 2-10, with the largest magnitudes (and statistical significance levels) for completion of Grades 6-8 (the end of primary school and the beginning of secondary school). Collectively, these findings provide causal evidence of the enduring consequences of early education quality on academic achievement over the longer term.

Our work is related to a body of prior research. The importance of human capital investments in early life has been extensively documented in the literature. Seminal work by Heckman (2006) emphasizes the critical role of early investments in human capital formation. A broad set of studies has shown the existence of "critical periods" – stages in life where health, economic, social, or other conditions have a persistent impact on later-life outcomes (Cunha et al. (2006), Maccini and Yang (2009), Almond and Currie (2011), and Currie and Almond (2011)). Studies such as the Perry Preschool Project (Schweinhart et al., 2005) and evaluations of Head Start (Ludwig and Miller, 2007) have demonstrated positive short-term effects of high-quality preschool programs. The persistence of these effects remains debated, with some research suggesting fade-out of initial gains (Bailey et al., 2020), while others find enduring impacts (Chetty et al., 2011).

Our work also contributes to the literature on education in developing countries that highlights unique challenges and policy considerations. Glewwe and Muralidharan (2016) emphasize the need for context-specific research and policy solutions on the economics of education in developing countries. In recent decades, developing countries have experienced a large expansion of schooling, with average years of formal education increasing from 2.0 years in 1950 to 7.2 years in 2010 (Barro and Lee, 2013). However, such gains in years of education do not always translate into learning or human capital gains (Pritchett, 2013; World Bank, 2018; Muralidharan et al., 2019). Because other studies do document the potential for schooling to have large returns (Duflo, 2001), prior research has studied the potential explanations for the inefficiency of schooling in developing countries. Common candidate explanations for inefficiency include low levels of spending associated with shortages in teaching materials and staff, over-ambitious or inappropriate curricula with students who fall behind never given the opportunity to catch up (Banerjee et al., 2016; Muralidharan et al., 2019), and teacher absenteeism associated with weak teacher incentives (Kremer et al., 2013; Mbiti et al., 2018).

We also provide novel insights on the impacts of mother tongue education policies in multilingual contexts. A number of studies have explored the role of the language of instruction in the human capital production function (Angrist and Lavy, 1997; Angrist et al., 2008; Argaw, 2016; Taylor and von Fintel, 2016; Ramachandran, 2017; Laitin et al., 2019). Mother tongue education policies are often motivated by the following causal chain: learning in the mother tongue may facilitate the acquisition of cognitive skills (both reading and numeracy skills) in early grades which may in turn improve the learning of a second language and the translation and expansion of such acquired skills in the second, dominant language (Taylor and von Fintel, 2016). It is such human capital gains in the second language that are expected to have the largest economic returns. The second link of this causal chain is the most controversial, namely the translation of skills into a second (dominant) language. Opponents worry that mother tongue instruction may actually *reduce* proficiency in the dominant language.[1] However, another important link, upstream in the causal chain, that is often overlooked, relates to the feasibility of teaching in the mother tongue and to the potential shock to education quality associated with a shift to instruction in local languages without adequate preparation. Bühmann and Trudell (2008) argue for the benefits of mother tongue education in improving learning outcomes, while Heugh (2012) highlights challenges in implementing such policies in developing countries. The complex linguistic landscape of the Philippines, as described by Tupas and Martin (2017), provides a relevant context for examining these issues.

[1]There is currently mixed evidence on this issue. For example, using quasi-random variation in Ethiopia, Argaw (2016) finds that mother tongue-based education leads to a 11 p.p. gains in reading skills and modest gains in labor market outcomes. In contrast, using a randomized experiment in Kenya, Piper et al. (2018) find no effect of mother tongue instruction on literacy skills in English and slightly negative impacts on numeracy skills.

Finally, this paper also highlights challenges of policy implementation at scale (Angrist et al., 2023; Angrist and Meager, 2023; List, 2022). Pritchett et al. (2013) discuss the complexities of policy implementation in developing countries, while Bold et al. (2018) provide evidence on how well-intentioned educational interventions can fail to deliver expected results at scale. Our work documents that a mother tongue education policy in a multilingual context, implemented nationwide—and associated with important implementation challenges—led to a sharp reduction in test scores and longer-run educational attainment.

Our study contributes to these strands of literature by leveraging a unique natural experiment in the Philippines to provide causal evidence on the long-term impacts of early education quality. In contrast to most prior work, we examine the effects of a system-wide change in educational quality in the first years of mainstream education, rather than the impacts of a targeted intervention in preschool years. Furthermore, the fact that we examine the impact of a *decline* rather than an improvement in educational quality is another distinctive feature of our study. Our findings thus serve as a cautionary tale encouraging policymakers to avoid disruptions to educational quality in early years, due to their longer-term negative consequences.

2 Philippine Education and the Mother Tongue Policy

In this section we provide background about public education in the Philippines, emphasizing its multilingual context. We then describe the mother tongue-based education policy, with key details about implementation challenges.

2.1 Education in the Philippines

As of school year 2020-2021, there were a total of 22.6 million students enrolled in public primary and secondary schools run by the Philippine government's Department of Education (DepEd), of whom 13.6 million were in primary schools (Kindergarten to Grade 6). Grade completion is high, with 90.3% of those aged 9-10 completing Grade 3, 87.1% of those aged 12-13 completing Grade 6, and 68.0% of those aged 16-17 completing Grade 10 in the 2020 Census. Basic Education in the Philippines is divided between Kindergarten and Elementary (Grades 1 to 6), taught in primary schools. Secondary schools cover Junior High School (Grades 7 to 10) and Senior

High School (Grades 11 to 12), taught in secondary schools (Brillantes et al., 2019).[2]

At the same time, Filipino students perform poorly in large-scale international assessments such as the 2018 PISA and the 2019 TIMSS (OECD, 2018; Mullis et al., 2020), and 9 out of 10 students at late primary age struggle to read and comprehend simple texts (World Bank, UNESCO, 2021). The Philippines ranks the lowest in expenditure per student among participants in the 2018 PISA survey, and 90% lower than the OECD average (OECD, 2018). The education budget per student averages approximately \$514 per student per year as of 2022 (compared to approximately \$15,000 in the U.S. (Hanson, 2024)).

Education in the Philippines takes place in a highly linguistically diverse context. The Ethnologue reports 184 distinct spoken languages in the Philippines (Ebernhard et al., 2023). Tagalog is the most widely spoken language, with 34.0% of primary school students declaring it as their mother tongue, closely followed by Cebuano/Bisaya/Binisaya at 25.3%. Other notable language groupings include Hiligaynon/Ilonggo at 7.4%, Ilocano at 6.7%, and Bikol at 5.7%. The top 19 languages (used in the mother tongue education policy that are the focus on this study) account for 94.8% of students nationwide.[3]

In 1973, in an effort to reconcile the Philippines' colonial history with its post-colonial nation-building objectives, the country adopted a bilingual education system, with both English and Filipino (a standardized form of Tagalog) as languages of instruction (Tupas and Martin, 2017; Monje et al., 2019). One of the disadvantages of the bilingual system of English and Filipino was frequent mismatch between a child's mother tongue and their school's languages of instruction. From the moment they started formal schooling, most students were taught in languages other than their mother tongue. Such language mismatch is associated with inequalities in access to learning in early childhood, stigma, and marginalization (UNESCO, 2010). Language of instruction was identified by the Philippine Department of Education of the Philippines (DepEd) as a key factor behind the country's relatively poor performance in international large-scale assessment studies such as the Trends in International Mathematics and Science Study (TIMSS) (DepEd, 2009).[4] In addition to perceived

[2]Note that prior to school year 2016-2017, high school ended in Grade 10 in the Philippines. Senior High School (Grades 11 and 12) was first implemented in school year 2016-2017.

[3]These percentages are authors' calculations from Philippine Department of Education student administrative data; see Section 3 and Appendix Table A1 for further details.

[4]The 2009 DepEd memo states, "Top performing countries in the Trends in International Math-

poor performance of the bilingual education system, other rationales for the move to mother tongue education included endorsements for the use of local languages from international organizations (Bühmann and Trudell, 2008; Ball, 2010) and a desire to promote cultural identities (Tupas and Martin, 2017).

2.1.1 Equivalence of Filipino and Tagalog

The term "Filipino" refers to the *national* language of the Philippines (it shares with English, the status of *official* language of the Philippines). Filipino was first formalized in 1937 when Tagalog was selected as its basis (Rubrico, 1998; Monje et al., 2019). Tagalog is a regional language spoken in the central region of Luzon (the Philippines' largest island) surrounding the capital Manila, contiguous provinces south of Manila, and large island provinces immediately south of Luzon such as Mindoro, Marinduque, Romblon, and Palawan. The Philippine government refers to the Tagalog-based national language as "Filipino", to disassociate it from the Tagalog ethnic group and encourage national identification with the language. Over time, Filipino has developed into a standardized form of Tagalog with some words incorporated from English and Spanish. However, in the Philippine population at large, the language is still largely referred to as "Tagalog" (Nolasco, 2007). Therefore, to avoid confusion, we will use the term "Tagalog" throughout the remainder of this paper to refer to both the official language as well as the mother tongue.[5]

2.2 The Mother Tongue Education Policy

Prior to the implementation of the mother tongue education policy in the 2012-13 school year, the status quo was as follows. The MOI in all public schools in all grades from K-10 was Tagalog and English (the two official languages of the Philippines).[6] Mathematics (which was taught starting in Grade 1) was taught in English, as was Science (taught starting in Grade 3), and the English language class (starting in Grade 1). Tagalog was used for all other subjects: Tagalog language class (*Filipino*), Social Studies (*Araling Panlipunan*), and Ethics/Humanities (*Edukasyon sa Pagpapakatao*).

ematics and Science Study (TIMSS) are those that teach and test students in science and math in their own languages" (DepEd, 2009).

[5] In the Philippine government, the practice is to refer to "Filipino" as the country's national language, to enhance national identification with the language, and to refer to "Tagalog" as the mother tongue in the context of DepEd's MTB-MLE policy.

[6] Use of Tagalog and English as MOI in public schools had been the policy from 1973 onwards.

Starting school year 2012-13, DepEd implemented the mother tongue-based education policy in primary schools nationwide (DepEd, 2012). The policy is known as Mother Tongue-Based Multilingual Education (MTB-MLE). The MTB-MLE policy required public primary schools to select a local mother tongue as the medium of instruction (MOI) for teaching in all non-language classes in Kindergarten to Grade 3. The MTB-MLE policy did not affect instruction in Grades 4-10, in which the MOI remained Tagalog and English. The policy remained in place throughout all periods of analysis in our data (and remains in place as of the writing of this paper in 2024).

Individual schools were free to choose their medium of instruction, but were mandated to select one (and only one). A list of 19 languages were officially recognized as MOI that schools could choose to teach in from Kindergarten to Grade 3, following the MTB-MLE guidelines. In practice, schools were encouraged to teach in the language that students knew best, i.e., to use learners' "first language (L1)" or their "native language" (DepEd, 2009).

Choice of medium of instruction under the MTB-MLE policy was made at the individual primary school level, potentially with input from school district officials and in consultation with staff and parents. Schools could also choose to teach in a language (usually Tagalog) even if it was not the most widely-spoken language among their students (for example, in linguistically diverse localities as a consensus choice).

Appendix Table A1 displays data on different languages, with shares of students speaking, shares of schools choosing the language as their MOI, and shares of students facing each MOI. Tagalog is the most widely-selected MOI, chosen by 33.5% of schools (resulting in 45.2% of G1-G3 students with Tagalog as MOI), followed by Cebuano/Bisaya/Binisaya with 27.4% and Ilocano with 9.2%. 1.5% of schools with MOI information chose a language from outside the list of 19 as their MOI.

2.3 Definition of Treatment and Control Schools

Because Tagalog was used as a medium of instruction pre-policy, the MTB-MLE policy affected the following two groups of schools differently: (1) **treated schools**: schools that changed their medium of instruction to a language other than Tagalog post-policy, and (2) **control schools**: schools that chose Tagalog as their MOI, and that therefore used Tagalog as MOI both pre- and post-policy.

As discussed above, prior to the MTB-MLE policy, English was used for teaching Mathematics (starting in Grade 1) and Science (introduced in Grade 3), while Tagalog was used for all other subjects. Consequently, even for schools in our control group that continued using Tagalog as their medium of instruction after the policy change, Tagalog was extended to Mathematics and Science. We consider these schools to be the control group because they were significantly less affected by the policy shift. Specifically, they did not need to change their medium of instruction for Tagalog (*Filipino* language class), Social Studies (*Araling Panlipunan*), and Ethics/Humanities (*Edukasyon sa Pagpapakatao*).

Because schools in the control group maintained Tagalog as MOI, they were also much more likely to have teachers capable of teaching Mathematics and Science in Tagalog, since literacy in Tagalog is a requirement for teacher certification in the Philippines (House of Representatives of the Philippines, 1994; Professional Regulation Commission, 2016). Additionally, Tagalog is recognized for having the most extensive literary tradition among Philippine languages (Tupas and Martin, 2017), with corresponding widespread availability of texts and other educational material. In sum, the policy led to significantly less disruption for students in schools where Tagalog remained the medium of instruction.

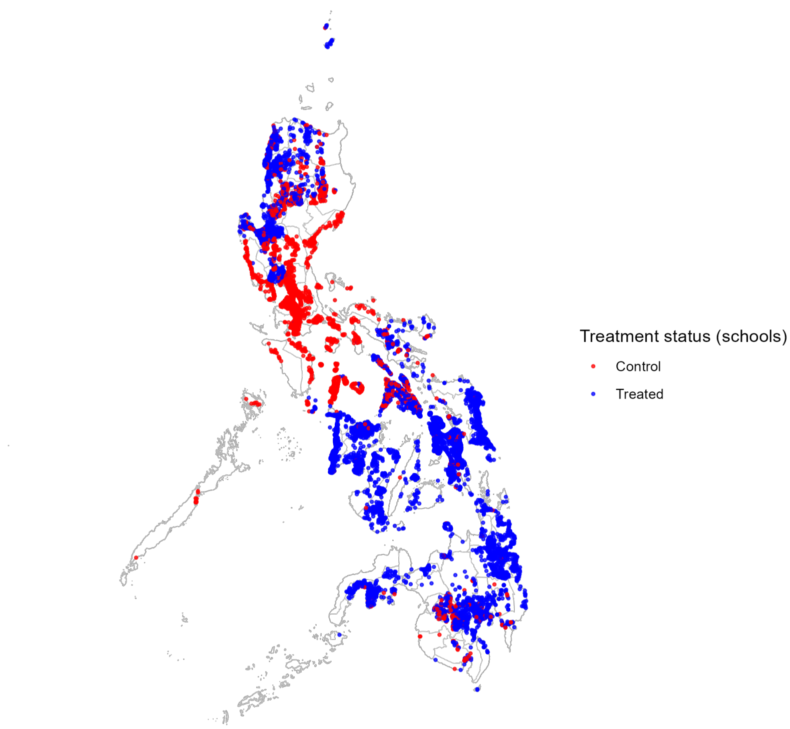
We show in Figure 1 a map of schools and their treatment and control status. Schools choosing Tagalog as their MOI are the control group, and schools choosing other languages as their MOI are the treatment group.

As an endogenous choice by schools, there are concerns about using school-level decisions as the basis of assignment to treatment and control status in our empirical analyses. While choice of MOI would certainly have been driven in part by relatively exogenous factors (such as local language composition), other factors may come into play that raise concerns about selection bias.

We consider it most likely that schools electing to be in the treatment group would be positively selected. Schools may be more likely to choose a non-Tagalog language as MOI (and thus enter the treatment group) if they have private knowledge (unobserved by us as researchers) that they will be differentially better able to provide quality education using a non-Tagalog MOI. Examples of such private knowledge are the availability of teachers on the school's staff who know the local (non-Tagalog) language, or the ability of school management to handle the transition to a new MOI. If choice of MOI is at least in part driven by such inside knowledge, then estimated effects of treatment would be biased in a positive direction.

To deal with selection bias, we take an instrumental variables approach. We use the composition of languages spoken by students in the school as instruments for the decision to choose a non-Tagalog MOI (i.e., the decision to be "treated"). This approach helps rule out concerns about positive selection into treatment (e.g., due to better management or teachers who speak the local language). We describe this approach further in Subsection 4.1.1 below.

Figure 1: Treatment and Control Groups for Analysis of MTB-MLE Policy



Note: Each point is the geographical location of a Philippine public school for which we have medium of instruction information in our sample. The Mother Tongue-Based Multilingual Education (MTB-MLE) policy mandated each school to choose a local language as medium of instruction (MOI). "Treatment group" schools, in blue, are those selecting a language other than Tagalog as their MOI. "Control group" schools are those that selected Tagalog as their MOI, and are shown in red. Light grey lines demarcate province borders.

2.4 Flawed Implementation of the Mother Tongue Education Policy

The implementation of the MTB-MLE policy has been viewed as flawed on multiple dimensions. Implementation problems have been documented in formal quantitative and qualitative process evaluations and assessments (Metila et al., 2016; Monje et al., 2019), and have also been widely reported in the news media (The Manila Times, 2020; CNN Philippines, 2022; Inquirer, 2023).

The policy is widely viewed as having been implemented without adequate advance notice and preparation. The directive to implement the policy nationwide was announced formally on February 17, 2012 with DepEd Order No. 16 (DepEd, 2012), with implementation to start in school year 2012-2013 in all public schools. Schools therefore had 3.5 months' advance notice to prepare for implementation before the school year began in June 2012.

A key issue is that few if any educational materials were available in local languages other than Tagalog. Recognizing this absence of materials, the MTB-MLE policy required schools to develop their own materials (DepEd, 2012, 2015). Monje et al. (2019) classify the preparation and development of learning resources into four activities: 1) develop their own learning resources such as writing classroom materials on language, literature and culture, 2) document the orthography and 3) the grammar of the language, and 4) develop a dictionary of the language. Such a set of expectations may be considered a very high bar for individual schools (with small staffs) to reach, in particular considering the lack of standardization and intellectualization of the orthography and grammar of local languages needed for instruction (especially in mathematics and science) (Metila et al., 2016). Indeed, few schools appear to have surpassed this very high bar for preparing their own instructional materials: only 9.0% of the 16,479 schools surveyed by Monje et al. (2019) report having carried out all four activities required by the MTB-MLE policy for successful implementation.[7]

Relatedly, there were major concerns that teachers were inadequately prepared for the new policy. Large shares of teachers did not actually speak and could not teach in local languages. Prior to the policy's implementation, public school teachers were recruited on a national basis, could have been posted in any part of the country, and were only required to show competence in Tagalog and English, not any local languages. Lack of teacher proficiency to teach in a school's MOI and insufficient teacher training were widely cited criticisms of the policy in formal process evaluations (Monje et al., 2019; Tupas and Martin, 2017; Metila et al., 2016).

A qualitative process evaluation conducted in 2014-15 by Metila et al. (2016) documented a wide range of implementation challenges, highlighting many of these same issues. The process evaluation emphasized the time-consuming requirement to produce instructional materials in local languages, the lack of financial support to pro-

[7]Given this 9.0% figure is self-reported by schools themselves, it is likely to be overstated, so the true figure is probably lower.

duce such materials, the absence of words in local languages for many academic terms used in instruction, and the lack of training for teachers and school administrators.

The policy was also criticized for its "one-size-fits-all" approach, since choosing one MOI for a school does not account for linguistic diversity within the classroom, or for the breadth of local dialects of the same language (Monje et al., 2019). The fact that many students in a school might have a different mother tongue than their school's chosen MOI was a common critique levied at the policy (Monje et al., 2019; Tupas and Martin, 2017; Metila et al., 2016).

In sum, due to the MTB-MLE policy's flawed implementation, it is unclear *a priori* whether the policy would have positive or negative effects on educational outcomes. Our empirical analyses below will help resolve this ambiguity.

3 Data and Summary Statistics

We combine administrative datasets obtained from the Philippine Department of Education (DepEd) at both the school and individual level together with survey and census data from the Philippine Statistics Authority (PSA). We summarize data sources below. Additional details are provided in Appendix A.1. The summary statistics for key variables are shown in Table 1.

3.1 Linguistic data (Mother Tongue & Medium of Instruction)

3.1.1 Linguistic data.

We use data on the mother tongue of the universe of elementary public school students to construct measures of the linguistic composition of each school's student body. These data were first collected by DepEd in SY 2012-2013, the same school year the MTB-MLE policy was first implemented.

Because the MTB-MLE policy could have changed the language composition of students in a school, we use data from *never treated* students who were in Grades 4 to 6 in 2012-13 (approximately 5 million of the 11 million elementary school students in that school year). We compute, for each school, the percentage of students speaking each of the 19 languages offered as media of instruction.[8] This approach ensures

[8] The correlation coefficients between the school-level linguistic variables constructed with all elementary grades (Grades 1 to 6) and those constructed with only never treated grades (Grades 4 to 6) are very close to 1. We use Gr 4 to Gr 6 students only to alleviate concerns about potential

that we are not constructing school-level linguistic composition data with students in grades (those up to Grade 3) that were subject to the MTB-MLE policy. We use this linguistic composition data to instrument for the school-level choice of the medium of instruction, which determines a school's treatment status.

3.1.2 Medium of instruction

Information on the medium of instruction adopted by each school as part of the MTB-MLE policy originates from two complementary data sources for which the sample is limited to a subset of public schools in the Philippines. The first is a DepEd-conducted survey of schools in 2022 profiling the medium of instruction adopted due to the MTB-MLE policy, which spans 20,430 schools. The second is a 2018-19 survey of 15,916 schools conducted by Monje et al. (2019) in their process evaluation study of the MTB-MLE policy implementation. Unifying these two data sources, we have MOI data for 24,529 out of the 34,807 public elementary schools in the Philippines with pre-policy linguistic composition data. We assign a school to the treatment group if it reported its medium of instruction to be non-Tagalog in either survey.

3.2 Standardized Test Scores

Our first main outcomes are nationally standardized test scores from the National Achievement Test (NAT) administered by DepEd. The test score data are repeated cross-sections of public school students in Grade 3 and Grade 6 from school year (SY) 2008-09 to SY 2017-18. DepEd provided us a 10% random sample of the universe of test score results for our empirical analyses.[9] Test subjects include English, Tagalog, and Mathematics in Grade 3 and Grade 6, as well as Science and History & Geography (referred to as "Hekasi") in Grade 6. An "overall" test score is computed as a simple average across subjects. We restrict our main sample to focus exclusively on public schools with medium of instruction information. This includes approximately 2 million test scores from students in 24,529 public schools in 1,482 municipalities and 84 provinces.

linguistic compositional changes for students in Grades 1 to 3 in 2012-2013.

[9] The sample was obtained via stratified random sampling on region, and school division conducted by the Bureau of Education Assessment at DepEd.

3.3 Enrollment, Teacher Counts, and Student-Teacher Ratios

We also examine public school administrative data on student enrollment in each grade from Grade 1 to Grade 6, and total primary school teacher counts. Data were provided by DepEd for the school years 2008-09 to 2017-18. We examine these outcomes themselves, and also calculate primary school student-teacher ratios by dividing student enrollment by teacher counts at the school level.

3.4 Longer-Run Outcomes: Grade Completion & Years of Education, Eight Years Post-Policy

It is also important to examine longer-run effects, as well as to conduct analyses less subject to selection bias concerns. Analyses of outcomes from DepEd administrative data described above (test scores, student enrollment, and teacher counts) are an important part of analysis of causal effects of the MTB-MLE policy, but they are open to concerns about selection bias. Movements of students between treatment and control schools may cause bias in estimated effects on test scores. Movements of students and teachers between treatment and control schools caused by the policy are a violation of the stable unit treatment value assumption (SUTVA), and thus may also cause bias in estimates of effects on test scores, enrollment, and teacher counts (Rubin, 1980).

We therefore analyze long-run impacts on educational outcomes in the Philippine Census. Inclusion in the Census is not conditional on one's school attendance characteristics (being in the public vs. private schooling system, being in a treatment or control school, or being in school at all). We use educational attainment data in the Census to calculate total years of education completed as well as indicators for completing specific educational levels in 2020, eight years after the initiation of the MTB-MLE policy.

We use data from the decennial 2010 and 2020 Philippine Census of Population and Housing (CPH), collected by the Philippine Statistics Authority (PSA). Use of the 2010 pre-MTB-MLE-policy Census (alongside the 2020 post-MTB-MLE-policy Census) allows us to conduct a triple-difference identification strategy (across cohorts, municipalities, and Census rounds).

Our analyses using Census data include approximately 35 million respondents aged 7 to 25 per Census round. We use data on municipality of birth, municipality of residence, age, and highest year of education completed. To analyze impacts in Census data, we match each individual using either their municipality of birth (for respondents in the 2020 census) or their municipality of residence (for respondents in the 2010 census) with measures of MTB-MLE treatment intensity calculated at the municipality level calculated using DepEd data. See Section 4.2.1 below for details of this treatment intensity calculation as well as the triple-difference identification strategy.

Table 1: Summary statistics

	Mean	SD	N
Student-level variables SY 2008-09 to SY 2017-18			
Grade 3 Overall Test Scores	0.059	0.779	856,735
Grade 6 Overall Test Scores	0.061	0.751	1,102,850
School-level variables SY 2008-09 to SY 2017-18			
Grade 1 Enrollment Count	66.75	97.6	241,583
Grade 2 Enrollment Count	62.14	90.4	239,470
Grade 3 Enrollment Count	61.04	89.4	241,583
Grade 4 Enrollment Count	59.73	88.2	241,583
Grade 5 Enrollment Count	58.58	87.1	238,665
Grade 6 Enrollment Count	55.69	84.2	238,665
Number of Elementary Teachers	11.29	14.36	239,217
Elementary Student-Teacher Ratio	33.71	10.71	236,585
School-level variables SY 2012-13			
Treatment status (Treat_s)	0.665	0.472	24,529
Pct. Tagalog (G1-G6) in 2012-2013	0.240	0.388	24,529
Pct. Tagalog (G4-G6) in 2012-2013	0.238	0.387	24,529
Census respondent-level variables (aged 7 to 25)			
Highest Grade Completed	7.274	3.264	73,267,484
Treatment intensity at the municipality level (Treat_m)	0.541	0.420	73,267,484

Note: Summary statistics (sample mean, standard deviation, and the number of observations) for individual-level and school-level outcomes from DepEd administrative data used in our main analysis, as well as respondent-level outcomes from the 2010 and 2020 census rounds (for respondents aged 7 to 25). Student- and school-level variables are from 24,529 public schools with MOI information and linguistic composition data pre policy. Treatment status (Treat_s) is a binary variable defined at the school level. A school is said to be treated if its medium of instruction post policy is **not** Tagalog (see Figure 1 and Section 2.3). Treatment intensity (Treat_m) is a continuous variable at the municipality level corresponding to the predicted percentage of treated students (see Figure 4 and Section 4.2.1).

4 Empirical Analyses

We aim to shed light on the impacts of the MTB-MLE policy "treatment" (switching medium of instruction to a language other than Tagalog) on a variety of education-related outcomes. First, we examine outcomes using DepEd administrative data for which treatment status is determined at the school level, such as standardized test scores, enrollment in primary school grades, primary school teacher counts and student-teacher ratios. Then, in analyses using Census data for which treatment intensity is defined at the municipality level, we examine impacts on respondents' years of completed education.

4.1 Analyses using DepEd Administrative Data: Test Scores, Enrollment, and Student-Teacher Ratios

4.1.1 Empirical Approach

To estimate the causal effect of the MTB-MLE policy on adminstrative data outcomes, we take a difference-in-differences approach. We start with a canonical two-way fixed effects (TWFE) *dynamic* specification which allows for differential treatment effects across time relative to a baseline school year. We take 2011-12, the school year before the initiation school year of the policy, as the baseline school year (labeled $t = -1$). We estimate the following regression equation :

$$Y_{ispt} = \alpha_s + \gamma_t + \eta_{pt} + \sum_{\substack{h=-4 \\ h\neq -1}}^{h=5} \tau_h \, \mathbf{1}\{t = h\} \times Treat_s + \varepsilon_{ispt}, \tag{1}$$

where Y_{ispt} is the outcome of individual i in school s, province p, and school year t. α_s and γ_t are school and school year fixed effects. η_{pt} are province-by-school year fixed effects; their inclusion ensures that we rely exclusively on within-school variation over time between treated and control schools within the same province, corresponding to deviations from province-specific time effects.

$Treat_s$ is a binary variable equal to 1 if school s switched its medium of instruction to a language other than Tagalog post-policy. τ_h for $h \geq 0$ (the post-policy years) are the parameters of interest, interpreted as the average treatment on the treated units (ATT) in period h. $\tau_h > 0$ would be interpreted as positive causal effects of treatment, which $\tau_h < 0$ would indicate negative effects.

The specification allows us to test for pre-trends before policy implementation. Estimates of τ_h that are small in magnitude and not statistically significantly different from zero for $h < -1$ would indicate absence of pre-trends.

The leads $h \in \{-4, -3, -2, -1\}$ correspond to school years 2008-09 to 2011-12, while the lags $h \in \{0, 1, 2, 3, 4, 5\}$ correspond to school years 2012-13 to 2017-18. For lagged outcomes such as Grade 6 test scores (3-year lag allowing students to reach Grade 6), these indexes will be shifted by -3, with the baseline (omitted) school year becoming 2014-15 and thus reducing the number of post periods to three.

Standard errors are clustered at the school level, the unit of treatment assignment. We also estimate a version of equation (1) aggregated up at the school level (suppressing the index i) for outcomes such as student enrollment and teacher counts.

Finally, we define as follows the average causal effect across all post-treatment periods, as a summary measure of the TWFE dynamic specification:

$$\tau_{\text{post}} = (1/\bar{T}) \sum_{h=0}^{\bar{T}} \tau_h \tag{2}$$

4.1.2 Identification

τ_h are identified under the *parallel trend* and *no anticipation* assumptions. Moreover, our context satisfies the following three conditions highlighted by De Chaisemartin and D'Haultfoeuille (2023) which ensure unbiasedness for the ATTs: (i) the treatment is an absorbing state, (ii) the treatment is binary, and (iii) there is no variation in treatment timing. This avoids the problem of negative weights which could arise from comparing newly treated units relative to already treated units in designs with variation in treatment timing (Callaway and Sant'Anna, 2021; Borusyak et al., 2024).

4.1.3 Instrumental Variables (IV) approach

We augment the standard dynamic TWFE specification (equation 1) with an instrumented difference-in-differences (IV-DID) approach. Although school fixed effects address some endogeneity or selection concerns by controlling for *time-invariant* school-level unobservables, they do not address selection on *time-varying* characteristics. A key time-varying characteristic is a school's ability to adapt to the new MTB-MLE policy. We may worry that schools better able to teach in the local mother tongue are more likely to select into treatment.[10]

[10]While a characteristic such as a school's ability to teach in the local mother tongue may be seen as time-invariant, the importance of that characteristic changes in SY2012-13, when MTB-MLE is implemented. So the interaction between a school's (time-invariant) ability to teach in the local mother tongue and a dummy variable for being in SY2012-13 or after is time-varying.

To address these concerns, we instrument the binary school-level treatment indicator $Treat_s$ (equal to 1 if a school switched to a medium of instruction other than Tagalog) with the percentage of SY 2012-2013 Grade 4 to Grade 6 learners (never treated cohorts) at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction, as well as a square and a cubic term in the percentage of Tagalog-speaking learners.

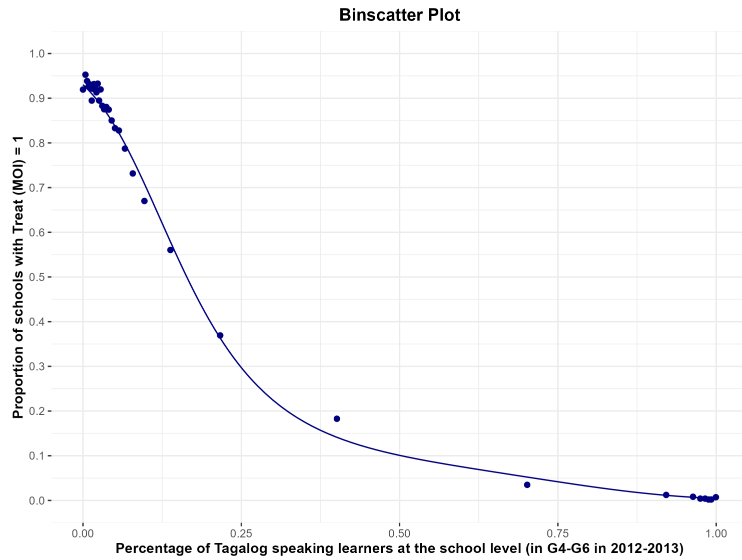
As discussed by Ye et al. (2023), the IV-DID approach is particularly attractive because it allows us to relax a key assumption of the standard IV. It is robust to violations of the exclusion restriction by allowing instruments to have a direct effect on the outcome. Instead, it only requires a weaker version of the exclusion restriction to hold: the instruments should have no direct impact on the *trend* in potential outcomes in the absence of treatment. Intuitively, in our context, as long as trends in outcomes are parallel for schools with different values of our instrumental variables if all schools were counterfactually not exposed to the MTB-MLE policy, then any observed nonparallel trends in outcomes post-policy between schools provides evidence for a causal impact of the policy.

Finally, we also report an "honest" confidence interval in our main tables for the estimated average causal effect across post-treatment periods τ_{post} using the robust inference methods developed by Rambachan and Roth (2023) for difference-in-differences designs where the parallel trends assumption may be violated. More specifically, we use their "smoothness restrictions" approach on non-parallel trends in pre-treatment periods assuming no change in slope for the post-treatment periods (which corresponds to the case where $\bar{M} = 0$ using the notation from their paper). This approach is akin to imposing a linear treatment-group-specific time trend estimated using only pre-treatment time periods. Intuitively, this method assumes that potential (linear) non-parallel trends would have persisted in the absence of the policy change, and thereby adjusts coefficient estimates to capture significant breaks from these potential pre-trends. For example, a null estimated impact in the presence of a positive pre-trend may actually correspond to a non-negligible break in the pre-treatment trend.

4.1.4 First Stage

In this subsection, we present the results from the first stage of our IV estimation strategy. While in practice, for the IV coefficient estimates presented in the following

Figure 2: School-level treatment status and percentage of Tagalog-speaking learners



Note: Binscatter plot with cubic fit illustrating relationship between treatment status at school level (a school is treated if its medium of instruction post policy is not Tagalog) and the percentage of Tagalog-speaking learners in Grades 4 to 6 during the 2012-13 school year. The optimal number of bins and the cubic fit were generated using the data-driven approach described in Cattaneo et al. (2024) with a starting choice of $n = 50$ bins.

subsections, we instrument all the interactions between the treatment variable and the individual year dummy variables ($\mathbf{1}\{t = h\} \times \text{Treat}_s$) presented in equation (1) with the interactions between our full set of IVs and the year dummy variables, Table 2 shows the first stage results from the static analog for simplicity of exposure.[11] This table presents the coefficient estimates from the regression of $Treat_s$ on all school-level linguistic composition variables used as instruments, as described above. Figure 2 shows non-parametrically, in a binscatter plot, the relationship between treatment status at the school level and the most predictive instrument, namely the percentage of a school's students whose mother tongue is Tagalog.

The figure clearly shows a strong, decreasing, and convex relationship between the percentage of Tagalog-speaking learners pre-policy and treatment status at the school level. The coefficient estimates in Table 2 confirms this graphical evidence, as indicated by the sign, magnitude and statistical significance of the linear, square, and cubic terms for the percentage of Tagalog-speaking learners.

The main takeaway from Table 2 is the very strong first stage with a F-statistic

[11] Note that this set of 19 linguistic variables are not collectively perfectly collinear, because students' mother tongue may be a language other than these 19 languages offered as media of instruction.

Table 2: Predicting school-level treatment status with student linguistic composition

	(1)		(2)		(3)
Tagalog	-2.205***	Hiligaynon	0.352***	Sambal	-0.090
	(0.098)		(0.015)		(0.129)
Tagalog squared	2.367***	Waray	0.390***	Akeanon	0.405***
	(0.242)		(0.014)		(0.018)
Tagalog cubed	-0.796***	Tausug	0.120*	Kinaray-a	0.424***
	(0.154)		(0.062)		(0.015)
Cebuano	0.366***	Maguindanaoan	-0.283***	Yakan	0.361***
	(0.014)		(0.057)		(0.043)
Kapampangan	0.293***	Maranao	0.269***	Surigaonon	0.413***
	(0.021)		(0.025)		(0.017)
Pangasinan	0.425***	Chabacano	0.401***	Obs. (Schools)	24,529
	(0.021)		(0.019)	R^2	0.662
Ilocano	0.180***	Ibanag	-0.067	F	10,807.9
	(0.016)		(0.055)	Prob. > F	0.000
Bikol	0.282***	Ivatan	0.415***		
	(0.016)		(0.028)		

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the results from the estimation of a linear probability model (LPM) corresponding to our **first stage** equation in which we regress the binary (school-level) treatment indicator Treat_s (equal to 1 if a school switched to a medium of instruction other than Tagalog) on the share of SY 2012-13 Grade 4 to 6 students (never treated cohorts) at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction, as well as a square and a cubic term in the share of Tagalog-speaking students. These variables account for each school's student body linguistic composition pre-policy. The binscatter plot in Figure 2 illustrates the (strongly predictive) decreasing relationship between treatment status and the percentage of Tagalog-speaking learners pre-policy.

of 10,807. This result empirically validates the assumption of trend relevance for our set of instruments. In this simple linear probability model, the linguistic composition variables explain 66.2% of the variation in treatment status. This suggests that when choosing whether or not to switch their medium of instruction, schools aimed to closely align their medium of instruction with the mother tongue of their students. Note that most coefficient estimates for the other languages offered as media of instruction are positive and statistically significant suggesting that a higher percentage of students speaking each of these languages increases the probability of treatment.[12]

[12] This is not the case for Ibanag nor Sambal which are estimated to have a null relationship with $Treat_s$, nor Manguindanaoan with a negative and statistically significant relationship.

4.1.5 Impacts on Standardized Test Scores

We first examine impacts of the MTB-MLE treatment (switching to a non-Tagalog MOI) on standardized test scores, based on estimation of regression equation 1 for Grade 3 and Grade 6 National Achievement Test scores. We report τ_{post} (the average causal effect across all post-treatment periods) in Table 3, and show the individual τ_h coefficient estimates in Figure 3.

τ_h coefficient estimates in Figure 3 correspond to the dynamic effects of the policy over time, as well as the pre-trend tests. We show OLS coefficient estimates in blue and IV coefficient estimates in green.

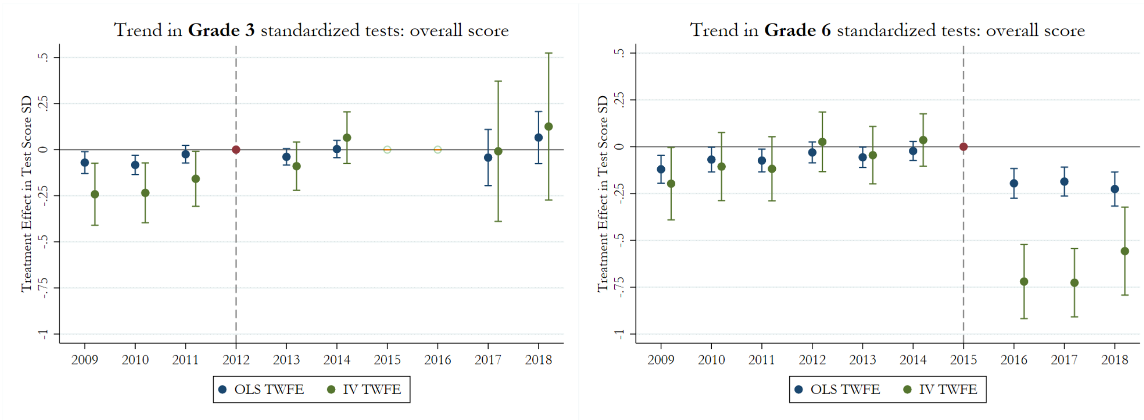
We first consider impacts on test scores in Grade 3. Tests are administered at the end of the academic year, so these results reveal the impacts on test scores at the very end of students' exposure to the MTB-MLE policy, after which they transition to being taught in Tagalog and English in Grade 4 and after. The point estimate in Table 3 for the average causal effect across all post-treatment periods (τ_{post}) is small in magnitude and not statistically significantly different from zero.

Panel (a) of the Figure 3 shows the dynamics of impacts on Grade 3 test scores, in which the null impact of the policy in post-treatment periods is also apparent.[13] There appears to be a positive pre-trend, with students in treated schools on an upward trajectory relative to those in control schools from 2009 to 2011. We therefore also report an honest CI for τ_{post} in Table 3 using the robust inference tools developed by Rambachan and Roth (2023) for which we assume persistence of the pre-treatment linear trend into post-periods. The "honest" 95% CI is considerably shifted leftward for Grade 3 overall test scores, with a lower bound of -0.378 and an upper bound of 0.101. This result for overall test scores is consistent across individual subjects in Grade 3 (see Appendix Table A2), with null coefficient estimates for English, Tagalog, and Mathematics. All told, there is no evidence of a positive impact of the MTB-MLE treatment on test scores in Grade 3.

We now turn to examining Grade 6 test scores, which measure longer-term learning once students transitioned back to instruction in Tagalog and English in Grade 4. In Table 3), the IV coefficient estimate for τ_{post} reveals a substantial decline of two-thirds of a standard deviation (-0.67). This large decline is also very clear in the dynamic treatment effect estimates of Panel (b) of Figure 3.

[13] Grade 3 test scores are missing in 2015 and 2016 because the test was not administered in those years.

Figure 3: Dynamic Impacts on Grade 3 and Grade 6 *Overall* test scores



(a) Grade 3 **(b)** Grade 6

Note: Coefficient estimates (with 95% confidence intervals) from estimation of equation (1). Regressions include school, school year, and province × year fixed effects. Dependent variable is Grade 3 *overall* test scores in Panel (a), and Grade 6 *overall* test scores in Panel (b). Year labels on x-axis indicate end year of each school year (e.g., "2009" indicates SY2008-09.) The pre-period is SY2008-09 to SY2011-12 for Grade 3 test scores, while it is SY2008-09 to SY2014-15 for Grade 6 test (accounting for a 3-year lag relative to SY2011-12). Vertical dashed red line indicates last school year before implementation of MTB-MLE policy (base year in regressions). Figure presents IV estimates where treatment status at the school level is instrumented with school-level linguistic composition variables (see Table 2 for first stage regression estimates.). Standard errors are clustered at the school level.

The decline in Grade 6 test scores holds across individual subjects. In Appendix Table A3 we show that students in treated schools score at least 0.5 standard deviations lower in each individual subject relative to students in control schools, and each of these treatment effects is highly statistically significantly different from zero at conventional levels.

The magnitudes of these negative coefficient estimates are large. In comparison, Evans and Yuan (2019) report that students learn between 0.15 and 0.21 standard deviation of literacy ability in a business-as-usual school year in a sample of low- and middle-income countries. If we extrapolate this to our setting, students are set back a little over three "equivalent years of schooling". Similarly, in a review of the literature, Evans and Yuan (2022) find average learning effect sizes of 0.15 standard deviations across quasi-experimental studies (0.18 standard deviations for reading, and 0.11 for mathematics) with the bottom percentile of -0.76 and a 90th percentile of 0.72. Our coefficient estimate of -0.67 standard deviations for Grade 6 test scores is therefore very near the lowest end of learning effect sizes in the literature on educational interventions in developing countries.

Appendix Table A3 also reports the OLS coefficient estimates for τ_{post} across sub-

Table 3: IV Estimates: Average Causal Effects on Grade 3 and Grade 6 *Overall* test scores

	Grade 3 Overall Score	Grade 6 Overall Score
	(1)	(2)
$\bar{\tau}_{\text{post}}$	0.023	-0.668***
	(0.086)	(0.093)
Honest CI (*smoothness restrictions*, $\bar{M} = 0$)	[-0.378, 0.101]	[-0.957, -0.532]
Control Mean, Pre Period	0.00	0.00
Year FE	Y	Y
School FE	Y	Y
Province × Year FE	Y	Y
Obs. (Students)	856,610	1,102,528
Clusters (Schools)	23,712	23,395

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates on Grade 3 overall test score in column (1) and Grade 6 overall test score in column (2) for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1) using the instrumented DID specification where treatment status as the school level is instrumented with school-level linguistic composition variables pre-policy, i.e., the percentage of learners at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction as well as a square and cubic term in the percentage of Tagalog-speaking learners (see Table 2). The pre-period is SY2008-2009 to SY2011-2012 for Grade 3 test scores, while it is SY2008-2009 to SY2014-2015 for Grade 6 test (accounting for a 3-year lag relative to SY2011-2012). See Figure 3 for per period coefficient estimates.

jects. Likewise, these estimates are negative and statistically significant but smaller in magnitude than the IV estimates, with a 0.2 standard deviation decline in overall Grade 6 test scores. This difference between IV coefficient estimates and OLS coefficient estimates can also be seen in Panel (b) of Figure 3. The fact that IV estimates are more negative than OLS estimates suggests the presence of positive selection into treatment, which leads the OLS coefficient to be positively biased ($\bar{\tau}^{\text{IV}}_{\text{post}} < \bar{\tau}^{\text{OLS}}_{\text{post}}$). Schools better able to teach in the local mother tongue may have been more likely to select into treatment, leading to positive bias in the OLS estimates.

4.1.6 Impacts on Enrollment

Next, we turn to grade by grade impacts on enrollment from Grades 1 to 6. Table 4 shows coefficient estimates for the average causal effect τ_{post} from our preferred IV specification while Appendix Table A4 reports OLS estimates. Strikingly, we find an interesting pattern across grades: enrollment in Grade 1 and Grade 2 increases considerably. This likely represents increased demand for instruction in the mother tongue in early grades. This may also reflect students increasingly repeating grades in early primary school. Moreover, enrollment significantly declines starting in Grade 3, and increasingly so in Grade 4, stabilizing in Grade 5 and Grade 6 with an esti-

Table 4: IV Estimates: Average Causal Effects on Grade 1 to Grade 6 Enrollment Counts

	Gr 1	Gr 2	Gr 3	Gr 4	Gr 5	Gr 6
	(1)	(2)	(3)	(4)	(5)	(6)
$\bar{\tau}_{\text{post}}$	23.604***	2.942**	-3.391***	-7.079***	-7.174***	-6.890***
	(2.728)	(1.298)	(1.248)	(1.543)	(1.281)	(1.148)
Honest CI (*smoothness restrictions*, $\bar{M} = 0$)	[18.513, 35.496]	[2.748, 12.406]	[-5.177, 3.835]	[-6.507, 0.953]	[-8.116, -2.081]	[-7.390, -2.545]
Control Mean, Pre Period	101.3	85.3	81.6	78.5	77.0	73.9
Year FE	Y	Y	Y	Y	Y	Y
School FE	Y	Y	Y	Y	Y	Y
Province × Year FE	Y	Y	Y	Y	Y	Y
Observations	241,572	239,459	241,572	241,572	238,654	238,654
Clusters (Schools)	24,527	24,527	24,527	24,527	24,234	24,234

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates on grade-level enrollment from Grade 1 to Grade 6 for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1) using the instrumented DID specification where treatment status as the school level is instrumented with school-level linguistic composition variables pre-policy, i.e., the percentage of learners at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction as well as a square and cubic term in the percentage of Tagalog-speaking learners (see Table 2). The pre-period is SY2008-2009 to SY2011-2012 for Gr 1 to 3, it is shifted by 1 year for Gr 4, two years for Gr 5, and three years for Gr 6 (allowing time for treated students to reach these grades).

mated decrease in enrollment of 7 students per school and per grade corresponding to approximately 9% of the pre-treatment control mean in Grades 4, 5 and 6. This suggests that as students get exposed to instruction in the mother tongue and the implementation challenges associated with the policy, and progress through grades, they either repeat grades, transfer out to control schools, or drop out of school altogether.

4.1.7 Impacts on Teacher Counts and Student-Teacher Ratios

In response to the MTB-MLE policy, teachers may also be induced to leave the public sector or move to control schools. We test this hypothesis in this subsection.

Table 5 shows coefficient estimates for the average post-treatment causal effect τ_{post} (equation 2) for the Grades 1 to 6 teacher count and the Student-Teacher Ratio from the estimation of equation (1). Appendix Table A5 reports OLS estimates. Appendix Figure A2 presents the corresponding per period dynamic effects and shows a clear and statistically significant reduction in the number of primary school teachers post-policy, with up to one teacher leaving treated schools, on average. This corresponds to a 7.5% decline relative to the control mean from SY 2008-2009 to SY 2011-2012 of 13.26 teachers per school.

Column 2 of the Table 5 displays the estimated impact on the student-teacher ratio (total students divided by total teachers at the school level).[14] The point estimate

[14]In our regression analyses, we drop student-teacher ratio values below 12, the 1st percentile, to exclude nonsensical outliers below 1, as well as values above 96, the 99th percentile, to exclude the few outliers greater than 1000. Results are robust to alternative specifications using either the raw

Table 5: IV Estimates: Average Causal Effects on Teacher Counts and Student-Teacher Ratio

	Teacher Count	Student-Teacher Ratio
	(1)	(2)
$\bar{\tau}_{\text{post}}$	-0.994***	0.379
	(0.192)	(0.644)
Honest CI (*smoothness restrictions*, $\bar{M} = 0$)	[-1.089, 0.100]	[-3.612, 1.966]
Control Mean, Pre Period	13.26	35.58
Year FE	Y	Y
School FE	Y	Y
Province × Year FE	Y	Y
Observations	239,140	236,494
Clusters (Schools)	24,461	24,431

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01. Coefficient estimates on elementary teacher count and the student-teacher ratio for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1), using IV DID specification where treatment status as the school level is instrumented with school-level linguistic composition variables (see Table 2 for first stage estimates). The pre-period is SY2008-09 to SY2011-12. For student-teacher ratio, values below the 1st and above the 99th percentiles were dropped. See Figure A2 for per period coefficient estimates.

is positive, but small in magnitude and not statistically significantly different from zero at conventional levels. There is no indication that on net (considering changes in student enrollment and changes in teacher counts) there was a substantial change in the student-teacher ratio as a result of the MTB-MLE policy.

4.2 Analyses using Census Data: Years of Completed Education

Empirical estimates presented so far – impacts of MTB-MLE on DepEd administrative data outcomes (test scores, student enrollment, and teacher counts) – are potentially subject to selection bias. Of particular concern is the evidence of changes in student enrollment numbers (increases in earliest grades, and declines in later grades), which is likely due to a combination of lower grade advancement (increased grade repetition) and movements of students across schools in response to treatment.

In this section we therefore turn to analyses that are immune from most selection bias concerns: analysis of outcomes in Philippine Census data, in which we construct a measure of treatment at the level of municipalities rather than schools. The outcome of interest will be years of completed education. We also overcome the potential selection bias concern of movements of students across municipalities by exclusively relying on municipality information that predates implementation of the MTB-MLE policy (in SY2012-2013) for assignment of treatment intensity, namely the municipal-

data or the winsorized data at the 1st and 99th percentile.

ity of *birth* of exposed cohorts in the 2020 census.

4.2.1 Empirical Approach

Our empirical approach for analysis of years of completed education in the Census data builds on birth cohort difference-in-difference (double-difference) approaches used in prior work such as Duflo (2001) and Shrestha (2017). In our context, an analogous double-difference approach would compare cohorts young enough to be treated at the time of MTB-MLE's implementation in SY2012-13 with older cohorts (the first difference), across more- vs. less-treated municipalities (the second difference). We extend this approach by adding a third difference across Census rounds: we compare the post-treatment Census round (2020) with the pre-treatment Census round (2010). This yields a triple-difference research design, where the differences are across cohorts, municipalities, and Census rounds.

Adding this third difference allows us to test as well as account for any potential violations of the parallel-trend assumption in the double-difference research design (i.e., the possibility of pre-existing differential trends across cohorts related to municipality-level treatment intensity). In addition to reporting the triple-difference regression results, we will also unpack the triple difference by showing double difference results separately for 2010 and 2020 Censuses (in which the 2010 Census analysis can be thought of as a "placebo" experiment.)

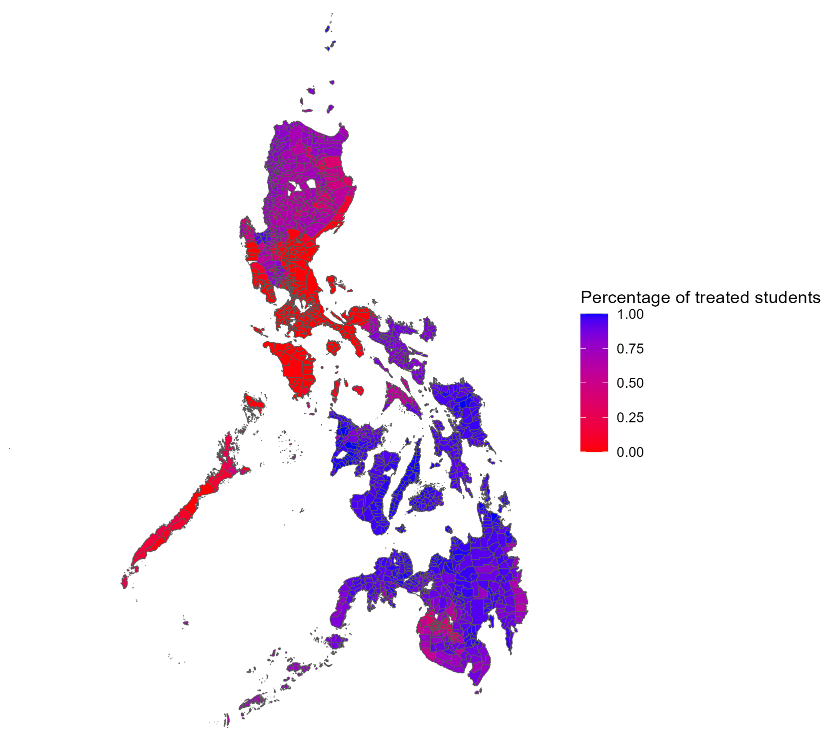
Municipality-Level Treatment Intensity. In our analysis of impacts on outcomes in the Census, the causal variable of interest is a measure of treatment intensity at the municipality level. We focus on a municipality-level treatment measure because a key aim of the analysis of Census outcomes is to ameliorate bias concerns related to self-selection of students across schools.[15]

We construct a municipality-level treatment intensity measure, $Treat_m$, by exploiting the first stage regression predicting a school's treatment status using student linguistic composition pre policy. For all public primary schools, we estimate the probability of treatment using information on the baseline linguistic composition of the school and first stage coefficient estimates from Table 2. We then aggregate the

[15] In any case, school information is not present in the Census, only municipality. For the 2020 Census we focus on respondents' municipality of birth so that municipality cannot be endogenous to the treatment intensity indicator (all cohorts of interest in our analyses were born before the MTB-MLE policy). For the 2010 Census, municipality of birth is not recorded, but as this is prior to the MTB-MLE policy, concerns about endogenous selection do not arise.

school-level predictions to the municipality level, weighting each school-level prediction with its baseline primary enrollment share in the municipality. This generates a municipality-level treatment intensity which varies between 0 and 1 for the 1,627 municipalities in our sample, and is interpreted as the predicted share of students in the municipality who are exposed to the MTB-MLE policy (i.e., who are taught in Kindergarten to Grade 3 in a non-Tagalog mother tongue). We provide further details on construction of $Treat_m$ in Appendix Section A.1.5. Figure 4 displays the spatial variation in municipality-level treatment intensity across the Philippines.

Figure 4: Municipality-level Treatment Variation Across Space



Note: This figure shows the geospatial variation of our treatment intensity variable defined at the municipality level. Treatment intensity varies from 0 to 1 and is defined as the predicted percentage of treated students at the municipality level. It is constructed using the predicted values from the first stage regression presented in Table 2 for all schools with linguistic composition data. It is then aggregated up at the municipality level weighting each school's predicted probability of treatment with the size of the Grades 1 to 6 student population in 2012-2013. Darker shades of red represent a lower treatment intensity while darker shades of blue correspond to higher treatment intensities. White shading indicates municipalities excluded from the analysis (for which either census data or linguistic composition data is missing).

Variation Across Birth Cohorts. The MTB-MLE policy only affected Kindergarten to Grade 3. Empirically, the modal age of children in Grade 3 is 9, and the modal age of children in Grade 4 is 10. We thus consider that individuals aged 18 or above in the 2020 Census were not exposed to the policy, because they were beyond the typical age for Grade 3 when the MTB-MLE policy was implemented in SY2012-13 (they were aged 10 or above in 2012). By contrast, we consider respondents aged 17 or below in the 2020 Census to be exposed to the policy as they were 9 or below in 2012, and thus were of age to be in Grade 3 or a lower grade in that year. (There is of course some fuzziness around these age cutoffs since children can be older or younger than the typical age for their grade. Such fuzziness will attenuate our treatment effect estimates.)

Regression Equation. Our triple-difference analyses of Census data exploit variation in treatment exposure exposure across age (birth cohort), municipalities, and Census rounds. We estimate the following regression equation:

$$\begin{aligned} Y_{iampr} = {} & \beta_{\mathrm{TD}}\, \mathbf{1}\{\mathrm{Age}_a < 18\} \times Treat_m \times \mathbf{1}\{\mathrm{Census}_r = 2020\} \\ & + \eta_{ar} + \eta_{mr} + \eta_{am} + \eta_{apr} + \varepsilon_{iampr}, \end{aligned} \tag{3}$$

where Y_{iampr} is the outcome of interest (primarily years of schooling completed) for respondent i from the age a cohort, born in municipality m, in province p, in Census round r.

$Treat_m$ is the municipality-level treatment intensity variable, the predicted share of students exposed to the MTB-MLE policy in the municipality. The coefficient of interest is β_{TD}, the triple-difference estimate: the impact of local treatment intensity on cohorts young enough to be treated (vs. older cohorts), differentially in the 2020 Census vs. the 2010 Census.

Fixed effects η_{ar}, η_{mr}, and, η_{am} are all double interactions between a, m and r. As in prior analyses, the inclusion of province $\times$ age $\times$ Census round fixed effects η_{apr} ensures that we focus exclusively on deviations from a province-specific time (Census round) effects for each age cohort. We cluster standard errors at the municipality level.

Identification. As discussed by Olden and Møen (2022), for this approach to have a causal interpretation, we must make a *relative* parallel trend assumption. In our setting, and with a binary interpretation, this requires that the relative outcome in 2020 of those born in treated municipalities vs. control municipalities to trend (across birth cohorts) in the same way as the relative outcome in 2010 of those born in treated municipalities vs. control municipalities in the absence of treatment.

Recall however that Treat_m varies continuously from 0 to 1. As Callaway et al. (2024) discuss for the difference-in-difference setting, for our estimates to have a causal interpretation we must make a stronger assumption: *generalized parallel trends* which involves potential outcomes under different doses of the treatment intensity. This assumes that the *observed* outcome changes for respondents in municipalities in each treatment intensity level reflect what would have happened – the counterfactual – for respondents in all other treatment intensity levels had they received that dose.

We also estimate the *dynamic* analog of equation (3) which allows for differential causal impacts across age cohorts relative to individuals aged 18 (which we take as the "base" cohort):

$$\begin{aligned} Y_{iampr} = & \sum_{\substack{h=7 \\ h \neq 18}}^{h=25} \tau_h \; (\mathbf{1}\{\text{Age}_a = h\} \times Treat_m \times \mathbf{1}\{\text{Census}_r = 2020\}) \\ & + \eta_{ar} + \eta_{mr} + \eta_{am} + \eta_{apr} + \varepsilon_{iampr}. \end{aligned} \tag{4}$$

This equation modifies equation (3) by estimating a different treatment effect for each cohort. The τ_h coefficients are the parameters of interest. All estimates are relative to the base cohort, 18-year-olds ($h = 18$), the cohort of age to have been in Grade 4 (one grade above the highest MTB-MLE grade) in the policy implementation school year (2012-13). For individuals young enough to be exposed to the policy ($h < 18$), τ_h are triple-difference treatment effects by cohort. For individuals of ages such that they should not have been exposed to the policy ($h \geq 19$), the cohort-specific effects provide tests of pre-trends (a partial test of the generalized parallel trend assumption).

4.2.2 Impact of MTB-MLE Policy on Years of Completed Education

Table 6 shows the coefficient estimate for the triple-difference treatment effect β_{TD} from estimation of equation (3) for years of completed education (column 1). Coefficient estimates for the component double-differences are also shown, for the 2010 (column 2) and 2020 Census (column 3). The 2010 Census double-difference corresponds to a placebo test.

We find a negative and statistically significant impact of the policy on years of completed education. The coefficient estimate in column 1 indicates that an individual

Table 6: Effect of MTB-MLE Policy on Years of Completed Education

	Full samp. Triple-diff.
	(1)
$Treat_m$ × (Age = 7-17) × (Census = 2020)	-0.334*** (0.079)
Mean Dep. Var. (Age = 7-17), Pre	5.05
Census × Age FE	Y
Municipality × Census FE	Y
Municipality × Age FE	Y
Province × Age × Census FE	Y
Observations	73,267,484
Clusters (Municipalities)	1,627

	2010 Double-diff.	2020 Double-diff.
	(2)	(3)
$Treat_m$ × (Age = 7-17)	0.0442 (0.204)	-0.353** (0.177)
Mean Dep. Var. (Age = 7-17)	5.05	5.74
Age FE	Y	Y
Municipality FE	Y	Y
Province × Age FE	Y	Y
Observations	35,025,793	38,241,691
Clusters (Municipalities)	1,585	1,627

Note: Standard errors clustered at municipality level. *** p<0.01, ** p<0.05, * p<0.01. Coefficient estimates for years of completed education for triple-difference treatment effect β_{TD} from estimation of equation (3) (column 1) and from the double-difference analogs for each Census round in columns (2) and (3). Treatment intensity at the municipality level ($Treat_m$) is predicted share of treated students, ranging from 0 to 1; see Figure 4 for spatial variation in $Treat_m$. The sample includes respondents aged 7 to 25 from the 2010 and 2020 Censuses. Treated cohorts are those aged 7 to 17 in the 2020 Census round.

in a fully-treated municipality (one with 100% of students treated) and young enough to be exposed to the policy has 0.33 fewer years of completed education on average.

Columns 2 and 3 of Table 6 decompose the triple difference estimate into two double differences. Column 2, the double difference for the 2010 Census, is a placebo test in the "false" or pre-treatment period. The coefficient estimate is close to zero and is not statistically significantly different from zero at conventional levels, providing no indication of worrying violations of the parallel trend assumption in the pre-treatment period.

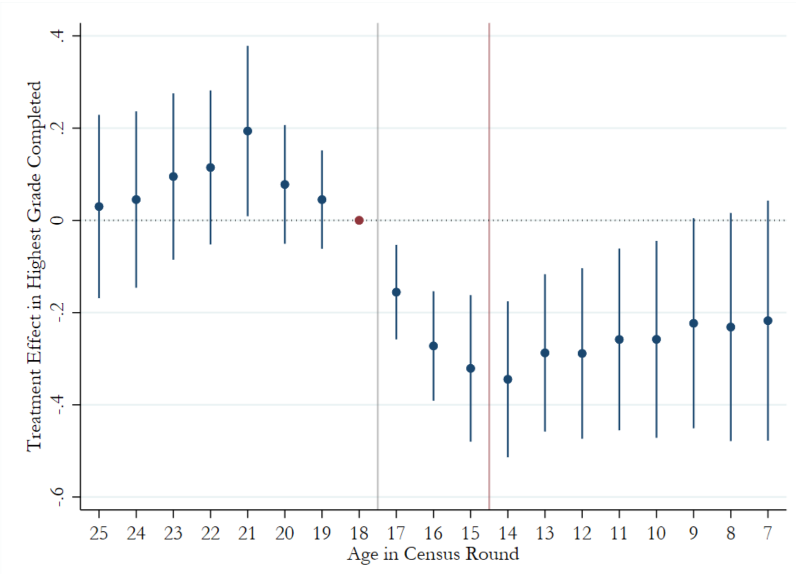
Column 3 is the double difference treatment effect estimate in the "true" treatment period, the 2020 Census. This coefficient estimate – a decline of a third of a year – is very similar to the triple-difference estimate of column 1 (consistent with the close-to-zero estimate of column 2).[16]

Figure 5 reports dynamic coefficient estimates (across age cohorts) from estimation of equation (4). For cohorts aged 17 and younger (to the right of the graph), estimates are causal effects of the policy for each individual age cohort relative to the omitted 18 years old cohort (at the time of the 2020 Census round). The age 18 cohort (the youngest untreated cohort) is the base or reference cohort. For cohorts aged 19 or

[16] It is of note that the standard error on the triple-difference coefficient estimate in column 1 is much smaller than that of the double-difference estimate of column 3. This increase in precision likely derives from inclusion of municipality × age fixed effects in the triple-difference regression equation, which absorb substantial residual variation. These fixed effects cannot be included in the double-difference regression equation because they would absorb all identifying variation.

above, the estimates provide a visual sense of the existence of pre-trends.

Figure 5: Dynamic Impacts (across age cohorts) on Highest Grade Completed



Note: Triple-difference coefficient estimates (with 95% confidence intervals) from estimation of equation (4). Respondents aged 18 at the time of the Census round (the youngest untreated cohort) are the base or reference age cohort. See Table 6 for other notes.

In interpreting the figure, it is useful to recall that exposure to the policy varies non-linearly with age in 2020. Students who were aged 9 and in Grade 3 in school year 2012-2013 are aged 17 in the 2020 Census. Those aged 17 in the 2020 Census typically have one year of exposure (in Grade 3) to MTB-MLE. Those aged 16 in 2020 have 2 years of exposure (Grades 2 and 3), while those aged 15 have 3 years of exposure (Grades 1, 2 and 3), and those aged 14 have 4 years of exposure (Kindergarten to Grade 3) typically.[17]

The profile of our dynamic triple-difference estimated coefficients across cohorts does reveal treatment effects becoming more negative for cohorts exposed to more years of the policy. The magnitude of the impact increases from those aged 17 to those aged 14 in 2020, with those aged 14 the most negatively affected by the policy with a 0.34 decline in the number of completed years of education. The effect then appears to stabilize for younger cohorts at around -0.25 completed years of education.[18]

[17]Exposure also decreases with age from age 9 to age 7 (in 2020) because students aged 7 or 8 have not yet reached Grade 3 in 2020.

[18]Treatment effects across cohorts may also vary if the policy's implementation quality improves over time. This could be another reason estimated effects appear to have a slight positive trend in the youngest cohorts. However, 95% confidence intervals are wide, so no strong statement about

Examining the coefficient estimates for the untreated, older cohorts (those aged 19 or older), there is no clear indication of any violation of parallel trends in the pre-period. Only one out of seven pre-period coefficient estimates is statistically significantly different from zero at conventional levels: the coefficient estimate for age 21, which appears to be an outlier. Having one out of seven coefficient estimates in the pre-period be statistically significantly different from zero is not too unlikely to arise by chance. There is no consistent upward or downward trend in the pre-period (coefficient estimates first rise from age 25 to 21, then fall afterwards).

Aside from examining years of completed education as an outcome, we can also examine as dependent variables indicators for completion of each different educational level in the data. Appendix Table A6 shows triple-difference coefficient estimates from estimation of equation (3), where the dependent variable in each regression is an indicator (binary) variable for completion of individual grades (Grades 1 to 10 in columns 1 to 10, respectively). We find large, statistically significant negative impacts in grade completion (ranging from 1-2 percentage points) in Grades 4 through 8. Point estimates for Grades 9 and 10 are as large in magnitude, but are not themselves statistically significant at conventional levels.

In sum, individuals in the 2020 Census (eight years after the policy was initiated) who were fully exposed to the MTB-MLE policy complete one-third fewer years of education on average. Negative effects of the policy extend to completion of levels of schooling after the grades during which the policy was actually implemented (Kindergarten to Grade 3). Most prominently, we find negative effects on completion of Grades 4 to 8 (the last years of primary school and the first years of secondary school). All told, the analyses in this section reveal negative impacts of the MTB-MLE policy on years of completed schooling at the population level.

5 Conclusion

In this paper, we exploit a unique natural experiment in the Philippines to examine the long-term consequences of early education quality. We find that the quality of education in the first years of schooling has substantial and enduring effects on academic achievement and educational attainment in later levels of schooling.

The unexpected decline in educational quality resulting from the flawed imple-

trends in treatment effect magnitudes is possible.

mentation of a mother tongue education policy allows us to isolate the causal impact of early education quality on later outcomes. We find that students exposed to lower quality early education experienced significant declines in Grade 6 test scores across all subjects. Corresponding declines in student enrollment and teacher retention provide additional evidence of lower education quality in treated schools. In addition, analysis of census data reveals long-lasting impacts: fully-affected cohorts completed one-third fewer years of schooling by 2020, including for students whose last exposure to the policy was up to eight years in the past.

These results have important implications for both theory and policy. From a theoretical perspective, our findings support the hypothesis that early educational experiences play a crucial role in shaping longer-term academic trajectories. The persistence of effects we observe underscores the complementarity between early and later human capital investments, as posited by Cunha and Heckman (2007). Our results also contribute to the ongoing debate about fade-out versus persistence of impacts of the early educational environment, providing evidence for persistence in the context of a broad, system-wide change in educational quality.

From a policy standpoint, our study highlights the critical importance of maintaining and improving the quality of early education. The substantial long-term costs associated with even temporary declines in educational quality suggest that policymakers should exercise extreme caution when implementing reforms that could potentially disrupt early learning environments. Furthermore, our findings emphasize the need for careful planning and piloting of educational reforms, particularly in multilingual contexts where language of instruction policies can have far-reaching consequences.

While our study focuses on the Philippines, the implications of our findings likely extend to other developing countries grappling with similar challenges in education policy implementation. The magnitude and persistence of the effects we observe underscore the high stakes involved in early education quality and the potential for significant losses from negative shocks to educational quality.

Future research could build on our findings by exploring the specific mechanisms through which early education quality affects long-term outcomes, and by investigating potential interventions to mitigate the negative impacts of temporary declines in educational quality. Additionally, longer-term follow-up studies could examine whether the effects we observe persist into adulthood, affecting labor market outcomes and other life circumstances.

In conclusion, our study provides robust causal evidence on the enduring impact of early education quality on academic achievement and educational attainment. These findings underscore the critical importance of prioritizing and protecting the quality of early education as a key component of human capital development.

References

ALMOND, D. AND J. CURRIE (2011): "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 25, 153–172.

ANGRIST, J., A. CHIN, AND R. GODOY (2008): "Is Spanish-Only Schooling Responsible for the Puerto Rican Language Gap?" *Journal of Development Economics*, 85, 105–128.

ANGRIST, J. AND V. LAVY (1997): "The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco," *Journal of Labor Economics*, 15, S48–S76.

ANGRIST, N., M. AINOMUGISHA, S. P. BATHENA, P. BERGMAN, C. CROSSLEY, C. CULLEN, T. LETSOMO, M. MATSHENG, R. M. PANTI, S. SABARWAL, ET AL. (2023): "Building resilient education systems: Evidence from large-scale randomized trials in five countries," *NBER Working Paper No. 31208*.

ANGRIST, N. AND R. MEAGER (2023): "Implementation matters: Generalizing treatment effects in education," *Available at SSRN 4487496*.

ARGAW, B. (2016): "Quasi-Experimental Evidence on the Effects of Mother Tongue-Based Education on Reading Skills and Early Labour Market Outcomes," *ZEW-Centre for European Economic Research Discussion Paper*.

BAILEY, D. H., G. J. DUNCAN, F. CUNHA, B. R. FOORMAN, AND D. S. YEAGER (2020): "Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions," *Psychological Science in the Public Interest*, 21, 55–97.

BALL, J. (2010): "Enhancing learning of children from diverse language backgrounds: Mother tongue-based bilingual or multilingual education in early childhood and early primary school years," *UNESCO*.

BANERJEE, A., R. BANERJI, J. BERRY, E. DUFLO, H. KANNAN, S. MUKHERJI, M. SHOTLAND, AND M. WALTON (2016): "Mainstreaming an Effective Intervention: Evidence from Randomized Evaluations of "Teaching at the Right Level" in India," *NBER Working Paper No. 22746*.

Barro, R. J. and J. W. Lee (2013): "A New Data Set of Educational Attainment in the World, 1950–2010," *Journal of Development Economics*, 104, 184–198.

Bold, T., M. Kimenyi, G. Mwabu, J. Sandefur, et al. (2018): "Experimental evidence on scaling up education reforms in Kenya," *Journal of Public Economics*, 168, 1–20.

Borusyak, K., X. Jaravel, and J. Spiess (2024): "Revisiting Event-Study Designs: Robust and Efficient Estimation," *The Review of Economic Studies*, rdae007.

Brillantes, K. D. B., A. C. Orbeta, K. A. Francisco-Abrigo, E. M. Capones, and J. B. B. Jovellanos (2019): "Status of Senior High School Implementation: A Process Evaluation," *PIDS Discussion Paper No. 2019-13.*

Bühmann, D. and B. Trudell (2008): *Mother Tongue Matters: Local Language as a Key to Effective Learning*, Paris: UNESCO.

Callaway, B., A. Goodman-Bacon, and P. H. Sant'Anna (2024): "Difference-in-differences with a continuous treatment," *NBER Working Paper No. 32117.*

Callaway, B. and P. H. C. Sant'Anna (2021): "Difference-in-Differences with Multiple Time Periods," *Journal of Econometrics*, 225, 200–230.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2024): "On binscatter," *American Economic Review*, 114, 1488–1514.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 126, 1593–1660.

CNN Philippines (2022): "DepEd to Remove Mother Tongue Subject in K-12 Curriculum," https://web.archive.org/web/20221017122623/https://www.cnnphilippines.com/news/2022/10/17/deped-mother-tongue-subject-k12.html, accessed: 2023-09-07.

Cunha, F. and J. Heckman (2007): "The Technology of Skill Formation," *American Economic Review*, 97, 31–47.

Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006): "Interpreting the Evidence on Life Cycle Skill Formation," *Handbook of the Economics of Education*, 1, 697–812.

Currie, J. and D. Almond (2011): "Human Capital Development Before Age Five," in *Handbook of Labor Economics*, Elsevier, vol. 4, 1315–1486.

De Chaisemartin, C. and X. D'Haultfoeuille (2023): "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey," *The Econometrics Journal*, 26, C1–C30.

DepEd (2009): "DepEd Order No. 74 s. 2009: Institutionalizing Mother Tongue-Based Multilingual Education (MLE)," *Republic of the Philippines*.

——— (2012): "DepEd Order No. 16 s. 2012: Guidelines on the Implementation of the Mother Tongue-Based Multilingual Education (MTB-MLE)," *Republic of the Philippines*.

——— (2015): "DepEd Order No. 55 s. 2015: Utilization of Language Mapping Data for Mother Tongue-Based Multilingual Education (MTB-MLE) Program Implementation," *Republic of the Philippines*.

Duflo, E. (2001): "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment," *American Economic Review*, 91, 795–813.

Ebernhard, D. M., G. F. Simons, and C. D. Fennig (2023): "Ethnologue: Languages of the World," http://www.ethnologue.com.

Evans, D. K. and F. Yuan (2019): "Equivalent years of schooling: A metric to communicate learning gains in concrete terms," *World Bank Policy Research Working Paper*.

——— (2022): "How Big Are Effect Sizes in International Education Studies?" *Educational Evaluation and Policy Analysis*, 44, 532–540.

Glewwe, P. and K. Muralidharan (2016): "Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications," in *Handbook of the Economics of Education*, Elsevier, vol. 5, 653–743.

HANSON, M. (2024): "U.S. Public Education Spending Statistics," https://educationdata.org/public-education-spending-statistics, Accessed online on 07/14/2024.

HECKMAN, J. J. (2006): "Skill formation and the economics of investing in disadvantaged children," *Science*, 312, 1900–1902.

HEUGH, K. (2012): "Theory and practice - language education models in Africa: research, design, decision-making and outcomes," in *Optimising Learning, Education and Publishing in Africa: The Language Factor: A Review and Analysis of Theory and Practice in Mother-Tongue and Bilingual Education in sub-Saharan Africa*, ed. by A. Ouane and C. Glanz, Hamburg: UNESCO, 105–156.

HOUSE OF REPRESENTATIVES OF THE PHILIPPINES (1994): "Republic Act No. 7836, December 16, 1994 : An Act to Strength the Regulation and Supervision of Teaching in the Philippines and Prescribing a Licensure Examination for Teachers and for Other Purposes," *Republic of the Philippines*.

INQUIRER (2023): "Gatchalian laments poor execution of mother tongue-based learning," https://newsinfo.inquirer.net/1799020/a-mere-experiment-gatchalian-laments-poor-execution-of-mother-tongue-based-learnin accessed: 2023-09-07.

KREMER, M., C. BRANNEN, AND R. GLENNERSTER (2013): "The Challenge of Education and Learning in the Developing World," *Science*, 340, 297–300.

LAITIN, D. D., R. RAMACHANDRAN, AND S. L. WALTER (2019): "The legacy of colonial language policies and their impact on student learning: Evidence from an experimental program in Cameroon," *Economic Development and Cultural Change*, 68, 239–272.

LIST, J. A. (2022): *The voltage effect: How to make good ideas great and great ideas scale*, Crown Currency.

LUDWIG, J. AND D. L. MILLER (2007): "Does Head Start improve children's life chances? Evidence from a regression discontinuity design," *The Quarterly Journal of Economics*, 122, 159–208.

Maccini, S. and D. Yang (2009): "Under the Weather: Health, Schooling, and Economic Consequences of Early-Life Rainfall," *American Economic Review*, 99, 1006–1026.

Mbiti, I., K. Muralidharan, M. Romero, Y. Schipper, C. Manda, and R. Rajani (2018): "Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania," *The Quarterly Journal of Economics*.

Metila, R., L. Pradilla, and A. Williams (2016): "Investigating best practice in Mother Tongue-Based Multilingual Education (MTB-MLE) in the Philippines, Phase 2 progress report: Patterns of challenges and strategies in the implementation of mother tongue as medium of instruction in the early years: A nationwide study," Tech. rep., Assessment, Curriculum and Technology Research Centre (ACTRC).

Monje, J. D., A. C. J. Orbeta, K. A. Francisco-Abrigo, and E. M. Capones (2019): "Starting Where the Children Are: A Process Evaluation of the Mother Tongue-Based Multilingual Education Implementation," *PIDS Discussion Paper No. 2019-06*.

Mullis, I. V., M. O. Martin, P. Foy, D. L. Kelly, and B. Fishbein (2020): "TIMSS 2019 international results in mathematics and science," .

Muralidharan, K., A. Singh, and A. J. Ganimian (2019): "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India," *American Economic Review*, 109, 1426–60.

Nolasco, R. M. (2007): "Filipino and Tagalog, Not So Simple," https://web.archive.org/web/20140522052247/http://svillafania.philippinepen.ph/2007/08/articles-filipino-and-tagalog-not-so.html, accessed: 2024-08-14.

OECD (2018): "Programme for International Student Assessment (PISA) Results: Country Note Philippines," .

Olden, A. and J. Møen (2022): "The Triple Difference Estimator," *The Econometrics Journal*, 25, 531–553.

Piper, B. et al. (2018): "Examining the Secondary Effects of Mother-Tongue Literacy Instruction in Kenya: Impacts on Student Learning in English, Kiswahili, and Mathematics," *International Journal of Educational Development*, 59, 110–127.

Pritchett, L. (2013): *The Rebirth of Education: Schooling Ain't Learning*, CGD Books.

Pritchett, L., M. Woolcock, and M. Andrews (2013): "Looking like a state: techniques of persistent failure in state capability for implementation," *The Journal of Development Studies*, 49, 1–18.

Professional Regulation Commission (2016): "Program of the Board Licensure Examination For Professional Teachers (BLEPT) on September 25, 2016," *Republic of the Philippines*.

Ramachandran, R. (2017): "Language Use in Education and Human Capital Formation: Evidence from the Ethiopian Educational Reform," *World Development*, 98, 195–213.

Rambachan, A. and J. Roth (2023): "A more credible approach to parallel trends," *Review of Economic Studies*, 90, 2555–2591.

Rubin, D. B. (1980): "Randomization analysis of experimental data: The Fisher randomization test comment," *Journal of the American Statistical Association*, 75, 591–593.

Rubrico, J. G. U. (1998): "The metamorphosis of Filipino as national language," *mimeo*.

Schweinhart, L. J., J. Montie, Z. Xiang, W. S. Barnett, C. R. Belfield, and M. Nores (2005): *Lifetime effects: The High/Scope Perry Preschool study through age 40*, Ypsilanti, MI: High/Scope Press.

Shrestha, S. A. (2017): "No man left behind: Effects of emigration prospects on educational and labour outcomes of non-migrants," *The Economic Journal*, 127, 495–521.

Taylor, S. and M. von Fintel (2016): "Estimating the Impact of Language of Instruction in South African Primary Schools: A Fixed Effects Approach," *Economics of Education Review*, 50, 75–89.

The Manila Times (2020): "Cordillera advocacy group urges rejection of Mother Tongue policy," https://www.manilatimes.net/2020/12/03/campus-press/cordillera-advocacy-group-urges-rejection-of-mother-tongue-policy/804196/, accessed: 2023-09-07.

Tupas, R. and I. P. Martin (2017): "Bilingual and Mother Tongue-Based Multilingual Education in the Philippines," in *Bilingual and Multilingual Education*, ed. by O. García, A. Lin, and S. May, Springer, Encyclopedia of Language and Education.

UNESCO (2010): *Education for All Global Monitoring Report 2010: Reaching the Marginalized*, UNESCO Paris.

World Bank (2018): "Learning to Realize Education's Promise," *World Development Report. The World Bank*.

World Bank, UNESCO (2021): "Philippines - Learning Poverty Brief - 2021 (English)," Tech. rep.

Ye, T., A. Ertefaie, J. Flory, S. Hennessy, and D. S. Small (2023): "Instrumented difference-in-differences," *Biometrics*, 79, 569–581.

A Appendix

A.1 Data Appendix

Here we share further details about variables used in the empirical analyses, and provide additional statistics.

A.1.1 Standardized Test Scores

Standardized test scores used as dependent variables in our analyses come from the National Achievement Test (NAT), administered by DepEd in Grade 3 and Grade 6.

We normalize NAT scores as follows. Both Grade 3 and Grade 6 test scores were originally raw scores graded on arbitrary scales varying by subject (e.g. out of 10, out of 20, or even out of 27). We normalize test scores, across test takers in our main sample of public schools with MOI information, in each school year, and for each subject, using the mean and the standard deviation of test scores of students from control schools (schools that chose Tagalog as their MOI).

Results for Grade 3 should be interpreted with caution because of changes in the test content and test language resulting from the policy itself. The test changed from *NAT G3* from SY 2008-09 to 2013-14, to *LAPG G3* in 2015-2016 which did not include a mathematics test, to *ELLNA G3* in 2016-17 and 2017-18. Moreover, the Grade 3 test has missing years with no or limited nationwide standardized testing (school years 2014-15 and 2015-16). Grade 6 test scores are a more attractive outcome to measure the impact of the policy because (i) they were not affected by these changes, (ii) they were consistent across the study period, and (iii) they measure longer-term learning once students transitioned back to the dominant language for instruction.

A.1.2 Years of Completed Education

Our key outcome in analysis of the Census data is years of completed education. We define this variable as follows. We use the Census questionnaire item "Highest Grade Completed" from the 2010 and 2020 rounds of the Philippine Census of Population and Housing. This item is asked in the individual roster section of the Census questionnaire of all household members who are age 5 and above. We convert respondents' answers to an integer years of completed education that ranges from 0 to 16.

For responses ranging from *Grade 1* to *Grade 10*, years of completed education is taken to be equal to the grade reported as highest grade completed.

For higher levels of educational attainment, we need to account for the Philippines' shift from K-10 to K-12 education in SY2016-17, at which point two years ("senior high school", Grades 11 and 12) were added to secondary school. Students who were in Grade 10 and below in SY2015-2016 had to complete two more years of high school (Brillantes et al., 2019) relative to older cohorts. For responses from the 2010 Census, before the shift to the K-12 system, secondary school concluded in Grade 10. In the 2010 Census we therefore encode the *1st Year of Post Secondary* or the *1st Year of College* as 11 years of completed education, following this logic up to the *6th Year of College or Higher* as 16 years of completed education. *Post Secondary Graduates* are assigned 12 years of completed education, while *Academic Degree Holders* are encoded as 14.

For responses from the 2020 Census, those reporting Grade 11 and Grade 12 are encoded as having 11 and 12 years of completed education. We encode *Post-Secondary Undergraduates* and *Short-Cycle Tertiary Undergraduates* as 11, *Post-Secondary Non-tertiary Graduates* and *Short-cycle Tertiary Graduates* as 12, *Bachelor's Degree Graduates* as 14, and *Master's Degree Graduates* and over as 16. This encoding reflects the fact that among respondents in the 2020 Census, those reporting completion of advanced degrees would have been older cohorts who would have completed secondary school under the prior K-10 system (completing secondary school at Grade 10) before the switch to K-12 in SY2016-17.

A.1.3 Municipalities and Provinces

Municipalities are administrative units which fully nest within provinces. In analyses of DepEd administrative data at the student or school level in which we restrict the sample to schools with medium of instruction information, there are a total of 1,482 municipalities within 84 provinces. In analyses using Census data from 2010 and 2020, there are a total of 1,627 municipalities within 87 provinces.

A.1.4 School-Level Data

Surveys Including Data on Medium of Instruction under MTB-MLE. First, we rely on a DepEd-conducted survey of schools in 2022 profiling the medium of instruction adopted due to the MTB-MLE policy. In our analyses, we use data for 19,260 public primary schools with linguistic composition data pre policy out of the 20,430 surveyed schools. We also use a 2018-19 survey of schools conducted by Monje et al. (2019). For this survey, we exploit data for 13,716 public primary schools with linguistic composition data pre policy out of the 15,916 surveyed schools. Combining these two sources results in a sample of 24,529 schools with both medium of instruction information and linguistic composition data pre policy. Treatment assignment (whether a school reported using a MOI different from Tagalog) does not coincide for 1,196 of the 8,447 schools with MOI information from both surveys. As discussed in the main text, we assign such small subset of schools with inconsistent information to the treatment group.[19]

[19]Reporting Tagalog as the MOI in one survey and then a non-Tagalog language in another survey is likely to reflect that a school switched back and forth between languages over time, and so likely reflects that the school was "treated" for at least some periods.

Table A1: Distribution of students and schools across mother tongues and MOI

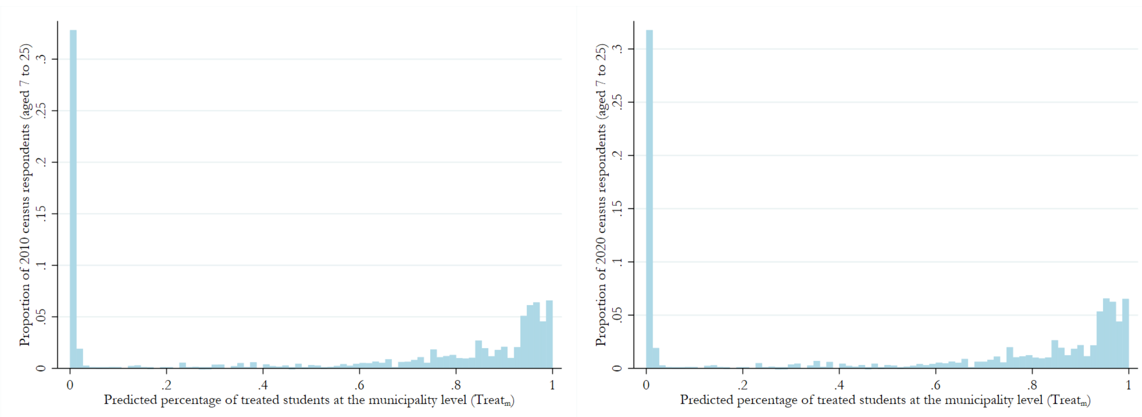
Language	Share of G1-G6 students by mother tongue	Share of G4-G6 students by mother tongue	Share of schools choosing MOI	Share of G1-G3 students facing MOI
Tagalog	33.98	35.36	33.53	45.22
Cebuano	25.27	24.52	27.40	24.79
Hiligaynon	7.37	7.28	7.52	7.16
Ilocano	6.69	7.03	9.20	5.00
Bikol	5.74	5.78	4.83	4.40
Waray	3.80	3.82	7.08	4.05
Kapampangan	2.58	2.77	1.81	2.41
Pangasinan	1.60	1.67	1.69	1.96
Maranao	1.54	1.19	0.74	0.53
Maguindanaoan	1.52	1.35	0.18	0.14
Tausug	1.20	1.05	0.17	0.16
Kinaray-a	0.93	1.02	1.50	0.76
Sinurigaonon	0.74	0.75	1.08	0.59
Akeanon	0.63	0.63	0.95	0.68
Chabacano	0.53	0.51	0.41	0.78
Ibanag	0.40	0.43	0.21	0.13
Yakan	0.17	0.13	0.13	0.09
Sambal	0.05	0.05	0.05	0.03
Ivatan	0.02	0.03	0.07	0.03
Other languages	5.24	4.63	1.48	1.07
Observations	11,094,240	4,811,287	24,529	4,407,559

Note: This table shows, for the 19 official languages offered as MOI by the MTB-MLE policy, the share of G1-G6 students from the universe of elementary public schools students in 2012-2013 reporting the language as their mother tongue, the share of G4-G6 students from the universe of elementary public schools students in 2012-2013 reporting the language as their mother tongue, the share of schools (with MOI information and pre-policy linguistic composition data) choosing the language as MOI as a result of the policy, and the estimated share of students facing that language as MOI post-policy (using the number of G1-G3 students in each school in 2012-2013 in a weighted average across schools).

A.1.5 Treatment Intensity at the Municipality Level

To construct our measure of treatment intensity at the municipality level, we first predict the probability of treatment for 34,807 public schools with DepEd administrative information on the linguistic composition of the school in 2012-2013 (using the universe of Grades 4 to 6 students with mother tongue information) and first-stage coefficient estimates from Table 2. Note that this generates both *in-sample* predictions for the 24,529 schools with MOI information used in our first-stage regression, and 10,278 *out-of-sample* predictions for public primary schools without MOI information. We then aggregate up the school-level predictions at the municipality level by weighting each school's predicted probability of treatment with the number of Grades 1 to 6 students in 2012-2013. This generates a municipality-level treatment intensity which varies between 0 and 1 for the 1,627 municipalities in our sample, and corresponds to the predicted percentage of treated students. Because this is a linear probability model, predictions are not bounded by 0 or 1 so we recode Treat_m to be equal to 1 for the 17 municipalities with values between 1 and 1.025, and to be equal to 0 for the 5 municipalities with values between -0.009 and 0. Figure A1 shows the resulting distribution of Treat_m across census respondents aged 7 to 25.

Figure A1: Distribution of census respondents across municipalities by treatment intensity



(a) 2010 Census **(b)** 2020 Census

Note: Histograms of the distribution of Census respondents aged 7 to 25 (2010 Census on left panel; 2020 Census on right panel) across values of treatment intensity. $Treat_m$ defined at municipality level for analyses of Census data. Number of bins set to 75. Mean value of $Treat_m$ across the 73,267,484 respondents in full sample (combining 2010 and 2020 respondents) is 54.1%, 25th percentile is 1.0%, median is 74.9%, 75th percentile is 94.1%, and standard deviation is 42.0%.

A.2 Additional Results

A.2.1 National Achievement Test Scores

Table A2: Impacts on Grade 3 test scores across subjects

	Overall		English		Tagalog		Math.	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\bar{\tau}_{\text{post}}$	-0.003	0.023	0.008	-0.001	-0.011	-0.039	0.002	0.119
	(0.034)	(0.086)	(0.025)	(0.065)	(0.025)	(0.062)	(0.036)	(0.095)
Honest CI	[-0.162, 0.019]	[-0.378, 0.101]	[-0.143, 0.025]	[-0.375, 0.098]	[-0.158, 0.005]	[-0.443, 0.000]	[-0.164, 0.021]	[-0.309, 0.177]
Control Mean, Pre Period	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
School FE	Y	Y	Y	Y	Y	Y	Y	Y
Province × Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Obs. (Students)	856,610	856,610	1,012,160	1,012,160	1,012,160	1,012,160	856,610	856,610
Clusters (Schools)	23,712	23,712	23,808	23,808	23,808	23,808	23,712	23,712

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1) using Grade 3 test scores as the dependent variables. Overall test scores in Grade 3 are the average of English, Tagalog and Mathematics test scores. Odd columns present estimates from the OLS specification while even columns show estimates from the instrumented DID specification where treatment status as the school level is instrumented with school-level linguistic composition variables pre policy, i.e., the percentage of learners at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction as well as a square and cubic term in the percentage of Tagalog-speaking learners (see Table 2). The pre-period is SY2008-2009 to SY2011-2012.

Table A3: Impacts on Grade 6 test scores across subjects

	Overall		English		Tagalog		Math.		Science		Hekasi	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$\bar{\tau}_{\text{post}}$	-0.203***	-0.668***	-0.180***	-0.568***	-0.156***	-0.645***	-0.170***	-0.535***	-0.163***	-0.503***	-0.169***	-0.563***
	0.038	0.093	0.034	0.083	0.032	0.085	0.036	0.088	0.035	0.087	0.035	0.089
Honest CI	[-0.316, -0.147]	[-0.957, -0.532]	[-0.287, -0.135]	[-0.843, -0.458]	[-0.261, -0.123]	[-0.869, -0.505]	[-0.269, -0.111]	[-0.813, -0.409]	[-0.274, -0.117]	[-0.785, -0.380]	[-0.251, -0.095]	[-0.777, -0.372]
Control Mean, Pre Period	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.01	0.01
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
School FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province × Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Obs. (Students)	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528	1,102,528
Clusters (Schools)	23,395	23,395	23,395	23,395	23,395	23,395	23,395	23,395	23,395	23,395	23,395	23,395

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1) using Grade 6 test scores as the dependent variables. Overall test scores in Grade 6 are the average of English, Tagalog, Mathematics, Science and Hekasi test scores. Odd columns present estimates from the OLS specification while even columns show estimates from the instrumented DID specification where treatment status as the school level is instrumented with school-level linguistic composition variables pre policy, i.e., the percentage of learners at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction as well as a square and cubic term in the percentage of Tagalog-speaking learners (see Table 2). The pre-period is SY2008-2009 to SY2014-2015.

A.2.2 Impacts on Enrollment Counts

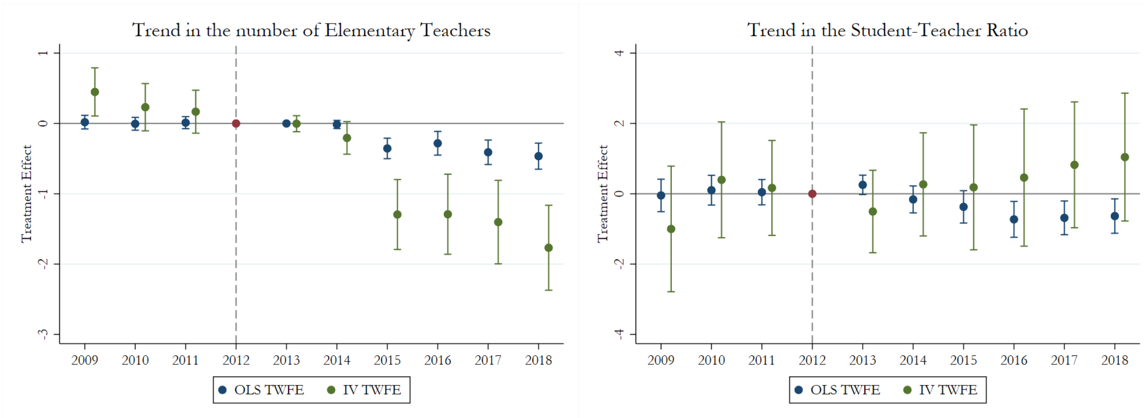
Table A4: OLS Estimates: Average Causal Effects on Grade 1 to Grade 6 Enrollment Counts

	Gr 1	Gr 2	Gr 3	Gr 4	Gr 5	Gr 6
	(1)	(2)	(3)	(4)	(5)	(6)
$\bar{\tau}_{\text{post}}$	3.373***	0.069	-0.583*	-2.119***	-1.501***	-1.646***
	(0.726)	(0.361)	(0.343)	(0.454)	(0.362)	(0.336)
Honest CI	[1.421, 6.401]	[-0.490, 2.434]	[-0.887, 1.758]	[-2.076, 0.084]	[-2.073, -0.357]	[-1.954, -0.557]
Control Mean, Pre Period	101.34	85.32	81.64	78.54	76.97	73.94
Year FE	Y	Y	Y	Y	Y	Y
School FE	Y	Y	Y	Y	Y	Y
Province × Year FE	Y	Y	Y	Y	Y	Y
Observations	241,572	239,459	241,572	241,572	238,654	238,654
Clusters (Schools)	24,527	24,527	24,527	24,527	24,234	24,234

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates on grade-level enrollment from Grade 1 to Grade 6 for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1) using the OLS specification. The pre-period is SY2008-2009 to SY2011-2012 for Gr 1 to 3. It is shifted by 1 year for Gr 4, two years for Gr 5, and three years for Gr 6 (allowing time for treated students to reach these grades).

A.2.3 Impacts on Teachers and the Student-Teacher Ratio

Figure A2: Dynamic Impacts on Teacher counts and the Student-Teacher ratio



(a) Teachers **(b)** Student-Teacher Ratio

Note: Coefficient estimates (with 95% confidence intervals) from the estimation of equation (1) using the specification with school, school year, and province × year fixed effects. The dependent variable is elementary teacher counts in Panel (a), and the Student-Teacher Ratio, for which values below the 1st and above the 99th percentiles were dropped, in Panel (b). The pre-period is SY2008-2009 to SY2011-2012 while the post period is SY2012-2013 to SY2017-2018. IV estimates correspond to the instrumented DID specification where treatment status as the school level is instrumented with school-level linguistic composition variables pre policy, i.e., the percentage of learners at the school level whose mother tongue corresponds to each of the 19 languages offered as media of instruction. Standard errors are clustered at the school level.

Table A5: OLS estimates: Average Causal Effects on Teacher counts and the Student-Teacher ratio

	Teachers	Student-Teacher Ratio
	(1)	(2)
$\bar{\tau}_{\text{post}}$	-0.255***	-0.388**
	(0.057)	(0.172)
Honest CI (*smoothness restrictions*, $\bar{M} = 0$)	[-0.396, -0.063]	[-1.164, 0.279]
Control Mean, Pre Period	13.26	35.58
Year FE	Y	Y
School FE	Y	Y
Province × Year FE	Y	Y
Observations	239,140	236,494
Clusters (Schools)	24,461	24,431

Note: Standard errors are clustered at the school level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates on elementary teacher count in column (1) and the Student-Teacher Ratio trimmed at the 1st and 99th percentiles in column (2) for the average causal effect across post-treatment periods (equation (2)) from the estimation of equation (1) using the OLS specification. The pre-period is SY2008-2009 to SY2011-2012. See Figure A2 for per period coefficient estimates.

A.2.4 Impacts on Grade Completion (Census)

Table A6: Triple Difference: Impacts on Grade Completion for each grade level

	Gr 1	Gr 2	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Gr 8	Gr 9	Gr 10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE
Treat × (Age = X-17) × (Census = 2020)	0.00102	-0.00216	-0.00694	-0.0108*	-0.0122*	-0.0172**	-0.0196**	-0.0167**	-0.0124	-0.0130
	(0.003)	(0.004)	(0.005)	(0.006)	(0.007)	(0.007)	(0.008)	(0.007)	(0.008)	(0.009)
Mean Dep. Var. (Age = X-17), Pre	0.96	0.93	0.90	0.87	0.85	0.80	0.73	0.67	0.59	0.50
Youngest cohort, X=	7	8	9	10	11	12	13	14	15	16
Census × Age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Municipality × Census FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Municipality × Age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Province × Age × Census FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	73,267,484	69,071,039	64,955,228	60,652,106	56,395,089	52,321,490	48,133,621	44,119,307	40,114,400	36,123,292
Clusters (Municipalities)	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627

Note: Standard errors are clustered at the municipality level. *** p<0.01, ** p<0.05, * p<0.01.
This table shows the coefficient estimates on grade completion (across grade levels) for the triple interaction term from the estimation of equation (3). Treatment intensity at the municipality level is defined as the percentage of treated student and varies continuously from 0 to 1. The sample includes respondents aged 7 to 25 from the 2010 and 2020 censuses. Treated cohorts are those aged 7 to 17 in the 2020 census round (with varying levels of treatment intensity based on their municipality of birth; see Figure 4).