

Abrigo, Michael Ralph M.; Estopace, Katha Ma-i M.; Lingatong, Edmar E.; Relos, Charlotte Marjorie L.

Working Paper

Linguistic mismatch and learning productivity: Evidence from mother tongue-based education in the Philippines

PIDS Discussion Paper Series, No. 2025-41

Provided in Cooperation with:

Philippine Institute for Development Studies (PIDS), Philippines

Suggested Citation: Abrigo, Michael Ralph M.; Estopace, Katha Ma-i M.; Lingatong, Edmar E.; Relos, Charlotte Marjorie L. (2025) : Linguistic mismatch and learning productivity: Evidence from mother tongue-based education in the Philippines, PIDS Discussion Paper Series, No. 2025-41, Philippine Institute for Development Studies (PIDS), Quezon City, <https://doi.org/10.62986/dp2025.41>

This Version is available at:

<https://hdl.handle.net/10419/339166>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Photo from the Philippine News Agency (<https://www.pna.gov.ph/photos/50646>)



Linguistic Mismatch and Learning Productivity: Evidence from Mother Tongue-based Education in the Philippines

*Michael R.M. Abrigo, Katha Ma-i M. Estopace,
Edmar E. Lingatong, and Charlotte Marjorie L. Relos*

The PIDS Discussion Paper Series constitutes studies that are preliminary and subject to further revisions. They are being circulated in a limited number of copies only for purposes of soliciting comments and suggestions for further refinements. The studies under the Series are unedited and unreviewed. The views and opinions expressed are those of the author(s) and do not necessarily reflect those of the Institute. The Institute allows citation and quotation of the paper as long as proper attribution is made.

This study is carried out with support from the Second Congressional Commission on Education (EDCOM II).


CONTACT US:

RESEARCH INFORMATION DEPARTMENT
Philippine Institute for Development Studies

18th Floor, Three Cyberpod Centris – North Tower
EDSA corner Quezon Avenue, Quezon City, Philippines

 <https://www.pids.gov.ph>

 publications@pids.gov.ph

 (+632) 8877-4000

Linguistic Mismatch and Learning Productivity: Evidence from
Mother Tongue-based Education in the Philippines

Michael R.M. Abrigo
Katha Ma-i M. Estopace
Edmar E. Lingatong
Charlotte Marjorie L. Relos

PHILIPPINE INSTITUTE FOR DEVELOPMENT STUDIES

December 2025

Abstract

We study how linguistic mismatch between students' home languages and the language of instruction shapes learning productivity in linguistically diverse societies. We develop a framework in which linguistic distance acts as a multiplicative tax on educational production and derive welfare-theoretic conditions for optimal language choice. We apply the framework to Philippine elementary schools during the 2009-2024 mother tongue-based multilingual education (MTB-MLE) reform, exploiting policy-induced variation in linguistic alignment using difference-in-differences and shift-share instrumental variables designs. We extend the Foster-Greer-Thorbecke family of deprivation measures, traditionally used to quantify income poverty, to network settings in order to capture spatial variations in degree of linguistic mismatch and estimate substantial school-based linguistic deprivation under a counterfactual pre-reform bilingual language regime, concentrated in non-Tagalog-predominant regions. We find that each unit of mismatch reduces achievement by 35-57% of potential, depending on curricular subject, yielding productivity losses comparable in magnitude to canonical estimates of spatial misallocation in firm production. Under a counterfactual pre-reform bilingual regime, an average student would have operated at 27-45% below learning potential, with losses concentrated in linguistically diverse regions. MTB-MLE reduced these losses by roughly two-thirds nationally, though systematic deviations from optimal language choice leave substantial efficiency gains unrealized. The reform also increased student continuation rates by 9-12 percentage points but temporarily strained classroom capacity, reflecting general-equilibrium resource pressures when demand-side constraints are relaxed. Overall, the results show that language policy is a first-order determinant of educational efficiency in multilingual settings and that measuring the quality of linguistic matching, rather than policy adoption alone, is central to evaluating language-in-education reforms.

Keywords: Language policy, multilingual education, linguistic distance, allocative efficiency, Philippines, policy evaluation

Table of Contents

1. Introduction.....	4
2. Philippine basic education language of instruction policy	8
2.1. 20 th century language-in-education policy evolution.....	8
2.2. Early 21 st century policy developments.....	9
3. Conceptual framework: Linguistic mismatch as tax on learning.....	11
3.1. Environment.....	12
3.2. Students' production of learning outcomes.....	12
3.2. Schools' choice of instructional language.....	15
4. Evaluation design.....	16
4.1. Data and sample construction	17
4.2. Measuring linguistic match.....	19
4.3. Difference-in-differences strategy	19
4.4. Instrumental variable Difference-in-differences	21
5. Results.....	23
5.1. School-based linguistic deprivation.....	23
5.2. MTB-MLE and learning outcomes.....	30
5.3. MTB-MLE and enrollment.....	45
5.4. MTB-MLE and school resources.....	55
5.5. Learning tax from linguistic mismatch.....	58
6. Policy implications.....	63
Appendix A. Linguistic distance and aggregate learning	72
Appendix B. Axiomatic foundations of network-adapted FGT index.....	76
Appendix C. Robustness analysis: IV-DID separate v. joint estimation.....	78
Appendix D. Shift-share IV relevance: NAT/ELLNA IV-DID.....	81
Appendix E. Shift-share IV relevance: Enrollment IV-DDD	83

List of Figures

Figure 1. Share of children aged 5 to 9 years by primary language used at home and residence location.....	25
Figure 2. School-level linguistic deprivation gap: Bilingual vs. MTB-MLE regimes	28
Figure 3. NAT G3/ELLNA by LOI intervention and year.....	31
Figure 4. MTB-MLE and G3 learning outcomes: DID and IV-DID estimates	34
Figure 5. MTB-MLE and G3 learning outcomes: Alternative LOI data, IV-DID model	38
Figure 6. Heterogeneity analysis: MTB-MLE and G3 learning outcomes by instructional language	42
Figure 7. MTB-MLE and municipal/city poverty rates: DID and IV-DID estimates.....	46
Figure 8. MTB-MLE and G1 enrollment.....	48
Figure 9. MTB-MLE and enrollment by grade level.....	50
Figure 10. Implied cohort survival function	54
Figure 11. MTB-MLE and school resources	56

List of Tables

Table 1. School-based linguistic deprivation: Philippines, 2017.....	26
Table 2. Baseline characteristics and covariate balance	29
Table 3. Linguistic mismatch learning tax.....	62

LINGUISTIC MISMATCH AND LEARNING PRODUCTIVITY: EVIDENCE FROM MOTHER TONGUE-BASED EDUCATION IN THE PHILIPPINES

Michael R.M. Abrigo, Katha Ma-i M. Estopace, Edmar E. Lingatong,
and Charlotte Marjorie L. Relos¹
(December 2025)

1. Introduction

A growing literature shows that low levels of foundational learning remain a major constraint on human capital accumulation in many low- and middle-income countries (World Bank, 2018; Pritchett and Beatty, 2015). Despite rapid increases in school participation over the past two decades, large shares of children complete the early grades without basic reading proficiency, and these early deficits compound into substantial learning shortfalls in later grades (Glewwe and Muralidharan, 2016). Research in education, linguistics, and cognitive psychology has long established that reading comprehension depends jointly on decoding, fluency, and the linguistic familiarity of instructional materials (Snow, 2002; August and Shanahan, 2006). Early theories argued that language mismatch between home and school directly impairs learning. Subsequent work showed a more complex relationship mediated by the level of first-language development students attain (Cummins, 1979; Benson, 2004). This mechanism ties language-of-instruction choices to the efficiency with which human capital is produced, especially in the grades when reading skills first form.

Recent assessments have documented a substantial learning crisis in Philippine basic education. Filipino students performed well below international benchmarks in recent international large-scale assessments, with average scores falling below minimum proficiency levels in reading, mathematics, and science (Orbeta, et al., 2025). When expected years of schooling are adjusted for learning quality, the Philippines faces an estimated learning gap of 5.5 years, suggesting that students spend considerable time in school but with limited learning productivity (Orbeta and Paqueo, 2025). Understanding how language-of-instruction policy contributes to this learning crisis provides critical insights for improving educational outcomes.

The use of languages unfamiliar to learners is widespread. Many countries maintain instructional languages chosen for political, historical, or administrative reasons rather than for pedagogical effectiveness, and large numbers of students begin schooling in a language they do not fully understand (UNESCO, 2016). The resulting misalignment should be expected to depress achievement, reduce progression, and generate inefficiencies that shape the long-run distribution of skills. These inefficiencies are conceptually distinct from the distortions emphasized in the macroeconomic literature on misallocation across firms (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008), but they share the basic feature that the productivity of an input can be diminished when it is deployed in a way that is poorly matched to the environment in which it operates. In the context of schooling, the relevant mismatch arises when the instructional language does not correspond to the linguistic capacities of students, and the consequences of that mismatch are magnified by the dynamic complementarities that govern early skill formation (Cunha and Heckman, 2007).

¹ Fellow II, Research Analyst II (former), and Project Technical Specialists, respectively, at the Philippine Institute for Development Studies (PIDS). The study has benefited from thoughtful discussions by Aniceto C. Orbeta, Jr., Brian Salvador C. Baran, and seminar participants at the PIDS Research Workshop Series and at the University of the Philippines School of Economics Friday Seminar Series. All remaining errors are by the authors.

The Philippine education system provides a particularly informative setting in which to examine these issues. With over 180 documented languages, the Philippines represents one of the most linguistically diverse countries globally and has conducted an extended experiment with language-in-education policy over the past century. The country cycled through five distinct regimes: English-only instruction during the American colonial period, limited vernacular education from 1957 to 1974, bilingual Filipino-English education from 1974 to 2008, mandatory mother tongue-based multilingual education from 2009 to 2023 (Monje, et al., 2021), and a recent shift toward conditional implementation beginning in 2024 driven by persistent implementation difficulties, weak stakeholder coalitions, and rising preference for early English instruction (Metila, et al., 2025). This history of policy experimentation is not unique among postcolonial states, but the 2009 reform stands out for its scale and ambition. The Mother Tongue-Based Multilingual Education (MTB-MLE) reform mandated that instruction in kindergarten through Grade 3 be delivered in one of 19 authorized local languages, making the Philippines the only Southeast Asian country to require mother tongue instruction nationwide in primary grades (Burton, 2013).

The reform created substantial variation in how schools implemented the policy. Schools chose languages from an authorized menu of instructional languages, but these choices did not always align with student linguistic composition. An earlier process evaluation of the MTB-MLE program found that students and teachers often lacked familiarity with their chosen medium of instruction, with less than 10% of schools completing the activities needed for proper implementation (Monje, et al., 2021). Some schools adopted languages spoken by most students, achieving good linguistic match. Others faced mismatches due to dialectal variation, teacher capability constraints, or inadequate preparation, leaving many students struggling to learn in unfamiliar languages. This variation in implementation quality, combined with the cohort-based rollout that exposed some students to mother tongue instruction while leaving earlier cohorts under the bilingual regime, creates conditions well-suited to identifying the causal effects of linguistic matching on learning outcomes.

Recent evaluations of the Philippine language-in-education reform have reported mixed or negative findings. Lloyd and Yang (2025) document negative effects on learning outcomes using a difference-in-differences approach that compares schools switching from Filipino or English to other Philippine languages against schools maintaining bilingual instruction. Igarashi et al (2024) find similarly discouraging results using a related identification strategy. These studies treat the institutional act of changing the language of instruction as the policy treatment. However, switching languages does not guarantee improved linguistic alignment. A school serving predominantly Cebuano-speaking students that switches from Filipino to Ilocano implements mother tongue policy but worsens linguistic match. Conversely, a school serving Tagalog speakers achieves good match whether it formally adopts mother tongue education or maintains Filipino under the bilingual regime. The critical distinction is between the policy mechanism of language switching and the underlying linguistic matching that directly governs comprehension and learning.

We argue that measuring matching quality directly provides a more informative evaluation of mother tongue-based education than estimating switching effects. This requires constructing measures of linguistic distance between students' home languages and their schools' instructional languages, accounting for partial mutual intelligibility across related languages, and tracking how match quality evolved with reform implementation. We develop a conceptual framework in which linguistic mismatch operates as a tax on learning, reducing the

productivity of both student effort and instructional resources. The framework formalizes mismatch as a multiplicative wedge in the education production function rather than an additive input, capturing the idea that poor linguistic alignment attenuates the returns to all other educational investments. This generates predictions not only about achievement effects but also about enrollment dynamics and resource allocation. Improved matching should raise continuation rates as learning barriers fall, but higher continuation expands enrollment and dilutes per-student resources when school capacity remains fixed.

We test these predictions using administrative panel data covering Philippine public elementary schools from 2008 to 2024. The data link student outcomes from national assessments to detailed information on school language choices, student and teacher language distributions, enrollment by grade, and resource availability. We construct linguistic match measures using phylogenetic distances between languages derived from the Ethnologue language classification system, combined with student language data from enrollment records and teacher language approximations from municipal population census distributions. This allows us to measure the degree of alignment between students, teachers, and the instructional language in each school and to track how this alignment changed with policy implementation.

The identification strategy exploits two sources of variation. First, the cohort-based rollout means that students entering Grade 1 in 2010 or later received school instruction in their mother tongues while earlier cohorts remained under bilingual education. Within schools, we compare outcomes for cohorts exposed to different regimes. Second, schools experienced differential improvements in linguistic matching depending on their baseline language ecology. Linguistically homogeneous schools where one local language dominated saw large match improvements when adopting that language, while diverse schools with fragmented language distributions achieved smaller gains. We use difference-in-differences specifications to estimate average treatment effects and instrumental variables to address endogeneity in school language choices. The instruments exploit shift-share variation combining predetermined language shares with contemporaneous neighborhood language distributions, isolating match variation driven by demographic composition rather than school decisions.

Three main findings emerge. First, improvements in linguistic matching generated substantial learning achievement gains across subjects. Each unit reduction in mismatch raises test scores by 25 to 35 percentage points, equivalent to 1.5 to 2.0 standard deviations of the outcome distribution. These effects translate to a mismatch tax of 35% in Filipino, 47% in English, and 57% in Mathematics, suggesting that complete linguistic mismatch in schools reduces learning productivity by roughly one-third to one-half of potential learning achievement. Extending this analysis to a bilingual education regime, an average Filipino student would have operated at 27 to 45 percent below their learning frontier. Mother tongue implementation reduced these potential losses by approximately two-thirds nationally, though more than one-quarter of students still face substantial mismatch because schools chose suboptimal languages or because local languages lack official authorization.

Second, improved linguistic matching raised continuation rates significantly. Schools experiencing larger match improvements saw enrollment expand by 9 to 12 percentage points from Grade 1 to Grade 6, indicating that linguistic barriers affect not only how much students learn but also whether they remain in school. This enrollment response generated resource pressure. Pupil-class ratios rose temporarily during 2017-2019 before infrastructure investment caught up, though the government partially offset this through aggressive teacher hiring during 2015-2018 that improved pupil-teacher ratios in better-matched schools. Similar dynamics

where successful demand-side interventions raise participation and require coordinated supply responses have been documented in conditional cash transfer programs (Schultz, 2004; Fiszbein and Schady, 2009), school feeding initiatives (Alderman, et al., 2012), and other education interventions where capacity constraints can dampen program impacts (Duflo, et al., 2015; Muralidharan and Sundararaman, 2015). The resource dilution mechanism operated as theorized but did not undermine the MTB-MLE policy because supply partially adjusted through the more flexible teacher hiring margin.

Third, many schools adopted instructional languages that did not minimize linguistic mismatch given their student populations. Current implementation achieves 28 percent national mismatch when 14 percent would be attainable by consistently selecting the modal student language from the 19 authorized options. These deviations were more common in linguistically heterogeneous communities and where teacher language backgrounds diverged from student compositions. The gap between actual and optimal implementation suggests that providing clearer guidance on language selection, improving information about student language distributions, and addressing teacher deployment constraints could substantially improve outcomes even without expanding the menu of authorized languages.

This study contributes to several literatures. Conceptually, we introduce linguistic mismatch as a form of allocative inefficiency in human capital production, paralleling factor misallocation frameworks that analyze how productivity losses arise when inputs are poorly matched to their deployment environment (e.g. Hsieh and Klenow, 2009; Restuccia and Rogerson, 2008). The application of Foster-Greer-Thorbecke deprivation measures (Foster, et al., 1984) to language policy provides a principled method for aggregating heterogeneous linguistic populations into a scalar mismatch index. This approach respects both the prevalence and severity of linguistic barriers while allowing schools to trade off administrative efficiency against match quality when mutual intelligibility permits use of regional lingua franca.

The production function we develop treats linguistic mismatch as a multiplicative efficiency parameter rather than an additive input. This specification formalizes the idea that comprehension barriers do not merely reduce learning in levels but instead attenuate the marginal returns to all educational investments. We derive analytical results characterizing optimal language choice and prove that when implementation capacity permits, offering instruction in multiple languages Pareto-dominates any single-language policy under realistic assumptions. These theoretical results provide quantitative guidance for policy design that existing frameworks lack.

Empirically, we provide the first large-scale evidence on how matching quality rather than policy adoption affects learning. This resolves the apparent contradiction between recent negative evaluations and positive findings from small-scale pilots. Schools that switched to languages poorly matched to their students experienced negative outcomes, while schools that improved actual linguistic alignment raised achievement substantially. The magnitude of productivity losses from poor matching, ranging from 35 to 57% of potential learning achievement depending on subject, proves comparable to estimates of losses from sectoral misallocation in firm production. This suggests that language policy operates as a first-order determinant of aggregate educational efficiency in multilingual societies.

Methodologically, the shift-share instrumental variables approach we employed demonstrates how demographic variation can provide identifying variation when policy choices are endogenous. This template applies broadly to evaluating place-based policies where treatment

intensity varies with predetermined local characteristics. The combination of reduced-form difference-in-differences and structural estimation of the mismatch tax parameter shows how theory can discipline empirical work while generating policy-relevant parameters.

For policy, we move beyond binary evaluations of whether mother tongue education “works” to provide quantitative implementation guidance. Our theory on optimal language choice specifies which language schools should select given their student distributions. The estimated tax parameters enable welfare comparisons across alternative regimes. The analysis of resource dilution effects illustrates why successful demand-side interventions require coordinated supply expansion. These contributions translate directly into actionable recommendations for policymakers navigating the complex tradeoffs inherent in multilingual education systems.

The findings also speak to debates about allocative efficiency in public service delivery. The systematic departures from optimal language choice documented here illustrate how decentralized implementation can generate inefficiencies even when broad policy goals align with evidence (Metila, et al., 2016; Monje, et al., 2021). Teacher preferences, incomplete information, political economy considerations, and capacity constraints all shape how schools select languages, and these forces do not always point toward choices that maximize student comprehension. Understanding these implementation dynamics is critical for translating research findings into effective policy.

The rest of the report is organized as follows. In Section 2, we provide a short summary of how Philippine language-in-education policy evolved over the past century, situating the 2009 mother tongue reform in historical and comparative context. In Section 3, we propose a conceptual framework that formalize linguistic mismatch as a tax on learning and derive welfare-theoretic predictions for learning achievement, student enrollment, and school resources. In the next section, we describe our data construction, measurement of linguistic distance and matching, and the empirical strategy. This is followed in Section 5 by a presentation and discussion of our results on learning outcomes, enrollment dynamics, resource effects, and structural parameter estimation. Finally, in Section 6, we summarize our findings and discuss policy implications.

2. Philippine basic education language of instruction policy

Language of instruction is a foundational education policy lever in multilingual countries, with significant implications for learning outcomes, cultural transmission, and equity. With over 180 documented languages, the Philippines represents one of the most extensive and longest-running national experiments in language-in-education reform globally. Over the past century, the country has navigated five major policy regimes: monolingual instruction (1901-1957), limited multilingual instruction (1957-1974), bilingual education (1974-2008), mandatory mother tongue-based multilingual education (2009-2023), and a recent shift toward conditional MTB-MLE implementation (2024-present). While these phases broadly reflect postcolonial patterns in education policy, they also exhibit distinctive features that make the Philippine experience particularly amenable to rigorous empirical analysis.

2.1. 20th century language-in-education policy evolution

The American colonial government that established the Philippine public school system introduced English as the sole medium of instruction through Act No. 74 in 1901, a policy that

persisted long after the US' recognition of Philippine independence in 1946. Although the 1935 Philippine Constitution called for the development of a national Filipino language based on native tongues, classroom instruction remained overwhelmingly conducted in English, with Filipino taught only as a subject. Executive issuances in the years following the 1935 Constitution further guided this process, ultimately shaping the modern Filipino language, which is largely based on Tagalog.

Several mother tongue-based education initiatives were introduced in the decades that followed, building on the 1935 Quezon Educational Survey that recommended controlled experiments comparing local-language and English instruction. An early pilot was the 1948-1954 Iloilo experiment, in which a subset of Grade 1 and 2 pupils used Hiligaynon, while others continued using English. The favorable outcomes in this experiment had encouraged similar pilots in other areas, including in Laguna and Misamis Oriental (Cena, 1958; Nolasco, 2008).

These initiatives paved the way for a limited shift towards multilingual education from 1957 to 1974 under the government's Vernacular Education Policy, the country's first formal multilingual instruction framework. The policy designated eight major Philippine languages as the medium of instruction for Grades 1 and 2, transitioning to English in later grades while retaining local vernaculars as auxiliary media. Filipino was introduced informally in Grade 1 and as a formal curricular subject in subsequent grades. However, implementation faced significant challenges, including the absence of systematic teacher training and instructional materials (Gonzalez, 2003).

The adoption of the 1974 Bilingual Education Policy through then Department of Education, Culture and Sports Order 25 marked an important nationwide, structured reform. Filipino became the language of instruction for social studies and related subjects, while English remained the medium for instruction for mathematics, science, and technology. Later revisions maintained the bifurcated structure and reauthorized regional languages only as auxiliary media in the early grades. In 2003, Executive Order 210 reinforced English's position in the Philippine education landscape, citing concerns about students' competitiveness in emerging and fast-growing industries.

This trajectory is not unique to the Philippines but aligns with broader postcolonial patterns. Many newly independent countries continued using colonial languages as primary media of instruction (Bamgbose, 1991; Migge and Leglise, 2007). Where national language promotion occurred, such as in Malaysia, Indonesia, and Tanzania, reforms typically elevated a single national language as the main medium (Sercombe and Tupas, 2014). The Philippine path diverged through early vernacular experimentation but maintained a hybrid model that preserved English's strong status while incrementally elevating Filipino, reflecting a combination of nation-building goals and economic incentives linked to English proficiency.

2.2. Early 21st century policy developments

Throughout the second half of the 20th century and into the early 21st century, experimentation with MTB-MLE continued through a series of small pilot initiatives. These included the second Iloilo experiment (1961-1964), the Rizal experiment (1960-1966), the First Language Component Bridging Program (1986-1993), and the Lingua Franca Education Pilot Study (1999-2002) (Gaylo, 2020), as well as various university-based studies (del Pilar, 2023). The most prominent among these was the Lubuagan experiment (1999-2003), which provided well-documented local evidence on the positive effects of MTB-MLE on student learning outcomes

(Dumatog and Dekker, 2003; Dekker and Young, 2005; Walter and Dekker, 2011). Nevertheless, all of these efforts remained limited in scale and operated as small pilot studies.

The most dramatic policy shift occurred with the 2009 Department of Education (DepEd) Order 74 which institutionalized MTB-MLE and launched a nationwide rollout supported by a three-year bridging plan beginning in school year 2009-2010. Implementation followed a cohort-based sequence wherein the policy was introduced first in kindergarten and expanded by one grade level per year, achieving full K-3 coverage by 2012-2013. Bilingual instruction remains for the upper grade levels. The 2013 K-to-12 Law further reinforced mandatory mother tongue instruction from kindergarten to Grade 3 as part of a broader reform of the basic education system. The initial full K-3 rollout in 2012-2013 covered only 12 local languages, but this was later expanded to 19 languages, including eight regional lingua franca and 11 local languages.

This period was exceptional not simply for adopting MTB-MLE but for doing so at a national scale. While the policy drew on global advocacy from UNESCO (2003) and Education for All frameworks, most countries did not attempt large-scale MTB-MLE implementation. Cambodia's Multilingual Education National Action Project covered only five languages in targeted provinces, Thailand and Vietnam maintained pilot programs, and Malaysia and Singapore preserved separate language-medium tracks rather than MTB-MLE models (Kosonen & Young, 2009; Wright et al., 2022). Beyond Southeast Asia, colonial-language dominance remained prevalent even where mother tongue instruction was officially endorsed (Brock-Utne, 2005; Trudell, 2016).

Burton (2013) notes that the Philippines became the only Southeast Asian country to require MTB-MLE nationwide in the primary years. The scale, speed, and complexity of this implementation across thousands of schools and diverse linguistic ecologies created conditions not previously observed in the comparative literature. As a result, the Philippines accumulated more than a decade of national implementation experience, offering rare data on how MTB-MLE performs outside controlled pilot environments.

The MTB-MLE national rollout faced several challenges and implementation deficits similar to those observed in earlier pilot interventions (e.g., Cena, 1958; Gonzalez, 2003). Burton (2013) documented early implementation issues that persisted in subsequent analyses (e.g. Metila, et al., 2016; Monje, et al., 2021), including inadequate teacher preparation, shortages of instructional materials, and the structural difficulties of implementing MTB-MLE in highly diverse linguistic environments. In this context, it is not surprising that recent evaluations (e.g., Lloyd and Yang, 2025; Igarashi, et al., 2024) report negative impacts of the policy on learning outcomes and enrollment. It is important to emphasize, however, that these studies assess the effects of switching from Filipino to another Philippine language, rather than the quality of linguistic matching within schools.

In 2024, legislated reform shifted the Philippines from a universal MTB-MLE mandate to a conditional policy determined by local language ecology and school capacity. Republic Act (RA) 12027 replaced the earlier compulsory model with default Filipino and English instruction, while allowing mother tongue instruction as an auxiliary medium only in contexts where monolingual conditions prevail and resource availability make implementation feasible.

This policy shift has few international parallels. No other Southeast Asian country has implemented a nationwide mandatory MTB-MLE policy and subsequently recalibrated it in response to practical implementation challenges. The closest comparison is Nigeria, whose

2022 mandate requiring mother tongue instruction through Primary 6 was reversed in 2025, reinstating English-only schooling at all levels. Unlike the Nigerian case, however, the Philippine reform retains MTB-MLE as an optional, context-dependent strategy rather than eliminating it entirely.

As a result, the Philippines represents a rare combination of nationwide mandatory MTB-MLE, extensive multilingual implementation, and systematic policy adjustment informed by decades of experimentation and field experience. The resulting variation across time, languages, and local contexts creates an unusually rich empirical setting for evaluating the effects of MTB-MLE on learning outcomes.

3. Conceptual framework: Linguistic mismatch as tax on learning

The Philippine experience with MTB-MLE offers rare empirical leverage to evaluate language-of-instruction policy at scale, but doing so requires a conceptual framework that addresses three fundamental questions. First, what is the welfare-theoretic basis for preferring one language over another as the medium of instruction? Second, how should we measure the costs of linguistic mismatch in a way that is both empirically tractable and normatively meaningful? Third, what determines the optimal language choice for a school serving a linguistically diverse student population? We develop a simple economic framework that formalizes language mismatch as a form of allocative inefficiency analogous to factor misallocation in production that reduces the productivity of educational investments. This framework yields testable predictions about the effects of MTB-MLE on learning outcomes and provides guidance for identifying causal effects in our empirical analysis.

We model linguistic mismatch in education as an economic distortion that reduces the productivity of student effort and school resources. Our framework serves three purposes. First, it formalizes the intuition that language barriers create learning frictions, providing a welfare-theoretic foundation for MTB-MLE policy. Second, it delivers an operational definition of “optimal” language choice based on minimizing aggregate mismatch, which we show is formally equivalent to minimizing a Foster-Greer-Thorbecke (FGT) poverty index applied to linguistic distances. Third, it generates an estimating equation that directly connects our theoretical mismatch tax parameter to empirically observable test score effects, enabling structural interpretation of reduced-form results.

Our approach differs from prior education production function studies in treating linguistic matching as a multiplicative efficiency parameter rather than an additive input. This distinction matters as mismatch doesn’t merely reduce learning in levels but instead attenuates the marginal returns to all other educational investments, creating complementarities between language policy and resource allocation.

Our model features two key mechanisms. First, linguistically mismatched students are directly taxed as they acquire less human capital from the same effort and resources. Second, improvements in linguistic matching raise learning and thereby increase continuation to higher education levels; when aggregate resources are fixed, this expansion dilutes per-student resources and partially offsets initial learning gains. Schools choose the language of instruction while anticipating these forces, selecting the option that maximizes overall learning outcomes given the potential feedback between linguistic matching and student achievement.

3.1. Environment

We consider an overlapping generations model in discrete time t in an economy populated by students and a single school. Each student lives for two periods. In the first period, the young students attend Level 1 education. Those who qualify in Level 1 attend Level 2 education in the second period. The lone school in the economy exclusively accepts all young cohorts and qualified old cohorts within its catchment area.

Each student i attending Level 1 in t has a mother tongue $\ell_{it} \in \{A, B, C, \dots\}$. At the start of any period, the school chooses a language of instruction for Level 1, denoted as $L_t \in \{A, B, C, \dots\}$. We define linguistic mismatch as the scaled distance between any student's mother tongue and the school language of instruction as

$$0 \leq m_{it} = I\{\ell_{it} \neq L_t\} \cdot \left[\frac{|\ell_{it} - L_t|}{\max_j \{|\ell_{jt} - L_t|\}} \right]^\rho \leq 1,$$

where $\rho > 0$ is a persistence parameter governing how mutual intelligibility decays with phylogenetic distance across languages. Higher values of ρ indicate slower decay, and thus greater intelligibility even across larger linguistic distances.

The stock of available school resources R_t evolves exogenously through

$$R_t = (1 - \delta)R_{t-1} + \Delta R_t,$$

where the first term represents the continuing, undepreciated value of the school's existing resource stock, while the second term captures exogenous resource infusion (e.g. from government budget). Let N_t^1 and N_t^2 denote total enrollment at Levels 1 and 2, respectively. Resources are shared across both levels, so resources per student are $r_t = R_t / (N_t^1 + N_t^2)$. Without adequate infusion, school resources per student decline when enrollment rises.

3.2. Students' production of learning outcomes

Student i at Level 1 produces learning outcome y_{it}^1 by combining schooling resources r_{it} and own effort e_{it} through a Cobb-Douglas production technology. However, linguistic mismatch reduces the productivity of both effort and school resources. We model this as a multiplicative distortion that scales down the effective learning technology proportionally to mismatch

$$y_{it}^1 = (1 - \tau m_{it}) [\phi_1 r_{it}^\alpha e_{it}^{1-\alpha} \exp(\varepsilon_{it})],$$

where $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ is an idiosyncratic ability and resource shock that captures random differences across students. The parameter $\phi_1 > 0$ is the baseline Level 1 learning outcome, while $\alpha \in [0, 1]$ and $(1 - \alpha)$ are respective output elasticities. If $\ell_{it} = L_t$, thus $m_{it} = 0$, students produce learning outcomes with full productivity. However, with linguistic mismatch, i.e., $m_{it} > 0$, production of learning is scaled down by a factor $(1 - \tau m_{it})$. The learning tax $\tau \in [0, 1]$ captures the cognitive burden, comprehension loss, and communication barriers from learning in a non-native language.

Students choose an effort level to maximize learning outcomes net of effort costs

$$\max_{e_{it} \geq 0} \{y_{it}^1 - c(e_{it})\}.$$

We assume a quadratic cost function $c(e_{it}) = (\kappa/2) \cdot e_{it}^2$, with $\kappa > 0$, capturing increasing cost of student effort. Optimality requires the first-order condition to be

$$\kappa e_{it} = (1 - \tau m_{it}) \left[\phi_1 (1 - \alpha) \left(\frac{r_{it}}{e_{it}} \right)^\alpha \exp(\varepsilon_{it}) \right],$$

from which optimal student effort may be derived as

$$e_{it}^* = \left[\phi_1 (1 - \alpha) \frac{(1 - \tau m_{it})}{\kappa} r_{it}^\alpha \exp(\varepsilon_{it}) \right]^{\frac{1}{1+\alpha}}.$$

The above equation suggests that poorly matched students exert less effort than better matched students. For any $m_{it} < m_{it}'$, it is easy to show that $e_{it}^*(m_{it}) > e_{it}^*(m_{it}')$. Because the marginal benefit to student effort is taxed, mismatched students optimally reduce own effort in learning.

Substituting optimal effort into the learning production function yields

$$y_{it}^{1*} = (1 - \tau m_{it}) [r_{it}^\alpha e_{it}^{*1-\alpha} \exp(\varepsilon_{it})] = \frac{\kappa}{1 - \alpha} \cdot \left[\phi_1 (1 - \alpha) \frac{(1 - \tau m_{it})}{\kappa} r_{it}^\alpha \exp(\varepsilon_{it}) \right]^{\frac{2}{1+\alpha}}.$$

Differentiating the optimal learning achievement y_{it}^{1*} with respect to the linguistic mismatch parameter m_{it} shows that learning loss from linguistic mismatch includes two components: (i) a direct tax effect and (ii) an effort response

$$\frac{dy_{it}^{1*}}{dm_{it}} = - \left[\underbrace{\tau}_{\substack{\text{direct} \\ \text{tax}}} + \underbrace{\frac{\tau(1 - \alpha)}{(1 + \alpha)}}_{\substack{\text{indirect} \\ \text{effort tax}}} \right] \cdot y_{it}^{1*}$$

The first component captures the direct tax of linguistic mismatch on learning productivity. The second term, on the other hand, is a second-round behavioral response wherein students reduce effort when its returns are taxed, amplifying the initial loss.

Young students who meet a certain threshold, i.e. $y_{it}^1 \geq \theta$, progress to Level 2, where they continue to accumulate human capital. Learning outcomes at this level are similarly produced through a Cobb-Douglas technology, but this time combining foundational achievement from Level 1 and second-period school resources as follows

$$y_{it+1}^2 = \phi_2 r_{t+1}^\beta (\lambda y_{it}^1)^{1-\beta},$$

where $\phi_2 > 0$ is the baseline Level 2 learning outcome, while β and $(1 - \beta)$ are respective output elasticities. The parameter $\lambda > 0$ captures how much Level 2 education builds on Level

1 learning outcomes. Crucially, we assume no linguistic mismatch tax in the Level 2 production function. By the second level, we assume that students have already adapted to the school language, or that pedagogy becomes less language-intensive.

Linguistically disadvantaged students in Level 1 are less likely to continue to Level 2. Since $y_{it}^{1*}(m_{it}) > y_{it}^{1*}(m'_{it})$ for any $m_{it} < m'_{it}$, then $P[y_{it}^1 \geq \theta | m_{it}] > P[y_{it}^1 \geq \theta | m'_{it}]$. That is, poorly matched students are less likely to continue schooling because they are less likely to clear the continuation threshold θ . This creates negative selection into Level 2. The Level 1 tax prevents some students from ever reaching Level 2 education.

While there is no direct mismatch penalty in Level 2, continuing students with similar second period school resources but had experienced greater linguistic mismatch at Level 1 enter Level 2 with a lower stock of learning achievement. Because Level 2 achievement is increasing and complementary in prior learning, the spillover loss is larger when the complementarity parameter λ is higher, reflecting stronger reliance on foundational skills accumulated at Level 1. The Level 1 linguistic mismatch tax thus has a persistent effect that spills over to Level 2, where students not instructed in their mother tongues trail behind and the initial disadvantage compounds over their educational careers.

In any given period, total school enrollment is composed of young students N_t^1 , which we take as exogenous, and continuing students from the previous period N_t^2 , where

$$N_{it}^2 = \sum_{i=1}^{N_{t-1}^1} I(y_{it-1}^1 \geq \theta) = N_{t-1}^1 P_{t-1}$$

and P_{t-1} is the continuation rate from the previous cohort. Without loss of generality, suppose we assume that cohort sizes are fixed at $N_t^1 = \bar{N}$ for any t .

Substituting the above enrollment dynamics into the resource constraint makes it clear that higher continuation rates dilute resources per student whenever aggregate resources are fixed

$$r_t = \frac{R}{N_t^1 + N_t^2} = \frac{R}{\bar{N}(1 + P_{t-1})}$$

This creates a negative feedback loop. Improving linguistic matching increases continuation rates, which dilutes resources per student, thereby reducing the learning productivity for both current Level 1 and Level 2 students. This resource-dilution effect partially offsets the efficiency gains from reducing linguistic mismatch, potentially dampening the long-run impact of language-matching reforms unless accompanied by commensurate resource infusion.

The strength of this feedback depends on how responsive continuation rates are to match quality (which depends on τ) and how productive school resources are in the education production function (captured by α and β). If resources are very important (α and β are large), or if LOI policy change induces large enrollment surges (through change in P_t) dilution can significantly reduce future net gains from better linguistic matching.

3.2. Schools' choice of instructional language

The lone school chooses a single language of instruction² L that maximizes current period average Level 1 student achievement. Under mild correlation between learning shocks and linguistic distances, it can be shown that the school's objective is approximately the same as minimizing the ρ -power mean of linguistic distances.³ Letting $\pi_{\ell t}$ be the fraction of students with mother tongue ℓ , the school's maximization problem is thus given by

$$\max_L \left\{ \frac{1}{N_t^1} \sum_i y_{it}^1 \right\} \cong \min_L \sum_k \left\{ \pi_{kt} \left[\frac{|\ell_k - L|}{\max_j \{|\ell_j - L|\}} \right]^\rho \cdot I\{\ell_j \neq L\} \right\}.$$

The above reformulation is useful for several reasons. First, it recasts the optimization of learning outcomes, which are not observable until ex post, as a linguistic-distance minimization problem that can be directly measured at the beginning of each period. Second, the resulting minimization problem closely resembles the structure of the classic Foster-Greer-Thorbecke poverty indices (Foster, et al., 1984, 2010), allowing linguistic mismatch to be interpreted as a form of measurable deprivation with well-defined axiomatic properties (see Appendix B). Taken jointly, this means that the learning-maximizing instructional language is approximately⁴ the one that minimizes linguistic deprivation. Finally, the solution to the linguistic-distance minimization problem provides clear guidance on the optimal instructional language to select.

Solutions to the above minimization problem are well studied in the literature (for application in Euclidean space, see Love, et al., 1988). For $\rho \rightarrow 0$, the solution is the modal language, minimizing the count of mismatches (e.g. Moulin, 1988). For $\rho = 1$, the optimal language is the weighted median on the phylogenetic tree, minimizing the sum of distances (e.g. Hakimi, 1964; Kariv and Hakimi, 1979). For $\rho = 2$, it corresponds to the Fréchet mean, minimizing squared distances (e.g. Karcher, 1977; Bacák, 2014). As $\rho \rightarrow \infty$, the solution converges to the minimax center, minimizing the maximum mismatch (e.g. Hakimi, 1965; Handler, 1973). General values of ρ interpolate between these cases, with larger ρ placing greater relative weight on linguistically distant students.

The school's optimal language of instruction is the language that minimizes the ρ -power mean of linguistic distances, which depends jointly on ρ , the linguistic distance matrix, and the population shares of each mother tongue. A school that optimizes only contemporaneous outcomes will choose this central language, effectively taxing all students whose mother tongues lie away from the linguistic center.

A more sophisticated school might internalize resource dilution by incorporating the discounted average learning outcomes of future cohorts into its decision rule as follows

² We relax this assumption in the Appendix.

³ The relationship is exact when learning shocks and induced linguistic distances are uncorrelated. For a mean-preserving dependence structure, average learning is maximized when induced linguistic distance is least favorably aligned with learning shocks, i.e., when the mismatch tax could contribute the least learning loss. See Appendix for detailed sketch of proof.

⁴ See Appendix for further qualifications.

$$\max_L \left\{ \sum_{s=t}^{\infty} \left(\frac{1}{1+r} \right)^{s-t} \left[\frac{1}{N_s^1} \sum_i y_{is}^1 \right] \right\},$$

where $r > 0$ is a discount factor, and subject to the constraint that the current language-of-instruction choice affects future resources through continuation rates.

If the school internalizes future resource dilution, it may rationally choose a less central instructional language to restrain continuation rates and avoid excessive enrollment growth. Such a choice reduces the number of students progressing to Level 2, thereby easing future pressure on per-student resources. This comes at the cost of lower contemporaneous learning efficiency, since a less central instructional language increases linguistic mismatch and raises the effective learning tax for many students. Modeling this intertemporal trade-off is beyond the present scope but represents a natural extension of the framework.

4. Evaluation design

The Philippine mother tongue-based multilingual education reform provides rare empirical leverage to evaluate language-of-instruction policy at scale. In the previous section, we argued that linguistic mismatch introduces a wedge on efficient learning, thereby shaping schooling outcomes through both achievement and subsequent enrollment. Mismatch reflects the joint choices of schools and students and is therefore inherently endogenous. More generally, schools with better linguistic matches may differ systematically from poorly matched schools along dimensions that independently influence achievement. Linguistically homogeneous communities may have stronger social cohesion, more effective local governance, or distinct preferences for educational investment. Schools that adopted local languages may have had more capable administrators, better-trained teachers, or greater community support that would raise achievement irrespective of mismatch. These considerations underscore the central identification challenge in assessing the impact of linguistic matching on schooling outcomes.

In 2009, the Philippines launched a large-scale natural experiment through a national, cohort-based rollout of MTB-MLE. The policy was first applied to kindergarten pupils in school year 2009-2010 and subsequently to all later cohorts as they progressed through Grade 3. The policy permitted schools to adopt local languages from an initial set of twelve, later expanded to nineteen, regional lingua franca and local languages as the medium of instruction in kindergarten to Grade 3. Although MTB-MLE was implemented nationwide, schools experienced very different reductions in linguistic mismatch because baseline language distributions vary sharply across communities. These distributions have been shaped by long-standing patterns of migration, trade, and settlement, and therefore generate substantial heterogeneity in how much the MTB-MLE reform realigned the language of instruction to students' mother tongues.

We exploit these policy-induced changes in linguistic mismatch over time to address potential endogeneity using two complementary approaches: a difference-in-differences (DID) design and an instrumental variables difference-in-differences (IV-DID) strategy implemented within an event-study framework. Both approaches draw on the reform's impact on language matching together with a rich school-level panel from 2009 to 2024. The DID strategy isolates departures from common trends that can be attributed to the reform, while the IV-DID strategy addresses

remaining concerns about endogenous treatment intensity by using shift-share instruments based on pre-policy and contemporaneous neighborhood language distributions.

This section describes our empirical approach to estimating the learning tax from linguistic mismatch. We proceed in four steps. First, we describe our data sources and construct the analytic sample linking student outcomes, linguistic composition, and school resources for a panel of schools spanning 2008 to 2024. Second, we operationalize our key explanatory variable by measuring the degree of linguistic match between students and their language of instruction, drawing on the conceptual framework from Section 3. Third, we present our difference-in-differences identification strategy, exploiting the MTB-MLE reform's differential impact across schools with varying baseline linguistic compositions. Fourth, we introduce our instrumental variables approach to address potential time-varying confounders, using shift-share measures of exogenous linguistic concentration to isolate causal effects of improved matching.

4.1. Data and sample construction

We are interested in assessing the impact of MTB-MLE on three key schooling outcomes, namely, learning achievement, student progression, and available school resources per student. These metrics align with results from our conceptual framework that better linguistic matching in schools should improve learning outcomes and raise likelihood of students continue schooling but, in the process, also contribute to school resource dilution in the absence of additional investments.

Our analysis draws on three primary data sources covering educational outcomes, linguistic composition, and school resources across the period 2008 to 2024. The extended panel allows us to observe schools both before and after the national MTB-MLE rollout that began in school year 2009-2010, with its first cohort completing the full MTB-MLE cycle in 2012-2013.

For learning achievement, we used mean percentage scores (MPS) from the National Achievement Test (NAT) for Grade 3, later known as the Early Language, Literacy, and Numeracy Assessment (ELLNA), an early grade assessment administered by DepEd to evaluate language and numeracy skills. The assessment covers multiple subject areas, including literacy in English and Filipino, as well as in the mother tongue with ELLNA, and numeracy. The ELLNA critically identifies the language of instruction used at each school through the mother tongue component of the test, allowing us to identify each school's language of instruction and measure its alignment with students' mother tongues.

The NAT/ELLNA had been administered almost annually between 2009 and 2024, with notable gaps in 2014-2015, 2018-2021, and 2023. The irregular administration schedule reflects planned assessment cycles, disruptions from policy changes, and the COVID-19 pandemic. Despite these gaps, the available years provide sufficient pre-policy observations for parallel trends assessment and extended post-policy coverage to trace dynamic treatment effects.

For student progression and school resources, we exploit the rich information from the Basic Education Information System (BEIS), DepEd's primary management information system used for its budget preparation, resource allocation, and performance monitoring. The system regularly collects reports on school enrollment, along with school statistics on teaching and non-teaching personnel and school infrastructure, among others.

We extract enrollment data from BEIS for the full 2008-2024 period to examine whether improved linguistic matching affects continuation rates and total school enrollment. For school resource measures, we construct indicators of pupil-class and pupil-teacher ratios where data availability permits. However, some BEIS modules are only available for public schools and coverage varies across years, limiting our panel's temporal extent for resource outcomes.

We construct measures of linguistic alignment that capture the extent to which the instructional language used in school is shared by both students and teachers. Our primary construct focuses on student linguistic match, defined as the proportion of primary school students whose self-reported mother tongue corresponds to the school's identified language of instruction. This measure captures the degree to which instruction is delivered in a language that is immediately accessible to students and therefore conditions their ability to absorb early-grade content.

We supplement NAT/ELLNA and BEIS data with information from the 2017 Student Profile collected at enrollment, which records students' self-reported mother tongues. This allows us to construct our primary treatment variable by matching students' mother tongues to the language of instruction identified in ELLNA. For robustness and to extend coverage, we also incorporate school-level language-of-instruction data from the 2018 PIDS school survey (Monje, et al., 2021) and the 2022 DepEd monitoring survey (c.f. Lloyd and Yang, 2025), both of which report the actual language of instruction used by teachers, in addition to media of instruction officially adopted by each school.

Using language of instruction measures derived from NAT/ELLNA yields 165,105 school-year observations for 28,278 schools representing 10,515,887 Grade 3 students in our main learning achievement sample. In contrast, LOI measures based on the PIDS/DepEd surveys cover 123,627 school-year observations for 21,216 schools representing 7,588,249 Grade 3 students over the same period. Nearly all observations come from public schools; private schools constitute less than one percent of the sample.

Because instructional effectiveness also depends on teachers' familiarity with the LOI, we develop a parallel teacher linguistic match measure by approximating teachers' home-language distributions using the municipality-level mother-tongue composition from the 2010 Census of Population and Housing. We map these distributions onto each school's location and compare them to the language of instruction recorded in ELLNA or in the PIDS and DepEd supplemental surveys. While this procedure does not identify individual teacher languages, it provides a consistent proxy for the linguistic alignment between a school's teaching workforce and its adopted medium of instruction.

To summarize the joint linguistic environment, we construct an integrated measure that combines student- and teacher-side alignment into a single index. Rather than imposing an additive structure that would mechanically overweight larger groups, we use the geometric mean of the two match rates. This aggregation captures the idea that successful classroom communication requires both students and teachers to share the instructional language, and that misalignment on either side diminishes the overall effectiveness of MTB-MLE.

For measuring phylogenetic distances between languages, we rely on the taxonomic hierarchy established in Ethnologue (Eberhard et al., 2024), which represents decades of comparative linguistic research. For Philippine languages specifically, the current scientific consensus reflected in Ethnologue draws on the comparative Austronesian work of Blust (1991, 2009) and Philippine subgrouping studies by Reid (1971, 1982) and Zorc (1977). These sources

provide hierarchical relationships among Philippine languages, allowing us to calculate tree distances that capture the degree of linguistic relatedness between language pairs.

4.2. Measuring linguistic match

A key parameter in our linguistic mismatch framework is ρ , which governs the rate at which mutual intelligibility decays as languages diverge along a linguistic phylogenetic tree. A smaller ρ value implies that intelligibility drops sharply even with modest linguistic distance, while larger values imply greater persistence. Back-of-the-envelope calculations using the correlations and derived tree-distance metrics reported in Gooskens, et al. (2017) indicate that this decay rate is relatively small, ranging from roughly 0.04 to 0.41 across three European language families. Applying the same procedure to Zorc's (1977) seminal reconstruction of Bisayan phylogeny yields a markedly wider range of 0.05 to 0.89 across three Bisayan subgroups. Igarashi et al. (2024), drawing on Fearon's (2003) cross-country linguistic distance data, adopt a somewhat more central value of $\rho = 0.50$ in their application on the Philippines.

Given the considerable uncertainty surrounding the appropriate value of ρ , our main analyses err on the conservative side by adopting $\rho = 0$. This choice treats linguistic mismatch analogously to an FGT headcount ratio, where only the extensive margin of mismatch enters the objective. Setting $\rho = 0$ therefore aligns the index with the most policy-relevant notion for MTB-MLE, i.e., whether students' mother tongues align with the school's language of instruction, regardless of the magnitude of linguistic distance among mismatched learners.

In our descriptive analysis, we also consider alternative values of ρ . Setting $\rho = 1$ corresponds to taking the simple average linguistic distance between students' mother tongues and the school's language of instruction, thereby capturing the typical degree of mismatch experienced within a school. In contrast, $\rho = 2$ places greater weight on students who are linguistically more distant from the LOI, highlighting distributional concerns and emphasizing the tail of the distance distribution. Together, these alternative parameterizations allow us to examine how sensitive our descriptive patterns are to assumptions about how linguistic mismatch aggregates across learners.

We normalize all phylogenetic distances relative to a maximum of 17 steps, reflecting the distance between Chavacano and other Philippine languages in the Ethnologue taxonomy. Chavacano is a Spanish-based creole that lies outside the Austronesian language family encompassing virtually all other Philippine languages. Because Chavacano has limited Austronesian substrate influence, we calculate its distance to other Philippine languages as a weighted combination of the substrate language distance and a fixed penalty for its creole status, specifically 0.15 times the distance to its nearest substrate (Tagalog, Sinugbuanong Binisaya, etc.) plus 0.85 times 20. This formula yields distances of approximately 17 steps between Chavacano and most Philippine languages, representing the maximum observed separation in our linguistic landscape. We use this maximum to scale all distances, dividing raw phylogenetic distances by 17 and raising the result to the power ρ . This ensures our mismatch indices range between 0 and 1 regardless of the chosen ρ value.

4.3. Difference-in-differences strategy

Our baseline DID specification estimates dynamic treatment effects in an event study framework

$$y_{st} = \eta_s + \eta_t + \sum_{k \neq 2012} \beta_k (c_s \cdot I[t = k]) + \epsilon_{st},$$

where y_{st} is the average outcome for school s in year t , c_s is our measure of school-average linguistic matching, which we define below, and ϵ_{st} is the model residual with standard stochastic properties for estimation. The school fixed effects η_s absorb time-invariant differences between linguistically homogeneous and heterogeneous schools, including baseline linguistic distribution, infrastructure quality, geography, and long-standing community characteristics. The year fixed effects η_t capture system-wide shocks such as curriculum changes, adjustments in assessment protocols, or nationwide reforms like K-to-12. We are interested in β_k , which trace out the dynamic evolution of the relationship between language matching and school outcomes, with year 2011, i.e., the year immediately preceding MTB-MLE implementation, serving as the reference period.

The above additive specification departs from the multiplicative structure of our conceptual framework. It also estimates the impact of linguistic match instead of mismatch. The two formulations are nevertheless fully concordant once we note that

$$E[\beta_k] = \frac{dE[y_{sk}]}{d(c_s \cdot I[t = k])} = \frac{dE[y_{sk}]}{d((1 - m_s) \cdot I[t = k])} = \frac{2\tau}{(1 + \alpha)} \cdot E[y_{sk}],$$

which shows that the estimated DID coefficients recover the average increase in total learning tax induced by a unit increase in linguistic mismatch.

The key identifying assumption is that schools with different baseline linguistic compositions would have followed similar trends in average schooling outcomes in the absence of MTB-MLE. While this is not directly testable, we assess its plausibility by examining whether the treatment variables predict differential pre-policy trends. Specifically, we test whether the pre-2012 event-study coefficients β_k are jointly equal to zero using F-tests. Statistically significant or economically meaningful deviations of these coefficients from zero would indicate differential pre-treatment trends, thereby casting doubt on the parallel trends assumption. To complement these tests, we plot the coefficients with 95% confidence intervals to visually assess the magnitude of any pre-trends relative to post-policy effects.

An important potential threat to parallel trends would be differential anticipation. If schools with certain linguistic profiles adjusted behavior before 2009 in expectation of the reform, this could create spurious pre-trends or mask true ones. MTB-MLE was implemented through a cohort-based rollout beginning in school year 2009-2010. The policy started with kindergarten and expanded by one grade level annually, reaching full K-3 implementation by 2012-2013. Critically, students already enrolled in Grade 1 or higher in 2009 continued under the previous bilingual education system and were never exposed to MTB-MLE. This cohort-based design eliminates concerns about anticipation since schools could not differentially adjust before implementation because treatment was determined by student grade level at policy introduction, not by school choice or preparedness.

A separate concern in many DID applications is staggered adoption, where variation in treatment timing can produce biased estimates under heterogeneous treatment effects (Callaway and Sant'Anna, 2021; Borusyak, et al., 2024). This issue does not arise in our current

application since MTB-MLE was introduced simultaneously for all schools, with treatment intensity varying by predetermined linguistic composition rather than by adoption timing.

Guided by our conceptual framework, we estimate the impact of improved linguistic matching on several key schooling outcomes. Specifically, we examine Grade 3 test scores, school-level enrollment, and learning resource intensity per student. For test scores, we align our school-level specifications with the structure of the student-level analysis used elsewhere (e.g., Lloyd and Yang, 2025) by weighting our regressions by Grade 3 enrollment to mirror the distribution of test takers across schools. A well-known challenge with group-data regression is that estimated standard errors tend to be overly optimistic, because aggregation masks within-school variation (Schweibert, 2014). To ensure that our inference reflects both the unobserved heterogeneity lost through aggregation and the heteroskedasticity inherent in school-level mean scores, we adjust estimated standard errors using information on school-level NAT/ELLNA standard deviations following the scaling procedure in Moineddin and Urquia (2014).

Across all models, we cluster standard errors at the schools-division level. This accounts for potential correlation in unobserved shocks among schools operating within the same administrative unit and over time, such as shared leadership, resource constraints, and implementation support. Clustering at this level yields more conservative inference while preserving meaningful cross-division variation in treatment intensity. Together, these estimation choices ensure that our results reflect policy-relevant aggregation and provide valid inference that is robust to general forms of within-division and temporal correlation.

4.4. Instrumental variable Difference-in-differences

While our DID specification addresses several identification concerns by differencing out the contribution of time-invariant school-level characteristics, it does not fully address selection on time-varying factors. A key time-varying concern is that schools with stronger linguistic alignment may experience differential trends in unobserved determinants of achievement. For example, linguistically homogeneous communities may grow more rapidly, invest more in complementary educational inputs, or attract higher-quality teachers in the post-policy period. To address these concerns, we augment the standard DID specification with an instrumented difference-in-differences (IV-DID) approach, in which actual language matching is instrumented using shift-share measures of exogenous linguistic composition.

Our shift-share instruments follow the general structure

$$z_s = \sum_{\ell} \theta_{s\ell}^{pre} \times \theta_{-s\ell}^{post},$$

where $\theta_{s\ell}^{pre}$ represents baseline (pre-policy) share of language ℓ speakers in school s or its narrow catchment area, while $\theta_{-s\ell}^{post}$ captures the contemporaneous share of language ℓ speakers in the school's broad catchment area, excluding the school itself. The shift-share resembles a Herfindahl index, here used to measure concentration in language distribution. The intuition is that schools in areas with more concentrated linguistic distributions, i.e., $z_s \rightarrow 1$, are more likely to switch to a local language of instruction and therefore achieve good language matches following MTB-MLE, independent of other school characteristics.

We exploit the idea that pre-policy language shares reflect longstanding, area-specific trends that existed prior to the introduction of MTB-MLE and are therefore unlikely to have been shaped by anticipatory responses to the reform. To ensure that the post-policy shifts are similarly insulated from school-level behavior, we define the broad catchment at the province level, which is large enough that no individual school can meaningfully influence aggregate language distributions. Not wanting in caution, we further implement a leave-one-out adjustment to the contemporaneous shares, removing the mechanical correlation that would otherwise arise from including school s in the construction of its own instrument.

Following Ye, et al. (2023), the IV-DID approach allows our instruments to have a direct effect on the outcome, as long as they do not affect the trend the outcome would have occurred without the policy. In our context, this means that linguistic concentration can influence achievement levels but should not drive differential growth trajectories across schools, thereby preserving parallel trends. The DID component is essential because it differences out the contribution of baseline differences and common shocks, accounting for pre-reform trends. Conditional on schools with different instrument values having followed parallel trends absent the policy, any post-policy divergence can be interpreted as the causal effect of MTB-MLE. This strategy relaxes the stringent exclusion restriction demanded by standard IV methods while still allowing consistent identification of the policy's impact.

We construct three shift-share instruments corresponding to our three linguistic match measures: student match, teacher match, and their geometric mean. For student linguistic match, the baseline shares come from the 2010 Census of Population and Housing (CPH) language distribution aggregated to the province level. The contemporaneous shares are constructed from the 2017 Student Profile, also aggregated at the province level but excluding the focal school's own contribution through a leave-one-out procedure.

For teacher linguistic match, we follow an analogous construction but adapt a geographic scope to reflect teacher labor markets. Teacher contemporaneous shares are approximated using 2010 CPH mother-tongue distributions at the municipality level with a leave-one-out adjustment, reflecting the fact that teachers are typically drawn from local communities and that municipal-level language distributions provide a reasonable proxy for the linguistic profiles of teaching staff available to each school. The baseline shares come from the 2000 CPH aggregated to the province level, providing a predetermined demographic anchor that predates both MTB-MLE and the specific hiring decisions of individual schools. Finally, for our integrated student-teacher linguistic match measure, we construct the instrument as the geometric mean of the student and teacher shift-share measures, preserving the multiplicative structure of our conceptual framework in which effective linguistic matching requires alignment on both sides of the classroom.

We complement our linguistic matching estimates by analyzing the effects of switching from bilingual instruction to other Philippine languages, using an identification strategy that is broadly parallel to Lloyd and Yang (2025). As in their study, we classify a school as a switcher if it adopted a language of instruction other than Filipino (Tagalog) or English, as recorded in ELLNA or the PIDS and DepEd school surveys. The decision to adopt a new instructional language may respond to unobserved changes in school quality, parental demand, or local linguistic politics, which makes the treatment endogenous. We therefore also employ an IV-DID approach, although our identifying restriction differs from that of Lloyd and Yang (2025), who rely on baseline language shares as a list of instruments.

To isolate plausibly exogenous variation in instructional language switching, we construct a shift-share instrument similar to the one used for student matching, except that Filipino is excluded from the calculation. This approach instead exploits exogenous variation in the concentration of non-Filipino languages in the school's catchment area, which provides a separate source of identifying variation and allows us to isolate the causal effect of switching from changes in linguistic matching. Intuitively, schools located in areas where one non-Filipino language is highly concentrated face lower coordination costs and likely stronger community pressure to adopt that language as the medium of instruction. These forces arise from broad demographic structure rather than school-specific shocks, satisfying the exclusion restriction insofar as baseline linguistic concentration affects student outcomes only through its influence on language choice.

This strategy allows us to identify the causal effect of instructional language switching independently of changes in linguistic matching. We argue that switching to a local language is neither necessary nor sufficient to improve student-school language concordance. Some schools switch to a language that few of their students speak, while others improve alignment without switching. The IV estimates therefore isolate the treatment effect of institutional language adoption itself. This distinction clarifies why studies that treat "switching" as the treatment may recover effects that differ from linguistic matching-based analysis. Switching captures the policy mechanism of changing the language of instruction, while matching captures the underlying linguistic alignment that directly governs comprehension and learning.

5. Results

This section presents empirical evidence on the effects of MTB-MLE implementation on Philippine schooling outcomes. We proceed in five parts. In Section 5.1, we document patterns of linguistic deprivation across schools based on different reference languages of instruction, establishing the extent of treatment variation and assessing covariate balance across comparison groups. In Section 5.2, we then estimate the impact of improved linguistic matching on Grade 3 learning outcomes using both difference-in-differences and instrumental variables approaches, examines robustness to alternative specifications and data sources, and explores treatment effect heterogeneity across language groups.

In Section 5.3, we analyze enrollment responses to MTB-MLE, tracing how reduced linguistic barriers affect student continuation rates and total school enrollment across grade levels. In Section 5.4, we investigate the resource dilution mechanism, testing whether enrollment expansion following improved linguistic matching reduced per-student availability of classrooms and teachers. Finally, in Section 5.5 we combine these estimates to recover the implied learning tax from linguistic mismatch and assess the net welfare effects of the reform accounting for both achievement gains and resource constraints.

5.1. School-based linguistic deprivation

We begin by documenting the extent of linguistic mismatch across Philippine schools with MTB-MLE implementation using different reference languages of instruction. Figure 1 maps the geographic distribution of children aged 5-9 by primary home language using based on the 2020 Census of Population and Housing. Panel A confirms that English is virtually absent as a mother tongue throughout the country, where fewer than one percent of children in nearly all municipalities report English as their primary home language. Panel B shows that Filipino

(Tagalog) speakers concentrate heavily in the National Capital Region (NCR) and surrounding provinces of Central Luzon, CALABARZON and MIMAROPA, with penetration exceeding 90% in core Mega Manila (NCR, Bulacan, Cavite, Laguna, and Rizal) cities and municipalities but falling below 25% across most of the Visayas and Mindanao. Panel C reveals that the 19 officially designated languages of instruction cover most Philippine children, though notable gaps remain in parts of the Cordillera Administrative Region, Masbate, Romblon, sections of Palawan and Cagayan Valley, and scattered areas throughout Mindanao where local languages lack formal MTB-MLE instructional language status.

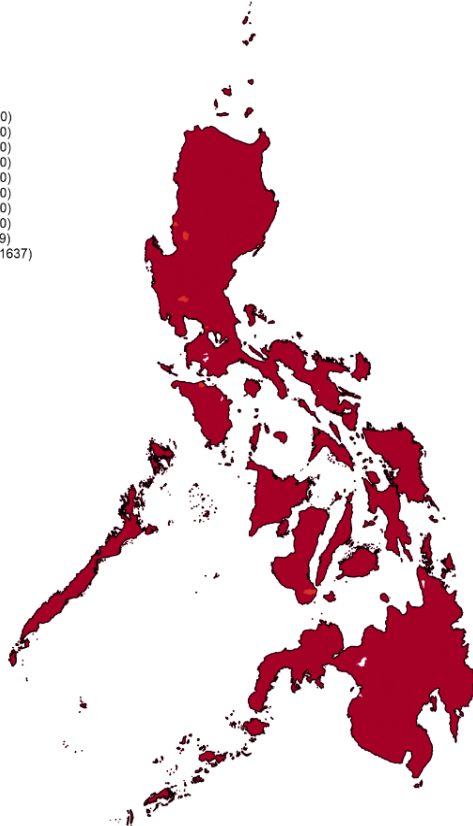
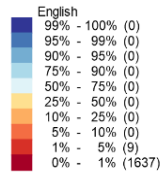
These geographic patterns translate into substantial variation in school-level linguistic matching. Table 1 quantifies aggregate linguistic deprivation under three scenarios, namely, the pre-reform bilingual education regime, MTB-MLE using schools' reported LOI, and MTB-MLE using the modal student language in each school. We measure deprivation using the Foster-Greer-Thorbecke family of indices with three parameterizations corresponding to the headcount ratio ($\rho=0$), gap ($\rho=1$), and severity ($\rho=2$) measures from our conceptual framework.

We begin by simulating linguistic deprivation under a counterfactual bilingual education regime using 2017 student language distributions. Panel A of Table 1 shows that linguistic deprivation would have been pervasive. The national headcount ratio would reach 0.79, indicating that nearly four in five students attended schools where the primary instructional language differed from their mother tongue. This aggregate masks sharp regional variation. The National Capital Region exhibited minimal mismatch (0.03 headcount ratio) because most students already spoke Filipino. In contrast, students in Visayas and Mindanao would have experienced near-universal school-based linguistic deprivation under a bilingual education regime, with headcount ratios of 0.99 in both regions. The gap and severity measures confirm this pattern. Mean linguistic distance stood at 0.24 nationally, rising to 0.29-0.30 in the Visayas and Mindanao, with squared distances similarly concentrated outside NCR.

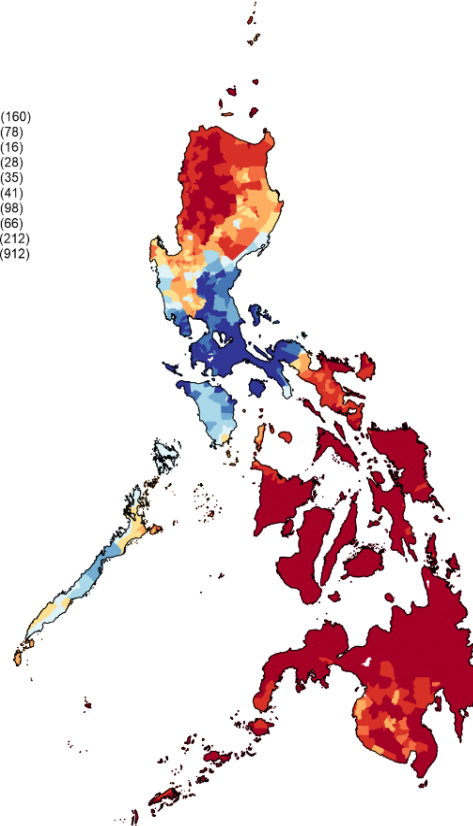
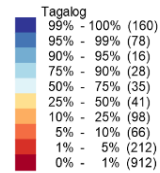
MTB-MLE implementation substantially reduced potential linguistic deprivation, though considerable mismatch persists. Using schools' reported LOI from ELLNA data, the national headcount ratio fell to 0.28, i.e., a 65% reduction from the bilingual regime. The gap measure declined by two-thirds to 0.08, while severity dropped by 63% to 0.03. These improvements occurred across all regions, including in the NCR where the reform had minimal impact on already well-matched students. The Visayas experienced the sharpest proportional gains, with headcount ratios falling from 0.99 to 0.21 and gap measures declining from 0.29 to 0.04. Mindanao saw more modest improvements, with headcount ratios dropping to 0.31 and gap measures to 0.09, reflecting the region's greater linguistic diversity and the incomplete coverage of local languages among the 19 official instructional languages.

Figure 1. Share of children aged 5 to 9 years by primary language used at home and residence location

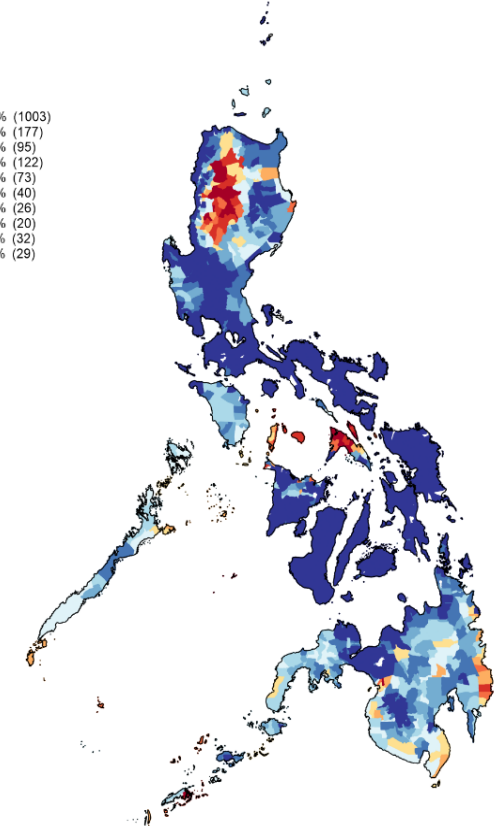
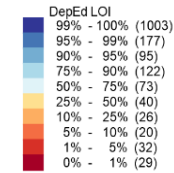
A. English



B. Tagalog



C. Any of the 19 DepEd official LOI



Source: Authors' estimates based on 2020 CPH. Note: The figure reports the distribution of primary home languages among children aged 5-9 years. For the 19 Department of Education-designated languages of instruction (LOIs), we classify all children whose primary home language corresponds to any of these 19 LOIs, irrespective of the LOI actually used in their school.

Table 1. School-based linguistic deprivation: Philippines, 2017

	Linguistic deprivation		
	Headcount ratio ($\rho = 0$)	Gap ($\rho = 1$)	Severity ($\rho = 2$)
A. Bilingual education			
Philippines	0.79	0.24	0.08
National Capital Region	0.03	0.01	0.00
Rest of Luzon	0.56	0.17	0.06
Visayas	0.99	0.29	0.09
Mindanao	0.99	0.30	0.10
B. MTB-MLE, reported language			
Philippines	0.28	0.08	0.03
National Capital Region	0.03	0.01	0.00
Rest of Luzon	0.32	0.10	0.04
Visayas	0.21	0.04	0.01
Mindanao	0.31	0.09	0.03
C. MTB-MLE, modal language			
Philippines	0.14	0.04	0.02
National Capital Region	0.03	0.01	0.00
Rest of Luzon	0.13	0.04	0.02
Visayas	0.10	0.02	0.00
Mindanao	0.19	0.06	0.03

Note: Linguistic deprivation is measured using the Foster-Greer-Thorbecke (1984) family of deprivation indices parameterized by ρ . Estimates combine school-level language-of-instruction (LOI) data from ELLNA with students' primary home languages from the 2017 Student Profile. Linguistic distance between a student's mother tongue and the reference language of instruction is computed using Ethnologue's phylogenetic classification of Philippine languages. We exclude schools with multiple ELLNA-identified LOI, corresponding to less than one percent of schools in our sample.

Residual mismatch under MTB-MLE reflects two distinct phenomena. First, some schools adopted media of instruction that did not align with their students' dominant mother tongues, either because the modal language lacked official instructional language status or because schools faced implementation constraints. Second, linguistic heterogeneity within schools means that any single language of instruction necessarily mismatches some fraction of students, even when schools choose optimally. The modal language scenario in Panel C provides a benchmark for these sources of mismatch by assigning each school the instructional language among the 19 MTB-MLE languages that minimizes its student headcount ratio. Under this counterfactual, national deprivation halves again to 0.14 headcount ratio and 0.04 gap measure. The Visayas approaches near-complete matching with just 0.10 headcount ratio and 0.02 gap, while Mindanao's more fragmented linguistic landscape yields 0.19 headcount and 0.06 gap, still representing a 62% improvement over reported MTB-MLE implementation.

To interpret these linguistic deprivation measures concretely, recall that the maximum nodal distance in the Philippine language phylogenetic tree is 17 steps, spanning languages as distant as Ilocano and Chavacano. Under bilingual education, the national gap measure of 0.24 implies

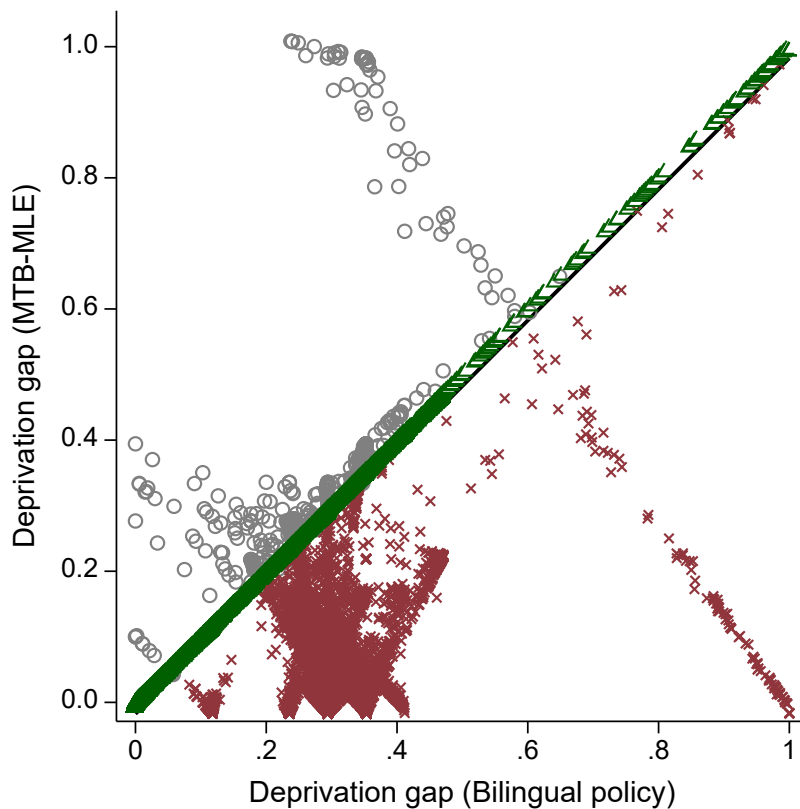
that the average mismatched student attended school in a language 4.1 phylogenetic steps removed from their mother tongue (0.24×17). In the Visayas and Mindanao, gap measures of 0.29-0.30 translate to average distances of 4.9-5.1 steps, roughly the distance separating closely related languages like Cebuano and Hiligaynon from more distant Visayan languages, or Tagalog from Central Luzon languages. MTB-MLE reduced these distances substantially. The post-reform national gap of 0.08 corresponds to 1.4 phylogenetic steps, while the Visayas gap of 0.04 represents just 0.7 steps, distances comparable to neighboring dialects within the same major language subgroup. The severity measures, which square these distances before averaging, fell even more dramatically from 0.08-0.10 under bilingual education to 0.01-0.03 under MTB-MLE. This compression indicates that MTB-MLE not only reduced average mismatch but disproportionately benefited the most linguistically distant students, eliminating the extreme mismatches that characterized the bilingual regime while concentrating remaining deprivation among students experiencing relatively modest linguistic distances from their school's instructional language.

While the MTB-MLE reform allowed schools to switch from bilingual Filipino-English instruction to one of 19 authorized local languages, this institutional change did not uniformly improve linguistic matching. Figure 2 displays the relationship between schools' linguistic deprivation gaps under a bilingual regime and under MTB-MLE implementation. Each point represents a school, with location relative to the 45-degree line indicating whether MTB-MLE improved, worsened, or left unchanged the school's linguistic alignment. Schools below the 45-degree line experienced reduced mismatch under MTB-MLE, while those above the line saw linguistic matching quality deteriorate despite switching languages. The concentration of points well below the diagonal confirms that most schools achieved substantial matching improvements, particularly those with high initial mismatch under the bilingual regime serving non-Tagalog populations.

However, the scatter plot reveals considerable heterogeneity in matching quality changes. Some schools that switched from bilingual instruction selected languages that remained poorly matched to their student populations, i.e., those located in the upper-right quadrant with persistently high deprivation gaps under both regimes, either because they chose a regional lingua franca that differed from most students' mother tongues or because teacher language capabilities, political considerations, or incomplete information led to suboptimal language selection (Monje et al., 2021). Other schools, clustered near the origin, achieved good matching under both regimes because they served predominantly Tagalog-speaking students for whom Filipino instruction already provided strong alignment. Still other schools, appearing above the 45-degree line, worsened their matching by switching to languages less aligned with student compositions than Filipino or English. This heterogeneity provides the identifying variation for our empirical analysis. Schools experienced differential changes in linguistic matching despite all formally adopting mother tongue education, allowing us to isolate the effects of actual alignment improvements from the institutional act of policy adoption.

Table 2 assesses covariate balance across treatment groups to evaluate whether schools experiencing different degrees of linguistic matching differ systematically along observable dimensions that could independently affect learning outcomes. We examine three comparison sets corresponding to our main empirical strategies: student-side language of instruction matching (Panel A), teacher-side linguistic matching (Panel B), and language switching from Filipino/English to other Philippine languages (Panel C).

Figure 2. School-level linguistic deprivation gap: Bilingual vs. MTB-MLE regimes



Notes: Each point represents a school's linguistic deprivation gap ($\rho=0$, capturing average phylogenetic distance between students' mother tongues and the school's language of instruction) under the bilingual regime (x-axis) and MTB-MLE regime (y-axis). The 45-degree line indicates no change in matching quality. Points below the line indicate improved matching under MTB-MLE; points above indicate worsened matching. Sample includes 35,182 schools with student mother tongues and school language of instruction data.

The first two comparisons capture variation in matching quality among all schools, including those serving predominantly Tagalog-speaking students that elected Filipino as their MTB-MLE language and thus experienced no change in instructional language. These comparisons assess whether better linguistic alignment correlates with baseline school characteristics, regardless of whether alignment improved under MTB-MLE or already existed pre-reform. The third comparison isolates schools that actually switched from Filipino/English to other Philippine languages, paralleling the treatment definition in Lloyd and Yang (2025). This captures the intensive margin of MTB-MLE implementation where instructional language changed discontinuously. For each comparison, we report 2010 baseline characteristics including school enrollment, pupil-class and pupil-teacher ratios, Grade 3 NAT mean percentage scores across subjects, and 2009 regional poverty rates matched to school locations.

Panel A splits schools at the median student linguistic match distribution. High-match schools are modestly larger (mean log enrollment of 5.4 versus 5.3) and have slightly higher average pupil-class ratios (36.5 versus 35.2), though standardized mean differences remain small at 0.18 and 0.10 respectively. Baseline test scores differ minimally across groups, wherein high-match schools score 1.7 percentage points higher in Filipino, 1.4 points in English, and 1.3 points in Mathematics, translating to standardized differences of just 0.11, 0.07, and 0.07.

Resource intensity measures show no meaningful imbalance, with average pupil-teacher ratios virtually identical at 33.7 versus 35.0. High-match schools locate in slightly less poor regions (30.3% versus 33.0% poverty rate), yielding a standardized difference of -0.19. Variance ratios across all covariates cluster near unity, ranging from 0.89 to 1.16, indicating similar within-group dispersion.

Table 2. Baseline characteristics and covariate balance

	Baseline characteristics by group				Covariate Balance	
	Reference		Comparison		Std. Mean Difference	Variance Ratio
	Mean	SD	Mean	SD		
A. LOI matching, students						
	Low match		High match			
School enrollment (in ln)	5.3	1.0	5.4	1.0	0.18	1.09
Pupil-class ratio	35.2	13.1	36.5	12.4	0.10	0.89
Pupil-teacher ratio	33.7	18.7	35.0	19.7	0.07	1.11
NAT G3, Filipino (MPS)	63.8	15.6	65.5	15.4	0.11	0.97
NAT G3, English (MPS)	59.0	18.7	60.4	19.0	0.07	1.03
NAT G3, Mathematics (MPS)	67.1	19.4	68.4	19.6	0.07	1.01
Poverty rate	33.3	15.2	30.3	16.3	-0.19	1.16
B. LOI matching, teachers						
	Low match		High match			
School enrollment (in ln)	5.3	0.9	5.4	1.1	0.17	1.35
Pupil-class ratio	35.9	13.7	35.8	11.9	-0.01	0.75
Pupil-teacher ratio	34.9	21.7	33.8	16.5	-0.06	0.58
NAT G3, Filipino (MPS)	63.8	15.5	65.5	15.4	0.11	0.99
NAT G3, English (MPS)	59.6	18.5	59.8	19.2	0.01	1.08
NAT G3, Mathematics (MPS)	67.6	19.3	68.0	19.7	0.02	1.05
Poverty rate	35.1	14.9	28.7	16.2	-0.41	1.19
C. Language switching						
	Filipino/English MT		With other MTs			
School enrollment (in ln)	5.5	1.0	5.3	1.0	-0.21	0.85
Pupil-class ratio	36.2	12.1	35.6	13.2	-0.05	1.19
Pupil-teacher ratio	34.6	16.6	34.1	20.7	-0.03	1.55
NAT G3, Filipino (MPS)	65.4	15.3	64.1	15.6	-0.08	1.05
NAT G3, English (MPS)	59.1	19.2	60.1	18.7	0.05	0.94
NAT G3, Mathematics (MPS)	67.2	19.7	68.1	19.4	0.05	0.97
Poverty rate	24.4	15.2	36.6	14.4	0.83	0.90

Note: Reference and comparison groups for student-level and teacher-level language of instruction (LOI) matching are constructed by splitting schools at the median of the respective mismatch distributions. The language-switching comparison contrasts schools where LOI remained bilingual Filipino or English with those where LOI switched to other Philippine languages based on LOI reported in ELLNA. All baseline covariates are measured prior to the rollout of MTB-MLE. School characteristics and NAT Grade 3 mean percentage scores (MPS) are from 2010, while municipal- and city-level poverty rates matched to school locations are from 2009. Standardized mean differences and variance ratios are reported to assess covariate balance across groups.

Panel B examines teacher-side matching using approximated teacher language distributions from municipal-level CPH data. The pattern closely parallels student-side comparisons. School size, classroom resources, and baseline achievement exhibit minimal differences, with standardized mean differences ranging from -0.06 to 0.17 across measures. The largest imbalance appears in municipal and city-level poverty rates, where high-match schools sit in areas with 28.7% poverty rate versus 35.1% for low-match schools, reflecting a standardized difference of -0.41 that exceeds conventional balance thresholds. This correlation between linguistic homogeneity and regional economic conditions motivates our instrumental variables strategy, which addresses potential confounding from time-varying local economic trajectories.

Panel C contrasts schools serving primarily Filipino/English speakers against those serving students with other mother tongues, corresponding to the intensive margin of MTB-MLE implementation. Filipino/English-dominant schools are larger (5.5 versus 5.3 log enrollment) and locate in substantially wealthier regions (24.4% versus 36.6% poverty), generating a standardized difference of 0.83 in local poverty rates. Baseline test scores and resource measures remain balanced, with standardized differences below 0.08 in magnitude. The sharp poverty differential reflects the geographic concentration of Filipino speakers in Metro Manila and surrounding provinces documented in Figure 1, underscoring the importance of controlling for regional economic conditions when estimating treatment effects.

Overall, the covariate balance assessment reveals two key patterns. First, schools experiencing different degrees of linguistic matching appear broadly similar along most observable dimensions, particularly baseline achievement and resource intensity. Second, systematic differences in municipal and city-level poverty rates across treatment groups, which are especially pronounced in teacher-side comparisons and language-switching analyses, highlight the potential for time-varying confounders related to local economic development trajectories. Our difference-in-differences specifications address this concern by absorbing time-invariant school characteristics and common time trends, while our instrumental variables approach further isolates exogenous variation in linguistic matching driven by predetermined demographic patterns rather than contemporaneous school or community choices.

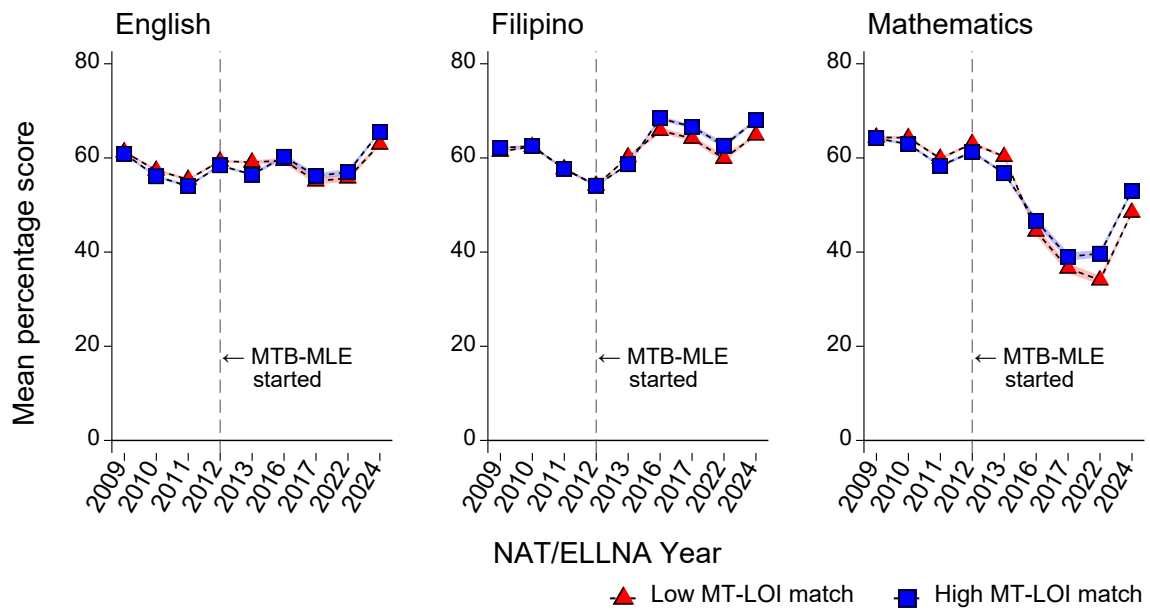
5.2. MTB-MLE and learning outcomes

Figure 3 plots mean NAT Grade 3/ELLNA mean percentage score by subject assessed, treatment group and year, providing a visual check of the parallel trends assumption underlying our difference-in-differences design. Reference and comparison groups for student, teacher, and integrated matching are formed by splitting schools at the median of their respective match distributions. The switching comparison contrasts schools that retained Filipino or English as their LOI with those that shifted to another Philippine language under MTB-MLE.

Prior to 2012, when the first MTB-MLE cohort reached Grade 3, treatment and comparison groups display broadly parallel trends in mean test scores, with only minor deviations that we explore more intently below. In the immediate post-implementation period, we observe a divergence in which better-matched schools initially perform slightly worse, while LOI-switching schools pull ahead of those that retained Filipino or English. This pattern reverses in later years. By 2024, better-matched schools exhibit higher MPS across subjects and match measures, whereas switchers fall behind schools that maintained bilingual instruction. The initial divergence coincides with the first cohort fully exposed to MTB-MLE in Grades 1-3, consistent with transitional adjustments during early implementation.

Figure 3. NAT G3/ELLNA by LOI intervention and year

A. LOI matching, students



B. LOI matching, teachers

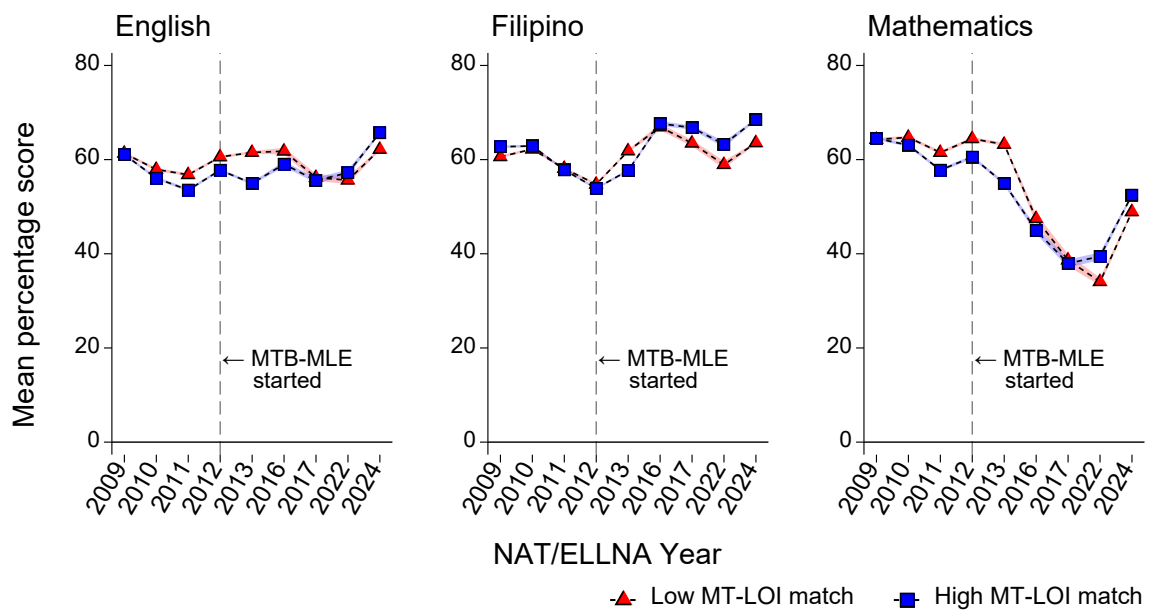
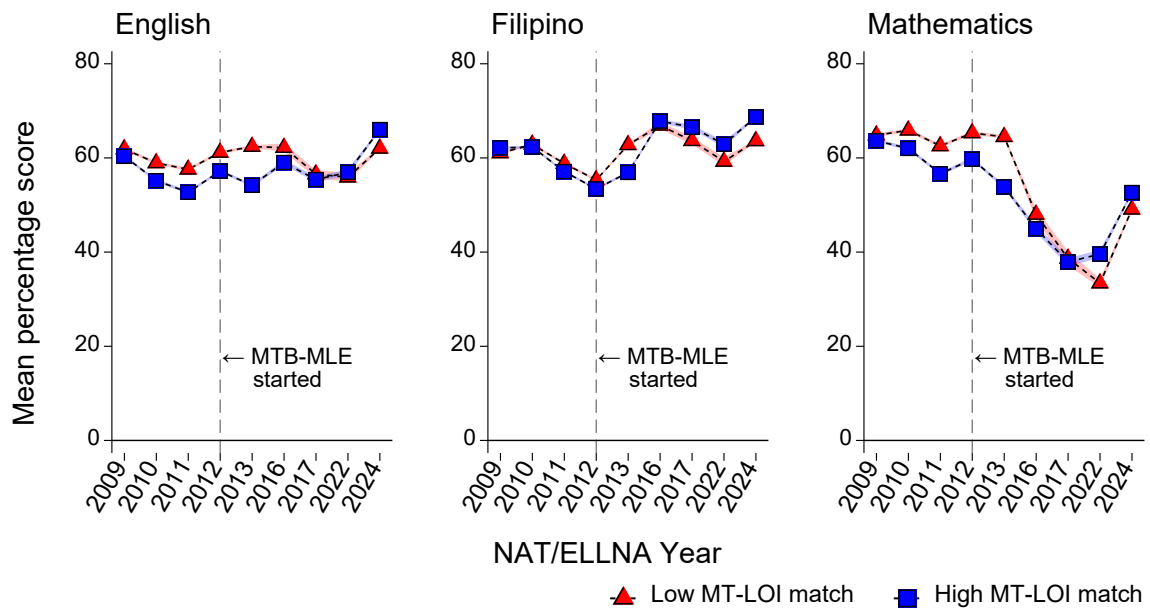
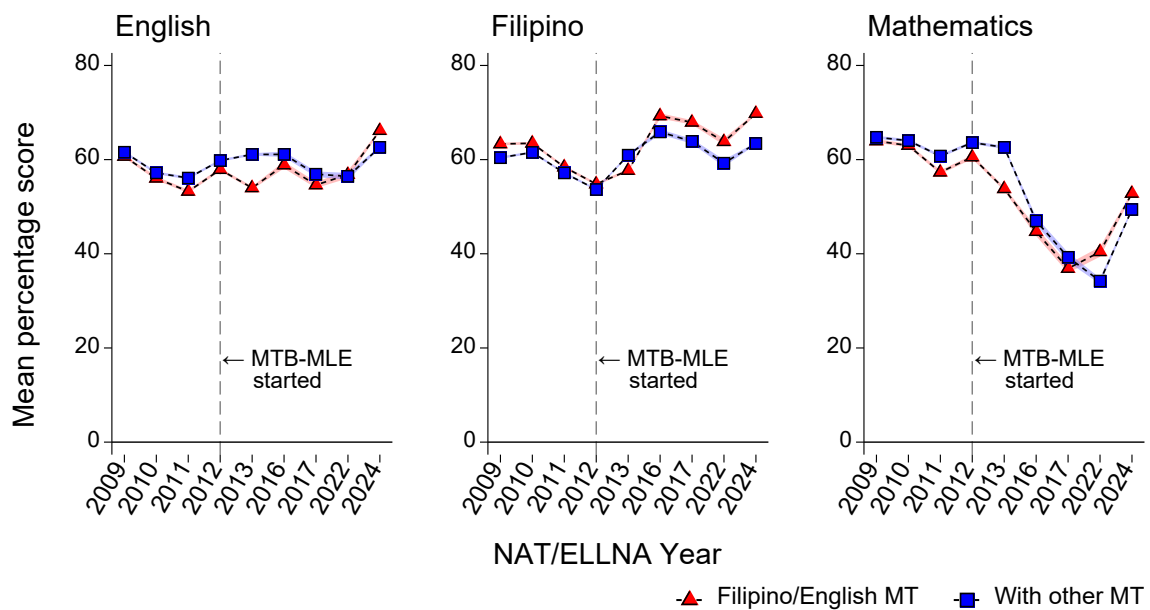


Figure 3. NAT G3/ELLNA by LOI intervention and year (continued)

C. LOI matching, interacted



D. LOI switching



Note: Reference (low match) and comparison (high match) groups for student-level and teacher-level language of instruction (LOI) matching are constructed by splitting schools at the median of the respective match distributions. The integrated student-teacher match uses the geometric mean of the two measures, also split at the median. The language-switching comparison contrasts schools where LOI remained bilingual Filipino or English with those where LOI switched to other Philippine languages based on LOI reported in ELLNA. NAT/ELLNA Grade 3 mean%age scores are plotted by year for each treatment group. The vertical line marks 2012, the year MTB-MLE achieved full K-3 implementation. Pre-2012 trends assess the parallel trends assumption, while post-2012 divergence identifies dynamic treatment effects.

5.2.1. Main results

Figure 4 presents our primary estimates of MTB-MLE effects on learning outcomes using event-study plots that trace standardized beta coefficients from 2009 to 2024 relative to the 2011 baseline. We report both DID and IV-DID specifications for each of the three matching measures, as well as for the language-switching comparison. To facilitate comparisons of effect sizes across outcomes, treatment variables, and specifications, we convert estimated coefficients into standardized betas by multiplying each coefficient by the ratio of the standard deviation (SD) of the treatment variable to the SD of the outcome variable.

By 2024, IV-DID standardized beta estimates indicate that improved linguistic matching raises average MPS substantially. A one SD increase in student linguistic match raises Filipino scores by 0.55 SD (standard error [SE] = 0.27), English scores by 0.70 SD (SE = 0.29), and Mathematics scores by 0.49 SD (SE = 0.19). For context, a one-SD increase in student match corresponds to approximately 30%age points improvement in match rates, while outcome standard deviations range from 16.4 to 17.6%age points, representing 24.5 to 34.7% of average MPS. Teacher-side matching and integrated student-teacher matching yield qualitatively similar results, with terminal estimated standardized beta effects ranging from 0.46 to 0.66 across subjects and measures. In absolute terms, the student matching effects translate to 25.3, 34.3, and 34.1 increase in MPS points respectively, equivalent to an effect size of 1.5, 2.0, and 1.9 SD of the outcome distributions.

These effect sizes are educationally meaningful. In low- and middle-income countries (LMICs), students typically gain only 0.15 to 0.21 standard deviations of literacy per school year (Evans and Yuan, 2019; c.f. for older students, see Avvisati and Givord, 2023). Benchmarking against global evidence, Evans and Yuan (2022) report that the 90th percentile of effect sizes from randomized controlled trials in LMICs reaches 0.50 SD for reading and 0.30 SD for mathematics. Our estimated matching effects meet or exceed these high-end benchmarks, indicating that MTB-MLE produces learning gains comparable to, or greater than, the most effective interventions documented in LMIC education contexts.

Additional comparisons from the broader education literature point in the same direction. Class-size reduction programs typically yield gains of 0.10 to 0.22 SD (Bowne et al., 2017), and technology-aided instruction in India generated about 0.60 SD (Muralidharan et al., 2019). The effects we recover exceed these prominent benchmarks. This underscores both the magnitude of the penalties created by linguistic mismatch and the substantial learning gains that arise when the language of instruction is aligned with students' mother tongues.

The IV-DID estimates consistently exceed their DID counterparts. For student matching in 2024, standardized beta DID estimates range from 0.11 to 0.17 SD across subjects, compared to 0.49 to 0.70 in IV-DID. It should be noted that this divergence reflects different estimands, and not necessarily measurement error correction. DID identifies average treatment effects under parallel trends, while IV-DID isolates local average treatment effects for schools whose matching status changes due to exogenous demographic shifts. The larger IV-DID estimates suggest that marginal schools induced into better matching by their linguistic composition experience stronger gains than the average treated school, consistent with linguistic match operating through continuous mechanisms where demographic variation creates more decisive shifts in instructional alignment for some schools than others.

Figure 4. MTB-MLE and G3 learning outcomes: DID and IV-DID estimates

A. Filipino

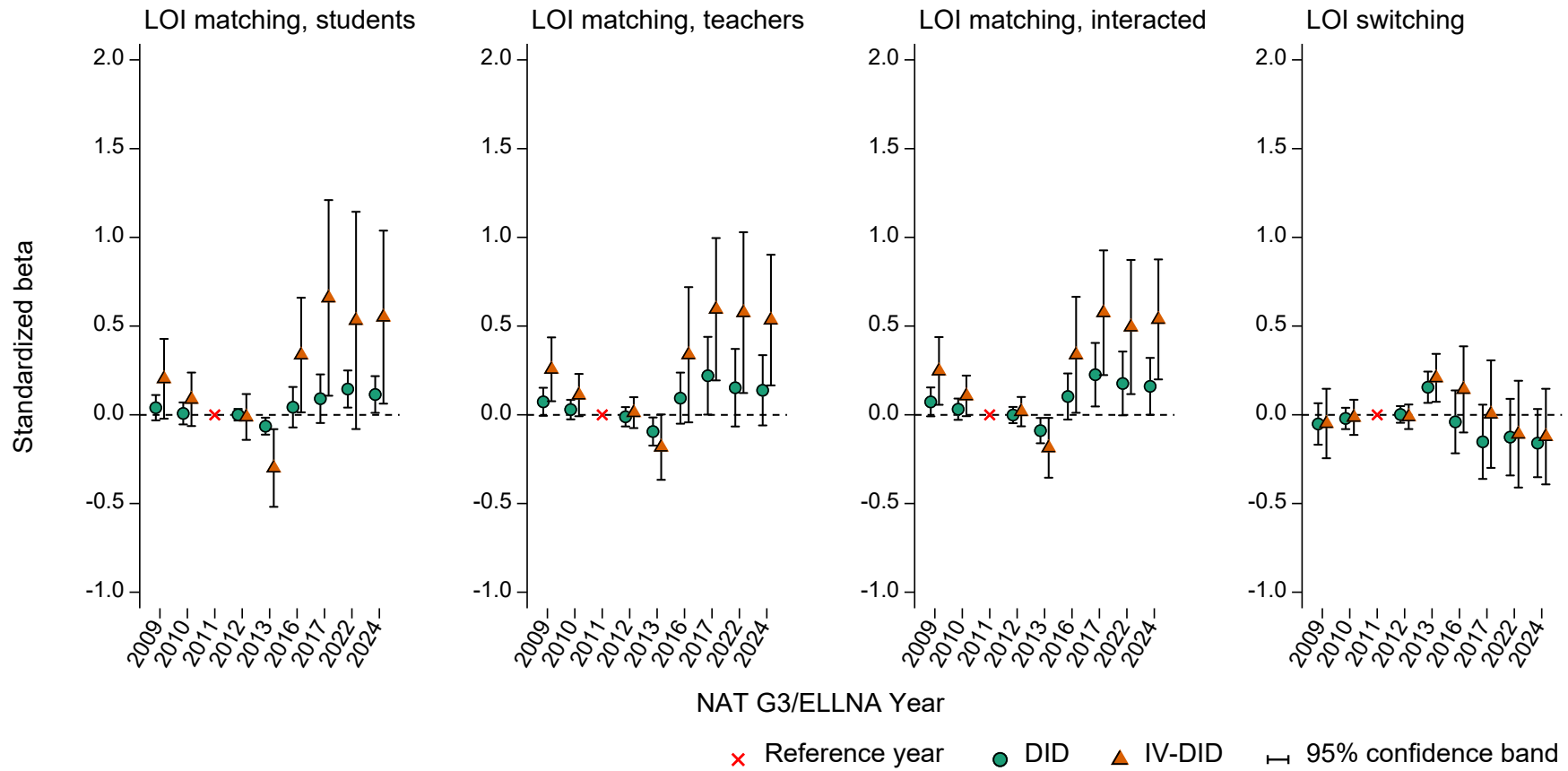


Figure 4. MTB-MLE and G3 learning outcomes: DID and IV-DID estimates (continued)

B. English

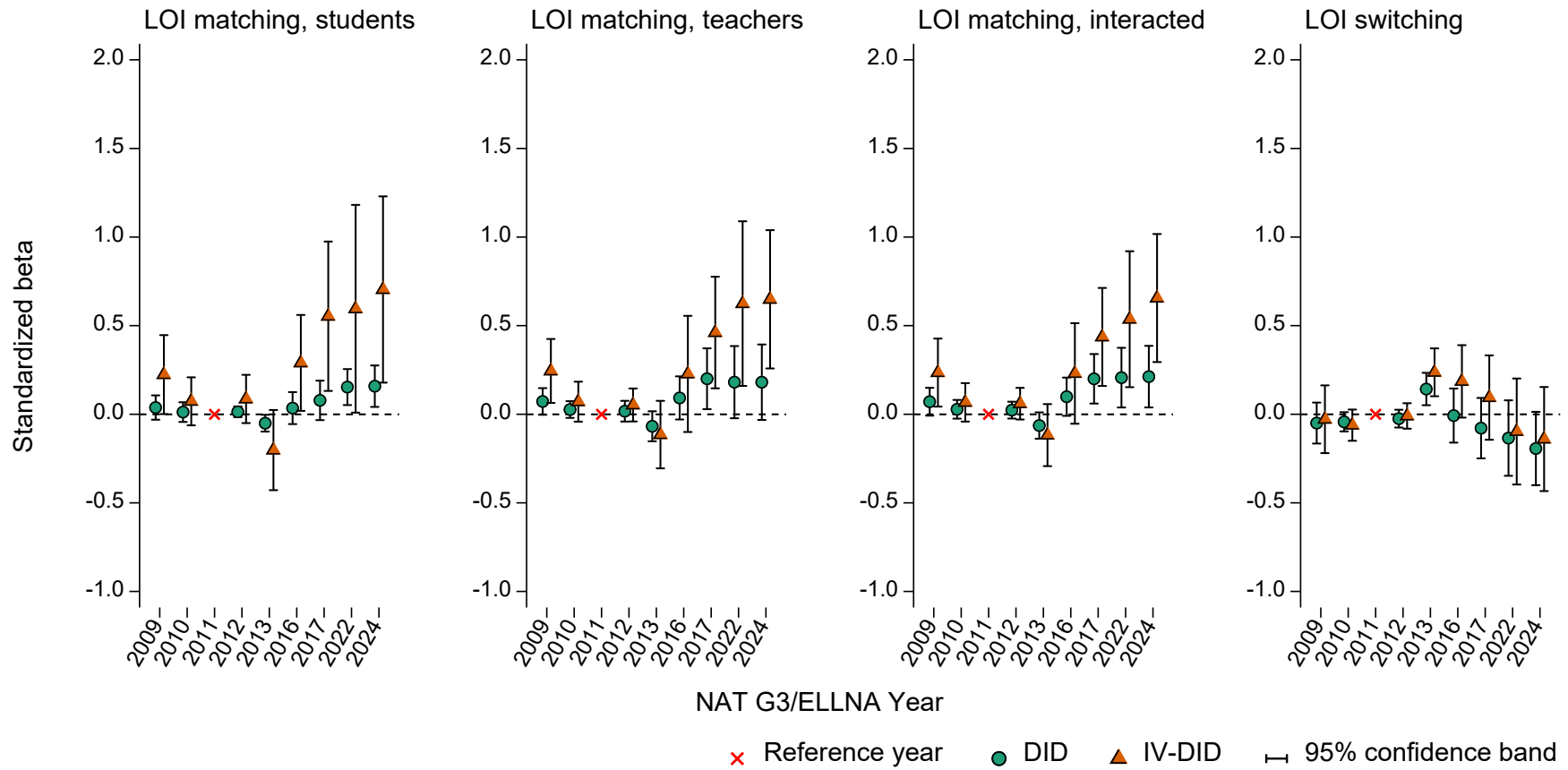
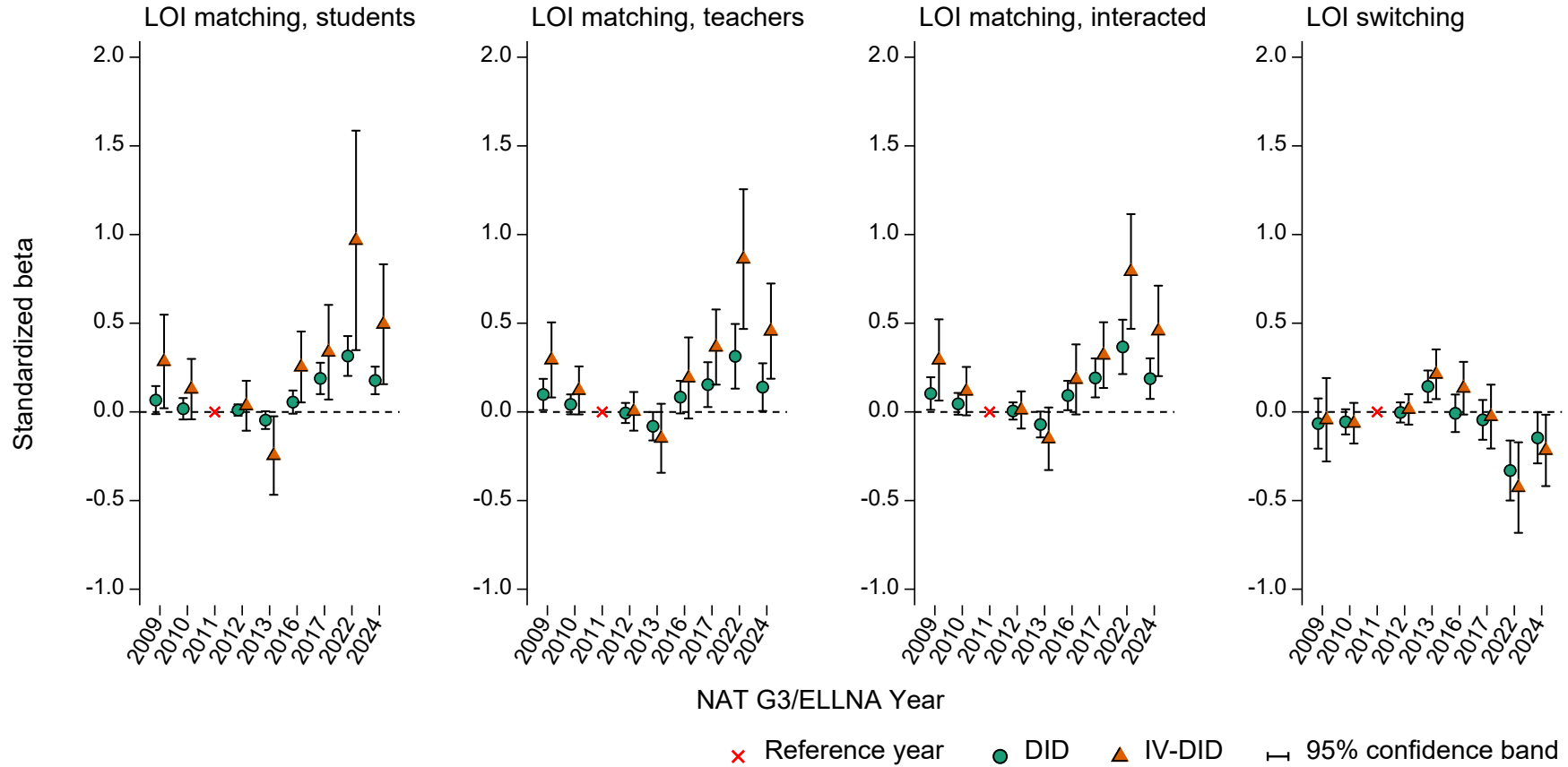


Figure 4. MTB-MLE and G3 learning outcomes: DID and IV-DID estimates (continued)

C. Mathematics



Notes: Figures display DID and IV-DID coefficients and 95% confidence intervals converted into standardized betas. The dependent variable is Grade 3 mean percentage scores in Filipino, English and Mathematics. IV-DID estimates instrument linguistic matching with shift-share calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school and year fixed effects. Standard errors are clustered at the schools-division level and scaled to include incorporate school-level variation following Moineddin and Urquia (2014).

For the switching comparison, estimated standardized betas remain close to zero through 2017 and subsequently turn negative, reaching -0.16, -0.19, and -0.15 SD for Filipino, English, and Mathematics, respectively, by 2024. These estimates are statistically indistinguishable from zero for Filipino and English, but the effect on Mathematics is marginally significant at conventional levels. Taken together, the pattern suggests that switching instructional language does not systematically yield learning gains and may impose transitional costs, particularly in Mathematics. This result aligns with an earlier study by Lloyd and Yang (2025), who document negative impacts of MTB-MLE using a related empirical approach but relying on a different identifying assumption.

Pre-reform trends merit careful attention as they inform the direction of potential bias. For matching treatments, pre-2012 coefficients are positive but declining toward zero as seen in Figure 3. In Filipino, student match coefficients fall from 0.20 (SE = 0.11) in 2009 to 0.09 (SE = 0.08) in 2010, approaching the 2011 reference point. Similar patterns appear for teacher matching and integrated matching, with 2009 coefficients of 0.26 and 0.25 respectively declining to 0.11 in 2010. This declining pre-treatment trend biases our estimates toward zero rather than away from it. If high-match schools were already experiencing relative score declines before the reform, then our post-2012 estimates understate the true effects.

The standardized pre-trend in 2009 ranges from 0.20 to 0.29 across subjects and matching measures. These values are smaller than the terminal effects in 2024, which range from 0.46 to 0.70. The downward trajectory of the pre-treatment coefficients therefore suggests that, under a linear continuation of pre-2012 trends, our main estimates represent lower bounds on the true treatment effects.

For switching, pre-trends are flat to slightly positive. The 2009 coefficients hover near zero, ranging from -0.03 to -0.05 across subjects, providing no evidence of differential pre-reform dynamics. Absent negative pre-trends, our switching estimates likely approximate true effects or represent upper bounds on null findings.

First-stage diagnostics reported in the Appendix support the validity of our instrumental-variables strategy. Kleibergen-Paap rank LM statistics range from 202 to 717 across specifications, strongly rejecting underidentification. The corresponding Kleibergen-Paap Wald F-statistics range from 202 to 718, far exceeding conventional weak-instrument thresholds (e.g., 10) as well as the more demanding Stock-Yogo critical values. These diagnostics indicate that linguistic homogeneity, as captured by our shift-share instrument, generates strong and plausibly exogenous variation in language-of-instruction alignment, providing confidence that our IV-DID estimates recover credibly causal effects.

5.2.2. *Robustness*

Figure 5 examines the robustness of our findings to alternative definitions of language of instruction. Our main analysis uses instructional language assignments from ELLNA testing protocols, which provide comprehensive national coverage but may contain classification error. As a robustness check, we re-estimate the models using languages of instruction reported in PIDS and DepEd surveys, which were also employed in Lloyd and Yang (2025). Although the survey-based measures cover fewer schools, they reflect teacher and administrator self-reports of actual instructional languages employed in classrooms, including those beyond the 19 authorized DepEd medium of instruction.

Figure 5. MTB-MLE and G3 learning outcomes: Alternative LOI data, IV-DID model

A. Filipino

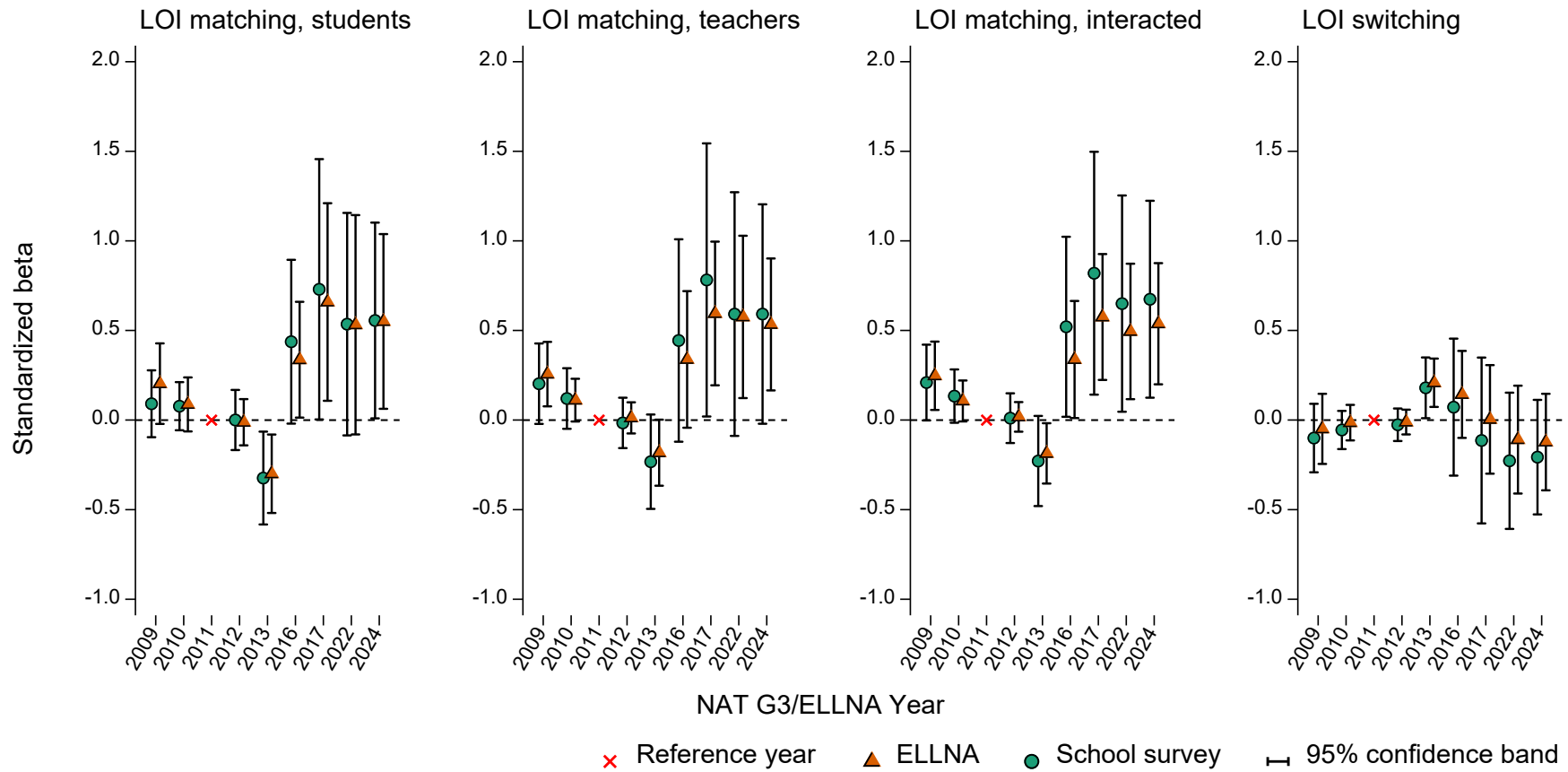


Figure 5. MTB-MLE and G3 learning outcomes: Alternative LOI data, IV-DID model (continued)

B. English

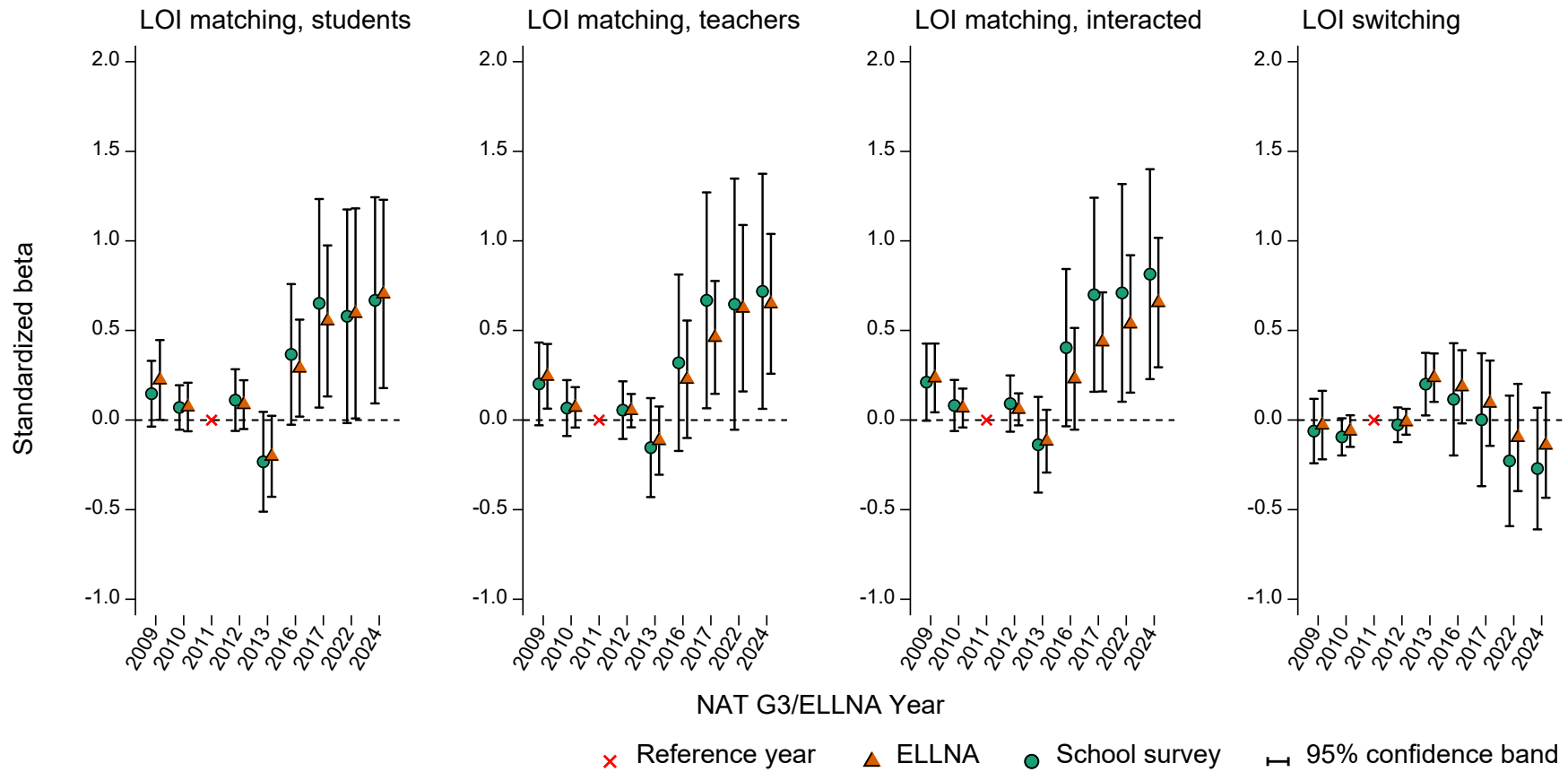
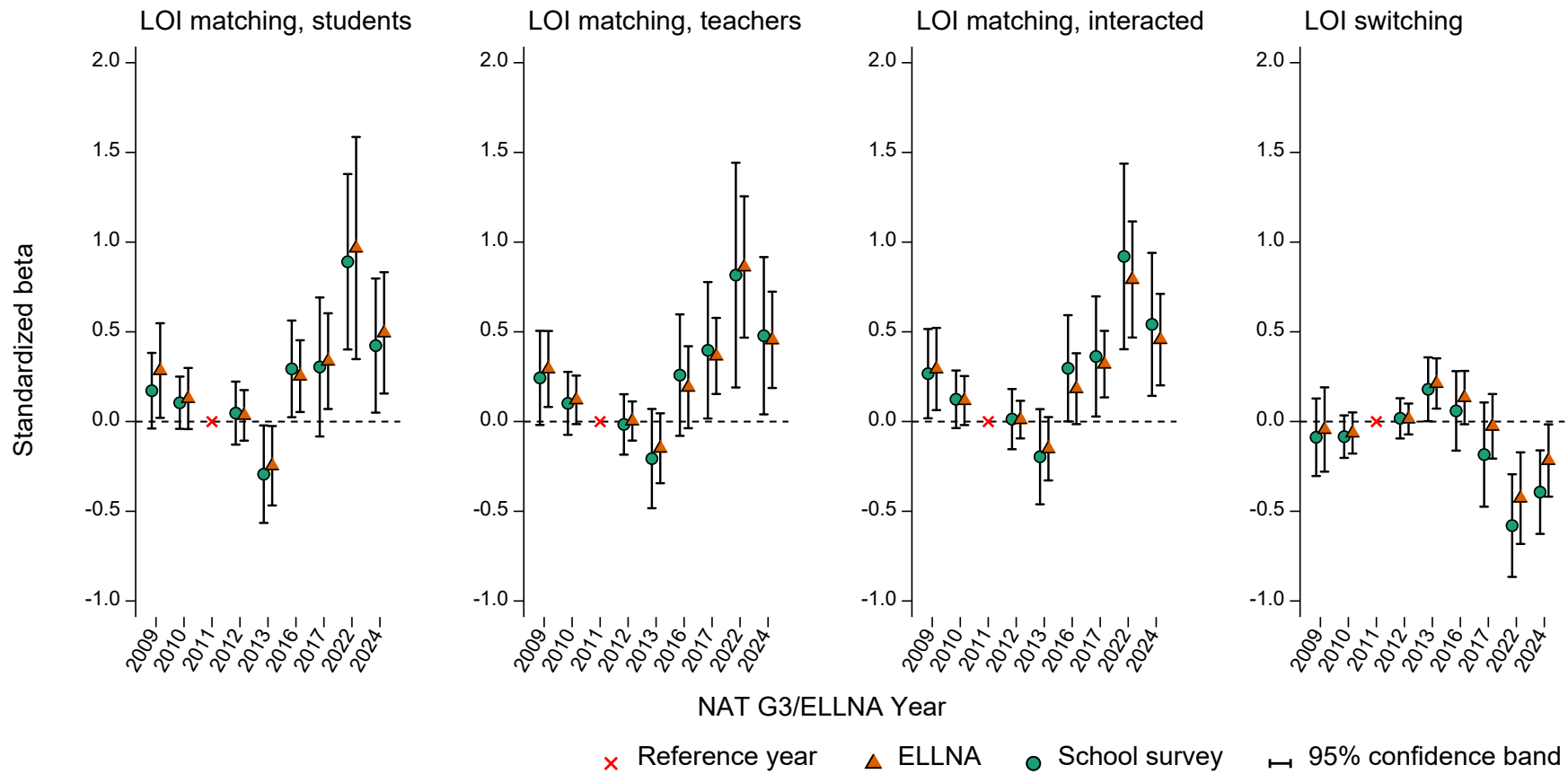


Figure 5. MTB-MLE and G3 learning outcomes: Alternative LOI data, IV-DID model (continued)
C. Mathematics



Notes: Figures display DID and IV-DID coefficients and 95% confidence intervals converted into standardized betas. The dependent variable is Grade 3 mean percentage scores in Filipino, English and Mathematics. IV-DID estimates instrument linguistic matching with shift-share calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school and year fixed effects. Standard errors are clustered at the schools-division level and scaled to include incorporate school-level variation following Moineddin and Urquia (2014).

The results are qualitatively similar. Terminal matching effects in 2024 remain positive and substantial. While confidence intervals widen due to smaller sample of schools with school survey data, point estimates closely track our baseline results. Switching effects remain near zero or negative. The consistency across instructional language definitions indicates that our findings do not hinge on any particular classification scheme but instead reflect underlying mechanisms of linguistic alignment.

We also compare estimates from joint models, where linguistic matching and instructional language switching enter simultaneously, to those from separate models where each matching measure is estimated independently (results reported in Appendix B). Coefficients are nearly identical across specifications. This stability suggests that matching and switching capture distinct dimensions of language-in-education policy implementation. Including both simultaneously does not attenuate either set of effects, providing little evidence of omitted-variable bias from excluding one treatment or the other. Moreover, schools that improve matching do not systematically differ in their switching behavior, confirming that our preferred specifications isolate independent causal pathways.

5.2.3. Heterogeneity analysis

In Figure 6, we investigate treatment effect heterogeneity by separating schools according to their chosen language of instruction. We estimate models for schools that use Filipino or English as well as for schools that use other regional or local languages.

Both groups exhibit positive linguistic matching effects, and the estimated magnitudes are broadly similar. Confidence intervals overlap substantially, so we cannot reject the hypothesis that effects are equal across language types. This suggests that the benefits of linguistic alignment are not tied to any particular instructional language but arise from the match between students' and teachers' language abilities and the language used in the classroom.

The results highlight that language operates primarily as a medium for learning rather than as content in itself. Mismatches between home language and school language reduce comprehension and slow learning regardless of which language is chosen. This finding implies that policy interventions should prioritize improving alignment between students' mother tongues and the language of instruction rather than favoring one language over another. It also underscores the potential of MTB-MLE to generate substantial learning gains across diverse linguistic contexts, showing that even schools using regional languages can achieve outcomes comparable to those using national languages when alignment is high.

5.2.4. Placebo test

We assess our identification strategy by examining whether our treatment variables predict municipal and city-level poverty rates—an outcome that should not respond directly to language policy. If our matching effects reflected local economic development rather than linguistic alignment, we should observe poverty declining in better-matched schools. Similarly, if switching estimates were uncontaminated by local economic trends, municipal- and city-level poverty rates should evolve similarly across switching and non-switching schools.

Figure 7 presents event-study estimates with local poverty rate as the outcome variable. Results reveal striking patterns that validate our identification strategy while highlighting the contamination of switching variation.

Figure 6. Heterogeneity analysis: MTB-MLE and G3 learning outcomes by instructional language

A. Filipino

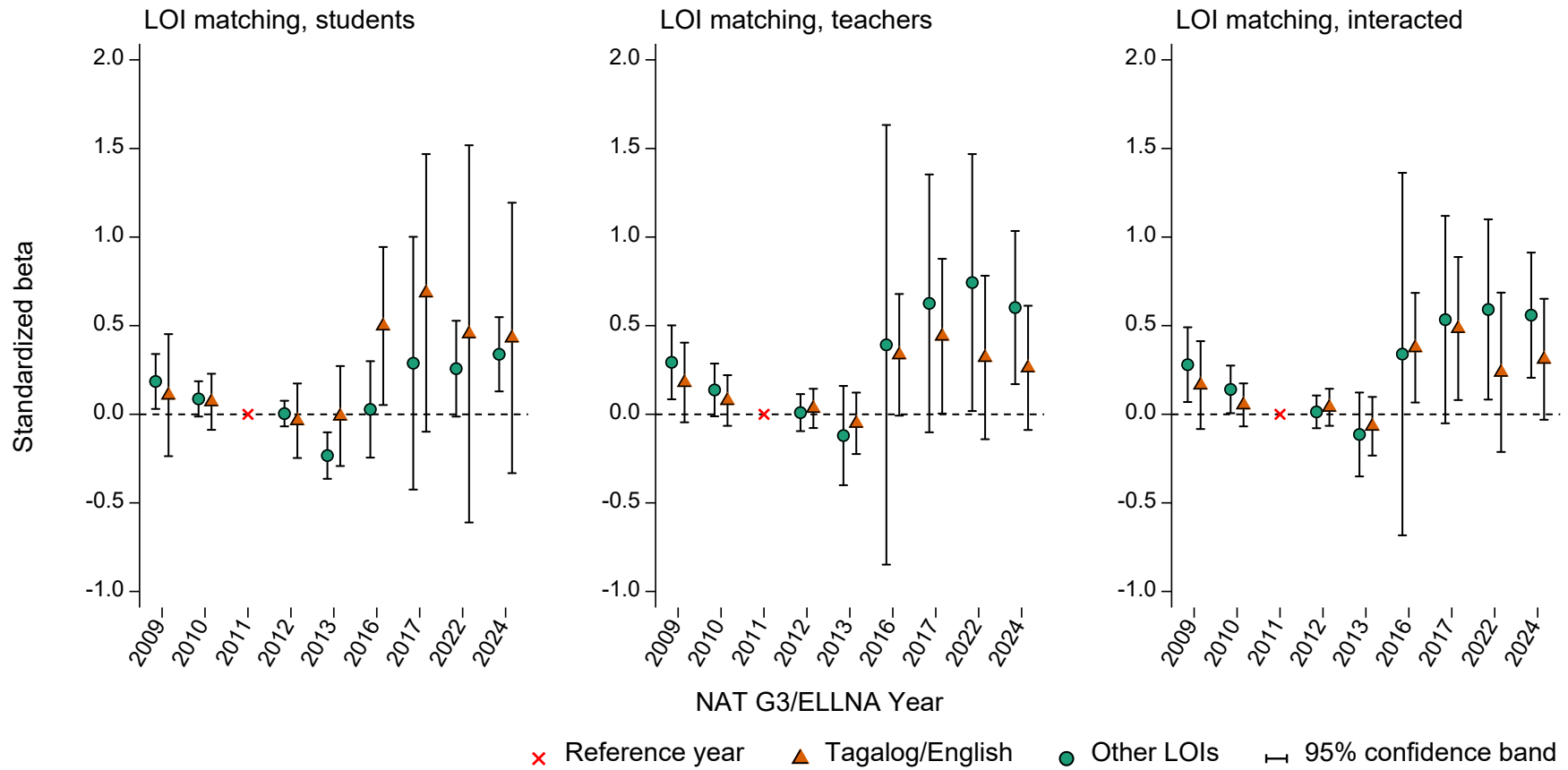


Figure 6. Heterogeneity analysis: MTB-MLE and G3 learning outcomes by instructional language (continued)

B. English

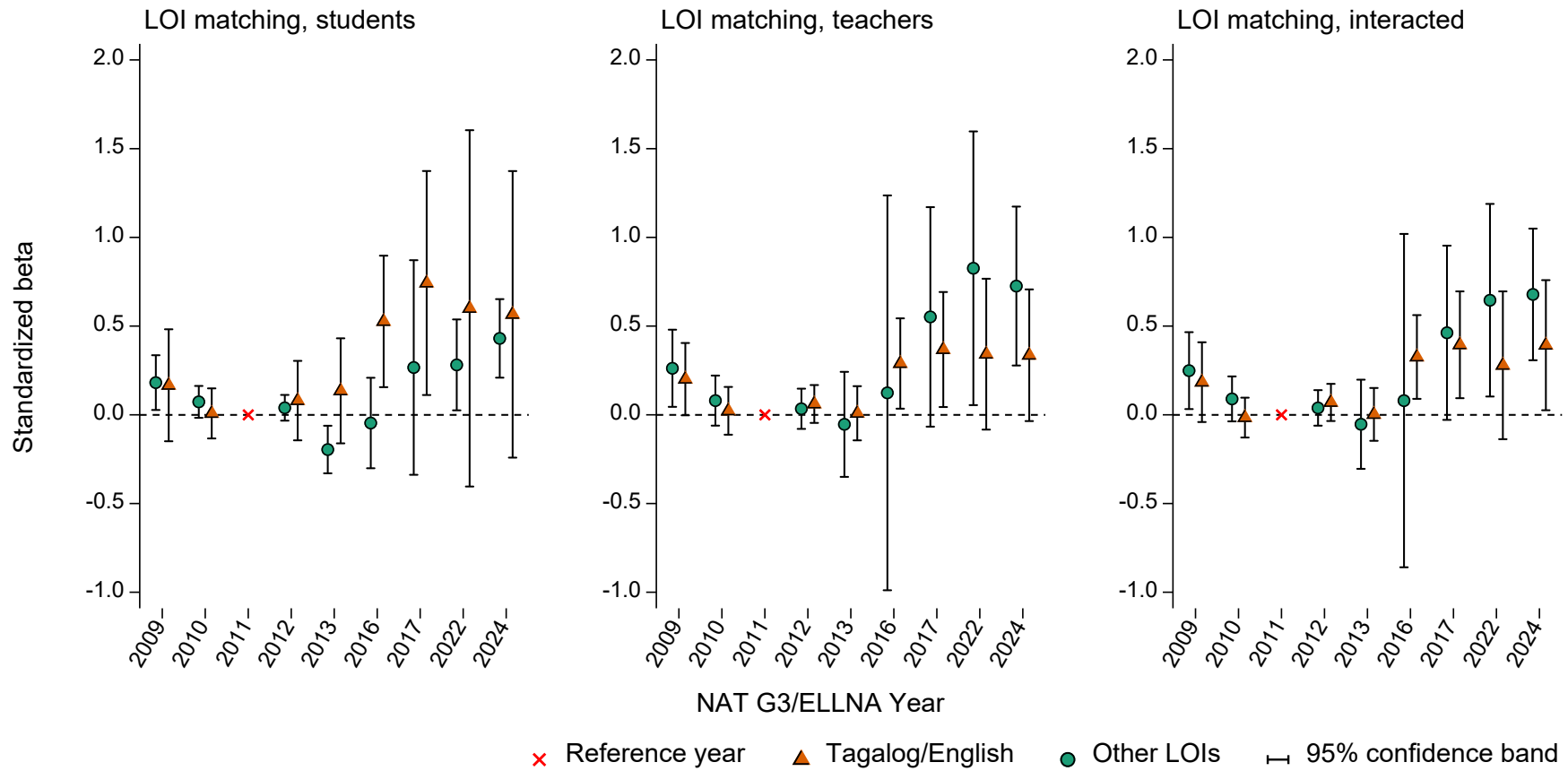
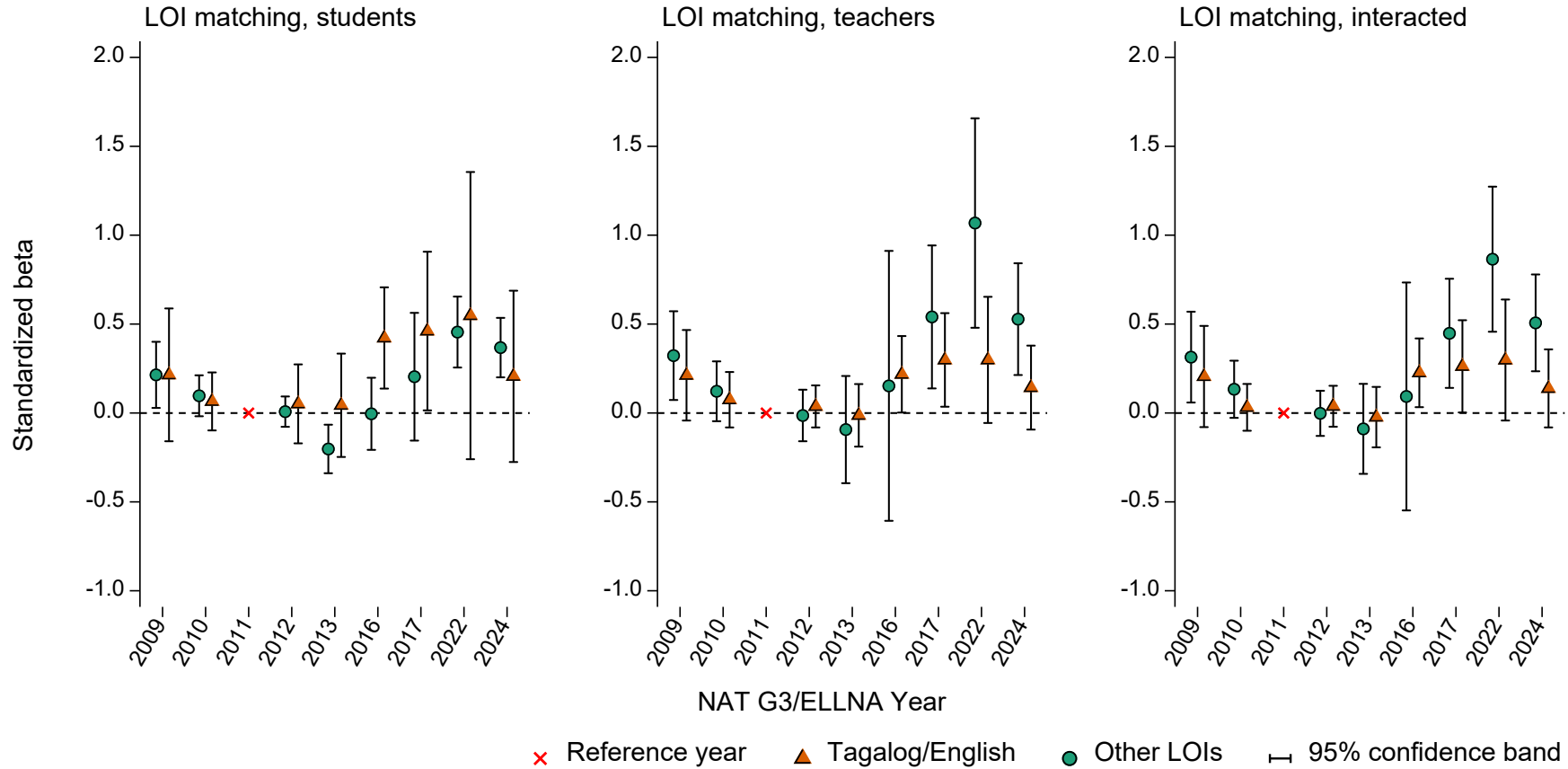


Figure 6. Heterogeneity analysis: MTB-MLE and G3 learning outcomes by instructional language (continued)
C. Mathematics



Notes: Figures display DID and IV-DID coefficients and 95% confidence intervals converted into standardized betas. The dependent variable is Grade 3 mean percentage scores in Filipino, English and Mathematics. IV-DID estimates instrument linguistic matching with shift-share calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school and year fixed effects. Standard errors are clustered at the schools-division level and scaled to include incorporate school-level variation following Moineddin and Urquia (2014).

Matching estimates show poverty rising in better-matched schools. By 2018-2021, better-matched schools experienced poverty increases of 0.26 to 0.42 SD relative to poorly-matched schools (Panel A). This pattern directly contradicts confounding from favorable local development, which would predict poverty declining in better-matched areas. Instead, the results indicate that linguistic matching improved achievement despite economic headwinds.

The positive correlation between matching and poverty implies our test score estimates represent lower bounds on the true causal effect of linguistic alignment. Rising poverty typically depresses achievement; that better-matched schools showed large test score gains while experiencing relative poverty increases suggests the actual linguistic matching effect exceeds our point estimates. Economic conditions worked against our matching results, making the observed achievement gains even more remarkable.

Switching estimates show poverty declining in schools that switched. Schools that adopted local languages experienced systematically lower poverty rates across all periods, with coefficients ranging from -0.10 to -0.24 SD (Panel D). This pattern indicates switching schools were located in economically developing regions.

The negative correlation between switching and poverty implies our switching estimates may be optimistically biased. Declining poverty typically raises achievement (c.f. Coleman, et al., 1966; Filmer and Pritchett, 1999); that switching schools showed null or negative test score effects despite favorable economic conditions suggests the true switching effect may be even more negative than estimated. Economic tailwinds should have helped switchers but didn't, revealing that the institutional act of changing instructional language provided no achievement benefit and possibly imposed costs.

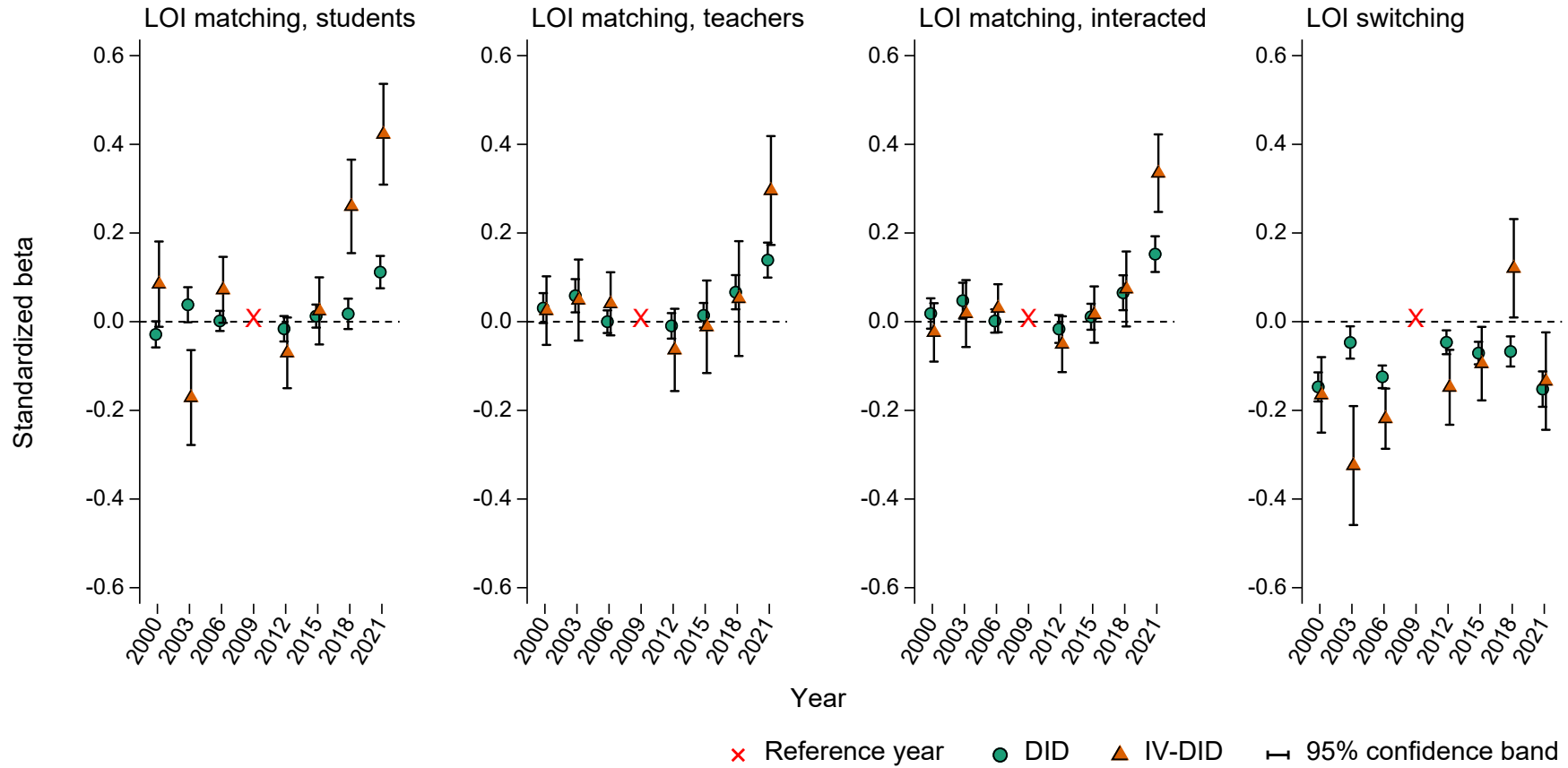
These placebo results validate three key conclusions. First, matching estimates identify linguistic alignment effects independent of regional development. Second, our point estimates are likely conservative lower bounds on true matching effects. Third, switching variation is contaminated by correlated economic trends, justifying our emphasis on matching as the primary identification strategy.

5.3. MTB-MLE and enrollment

Our conceptual framework implies that stronger linguistic matching in the early grades should raise learning outcomes and, through this channel, increase the likelihood that students progress to higher grade levels. The enrollment response therefore provides an additional testable implication of the model that is distinct from the direct achievement effects documented in Section 5.2.

We evaluate this prediction by estimating triple-difference (DDD) models that examine how MTB-MLE implementation influences enrollment patterns across grade levels. The design draws on three sources of variation: (i) cross-sectional differences in treatment intensity across schools, (ii) time-series variation associated with policy introduction, and (iii) differential exposure across cohorts. Because the reform was implemented on a cohort basis, students who had already advanced beyond the affected grades at the time of rollout were never exposed to mother tongue instruction. This feature generates sharp within-school variation in exposure between adjacent cohorts and provides a credible source of identifying variation for the triple-difference specification.

Figure 7. MTB-MLE and municipal/city poverty rates: DID and IV-DID estimates



Notes: Figures display DID and IV-DID coefficients and 95% confidence intervals converted into standardized betas. The outcome variable is small-area municipal and city-level poverty rate estimates by the Philippine Statistics Authority (earlier years by the National Statistical Coordination Board). IV-DID estimates instrument linguistic matching with shift-shares calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school and year fixed effects. Standard errors are clustered at the schools-division level.

Let w_{sgt} denote the year a student cohort entered Grade 1, constructed as $w_{sgt} = t - g + 1$ where t is the calendar year and g is the grade level. We define affected cohorts as those entering Grade 1 in 2010 or later, when the first MTB-MLE kindergarten cohort advanced to Grade 1. The affected cohort indicator is $I(w_{sgt} \geq 2010)$. Cohorts entering Grade 1 before 2010 were never exposed to MTB-MLE and serve as within-school controls.

The empirical specification takes the form

$$\begin{aligned} \ln(N_{sgt}) = & \eta_s + \eta_t + \sum_{h>1} \xi_h I[g = h] + \sum_{h>1} \vartheta_h (c_s \cdot I[g = h]) \\ & + \sum_{\substack{k \neq 2010 \\ k \neq 2010}} \zeta_k (c_s \cdot I[w_{sgt} = k]) + \sum_{h>1} \gamma_h (c_s \cdot I[g = h] \cdot I[w_{sgt} \geq 2010]) \\ & + \epsilon_{sgt}, \end{aligned}$$

where $\ln(N_{sgt})$ is log enrollment at school s , grade level g , in year t . School fixed effects η_s absorb time-invariant differences in school size, while year fixed effects η_t control for system-wide enrollment trends affecting all schools and grade levels equally.

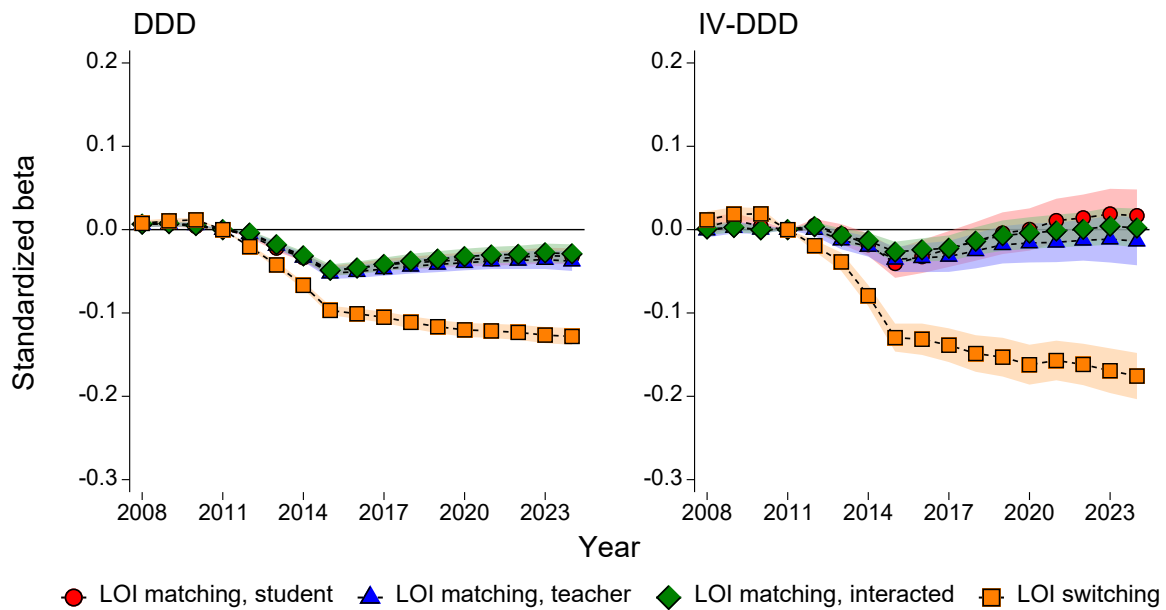
The coefficient ξ_h captures the baseline grade enrollment gradient in untreated schools (where $c_s = 0$), measuring how much smaller enrollment is at grade h relative to Grade 1 due to standard grade-to-grade attrition. The coefficient ϑ_h identifies how treatment affects the enrollment gradient for unaffected cohorts, i.e., those who entered Grade 1 before 2010 and were never exposed to MTB-MLE. The coefficient ζ_k captures treatment effects on Grade 1 enrollment over time, testing whether better linguistic matching altered initial enrollment decisions. The key parameter γ_h recovers the triple-difference estimand, which captures the additional enrollment effect at grade h for affected cohorts in treated schools, relative to unaffected cohorts in the same schools. This isolates the causal effect of MTB-MLE exposure on continuation rates.

Our treatment variables follow the definitions from Section 5.2 for student linguistic match, teacher linguistic match, integrated student-teacher match, and language switching. To facilitate comparison across these measures, which have different scales and variances, we standardize all treatment coefficients by multiplying by the ratio of standard deviation of the respective treatment variable relative to that of the outcome variable. This converts coefficients into standardized betas, representing the enrollment effect of a one standard deviation increase in treatment intensity. Grade level main effects ξ_h are left unstandardized as they describe baseline enrollment patterns in log points rather than treatment effects.

We estimate both DDD specifications and IV-DDD specifications using shift-share instruments to address potential endogeneity from time-varying factors correlated with linguistic homogeneity. The IV approach instruments treatment interactions with shift-share measures of exogenous linguistic concentration, following the strategy described in Section 4.4.

Figure 8 presents standardized beta estimates of MTB-MLE effects on Grade 1 enrollment over time from both DDD and IV-DDD specifications. The patterns reveal significant Grade 1 enrollment responses to MTB-MLE implementation, though the mechanisms driving these responses remain ambiguous.

Figure 8. MTB-MLE and G1 enrollment



Notes: Figure displays DDD and IV-DDD coefficients and 95% confidence intervals from the policy-year interaction terms, converted into standardized betas. The dependent variable is natural logarithm of school enrollment by grade level. The IV-DDD instruments linguistic matching with shift-shares calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school, year, and grade fixed effects. Standard errors are clustered at the schools-division level.

Switching exhibits the most pronounced pattern. Schools that adopted instructional languages other than Filipino (Tagalog) show positive and significant Grade 1 enrollment effects in the pre-reform period (2008-2010), with IV-DDD estimates ranging from 0.01 to 0.02 standardized betas. This reverses sharply post-implementation, turning significantly negative by 2012 and declining steadily to -0.18 standardized betas by 2024. These effects are highly significant throughout, with t-statistics exceeding 5.0 in most years. The persistent negative trend suggests that parents systematically avoided schools that switched to non-Filipino instructional languages, though we cannot rule out confounding from differential fertility trends, migration patterns, or changes in school choice behavior in areas where switching occurred.

Matching measures display more complex temporal dynamics. Student, teacher, and integrated matching show small positive effects in 2009 (significant only for student matching), turn negative around 2012-2013, reach significantly negative values during 2014-2017 (ranging from -0.01 to -0.04 standardized betas depending on the measure), then return toward zero and lose statistical significance by 2018. The mid-period negative effects are particularly puzzling given that these same schools experienced positive achievement gains for affected cohorts (see Section 5.2). Multiple mechanisms could generate declining Grade 1 enrollment in high-match schools during this period, including fertility declines in linguistically homogeneous areas, parental sorting away from schools perceived as lower quality, compositional shifts in formal schooling enrollment, or changes in grade repetition practices that affect cohort entry sizes. We cannot definitively identify which factors drive the observed patterns. The return to statistical insignificance by 2018 suggests these represent temporary disruptions rather than permanent structural changes in enrollment behavior.

The presence of significant Grade 1 enrollment effects across all treatment measures has important implications for our identification strategy. These patterns demonstrate that schools with different linguistic matching experienced differential Grade 1 cohort sizes over time. The negative coefficients for matching measures in 2014-2017 indicate that schools with better post-policy linguistic alignment enrolled systematically smaller Grade 1 cohorts during this period, while the persistent negative effects for switching indicate that schools adopting non-Filipino (Tagalog) languages experienced declining Grade 1 enrollment throughout the post-policy period. A naive difference-in-differences comparison of high-match versus low-match schools would confound these cohort size differences with the causal effects of MTB-MLE on student outcomes.

Our triple-difference specification explicitly addresses this identification threat. The Grade 1 enrollment effects capture differences in cohort size, i.e., how many students enter across the treatment gradient, but our analysis focuses on continuation rates, specifically what fraction of entrants progress to upper grades. The DDD compares enrollment gradients from Grade 1 to Grade 6 for affected versus unaffected cohorts within the same schools. A cohort entering with lower initial enrollment due to the negative Grade 1 effects documented in Figure 8 could still exhibit the same continuation rate, the same proportional decline from Grade 1 to Grade 6, as cohorts entering under different demographic conditions. The triple-difference isolates whether MTB-MLE exposure altered continuation rates, which is independent of initial cohort size. By differencing across cohorts within schools, we net out any confounding variation in absolute enrollment levels and identify the causal effect of mother tongue instruction on the probability that students, conditional on entering Grade 1, persist through elementary education.

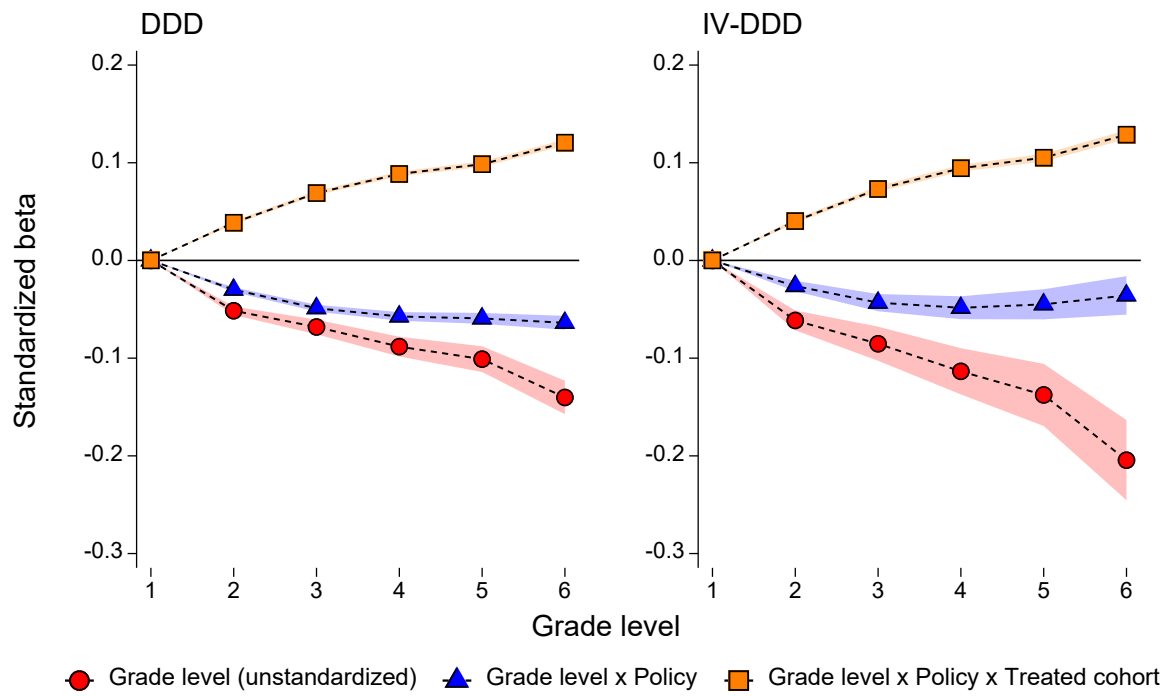
Figure 9 shifts focus to enrollment effects across the full grade distribution, plotting coefficients from the triple-difference specification. We present three sets of estimates: (1) grade level main effects ξ_h , (2) treatment-by-grade interactions ϑ_h for unaffected cohorts, and (3) treatment-by-grade-by-affected cohort interactions γ_h . Each panel displays both DDD and IV-DDD results for one of our four treatment measures.

The grade level main effects decline monotonically from Grade 2 to Grade 6, capturing the baseline enrollment gradient in untreated schools. Grade 6 enrollment is 0.20 log points below Grade 1 in the IV-DDD specification, indicating roughly 18% attrition from the initial cohort size by the final year of elementary education. This gradient is nearly identical across all four treatment comparisons, reflecting the well-documented pattern of grade-to-grade dropout in Philippine basic education.

The treatment-by-grade interactions for unaffected cohorts reveal a striking pattern. These coefficients are uniformly negative and highly statistically significant across all grades and treatment measures. For student matching in the IV-DDD specification, the estimates decline from -0.026 standardized betas (SE = 0.003) at Grade 2 to -0.036 (SE = 0.010) at Grade 6. Teacher matching follows a similar pattern, reaching -0.025 standardized betas (SE = 0.008) by Grade 6. Integrated matching yields -0.031 (SE = 0.007) at Grade 6, while switching shows the steepest gradient at -0.091 (SE = 0.009).

Figure 9. MTB-MLE and enrollment by grade level

A. LOI matching, students



B. LOI matching, teachers

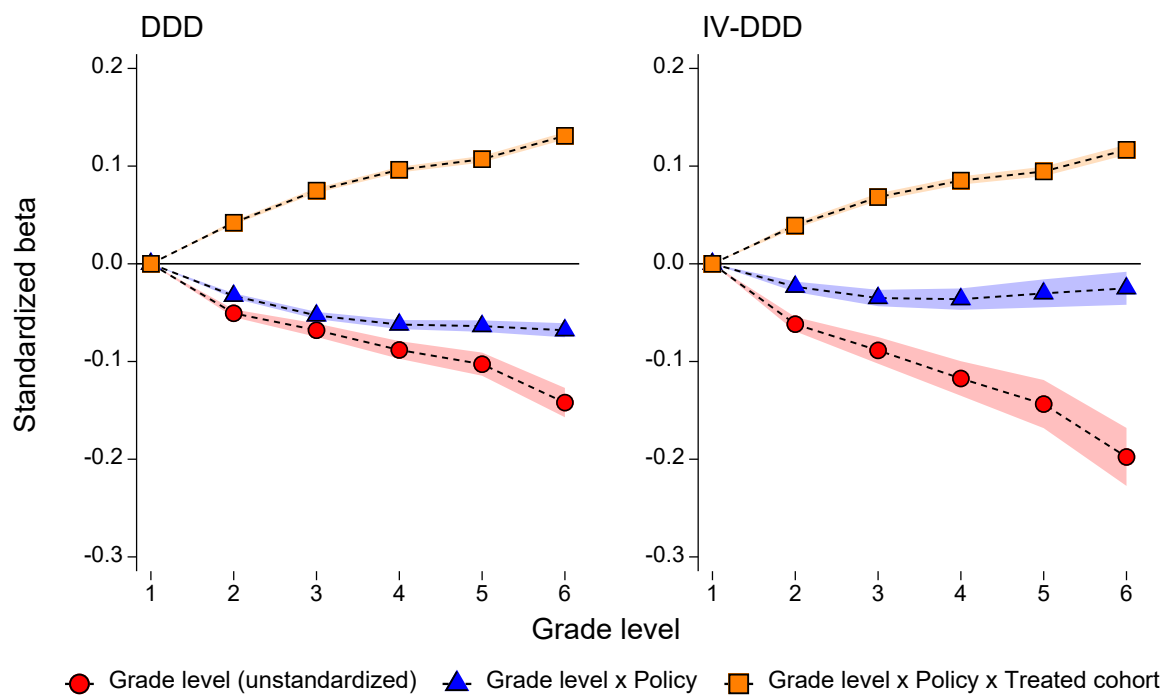
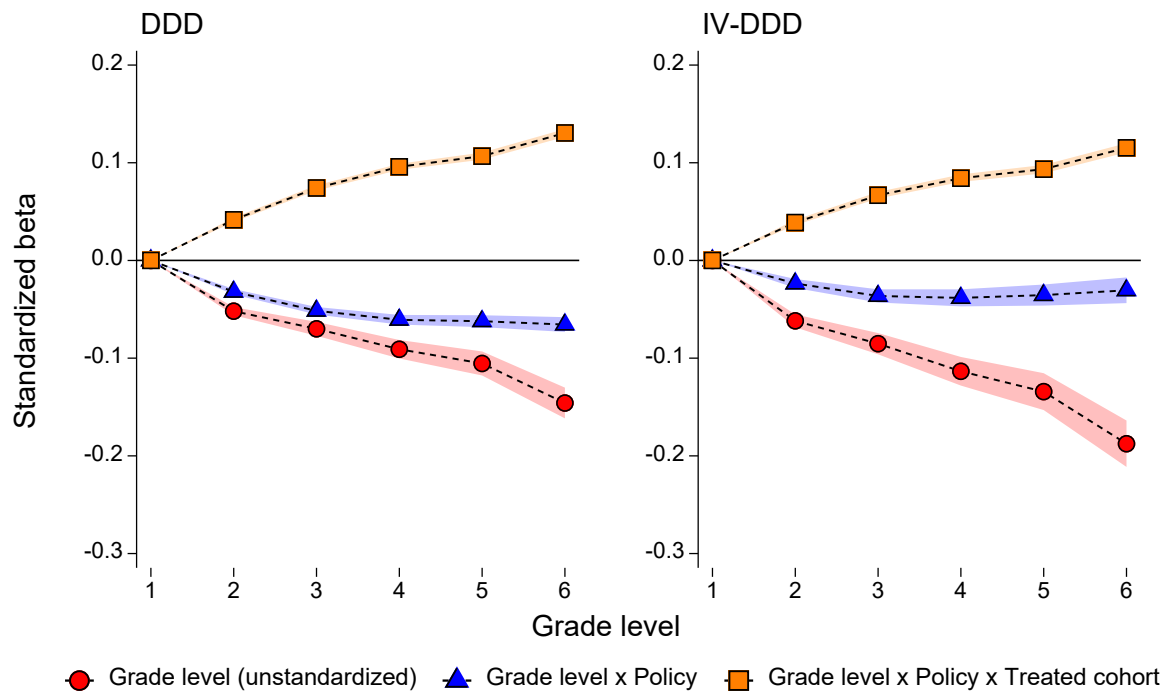
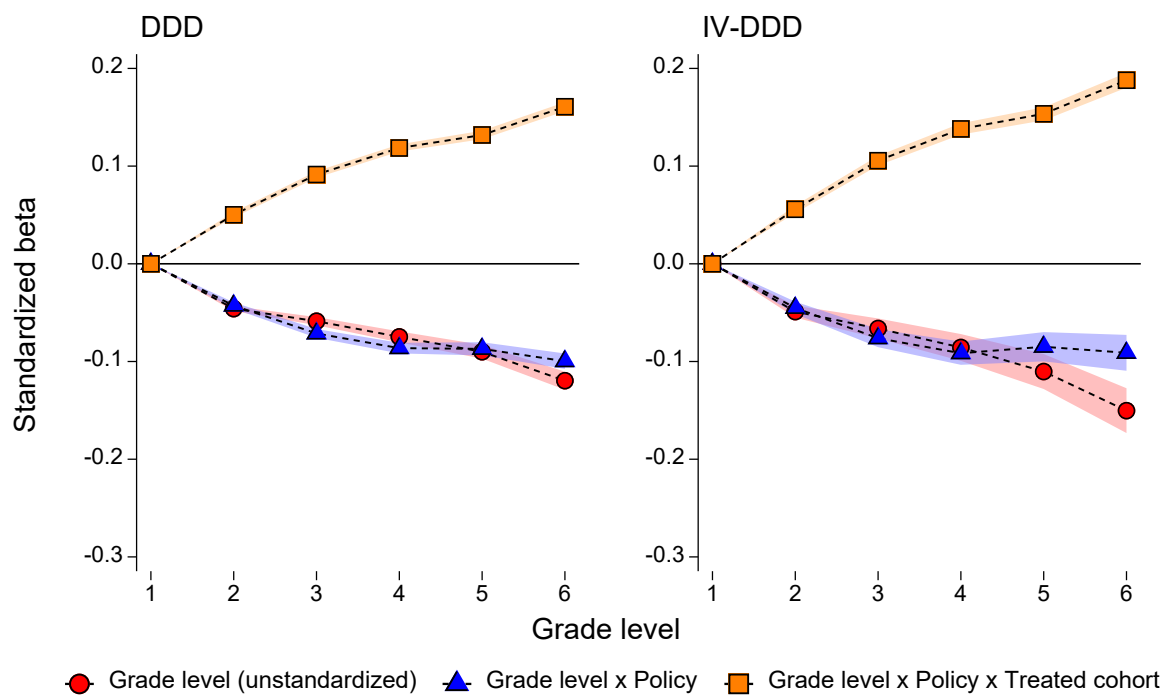


Figure 9. MTB-MLE and enrollment by grade level (continued)

C. LOI matching, interacted



D. LOI switching



Notes: Figure displays DDD and IV-DDD coefficients and 95% confidence intervals, converted into standardized betas. The dependent variable is natural logarithm of school enrollment by grade level. IV-DDD instruments linguistic matching with shift-shares calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school, year, and grade fixed effects. Standard errors are clustered at the schools-division level.

These negative coefficients indicate that, among cohorts never exposed to MTB-MLE, schools with stronger baseline linguistic matching or those that would later adopt a language switch experienced systematically lower continuation rates than comparison schools. For example, a one standard deviation increase in student linguistic match is associated with a 3.6 percentage point reduction in Grade 6 enrollment relative to initial Grade 1 enrollment for cohorts entering Grade 1 before 2010. This effect is additional to the typical grade-to-grade attrition captured through ξ_h . The pre-existing enrollment disadvantage persists after controlling for school and year fixed effects, suggesting that areas where MTB-MLE would later improve linguistic alignment were already experiencing lower continuation rates prior to the reform.

The triple-difference interactions show the causal effect of MTB-MLE exposure. These γ_h coefficients are uniformly positive and increasing across grades, directly offsetting the negative ϑ_h estimates. For student matching in the IV-DDD specification, the affected cohort coefficients rise from 0.040 standardized betas (SE = 0.001) at Grade 2 to 0.13 (SE = 0.002) at Grade 6. Teacher matching reaches 0.117 (SE = 0.003) at Grade 6, while integrated matching attains 0.115 (SE = 0.002). Switching exhibits the largest affected cohort effects, climbing to 0.188 standardized betas (SE = 0.004) by Grade 6. The small standard errors relative to point estimates reflect the large sample size and precise estimation enabled by the panel structure.

The net effect of MTB-MLE for affected cohorts combines the baseline treatment-by-grade coefficient with the additional affected cohort interaction: $\vartheta_h + \gamma_g$. For student matching, this sum equals +0.093 standardized betas at Grade 6, indicating that affected cohorts in treated schools achieved continuation rates approximately 9 percentage points higher than their Grade 1 enrollment baseline, compared to unaffected cohorts in the same schools. For teacher matching, the net effect is +0.092 while integrated matching yields +0.084. Switching produces the largest net positive effect of +0.097 at Grade 6.

These results reveal that MTB-MLE not only reversed the pre-existing enrollment disadvantage experienced by unaffected cohorts in linguistically better matched schools but generated substantial net improvements in continuation rates for affected cohorts. Rather than merely achieving parity with untreated schools, students exposed to mother tongue instruction in schools with better linguistic alignment exhibited systematically higher continuation rates than comparable students in schools with poorer matching. The pattern is consistent across all three matching measures and, notably, reaches its largest magnitude for switching, where schools that adopted non-Filipino languages saw affected cohorts overcome a baseline disadvantage of -0.091 to achieve a net positive continuation effect of +0.097.

The increasing gradient of the γ_h coefficients from Grade 2 through Grade 6 aligns with our theoretical prediction that MTB-MLE effects compound over time. Students exposed to better linguistic matching in Grades 1-3 achieved higher learning outcomes (see Section 5.2), which raised their probability of continuing through each subsequent grade. The enrollment effects emerge at Grade 2 and grow steadily through Grade 6, consistent with MTB-MLE improving early-grade learning that then facilitates progression through the elementary education cycle. The cumulative nature of continuation, wherein students must successfully complete each prior grade to appear in later-grade enrollment counts, amplifies small differences in annual continuation rates into substantial differences in Grade 6 survival.

The results for instructional language switching present a puzzle. Schools that adopted non-Filipino languages experienced large, persistent declines in Grade 1 enrollment post-implementation (see Figure 6), suggesting parental avoidance of unfamiliar instructional

languages. Yet these same schools exhibit the strongest continuation effects, with affected cohorts overcoming a baseline disadvantage of -0.091 to achieve a net positive effect of +0.097 at Grade 6, representing the largest among all treatment measures.

This pattern is difficult to reconcile with the achievement results from Section 5.2, where instructional language switching is documented to have near-zero or negative effects on test scores. If switching neither improved learning outcomes nor attracted more students at entry, why would it generate the strongest continuation effects? Several mechanisms could potentially explain this apparent contradiction.

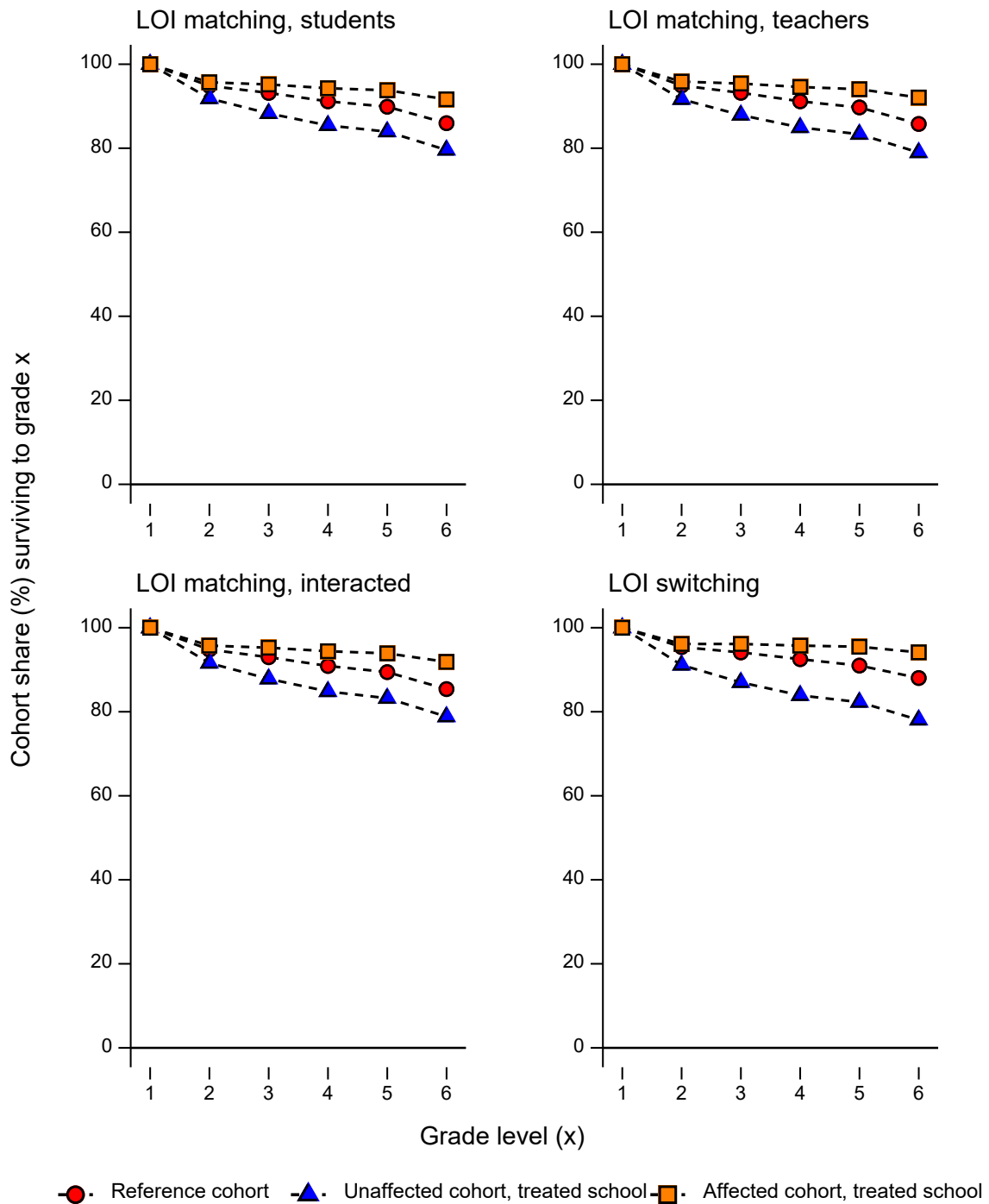
First, compositional selection at the entry margin may have concentrated more motivated or resilient families in switching schools, though this would not explain continuation improvements relative to unaffected cohorts in the same schools facing similar selection. Second, measurement error in our switching indicator may obscure heterogeneity as some schools may have switched to well-matched languages while others switched poorly, with continuation responding to actual alignment rather than the institutional act of switching itself. Third, continuation decisions may respond to factors beyond test score gains, such as teacher effort, parental involvement, or community support mobilized around the language transition, even when these factors do not translate into measurable achievement advantages. We cannot definitively resolve this puzzle with available data, and the switching continuation effects should be interpreted with appropriate caution given their tension with both the achievement and Grade 1 enrollment patterns.

To visualize these dynamics, Figure 10 constructs implied cohort survival functions using the estimated enrollment effects. We plot survival rates for three groups: (i) untreated schools ($c_s = 0$), (ii) unaffected cohorts in treated schools ($c_s > 0$, $w_{sgt} < 2010$), and (iii) affected cohorts in treated schools ($c_s > 0$, $w_{sgt} \geq 2010$). Survival rates are calculated by exponentiating cumulative log enrollment changes relative to Grade 1.

For untreated schools, the survival rate from Grade 1 to Grade 6 is approximately 87%, reflecting the baseline ξ_h coefficients. Unaffected cohorts in high-match schools (evaluated at one standard deviation above the mean treatment intensity) follow a steeper decline, reaching 78% survival by Grade 6 for student matching, 79% for teacher matching, and 78% for integrated matching. Switching shows the sharpest disadvantage for unaffected cohorts at 74% survival, capturing the severe pre-existing enrollment gradient in schools that would later adopt non-Tagalog languages.

Affected cohorts in high-match schools reverse and exceed these baselines. For student matching, survival rates climb to 90% by Grade 6, representing a 12-percentage point gain over unaffected cohorts in the same schools and a 9-percentage point gain over untreated schools. Teacher matching yields 90% survival, integrated matching reaches 89%, and switching attains 91%. These convergence patterns demonstrate that MTB-MLE enabled students in linguistically matched schools to achieve continuation rates substantially above both their own pre-policy baselines and the system-wide average for untreated schools.

Figure 10. Implied cohort survival function



Notes: Figure displays implied cohort survival rates constructed from estimated enrollment effects. Survival rates are calculated by exponentiating cumulative log enrollment changes relative to Grade 1. Three groups plotted: (1) untreated schools, (2) unaffected cohorts in treated schools, and (3) affected cohorts in treated schools. Each panel shows one treatment measure. IV-DDD instruments linguistic matching with shift-shares calculated from predetermined language shares and contemporaneous neighborhood language distributions.

The magnitude of these survival gains is economically meaningful. In Philippine elementary education, progression from Grade 1 to Grade 6 is not automatic as students face academic requirements, financial pressures, and family circumstances that can interrupt schooling at any grade. A 9-12 percentage point increase in six-year survival represents a substantial reduction in dropout risk. For a school with 100 students entering Grade 1, the difference between 78 and 90% survival translates to 12 additional students completing elementary education, a 15% increase in the number of Grade 6 graduates.

These enrollment gains carry important implications for school resource constraints. Even modest increases in continuation rates compound over multiple grades, expanding total enrollment and potentially diluting per-student availability of classrooms, teachers, and instructional materials when aggregate capacity remains fixed. We examine this resource dilution mechanism in Section 5.4, testing whether the enrollment expansion documented here translated into measurable pressure on school inputs.

5.4. MTB-MLE and school resources

Did MTB-MLE strain school resources? Our conceptual framework predicts that improved linguistic matching raises continuation rates, which expands total enrollment and dilutes per-student resources when aggregate capacity is fixed. However, rational policymakers may respond to anticipated enrollment growth by expanding capacity. We therefore examine two resource intensity measures, namely, pupil-class ratios (PCR) and pupil-teacher ratios (PTR), to test whether MTB-MLE strained school resources and whether the education system adjusted through compensatory investments.

To test for resource dilution, we estimate effects on PCR and PTR using the same DID and IV-DID specifications described in Section 5.2. As before, we report standardized beta coefficients to facilitate comparison across treatment dimensions; specifically, we rescale each estimate by the ratio of the standard deviation of the policy variable to that of the outcome. In our discussion, we focus on the IV-DID results. Figure 11 presents standardized betas using the student-matching measure, with results based on teacher matching and integrated matching yielding similarly patterned estimates.

Panel A reports small positive effects on PCR during 2017-2019 based on our IV-DID model: 0.090 SD (SE = 0.049) in 2017, 0.133 SD (SE = 0.050) in 2018, and 0.091 SD (SE = 0.055) in 2019 under student matching. These estimates suggest modest increases in the number of pupils per class in better-matched schools during this period, although the coefficients are mostly only marginally statistically significant. By 2020-2022, the effects attenuate toward zero.

The timing of these transitory increases is consistent with documented public school infrastructure cycles. Total classroom stock expanded from 509.8 thousand in 2011 to 699.2 thousand by 2017, and further to 836.0 thousand by 2023 (Abrigo, et al., 2025). However, classroom construction typically requires about three years from budget appropriation to completion. Although infrastructure appropriations peaked in 2017 at PHP 128.6 billion with a 95.6% utilization rate, much of this spending merely replaced dilapidated facilities amid rising enrollments. Appropriations made during 2014-2017 were realized as completed classrooms in 2017-2020, generating a temporary rise in PCR in the interim and a subsequent return to pre-policy levels as the additional classrooms were completed and became operational.

Figure 11. MTB-MLE and school resources

A. Pupil-class ratio

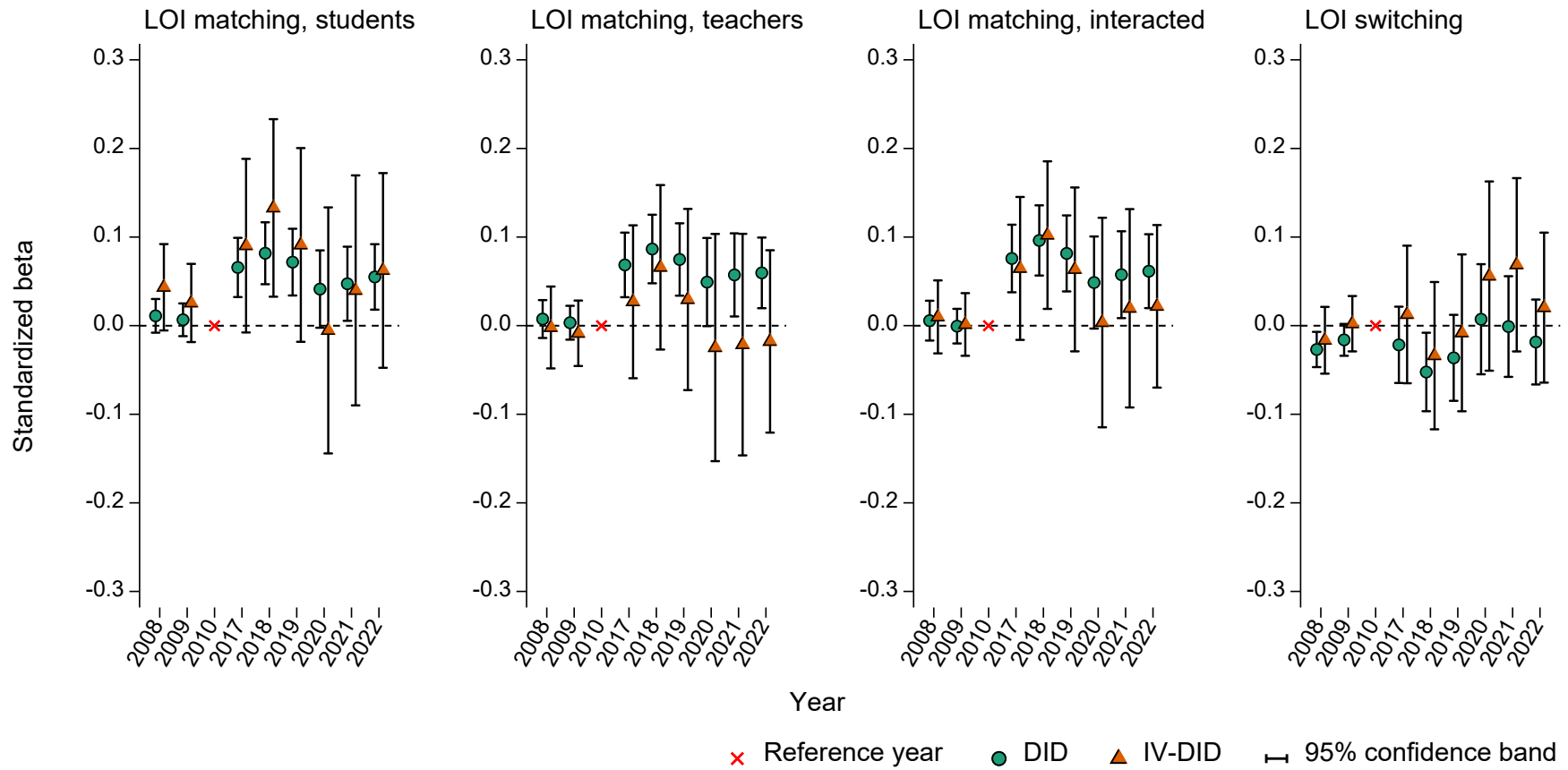
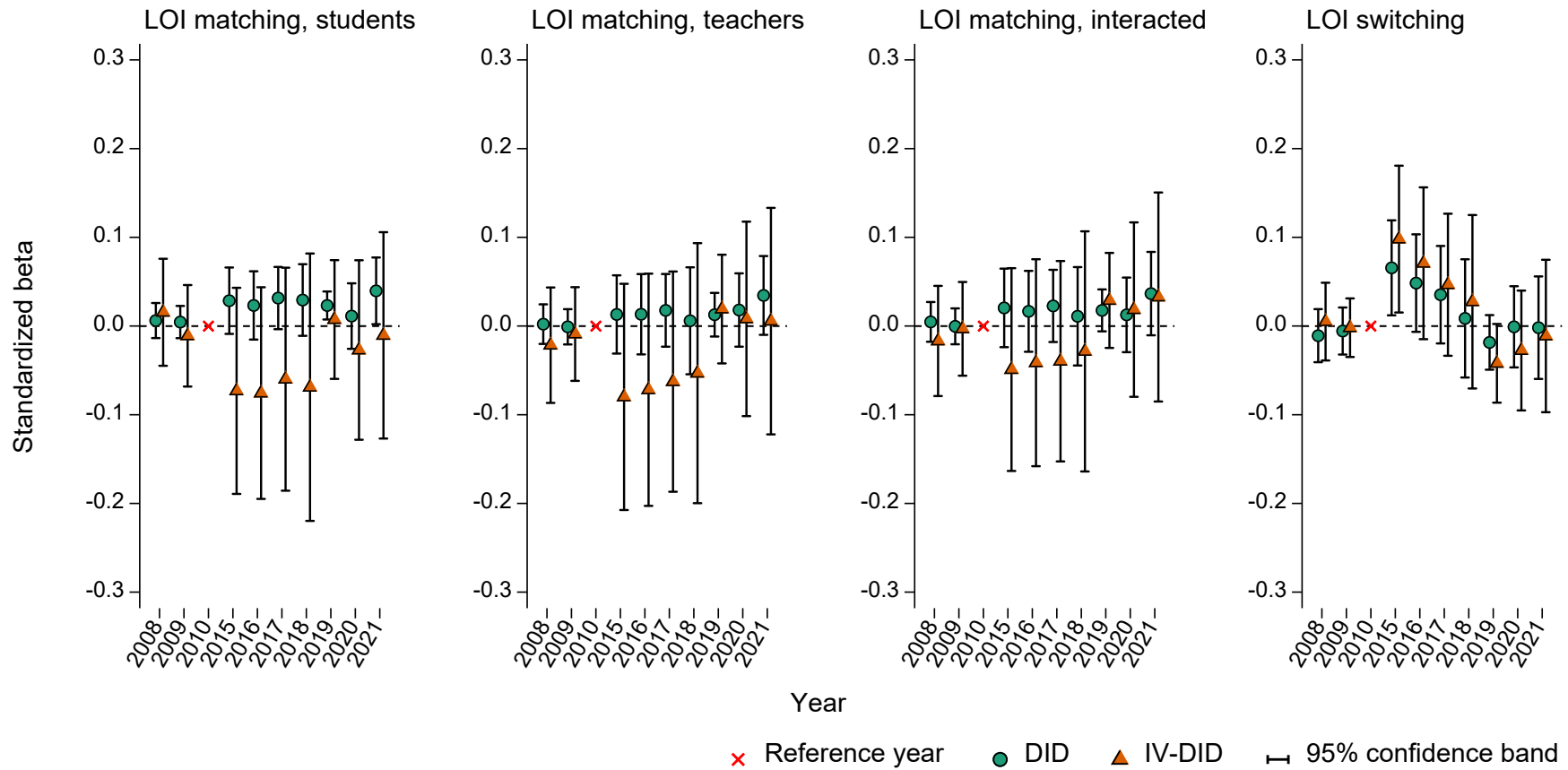


Figure 11. MTB-MLE and school resources (continued)

B. Pupil-teacher ratio



Notes: Figures display DID and IV-DID coefficients and 95% confidence intervals converted into standardized betas. IV-DID estimates instrument linguistic matching with shift-shares calculated from predetermined language shares and contemporaneous neighborhood language distributions. Regressions include school and year fixed effects. Standard errors are clustered at the schools-division level.

Panel B shows the opposite pattern for PTR. Under student-matching, we estimate small negative effects of -0.06 to -0.08 SD, consistent with modest improvements in pupil-teacher ratios in better matched schools but individually insignificant. When we shift to the alternative, but smaller LOI database based on the PIDS/DepEd school surveys, the estimated effects become substantially larger: -0.166 SD (SE = 0.053) in 2015, -0.339 SD (SE = 0.095) in 2016, -0.323 SD (SE = 0.111) in 2017, and -0.246 SD (SE = 0.087) in 2018. These negative coefficients indicate fewer pupils per teacher, i.e., improved PTRs, in better-matched schools.

Rather than contradicting the resource dilution mechanism, these negative coefficients may reveal anticipatory and responsive policy adjustments. National trends are consistent with rapid teacher deployment during this period. Elementary schools added 43,444 teacher positions in 2013-2014 alone, improving the national PTR from 1:35 to 1:32 within a single year (Philippine Statistics Authority, 2014). This expansion reflected DepEd's 128,000-teacher hiring target for 2010-2014 and was reinforced by allocation rules that implicitly favored linguistically matched schools. The reinstatement of teacher localization in 2013 prioritized provincial residents for teaching posts, while MTB-MLE's L1 proficiency requirements naturally sorted teachers toward linguistically aligned communities. Because teacher hiring draws on recurrent operating budgets and allows rapid deployment, its effects manifested much sooner than those for classrooms, which depend on multi-year capital outlays.

In Section 5.2 we document substantial achievement gains from improved linguistic matching. However, in this section, we also find that better-matched schools experienced improvements in teacher ratios during MTB-MLE implementation, which raises the concern that our earlier achievement estimates might partly reflect more favorable resource conditions rather than linguistic alignment. If better-matched schools attracted additional teachers for reasons unrelated to language policy, our estimates could combine two distinct mechanisms. To address this, we quantify how much of the observed achievement gains could plausibly be attributed to improved PTRs. We use the larger PTR effects estimated with the alternative LOI database based on the PIDS and DepEd school surveys in order to set an upper bound on the contribution of improved supply of school resources per student. This approach deliberately maximizes the role assigned to improved staffing ratios and therefore provides the strictest limit on the portion of the effect that can be attributed to linguistic matching.

Krueger (1999) reports that reducing class size by about seven students in the Tennessee STAR experiment produced test score gains of approximately 0.25 SD, which implies scores improvement close to 0.50D for every percent drop in class size in their study. Applying this benchmark to the 0.25 SD reduction in PTRs in our alternative sample yields a predicted achievement gain of roughly 0.15 SD. This amount represents about 20-30% of the 0.49-0.70 SD student-matching effects reported in Section 5.2. The remaining 70-80% reflects the contribution of linguistic matching itself. Better teacher ratios contributed, but they account for only a modest share of the overall achievement gains.

5.5. Learning tax from linguistic mismatch

The reduced-form estimates in Section 5.2 show that improving linguistic match raises test scores by large margins, while the descriptive patterns in Section 5.1 indicate that MTB-MLE implementation substantially reduced linguistic mismatch across most Philippine schools. We now use these results to recover the structural parameter that links linguistic mismatch to student achievement. In particular, we estimate the total mismatch tax rate, which incorporates both the direct effect of mismatch on learning and the indirect behavioral responses it induces.

This parameter captures the proportional learning productivity loss associated with one unit of linguistic mismatch, expressed as a share of the achievement that would have been realized under perfect linguistic alignment.

This structural approach serves three purposes. First, it provides an interpretable magnitude for the learning penalty imposed by linguistic barriers. Second, it enables counterfactual policy evaluation by simulating outcomes under alternative language-of-instruction regimes. Finally, it yields welfare-relevant measures of the gains from improved linguistic matching and the costs of returning to bilingual instruction under the recent policy reversal.

We proceed by first estimating frontier learning outcomes, i.e., the achievement levels students would have had attained if school instruction occurred in their mother tongue with no linguistic barriers. We then calculate the mismatch tax rate as the ratio of the IV-DID coefficient reported in Section 5.2 to this frontier. Using this parameter together with observed mismatch levels under different policy regimes shown in Section 5.1, we then quantify learning losses under bilingual education, current MTB-MLE implementation, and an optimized MTB-MLE system where each school adopts the modal student language from among the 19 authorized languages of instruction.

5.5.1. Estimating frontier learning outcomes

The test score frontier \bar{y} represents the achievement students would have reached if school instructions had been perfectly aligned with their mother tongue, holding constant all other inputs including student effort, school resources, and teaching quality. We cannot observe this directly because most students face some degree of linguistic mismatch even under MTB-MLE. We therefore estimate it using a bottom-up approach that builds on the IV-DID coefficients from Section 5.2.

Our IV-DID estimates for 2024 showed that a one-unit increase in student linguistic match raises test scores by 25.0 MPS points for Filipino, 34.6 for English, and 34.6 for Mathematics. These effects measure the productivity gain from eliminating mismatch entirely. Because observed 2024 scores reflect achievement under the current MTB-MLE implementation with mean mismatch of 0.28 nationally, we can add back the loss from this residual mismatch to recover the frontier as follows:

$$\bar{y} = y^0 + \beta \cdot m_{MT},$$

where y^0 is the observed baseline mean score in 2024, β is the IV-DID coefficient, and m_{MT} is the observed country-wide average linguistic mismatch level of 0.28 from Table 1, Panel B. This calculation assumes the linear approximation from the DID specification holds over the relevant range, which is reasonable given that we are extrapolating only 28 percentage points of additional linguistic match improvement rather than the full zero-to-one range.

For Filipino literacy, the national baseline score is 66.1 and the estimated coefficient is 25.0 MPS points per unit mismatch. Together with the national-level linguistic mismatch of 28%, this translates to a frontier score of 73.3. Similar calculations for English and Mathematics yield frontiers of 73.7 and 60.1 respectively. Regional frontiers vary modestly, ranging from 75.0 in the National Capital Region for Filipino to 56.6 in Visayas for Mathematics. These differences reflect variation in school resources, poverty levels, and other structural determinants of achievement potential that operate independently of linguistic mismatch.

5.5.2. Recovering mismatch tax rate

The total mismatch tax rate τ' is the fundamental parameter governing productivity losses from linguistic barriers. We define it as the proportional reduction in achievement per unit of mismatch:

$$\tau' = \frac{\beta}{\bar{y}}.$$

This formulation treats the IV-DID coefficient as an absolute productivity loss measured in MPS points, following our exposition in Section 4, then expresses it relative to potential achievement to obtain a proportional tax rate. The resulting parameter τ' captures how severely linguistic mismatch disrupts the learning process, combining both the direct comprehension barriers and the indirect effects through reduced student effort documented in Section 3.

Table 3 reports estimated tax rates and their implications for learning losses under alternative policy regimes. For Filipino, the national tax rate is 34.6%. This means that each unit of linguistic mismatch reduces learning productivity by roughly one-third of potential achievement. English shows a substantially higher tax at 46.5%, while Mathematics exhibits the steepest penalty at 56.7% per unit mismatch.

The cross-subject variation contradicts conventional intuition that mathematics is less language-dependent than literacy subjects. Several factors may explain why mathematical learning suffers particularly severe penalties from linguistic mismatch. Elementary mathematics instruction relies heavily on word problems that require verbal comprehension of quantitative relationships. Conceptual understanding of foundational ideas such as place value, fractions, and measurement requires clear verbal explanation from teachers and precise mathematical language from students. Procedural learning depends on students accurately interpreting multi-step instructions, where misunderstanding a single word can derail an entire solution strategy. The ELLNA mathematics assessment itself contains substantial linguistic load through word problems and written explanations, meaning that linguistic barriers affect both learning and demonstration of knowledge.

The high English tax of 46.5% reflects the double burden faced by non-native speakers who must simultaneously acquire English proficiency and learn academic content through a language they are still mastering. Filipino shows the lowest tax at 34.6%, likely because most Philippine languages share Austronesian linguistic roots that create partial mutual intelligibility even across substantial phylogenetic distances. A Cebuano-speaking student learning in Filipino faces comprehension challenges, but cognate vocabulary and parallel grammatical structures reduce the barrier compared to learning in English, which sits outside the Austronesian family entirely.

We estimate a single national tax rate for each subject rather than region-specific parameters because our IV-DID specification pools all schools into one regression. The tax rate represents a technology parameter that captures the fundamental severity of linguistic barriers to learning. This parameter should be roughly constant across regions because the cognitive and linguistic mechanisms that make mismatch costly operate similarly regardless of geographic location. Regional variation in school quality, teacher training, or implementation fidelity might create some heterogeneity in τ' , but we cannot identify such differences with our current empirical

design. Attempting to calculate regional tax rates by dividing the national coefficient by region-specific frontiers would mechanically generate spurious variation, with poorer regions showing artificially inflated tax rates simply because their achievement potential is lower due to factors unrelated to linguistic mismatch.

5.5.3. Quantifying learning losses under alternative policy regimes

The mismatch tax rate enables us to quantify learning productivity losses under different language-of-instruction policies. For any given regime, the learning loss as a share of frontier achievement equals the tax rate multiplied by the regime's average mismatch level:

$$\Delta y = \tau' \cdot m_L.$$

Table 3 presents these calculations for three policy scenarios: the pre-2009 bilingual education system, current MTB-MLE implementation with reported languages of instruction, and an optimized MTB-MLE system where each school adopts the modal student language from among the 19 authorized options.

Under a counterfactual bilingual education, linguistic deprivation would have been substantial if the 2017 student population language distribution is applied. The national school-based linguistic deprivation headcount ratio would reach 0.79, indicating that 79 percent of students attend schools where the primary language of instruction differed from their mother tongue. Applying the estimated tax rate of 34.6 percent for Filipino to this level of mismatch implies a productivity loss of 27.2 percent of frontier achievement. Losses in English and Mathematics were even larger, at 36.6 and 44.7 percent respectively. In absolute terms, students under bilingual education scored roughly 20 to 27 MPS points below their potential.

These aggregate losses masked profound geographic inequality. The National Capital Region would experience minimal penalties because most students already spoke Filipino natively, with losses below 2% of frontier across all subjects. In contrast, Visayas and Mindanao students face learning productivity losses exceeding 35% in Filipino, 47% in English, and 56 to 59% in Mathematics. A return to bilingual education regime would effectively impose a massive implicit tax on learning for students in linguistically diverse regions while leaving Metro Manila students largely unaffected.

MTB-MLE implementation substantially reduced this burden. National mismatch fell from 0.79 to 0.28, cutting learning losses by approximately two-thirds. Filipino productivity loss dropped to 9.8% of frontier, English to 13.2%, and Mathematics to 16.1%. Regional disparities narrowed considerably, with Visayas losses falling to 7.5 to 12.4% and Mindanao losses to 11.3 to 17.8% across subjects. The National Capital Region remained near-perfectly matched with losses below 2%, but the gap between the best- and worst-served regions compressed dramatically.

Considerable optimization potential remains even within the constraint of the current 19 authorized languages. The third scenario in Table 3 simulates an optimized system where each school selects the modal student language from this menu. This reduces national mismatch to 0.14, implying that 86% of students would receive instruction in their mother tongue. Learning losses under this scenario fall to 4.8% for Filipino, 6.4% for English, and 7.8% for Mathematics. The Visayas approaches near-complete optimization with losses of just 3.8 to 6.3%, reflecting the region's linguistic homogeneity and the availability of Cebuano,

Hiligaynon, and Waray among the authorized languages. Mindanao shows larger residual losses of 6.8 to 10.6%, consistent with the region's greater linguistic fragmentation and the incomplete coverage of local languages among the 19 official options.

The gap between current implementation and the optimized scenario indicates that many schools chose suboptimal languages of instruction. This may reflect incomplete information about student language distributions, political economy considerations in heterogeneous schools where no single language commands a clear majority, or implementation constraints related to teacher availability and materials development. The remaining mismatch under the optimized scenario represents a lower bound given the current menu of authorized languages. Further reductions would require expanding this menu to include additional local languages, though doing so may increase coordination costs and materials development expenses.

Table 3. Linguistic mismatch learning tax

	Average test scores (MPS)		Linguistic mismatch tax rate (%)	Implied learning loss (% of frontier)		
	Baseline	Frontier		Bilingual LOI	MTB-MLE LOI	
					Actual	Modal
A. Filipino						
Philippines	66.1	73.3	34.6	27.2	9.8	4.8
NCR	74.3	75.0	...	0.9	0.9	0.9
Rest of Luzon	67.5	75.7	...	18.7	10.7	4.4
Visayas	64.6	69.8	...	35.8	7.5	3.8
Mindanao	62.0	70.0	...	35.8	11.3	6.8
B. English						
Philippines	64.0	73.7	46.5	36.6	13.2	6.4
NCR	72.3	73.3	...	1.3	1.3	1.3
Rest of Luzon	63.4	74.4	...	25.8	14.8	6.0
Visayas	64.4	71.4	...	47.3	9.9	5.0
Mindanao	61.6	72.4	...	46.9	14.8	8.9
C. Mathematics						
Philippines	50.4	60.1	56.7	44.7	16.1	7.8
NCR	57.7	58.7	...	1.6	1.6	1.6
Rest of Luzon	49.8	60.7	...	31.4	18.0	7.3
Visayas	49.6	56.6	...	59.4	12.4	6.3
Mindanao	49.5	60.2	...	56.1	17.8	10.6

Notes: Baseline average test scores show 2024 observed mean percentage scores (MPS) under MTB-MLE. Frontier estimates provide potential MPS under perfect linguistic alignment (zero mismatch). Mismatch tax measures proportional learning productivity loss per unit linguistic distance as share of frontier. Learning loss columns show achievement reductions (% of frontier) under three regimes, calculated as the product of linguistic mismatch tax and linguistic mismatch. National linguistic mismatch tax applied to all regions; regional loss variation reflects mismatch differences.

Our theoretical exposition in Appendix A shows that expanding the language menu can both reduce average linguistic mismatch and improve equity for minority-language students. While schools serving linguistically diverse populations face a tradeoff between minimizing mean mismatch and aligning instruction with student learning potential, implementing multiple languages of instruction with sufficient linguistic coverage can raise overall learning outcomes while protecting disadvantaged students. This insight underscores the potential gains from carefully designed multilingual education policies and sets the stage for the discussion of practical implementation strategies.

6. Policy implications

This study exploits the nationwide rollout of mother tongue-based multilingual education in the Philippines to estimate the learning tax imposed by linguistic mismatch between students' home languages and school instruction. We find that each unit of mismatch reduces learning productivity by 35 percent in Filipino, 47 percent in English, and 57 percent in Mathematics relative to potential achievement under perfect linguistic alignment. With a return to a bilingual education regime that prevailed through 2008, the average Philippine student would have operated at 27 to 45 percent below their learning potential depending on subject, with losses concentrated heavily in Visayas and Mindanao where local languages differ substantially from Filipino and English. MTB-MLE implementation reduced these potential losses by approximately two-thirds, bringing students to within 10 to 16 percent of their frontier achievement. However, more than one-quarter of students still face substantial mismatch under current implementation, and the 2024 policy reversal in Republic Act 12027 risks returning many schools to the high-tax bilingual regime.

These findings contribute to several literatures. The magnitude of productivity losses from linguistic mismatch parallels estimates of allocative inefficiency from factor misallocation across firms, suggesting that language-of-instruction policy represents a first-order determinant of aggregate human capital formation in multilingual societies. The enrollment responses documented in Section 5.3 demonstrate that reducing barriers to learning increases educational attainment through both achievement and continuation channels, consistent with theoretical models where human capital investment responds to returns. The resource dilution mechanism in Section 5.4 illustrates how successful interventions can generate general equilibrium effects that partially offset direct gains, a pattern observed across diverse education programs from conditional cash transfers to school feeding initiatives. Methodologically, the shift-share instrumental variables approach addresses endogeneity in language choice by exploiting predetermined demographic variation, offering a template for evaluating place-based policies where treatment intensity varies with local characteristics.

Linguistic mismatch reflects allocative inefficiency, not just pedagogical challenge. The learning losses we document can be understood as resource misallocation analogous to factor misallocation in production. When students receive instruction in languages poorly matched to their comprehension, learning productivity falls just as output falls when inputs flow to low-productivity uses. The geographic concentration of losses in Visayas and Mindanao while Metro Manila remained unaffected represents spatial misallocation of educational opportunity. MTB-MLE reallocated instructional languages toward better alignment with student populations, raising aggregate productivity. This framing suggests that language policy in multilingual countries operates as a critical determinant of how effectively educational investments translate into human capital.

Successful programs require complementary resource expansion. The temporary resource strain documented in Section 5.4 reflects a general pattern in education policy. Interventions that improve outcomes often increase participation. Conditional cash transfers raise enrollment substantially (Schultz, 2004; Schady, et al., 2009), while school feeding programs improve attendance (Alderman et al., 2012). When these demand expansions occur without parallel supply increases, per-student resources deteriorate (Duflo et al., 2015). MTB-MLE raised continuation rates by 9 to 12 percentage points, creating enrollment pressure that elevated pupil-class ratios during 2017-2019 before infrastructure investment caught up. More generally, interventions operating through demand channels generate equilibrium effects that can dampen or even reverse program impacts when supply constraints bind (Muralidharan and Sundararaman, 2015). The implication extends beyond language policy. Policymakers should anticipate demand responses when implementing successful interventions, front-load infrastructure investments, and monitor resource intensity alongside coverage and outcomes.

Schools should minimize linguistic deprivation using available language menus. Current implementation achieves 28 percent national mismatch when 14 percent would be attainable by selecting the modal student language from the 19 authorized options. The conceptual framework in Section 3 provides formal welfare-theoretic guidance. When mutual intelligibility decays rapidly with distance, the optimal choice minimizes the headcount of mismatched students by selecting the modal language. If intelligibility persists across moderate distances, schools may select a closely related regional lingua franca offering implementation advantages without substantially increasing the learning tax. The Foster-Greer-Thorbecke linguistic deprivation indices from Section 5.1 provide a practical metric for evaluating language choices and identifying where menu expansion would help most.

Implementation strategies should vary with linguistic composition. Linguistically homogeneous schools where one language represents more than 70 percent of students should implement monolingual instruction in that local language through Grade 3. Student language data from student profiles collected at enrollment enable evidence-based choices. Moderately heterogeneous schools should adopt the modal language if authorized, or a regional lingua franca linguistically closer to most students. Where feasible, separate classes by mother tongue within grades implement the multiple-language approach that Pareto-dominates single-language instruction. Highly heterogeneous schools with no dominant language should select the regional lingua franca closest to the plurality, supplement with peer tutoring, and accept some residual mismatch as unavoidable without prohibitive coordination costs.

Choosing one language creates winners and losers in a multilingual school setting. In schools implementing monolingual or modal language instruction, the magnitude of required support scales with both the degree of linguistic distance and the size of the excluded linguistic minority group. Where the dominant language covers a substantial share of students, targeted interventions for the remaining students, such as structured peer tutoring, supplementary instruction in transitional periods, modified pacing, and mother tongue-based materials, can improve learning outcomes without sacrificing equity goals. Schools should explore opening separate language-of-instruction classes when a linguistic minority group reaches a viable class size, since compensatory learning interventions for large minority shares may become prohibitively expensive as the population of excluded minority language speakers increase.

Teacher language alignment requires systematic data and targeted deployment. Effective MTB-MLE requires teachers who speak the instructional language natively or fluently. Teacher

language data should be collected systematically. Localization policies help by channeling teachers to their home provinces where they likely speak local languages. Indigenous Peoples communities face particular challenges given limited local teacher supply. Scholarship programs targeting indigenous youth for teaching careers and language immersion training for non-indigenous teachers assigned to these communities could address this gap. Teacher training curricula should include modules on adapting instruction to local linguistic contexts, drawing on the broader literature on culturally responsive pedagogy.

Switching languages does not guarantee improved matching. We showed in Section 5.2 that schools switching from Filipino or English to other Philippine languages experienced near-zero or negative achievement effects despite improved continuation rates. Our results suggest that what matters is actual alignment between students' languages and instruction, not the institutional act of changing the language of instruction. Schools choosing languages for political reasons, teacher preferences, or historical practice rather than student linguistic composition may worsen outcomes. Sub-optimal implementation with the wrong language, untrained teachers, or inadequate materials may produce worse results than maintaining bilingual instruction.

Republic Act 12027 creates space for optimization if criteria focus on linguistic ecology. The law's conditional approach allows local adaptation rather than mandating uniform implementation. This flexibility could improve overall student welfare if feasibility criteria emphasize linguistic composition and implementation capacity. Schools with high homogeneity, authorized language availability, and teacher capacity should implement monolingual local-language instruction. Schools with extreme diversity, unauthorized modal languages, or severe teacher shortages may reasonably use regional lingua francas or Filipino. The risk is that schools revert to bilingual instruction by default, losing MTB-MLE gains. Clear guidance on when conditions favor local-language instruction would help schools make welfare-improving choices.

Expanding linguistic coverage beyond 19 languages could further reduce mismatch. The Appendix theorem demonstrates that expanding linguistic coverage by including more instructional languages Pareto-dominate restricted menus by lowering mismatch variance and raising average outcomes. Current residual mismatch of 14 percent under optimized use of the 19-language menu concentrates in linguistically fragmented areas of Mindanao and indigenous communities. However, it must be emphasized that adding languages requires balancing students served against materials development costs. Prioritizing languages that reduce school-based linguistic deprivation in currently worst-served areas could maximize welfare gains per implementation peso.

Critical research gaps limit precision of policy guidance. Mutual intelligibility across Philippine language pairs requires empirical validation beyond phylogenetic distances. Field studies measuring actual comprehension across language pairs would calibrate the persistence parameter and improve guidance on when regional lingua francas substitute effectively for strict modal languages. Implementation cost data would enable rigorous cost-benefit analysis of menu expansion. Long-run tracking of MTB-MLE cohorts into secondary school would test whether gains from early grade levels persist, compound through continuation channels as the model predicts, or fade.

MTB-MLE generates substantial welfare gains when schools select appropriate languages, align teacher deployment, and receive adequate resources. The two-thirds reduction in learning

losses demonstrates that language policy represents a high-return investment in human capital formation. Gains concentrate in already-disadvantaged regions, suggesting that effective implementation promotes both efficiency and equity. Reverting broadly to bilingual instruction would impose costs equivalent to multiple years of learning, falling disproportionately on students in Visayas and Mindanao. Conditional implementation can optimize the efficiency-equity tradeoff if criteria focus on observable linguistic composition and implementation capacity rather than political economy considerations. Getting language policy right requires accurate data, appropriate choice rules, teacher alignment, and resource planning, but the potential returns justify these investments.

References

- Abrigo, M.R.M., T.D.S. Perez, N.A.M. Quejada, E.D.J.C. Daga, K.M. Estopace, E.E. Lingatong, C.M.L. Relos, Relos, J.L. Paquibot, D.S.A. Gavan (2025). Low fertility, ageing buildings, and school congestion in the Philippines: Tailwinds, headwinds and some policy options. PIDS Discussion Paper Series No. 2025-10. Quezon City: Philippine Institute for Development Studies.
- Alderman, H., D.O. Gilligan, and K. Lehrer (2012). The impact of food for education programs on school participation in northern Uganda. *Economic Development and Cultural Change*, 61(1), 187-218.
- August, D., and T. Shanahan (Eds.) (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Avvisati, F., and P. Givord (2023). The learning gain over one school year among 15-year-olds: An international comparison based on PISA. *Labour Economics*, 84, 102365.
- Bacak, M. (2014). Computing medians and means in Hadamard spaces. *SIAM Journal on Optimization*, 24(3), 1542-1566.
- Bamgbose, A. (1991). Language and the Nation. *The language question in Sub-Saharan Africa*. England: Englewood University Press.
- Benson, C. (2004). The importance of mother tongue-based schooling for educational quality. Background paper for Education for All Global Monitoring Report 2005. Paris: UNESCO.
- Blust, R. (1991). The Greater Central Philippines hypothesis. *Oceanic Linguistics*, 30(2), 73-129.
- Blust, R. (2009). *The Austronesian languages*. Canberra: Pacific Linguistics.
- Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6), 3253-3285.

- Bowne, J.B., K.A. Magnuson, H.S. Schindler, G.J. Duncan, and H. Yoshikawa (2017). A meta-analysis of class sizes and ratios in early childhood education programs: Are thresholds of quality associated with greater impacts on cognitive, achievement, and socioemotional outcomes? *Educational Evaluation and Policy Analysis*, 39(3), 407-428.
- Brock-Utne, B. (2005). Language-in-education policies and practices in Africa with a special focus on Tanzania and South Africa – Insights from research in progress. In A.M.Y. Lin and P.W. Martin (Eds.), *Decolonisation, globalization: Language-in-education policy and practice*. Clevedon, UK: Multilingual Matters.
- Burton, L. A. (2013). Mother tongue-based multilingual education in the Philippines: Studying top-down policy implementation from the bottom up. Doctoral dissertation. Minneapolis, MN: University of Minnesota.
- Callaway, B., and P.H.C. Sant'Anna (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200-230.
- Cena, D.A. (1958). The local dialects as the medium of instruction in the primary grades. *Philippine Studies*, 6(1), 115-120.
- Coleman, J.S., E.Q. Campbell, C.J. Hobson, J. McPartland, A.M. Mood, F.D. Weinfeld, F.D., and R.L. York (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education; U.S. Government Printing Office.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222-251.
- Cunha, F., and J.J. Heckman (2007). The technology of skill formation. *American Economic Review*, 97(2), 31-47.
- del Pilar, G.E.H. (2023). Isang rebista ng mga empirikal na pag-aaral sa pinagkumparang bisa ng Ingles at Wikang Pilipino bilang midyum ng pagtuturo. *Mabini Review*, 12, 1-30.
- Dekker, D., and C. Young (2005). Bridging the gap: The development of appropriate educational strategies for minority language communities in the Philippines. *Current Issues in Language Planning*, 6(2), 182-199.
- Duflo, E., P. Dupas, and M. Kremer (2015). School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics*, 123, 92-110.
- Dumatog, R.C., and D.E. Dekker (2003). First language education in Lubuagan, Northern Philippines. Paper presented at the Conference on Language Development, Language Revitalization and Multilingual Education in Minority Communities in Asia held on 6-8 November 2003 in Bangkok, Thailand. Accessed on 15 June 2025 from SIL International website.
- Eberhard, D.M., G.F. Simons, and C.D. Fennig (Eds.) (2024). *Ethnologue: Languages of the world*, 27th ed. Dallas, TX: SIL International.

- Evans, D.K., and F. Yuan, F. (2019). Equivalent years of schooling: A metric to communicate learning gains in concrete terms. Policy Research Working Paper No. 8752. Washington, D.C.: The World Bank.
- Evans, D.K., and Yuan, F. (2022). How big are effect sizes in international education studies? *Educational Evaluation and Policy Analysis*, 44(3), 532-540.
- Fearon, J.D. (2003). Ethnic and cultural diversity by country. *Journal of Economic Growth*, 8(2), 195-222.
- Foster, J., J. Greer, and E. Thorbecke (1984). A class of decomposable poverty measures. *Econometrica*, 52(3), 761-766.
- Foster, J., J. Greer, and E. Thorbecke (2010). The Foster-Greer-Thorbecke poverty measures: 25 years later. *Journal of Economic Inequality*, 8, 491-524.
- Gaylo, P.J.B. (2020). Implementing mother tongue-based multilingual education (MTB-MLE): Outcomes and challenges. *International Social Science*, 2(1), 148-183.
- Glewwe, P., and K. Muralidharan (2016). Improving education outcomes in developing countries: Evidence, knowledge gaps, and policy implications. In E.A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education*, Vol. 5. Amsterdam: Elsevier.
- Glewwe, P., and M. Kremer (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education*, 2, 945-1017.
- Gonzalez, A. (2003). Language planning in multilingual countries: The case of the Philippines. Paper presented at the Conference on Language Development, Language Revitalization and Multilingual Education in Minority Communities in Asia held on 6-8 November 2003 in Bangkok, Thailand. Accessed on 15 June 2025 from SIL International website.
- Gooskens, C., V.J. van Heuven, J. Golubović, A. Schuppert, F. Swarte, and S. Voigt (2017). Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism*, 15(2), 169-193
- Hakimi, S. L. (1964). Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research*, 12(3), 450-459.
- Hakimi, S. L. (1965). Optimum distribution of switching centers in a communication network and some related graph theoretic theorems. *Operations Research*, 13(3), 462-475.
- Hsieh, C.-T., & Klenow, P.J. (2009). Misallocation and manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4), 1403-1448.
- Igarash, T., S. Maulana, and D. Suryadaram (2024). Mother-tongue based education in a diverse society and the acquisition of foundational skills: Evidence from the Philippines. *Labour Economics*, 91, 102641.

- Igarashi, T., S. Maulana, and D. Suryadarma, D. (2024). Mother-tongue based education in a diverse society and the acquisition of foundational skills: Evidence from the Philippines. *Labour Economics*, 91, 102641.
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5), 509-541.
- Karcher, H. (1977). Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5), 509-541.
- Kariv, O., and S.L. Hakimi (1979). An algorithmic approach to network location problems. II: The p-medians. *SIAM Journal on Applied Mathematics*, 37(3), 539-560.
- Kosonen, K., and C. Young (Eds.) (2009). *Mother tongue as bridge language of instruction: Policies and experiences in Southeast Asia*. Bangkok, Thailand: SEAMEO.
- Krueger, A.B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, 114(2), 497-532.
- Lloyd, T., and D. Yang (2025). The long shadow of early education: Evidence from a natural experiment in the Philippines. NBER Working Paper No. 33600. Cambridge, MA: National Bureau of Economic Research.
- Lloyd, T., and D. Yang (2025). The long shadow of early education: Evidence from a natural experiment in the Philippines. NBER Working Paper 33600. Cambridge, MA: National Bureau of Economic Research.
- Love, R. F., J.G. Morris, J. G., and G.O. Wesolowsky (1988). *Facilities location: Models & methods*. New York: North-Holland.
- Metila, R.A., L.A.S. Pradilla, and A.B. Williams (2016). The challenge of implementing mother tongue education in linguistically diverse contexts: The case of the Philippines. *The Asia-Pacific Education Researcher*, 25(5), 781-789.
- Metila, R.A., T.J. D'Agostino, and E. Iwasaki (2025). Why the Philippines reversed its mother-tongue instruction policy: Tracing erosion of political priority for first language-based education in the Philippines. *International Journal of Educational Development*, 118, 103408.
- Migge, B. and I. Léglise (2007). Language and colonialism: Applied linguistics in the context of creole communities. In M. Hellinger and A. Pauwels (Eds.), *Language and communication: Diversity and Change. Handbook of Applied Linguistics*, 297-338.
- Moineddin, R., and M.L. Urquia (2014). Regression analysis of aggregate continuous data. *Epidemiology*, 25(6), 929-930.
- Monje, J.D., A.C. Jr. Orbeta, K.A. Francisco, and E.M. Capones (2021). 'Starting where the children are': *Mother-tongue based multilingual education program implementation*. Research Paper Series No. 2021-02. Quezon City, Philippines: Philippine Institute for Development Studies.

- Moulin, H. (1988). *Axioms of Cooperative Decision Making*. Cambridge, UK: Cambridge University Press.
- Muralidharan, K., A. Singh, and A.J. Ganimian, A.J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4), 1426-1460.
- Muralidharan, K., and V. Sundararaman (2015). The aggregate effect of school choice: Evidence from a two-stage experiment in India. *Quarterly Journal of Economics*, 130(3), 1011-1066.
- Nolasco, R.M. (2008). The prospects of multilingual education and literacy in the Philippines. Paper presented at the International Conference on Language Development, Language Revitalization and Multilingual Education in Ethnolinguistic Communities held on 1-3 October 2008 in Bangkok, Thailand. Accessed on 15 June 2025 from the SEAMEO website.
- Orbeta, A.C. Jr., and V.B. Paqueo (2025). Synthesis of Philippine education experience: Legacy issues, emerging challenges, and ways forward. In A.C. Orbeta Jr., M.R.M. Abrigo, and V.B. Paqueo (Eds.), *Making Philippine Education Reforms and Interventions More Transformative*. Quezon City: Philippine Institute for Development Studies.
- Orbeta, A.C. Jr., M.R.M. Abrigo, and V.B. Paqueo (Eds.) (2025). *Making Philippine Education Reforms and Interventions More Transformative*. Quezon City: Philippine Institute for Development Studies.
- Philippine Statistics Authority (2014). 2013-2014 Philippine Statistical Yearbook. Manila: Philippine Statistics Authority.
- Pritchett, L., and A. Beatty (2015). Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development*, 40, 276-288.
- Reid, L.A. (1971). Philippine minor languages: Word lists and phonologies. Oceanic Linguistics Special Publications No. 8. Honolulu: University of Hawaii Press.
- Reid, L.A. (1982). The demise of Proto-Philippines. In A. Halim, L. Carrington, & S.A. Wurm (Eds.), *Papers from the Third International Conference on Austronesian Linguistics*. Canberra: Pacific Linguistics.
- Restuccia, D., and R. Rogerson (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic Dynamics*, 11(4), 707-720.
- Schady, N., A. Fiszbein, F.H.G. Ferreira, N. Keleher, M. Grosh, P. Olinto, and E. Skoufias (2009). Conditional cash transfers: Reducing present and future poverty. World Bank Policy Research Report. Washington, DC: The World Bank.
- Schultz, T.P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, 74(1), 199-250.

- Schweibert, S. (2014). Computing regression statistics from grouped data. *Journal of Economic and Social Measurement*, 39(4), 283-303.
- Sercombe, P. and R. Tupas, eds. (2014). *Language, education and nation-building: Assimilation and shift in Southeast Asia*. London: Palgrave Macmillan.
- Snow, C.E. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.
- Trudell, B. (2016). *The impact of language policy and practice on children's learning: Evidence from Eastern and Southern Africa*. Nairobi, Kenya: UNICEF Eastern and Southern Africa Regional Office, Basic Education and Gender Equality Section.
- United Nations Educational, Scientific and Cultural Organization [UNESCO] (2016). If you don't understand, how can you learn? Global Education Monitoring Report Policy Paper No. 24. Paris: UNESCO.
- United Nations Educational, Scientific and Cultural Organization [UNESCO] (2003). *Education in a multilingual world: UNESCO Education Position Paper*. Paris: UNESCO.
- Walter, S.L., and D.E. Dekker (2011). Mother tongue instruction in Lubuagan: A case study from the Philippines. *International Review of Education*, 57(5), 667-683.
- World Bank (2018). *World Development Report 2018: Learning to realize education's promise*. Washington, DC: The World Bank.
- Ye, T., A. Ertefaie, J. Flory, S. Hennessy, and D.S. Small (2023). Instrumented difference-in-differences. *Biometrics*, 79(2), 569-581.
- Zorc, R.D.P. (1977). *The Bisayan dialects of the Philippines: Subgrouping and reconstruction*. Canberra: Pacific Linguistics.

Appendix A. Linguistic distance and aggregate learning

Proposition (Average-distance sufficiency). *In environments where linguistic mismatch enters learning multiplicatively and smoothly, maximizing mean learning is, to a first-order approximation, equivalent to minimizing mean linguistic distance. The optimal language-of-instruction problem reduces to an average-distance minimization problem. That is,*

$$\max_L \left\{ \frac{1}{N^1} \sum_i y_i^1 \right\} \cong \min_L \sum_k \left\{ \pi_k \left[\frac{|\ell_k - L|}{\max_j \{|\ell_j - L|\}} \right]^\rho \cdot I\{\ell_j \neq L\} \right\}.$$

Linguistic mismatch acts like a proportional wedge on the productivity of effort and instructional inputs, such that changes in mismatch affect average learning primarily through their effect on the mean size of this wedge. When the distortion varies smoothly with linguistic distance, the aggregation problem becomes effectively one-dimensional. The policymaker's objective depends on the average attenuation of learning inputs, not on the full distribution of mismatch. The practical implication is that, to a first-order approximation, selecting the language of instruction is equivalent to choosing the option that minimizes average linguistic distance across students, making mean distance a sufficient statistic for evaluating alternative language configurations.

Proof. Substitute y_i^1 by the optimal y_i^{1*} in the school's optimization problem

$$\max_L \left\{ \frac{1}{N^1} \sum_i y_i^1 \right\} = \max_L \frac{1}{N^1} \sum_i \left\{ \frac{\kappa}{1-\alpha} \cdot \left[\phi_1(1-\alpha) \frac{(1-\tau m_i)}{\kappa} r_i^\alpha \exp(\varepsilon_i) \right]^{\frac{2}{1+\alpha}} \right\}$$

where m_i is linguistic distance, a function of L (see main paper). Since students are in same school $r_i = r$, then the optimization simplifies to

$$\max_L \frac{1}{N_t^1} \sum_i \left\{ \exp(\varepsilon_{it}) \cdot (1-\tau m_{it})^{\frac{2}{1+\alpha}} \right\}$$

by factoring out the constants insides the summand and noting that the solution does not change by excluding this common constant in the optimization.

For any $\gamma, \tau > 0$, the function $(1-\tau m_{it})^\gamma$ is strictly increasing in m_{it} . By the monotonicity of this transformation, the solution to the above simplified maximization problem is the same as the solution to the minimization problem below

$$\min_L \frac{1}{N_t^1} \sum_i \{ \exp(\varepsilon_{it}) \cdot m_{it} \}$$

By the Law of Large Numbers (LLN)

$$\min_L \frac{1}{N_t^1} \sum_i \{ \exp(\varepsilon_{it}) \cdot m_{it} \} \rightarrow \min_L E[\exp(\varepsilon_{it}) \cdot m_{it}]$$

Expanding the expectation, we get

$$\min_L E[\exp(\varepsilon_{it}) \cdot m_{it}] = \min_L E[\exp(\varepsilon_{it})]E[m_{it}] + cov(\exp(\varepsilon_{it}), m_{it})$$

Note that

$$E[\exp(\varepsilon_{it}) \cdot m_{it}] \cong E[\exp(\varepsilon_{it})]E[m_{it}]$$

for

$$|\rho_{xm}| \cdot \frac{\sigma_x}{\mu_x} \cdot \frac{\sigma_m}{\mu_m} < \epsilon,$$

where $x = \exp(\varepsilon_{it})$, the ratio σ_j/μ_j is the coefficient of variation of $j = \{x, m\}$, ρ_{xm} is their correlation, and ϵ is some small number.

Since $E[\exp(\varepsilon_{it})]$ is a constant, minimizing $E[\exp(\varepsilon_{it})]E[m_{it}]$ with respect to m_{it} is the same as minimizing $E[m_{it}]$. By the LLN, we can replace $E[m_{it}]$ by its sample analogue. Thus, the minimization simplifies to

$$\min_L \frac{1}{N_t^1} \sum_i \{\exp(\varepsilon_{it}) \cdot m_{it}\} \cong \min_L \frac{1}{N_t^1} \sum_i m_{it}$$

Replacing m_{it} by the linguistic distance formula, we get

$$\min_L \sum_k \left\{ \pi_{kt} \left[\frac{|\ell_k - L|}{\max_j \{|\ell_j - L|\}} \right]^\rho \cdot I\{\ell_j \neq L\} \right\},$$

thereby completing the proof.

Corollary 1 (Covariation principle for instructional language). *Fix a mean-preserving distribution of linguistic distances. Among all admissible language-of-instruction choices that generate the same mean linguistic distance but differ in their dependence structure with idiosyncratic learning shocks, average learning is maximized when linguistic distance is least favorably aligned with those shocks, i.e., when distance and learning shocks covary most negatively. Conversely, any instructional language that induces positive covariation between distance and shocks reduces average learning.*

When the mean level of linguistic mismatch is held fixed, what matters for aggregate learning is where that mismatch occurs. Because learning shocks raise or lower the productivity of instructional inputs, the same average distance can be more or less damaging depending on how it covaries with those shocks. Average learning is maximized when mismatch is concentrated in low-productivity states, i.e., where the distortion does the least harm, yielding a negative covariance between distance and shocks. Conversely, if mismatch tends to be higher precisely when students are most productive, overall learning is disproportionately suppressed. The implication is that dependence structure, not just average mismatch, shapes the learning

consequences of any language-of-instruction choice, and that empirical evaluations should pay attention to how linguistic distance is distributed across high- and low-productivity learners.

Proof. This follows directly from the minimization equivalence

$$\min_L E[\exp(\varepsilon_{it}) \cdot m_{it}] = \min_L E[\exp(\varepsilon_{it})]E[m_{it}] + \text{cov}(\exp(\varepsilon_{it}), m_{it}).$$

Corollary 2 (Efficiency-equity frontier with single instructional language in multilingual schools). *In linguistically heterogeneous, multilingual schools where language distribution shares vary positively with pre-existing educational advantage, there generally exists no single language of instruction that simultaneously minimizes the average linguistic distance and aligns mismatch perfectly with student learning potential.*

Mean-minimizing languages, e.g. modal language, reduce the average barrier for most students but tend to increase mismatch for minority-language or disadvantaged students, creating positive covariance with productivity. Covariance-minimizing languages, on the other hand, reduce mismatch for lower-potential students but raise the mean linguistic distance for the rest. This trade-off generates an efficiency-equity frontier, wherein improving distributional alignment, leading to more negative covariance, requires accepting a higher average distance, and vice versa.

Proof. Suppose L^* minimizes $E[m]$ and L^{**} minimizes $\text{cov}(\exp(\varepsilon_{it}), m_{it})$. In a heterogeneous school where disadvantaged students speak minority languages, $L^* \neq L^{**}$. No single L achieves both optima except in the knife-edge case of complete linguistic homogeneity.

Theorem (Pareto dominance of multilingual instruction with sufficient linguistic coverage). *In linguistically heterogeneous schools, multiple languages of instruction with sufficient linguistic coverage Pareto-dominate any single-language regime, simultaneously improving both efficiency and equity. Let $\mathcal{L} = \{L_1, \dots, L_K\}$ denote K instructional languages with students optimally assigned to minimize individual mismatch. Then:*

- (i) $E[m_i^{\mathcal{L}}] < E[m_i^L]$ for any single L (lower mean distance)
- (ii) $\text{cov}(\exp(\varepsilon_i), m_i^{\mathcal{L}}) \leq \text{cov}(\exp(\varepsilon_i), m_i^L) \leq 0$ (improved equity)
- (ii) $E[y_i^{\mathcal{L}}] > E[y_i^L]$ for sufficiently comprehensive K (higher average learning outcome)

This result holds provided implementation is feasible and that assignment optimizes individual linguistic matching.

Proof. We provide proof to the above in parts. Suppose throughout that implementation of multiple languages of instruction is feasible.

(i) Lower mean distance. Under multiple language of instructions with multiple assignment, each student i is matched to the closest available language

$$m_i^{\mathcal{L}} = \min_{k \in \{1, \dots, K\}} m_i^{L_k}.$$

By definition, the minimum linguistic distance from using \mathcal{L} over K options cannot exceed any single option L . That is,

$$m_i^{\mathcal{L}} \leq m_i^L \forall L.$$

Taking expectations:

$$E[m_i^{\mathcal{L}}] \leq E[m_i^L].$$

Strict inequality holds when some students are strictly closer to alternative L_k than to any single L . In linguistically heterogeneous schools, this occurs generically.

(ii) Improved equity. Under a single optimal L , when disadvantaged students, i.e., with low ε_i , speak minority languages, they experience high mismatch, inducing negative covariation

$$\text{cov}(\exp(\varepsilon_{it}), m_{it}) < 0.$$

In an environment with multiple languages of instruction that sufficiently cover the underlying language distribution, disadvantaged minority speakers can be assigned to their mother tongue, yielding $m_i^{\mathcal{L}} \approx 0$. This breaks the correlation between low productivity and high mismatch. Formally, since $m_i^{\mathcal{L}} = \min_k \{m_i^{L_k}\}$ compresses the range of mismatch values and reduces variation that correlates negatively with $\exp(\varepsilon_{it})$. That is,

$$\text{cov}(\exp(\varepsilon_i), m_i^L) \leq \text{cov}(\exp(\varepsilon_i), m_i^{\mathcal{L}}) \leq 0$$

(iii) Higher average learning outcome. From the Proposition, we have

$$E[\exp(\varepsilon_i) \cdot m_i] = E[\exp(\varepsilon_i)]E[m_i] + \text{cov}(\exp(\varepsilon_i), m_i).$$

As \mathcal{L} covers greater proportion of the language distribution, more students receive instruction close to their mother tongue. From (i), $E[m_i^{\mathcal{L}}] \rightarrow 0$. From (ii), as mismatch becomes uniformly small across all students regardless of ε_i then $\text{cov}(\exp(\varepsilon_i), m_i^{\mathcal{L}}) \rightarrow 0$. Combining these results, for \mathcal{L} that provide sufficient linguistic coverage,

$$E[\exp(\varepsilon_i) \cdot m_i^{\mathcal{L}}] < E[\exp(\varepsilon_i) \cdot m_i^L],$$

implying that

$$E[y_i^{\mathcal{L}}] > E[y_i^L].$$

Appendix B. Axiomatic foundations of network-adapted FGT index

Let \mathcal{L} be a finite set of nodes in a network graph. Each node $l \in \mathcal{L}$ has a population weight $p_l \geq 0$, with the total population given by $P = \sum_l p_l$. Fix a reference node $k \in K$. Define the normalized nodal distance as

$$d_i := d(l, k) = I(l \neq k) \cdot \frac{\min_{k \in K} |l - k|}{\max_{j \in \mathcal{L}} |j - k|},$$

where $d(l, k)$ is the shortest path, i.e., minimum number of steps, between two nodes l and k . The normalization $\max_{j \in \mathcal{L}} |j - k|$ ensures that $0 \leq d(l, k) \leq 1$ for every l . The reference node need not be unique. For $\rho \geq 0$, define the network-adapted FGT, henceforth FGTA, index as

$$N_\rho(k) = \frac{1}{P} \sum_{l \in \mathcal{L}} p_l d_l^\rho.$$

We show that FGTA satisfies the standard FGT axioms, i.e., invariance (symmetry, replication, scale, focus, continuity), dominance (monotonicity, transfer, transfer sensitivity) and subgroup (decomposability, subgroup consistency) axioms (Foster, et al., 1984, 2010). Note that many of the axioms follows directly from the definition of $N_\rho(k)$ above.

(i) Invariance axioms

1. Symmetry. $N_\rho(k)$ depends only on the multiset $\{(p_l, d_l)\}$ and not on labels. Permuting node labels leaves the sum $N_\rho(k)$ unchanged.
2. Replication invariance. Replicating the population weight at each node by a constant factor r does not change $N_\rho(k)$. This may be obvious by factoring out r from the summation and noting that the total population is also scaled by r , leaving FGTA unchanged.
3. Scale invariance. Multiplying all nodal distances $|l - k|$ by a constant factor s does not change $N_\rho(k)$. Since all nodal distances are multiplied by s , so does the maximum distance, thereby leaving FGTA unchanged.
4. Focus. Only nodes away from the reference node, i.e., $l \neq k$, contribute to $N_\rho(k)$. This follows directly from the indicator function $I(l \neq k)$, but also from the nodal distances $|l - k|$ since $|k - k| = 0$ by definition.
5. Continuity. For any $\rho \geq 0$, the index $N_\rho(k)$ is a continuous function of the normalized distance d_l . Small changes in d_l cannot cause jumps in $N_\rho(k)$. Note that continuity comes from the continuity of normalized d_l and not the discrete set-up of the network graph.

(ii) Dominance axioms

1. Monotonicity. If a nodal distance increases, keeping all others the same, the FGTA index either stays the same or increases. Note that for any $\rho \geq 0$, the FGTA index is non-decreasing in d_l . If $d'_l \geq d_l$, holding p_j for all $j \in \mathcal{L}$ fixed, then $N'_\rho(k) \geq N_\rho(k)$, with strict inequality holding for any $\rho > 0$.
2. Pigou-Dalton transfer. For any $\rho \geq 1$, shifting population weight towards farther node l from closer node k , i.e., $d_l > d_k$, increases overall FGTA, or at least does not decrease it. Formally, if $p'_l = p_l + \epsilon$ and $p'_k = p_k - \epsilon$ with $\epsilon > 0$, then

$N'_\rho(k) \geq N_\rho(k)$ with strict inequality for $\rho > 1$ and both nodes have non-zero population weights.

3. Transfer-sensitivity. For any $\rho \geq 2$, increasing ρ makes the FGTa more sensitive to nodes that are farther away. Note that each node contributes $p_l d_l^\rho / P$ to the index. Raising ρ amplifies the largest distances relative to smaller ones. Consider two nodes with $d_l > d_k$, for any $\rho' > \rho \geq 2$ it follows that

$$\frac{d_l^{\rho'}}{d_k^{\rho'}} > \frac{d_l^\rho}{d_k^\rho}$$

Suggesting that the contribution of farther nodes growing faster than closer nodes.

(iii) Subgroup axioms

1. Decomposability. Partition \mathcal{L} into disjoint groups $\{\mathcal{g}\}$ with subgroup population $P_\mathcal{g} = \sum_{l \in \mathcal{g}} p_l$. The index $N_\rho(k)$ may be decomposed as a weighted sum of the subgroup-specific $N_{\rho,\mathcal{g}}(k)$. In our application to linguistic distance, we assume the same scaling factor $\max_{j \in \mathcal{L}} |j - k|$, which leads to subgroup weight of $P_\mathcal{g}/P$. More generally, the weight is given by

$$w_\mathcal{g} = \left(\frac{P_\mathcal{g}}{P}\right) \left(\frac{M_\mathcal{g}}{M}\right)^\rho,$$

where $M_\mathcal{g}$ is the subgroup-specific maximum distance and M is the global maximum. That is,

$$N_\rho(k) = \sum_x w_\mathcal{g} N_{\rho,\mathcal{g}}(k).$$

2. Subgroup consistency. If one subgroup has an increase in $N_{\rho,\mathcal{g}}(k)$ while all other subgroups remain the same, then the weighted sum above strictly increases because the relative weight $w_\mathcal{g} > 0$, thus $N_\rho(k)$ increases.

The index $N_\rho(k)$ satisfies the full set of Foster-Greer-Thorbecke axioms (invariance, dominance, subgroup decomposability) when network distances are normalized to $[0,1]$. The family $\{N_\rho(k); \rho \geq 0\}$, here denoted as FGTa, therefore constitutes a class of decomposable, network-adapted FGT measures suitable for measuring population-weighted deprivation on network structures.

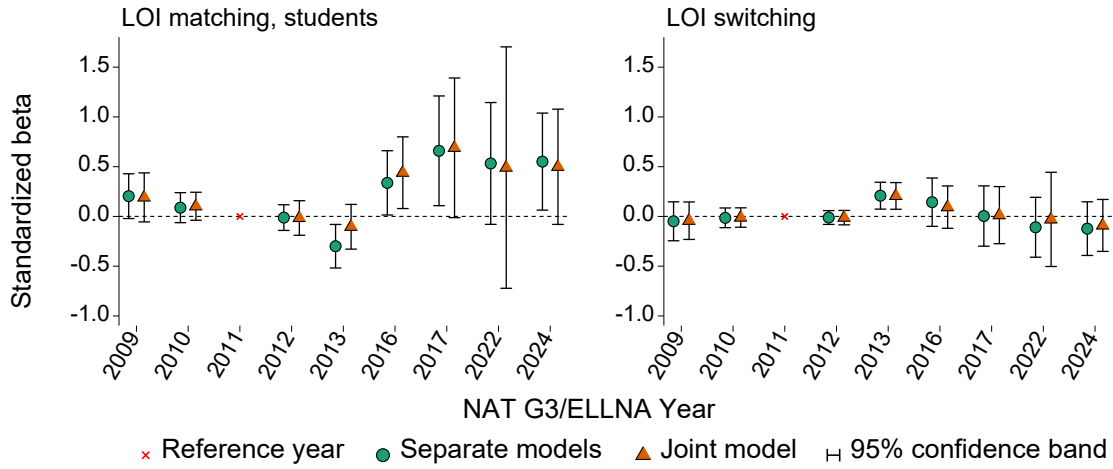
FGTa nests the classical Foster-Greer-Thorbecke family of poverty indices as a special case. This becomes clear when we treat each income (or expenditure) level as a separate node weighted by the respective population and define nodal distance relative to a poverty line. Individuals with income at or above the poverty line form the reference node with zero distance, while those below the line have distance equal to their normalized shortfall. Under this mapping, the FGTa index reduces exactly to the classical FGT measures.

Extensions of FGTa to n -factor graphs using a generalized distance formula is rather straightforward, allowing multidimensional distance calculations. However, standard decomposability does not generally carry over.

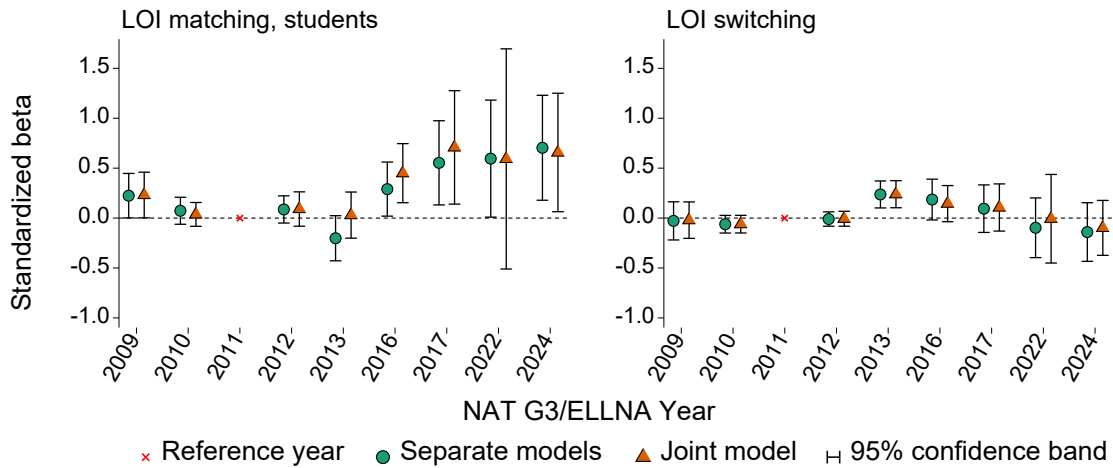
Appendix C. Robustness analysis: IV-DID separate v. joint estimation

A. LOI matching (students) and LOI switching

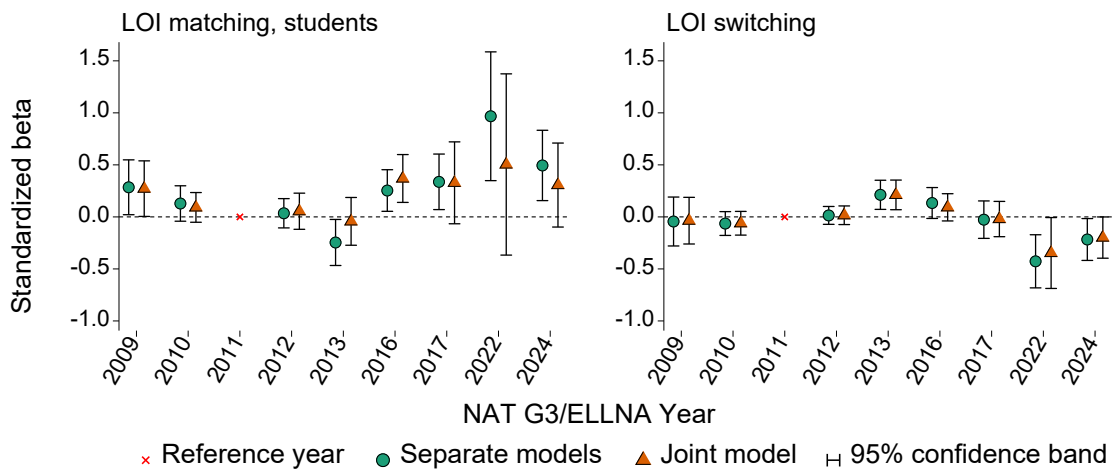
A.1. Filipino



A.2. English



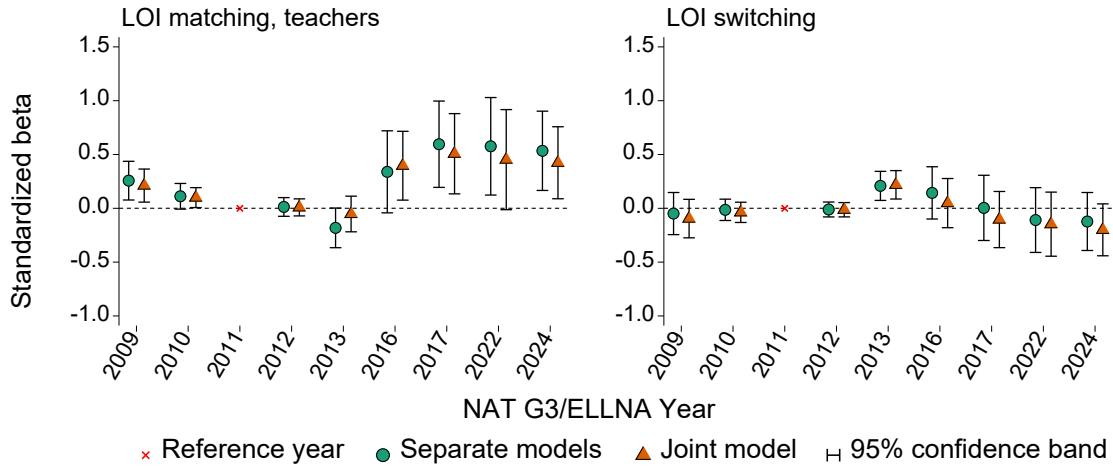
A.3. Mathematics



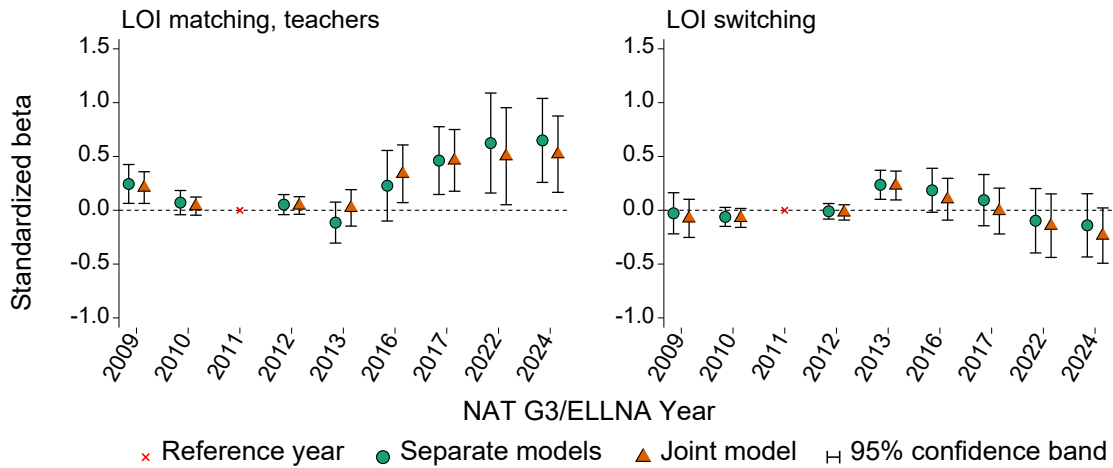
Appendix C. Robustness analysis: IV-DID separate v. joint estimation (continued)

B. LOI matching (teachers) and LOI switching

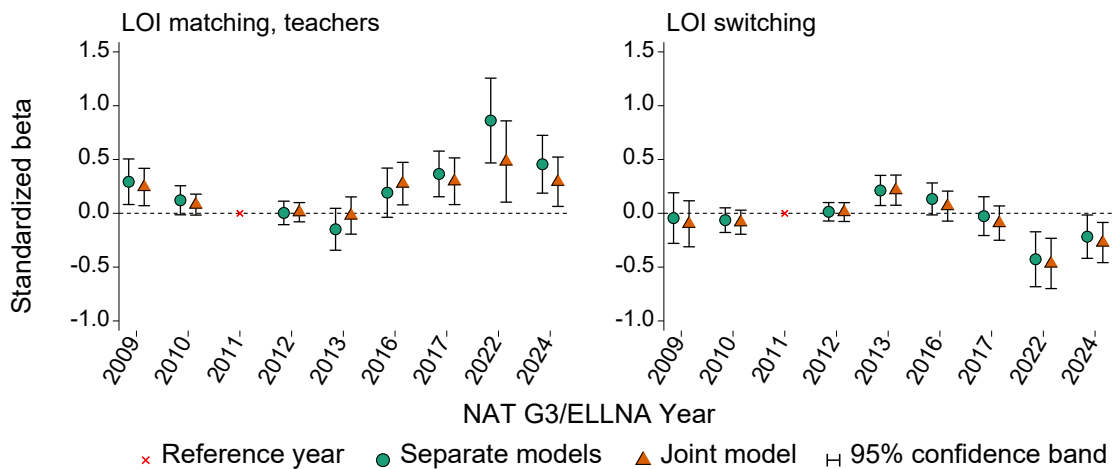
A.1. Filipino



A.2. English



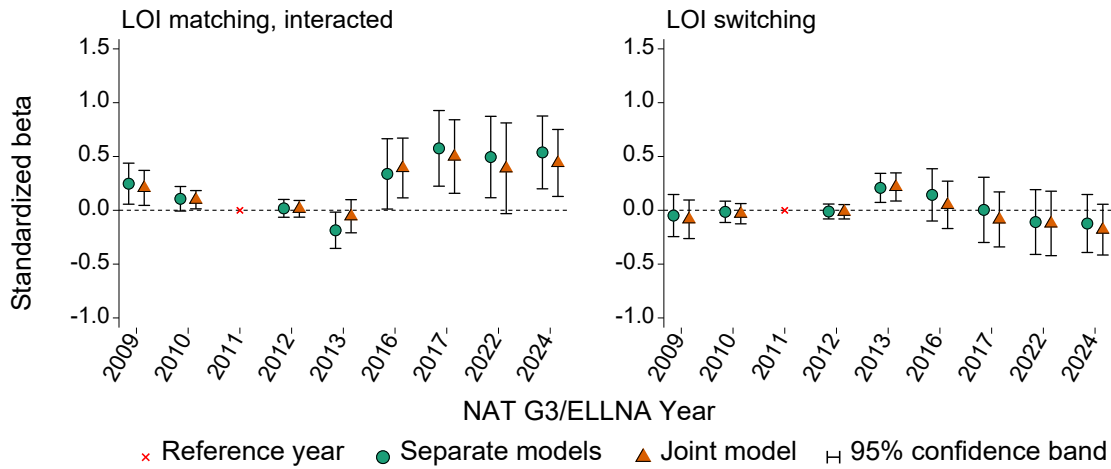
A.3. Mathematics



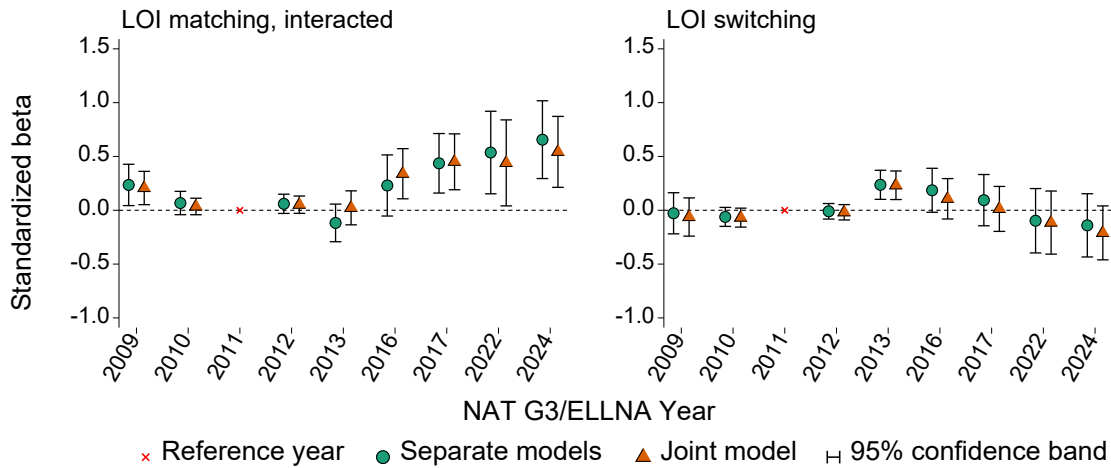
Appendix C. Robustness analysis: IV-DID separate v. joint estimation (continued)

C. LOI matching (interacted) and LOI switching

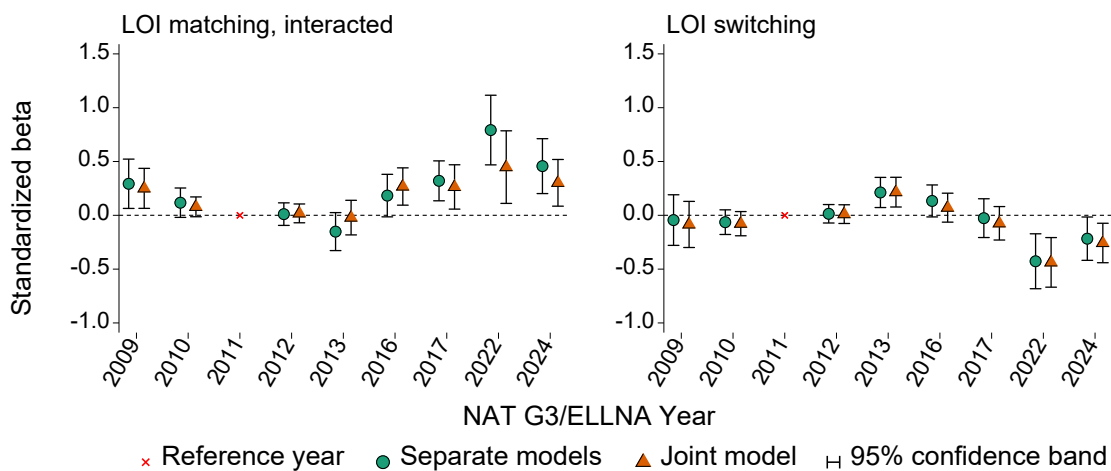
C.1. Filipino



C.2. English



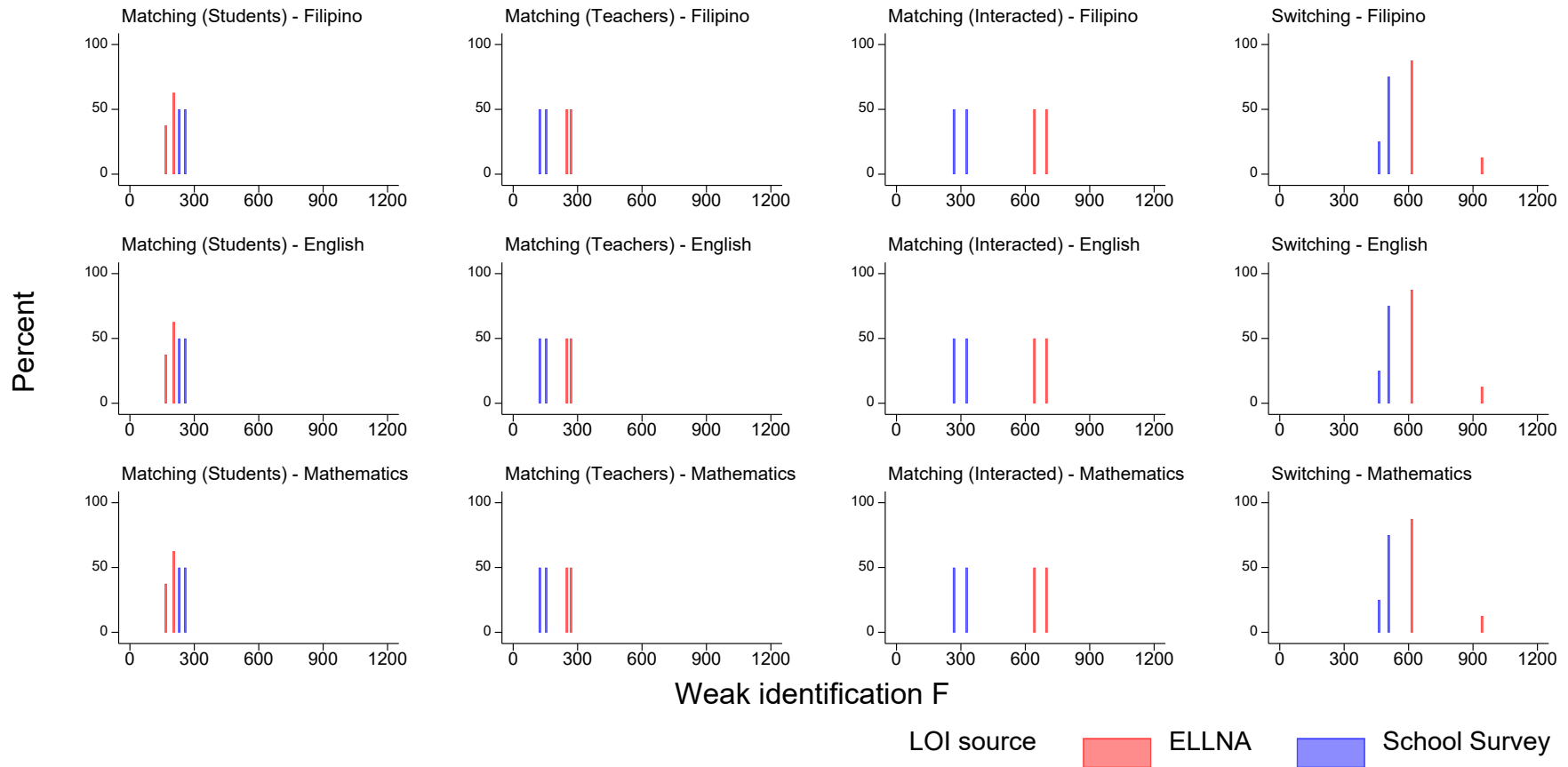
C.3. Mathematics



Notes: Figures display DID and IV-DID coefficients and 95% confidence intervals converted into standardized betas. Standard errors are clustered at the schools-division level and scaled to include incorporate school-level variation following Moineddin and Urquia (2014).

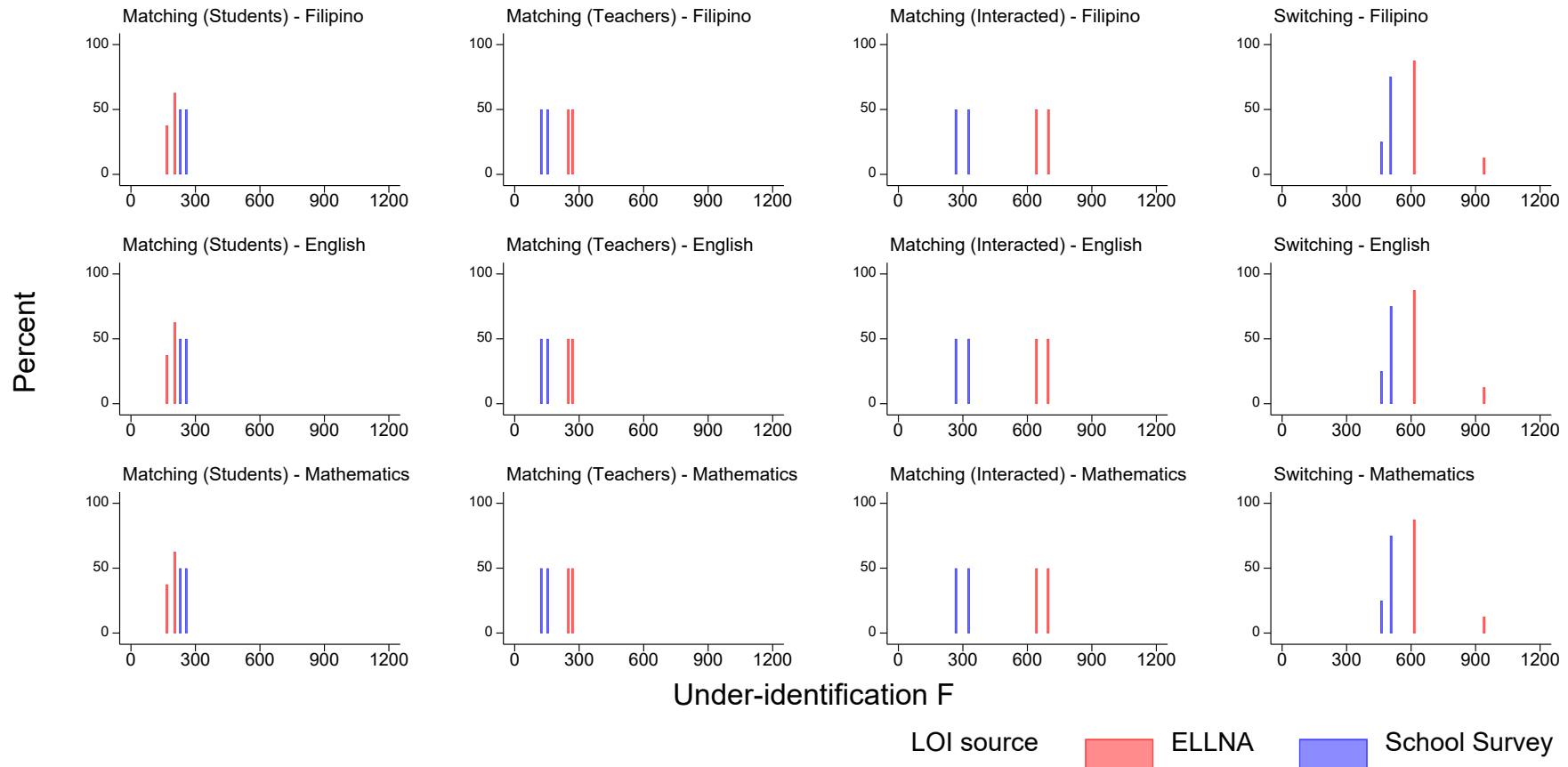
Appendix D. Shift-share IV relevance: NAT/ELLNA IV-DID

A. Weak identification



Appendix D. Shift-share IV relevance: NAT/ELLNA IV-DID (continued)

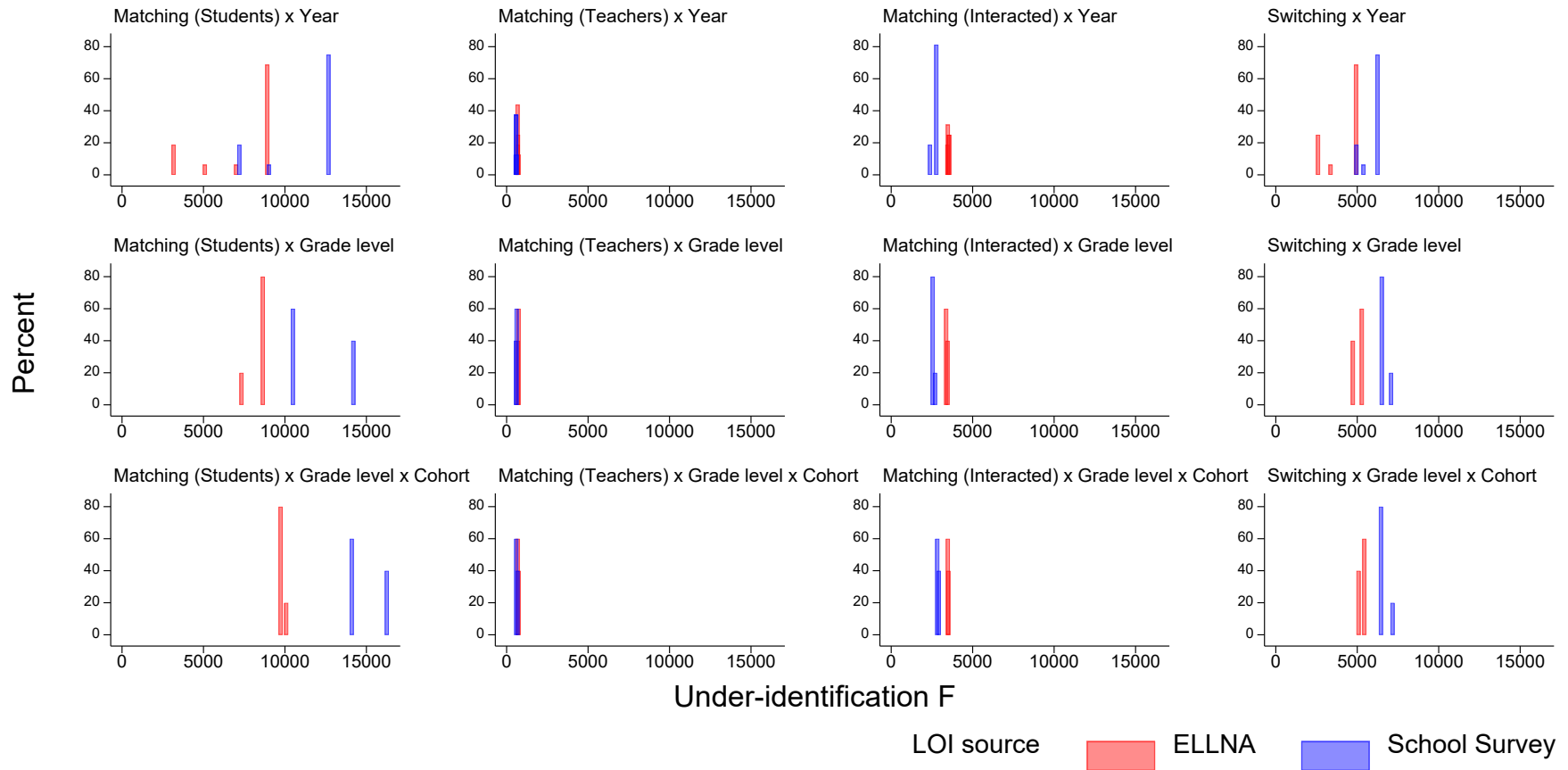
B. Under-identification



Notes: Figure displays Kleibergen-Paap rk Wald F-statistics and Kleibergen-Paap rk LM statistics testing instrument strength and identification from shift-share instrumental variables regressions. Each pair of bars represents statistics from a separate first-stage regression for a given instructional language intervention and language of instruction (LOI) database. The rk Wald F-statistic tests for weak identification; the rk LM statistic tests for underidentification. The horizontal dashed line indicates the conventional threshold of $F = 10$ for weak instrument concerns (Stock and Yogo, 2005). All first-stage regressions include school and year fixed effects.

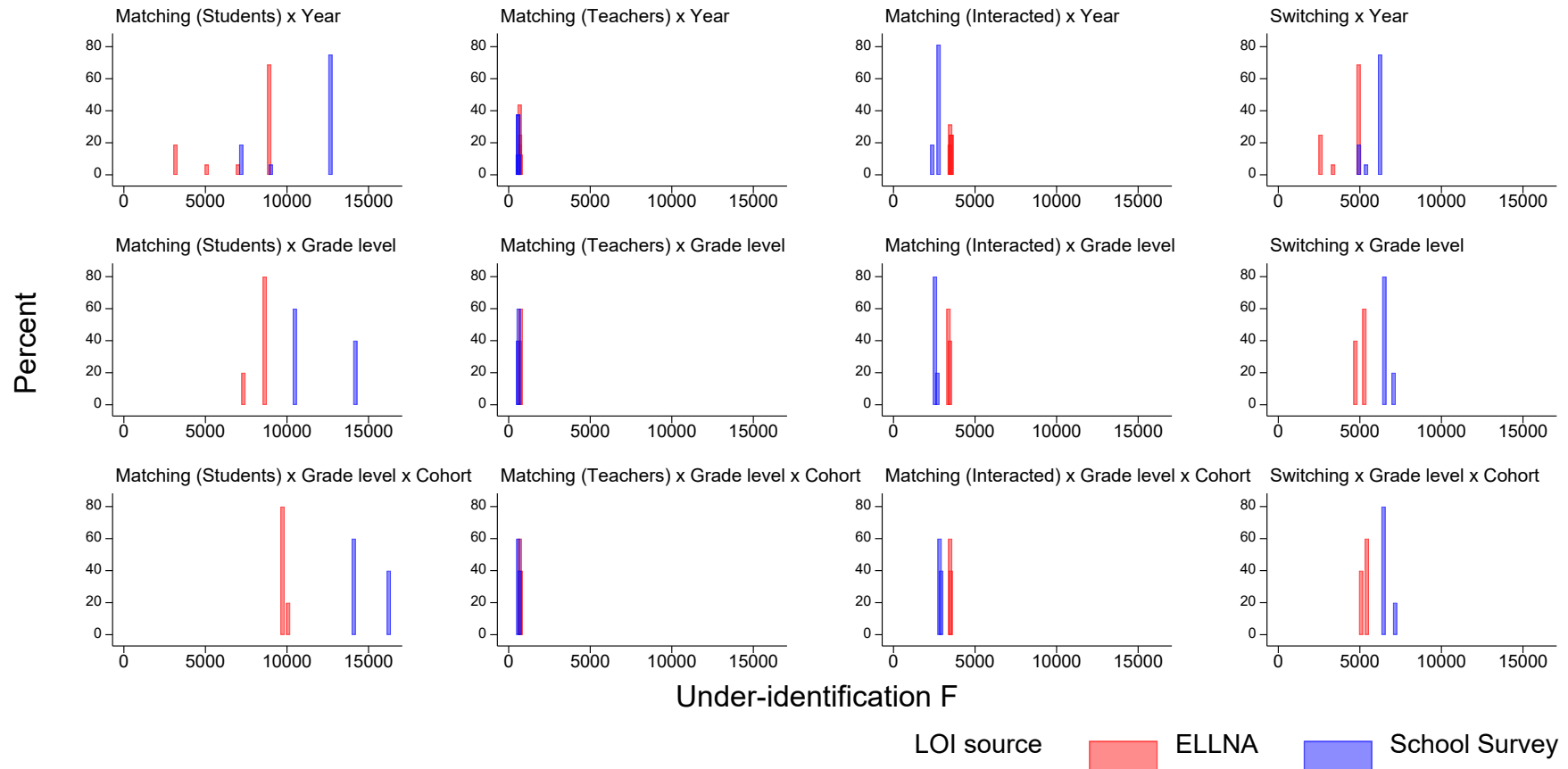
Appendix E. Shift-share IV relevance: Enrollment IV-DDD

A. Weak identification



Appendix E. Shift-share IV relevance: Enrollment IV-DDD (continued)

B. Under-identification



Notes: Figure displays Kleibergen-Paap rk Wald F-statistics and Kleibergen-Paap rk LM statistics testing instrument strength and identification from shift-share instrumental variables regressions. Each pair of bars represents statistics from a separate first-stage regression for a given instructional language intervention and language of instruction (LOI) database. The rk Wald F-statistic tests for weak identification; the rk LM statistic tests for underidentification. The horizontal dashed line indicates the conventional threshold of $F = 10$ for weak instrument concerns (Stock and Yogo, 2005). All first-stage regressions include school and year fixed effects.