<u>Better educated or better educators? The effects of hiring standards on teacher and student outcomes in Pakistan</u>

Saher Asad and Freeha Fatima

This paper evaluates the impact of teacher hiring reforms on educational outcomes in Pakistan. Prior to 2016, public school teachers were required to hold a bachelor's or master's degree in education for recruitment. In 2016, all provinces implemented a policy mandating standardized testing through the National Testing Service (NTS) for new hires, focusing on STEM subject matter knowledge. Using the first provincially representative data from the Global Education Policy Dashboard Survey (2023) and administrative census data, applying a regression discontinuity to year of hiring, we find that teachers recruited under the new regime are likely to be more educated, and this finding holds true particularly among female teachers. Furthermore, these female teachers are more likely to have majored in science, resulting in enhanced mathematical content knowledge. While the reforms succeeded in raising the educational qualifications of new hires, they did not translate into better teaching quality measured through classroom practices or higher student achievement measured through student assessments. These findings underscore the importance of the substance of educational standards. When countries rely solely on STEM focused content-heavy assessments, they risk unintended consequences that may undermine teaching quality. Therefore, teacher selection standards need to prioritize candidates' pedagogical skills, not just their subject-matter knowledge.

I. Introduction

Teacher hiring practices in developing countries have emerged as a pivotal focus of educational policy, particularly in South Asia, where systemic challenges in the education sector often hinder progress. The significance of effective teacher recruitment cannot be overstated; in many of these countries, once teachers are integrated into the public school system, their positions become highly secure, making it virtually impossible to remove them for underperformance. This entrenched system of job security underscores the critical importance of ensuring that only high-quality teachers are inducted to begin with, as they can impact the educational trajectories of students for decades.

The implications of teacher hiring policies extend beyond individual classrooms; they are intricately linked to broader educational outcomes and human capital development. Quality education is a fundamental driver of human capital formation, and in developing contexts, the recruitment and retention of capable teachers are essential to improving learning outcomes and fostering better standards in teaching. Yet, the processes and criteria employed in teacher hiring often remain contentious and policy makers continue to struggle with the best practices needed in their specific context.

Standardized testing is becoming a widespread policy tool intended to enhance teacher and student quality and learning outcomes globally.[1] These tests can be administered both ex ante and ex post and their impacts on educational outcomes can vary substantially across country contexts.[2] Moreover, the outcomes can vary by the types of tests administered, the grade-level at which the students are evaluated, and the type of data used to evaluate the reforms. In addition to the intention of hiring better quality teachers, standardized tests are commonly used to improve teacher hiring practices through a consistent and non-politicized way of hiring public sector employees. The latter however may not be achieved if such testing becomes an avenue for nepotism and favoritism in teacher hiring or if the system becomes highly politicized due to possible opposition from teacher unions and affected candidates. Consequently, researchers have found mixed results when evaluating their impacts on teacher and student outcomes.

The current literature focuses on two types of effects of standardized teacher hiring policies; teacher composition and student outcomes, both of which speak closely to our analysis. In terms of student achievement, for instance, Brutti and Torres (2022) find positive effects of standardized examinations on student outcomes in Colombia.[3] They find that quality-screened teachers through standardized exams improve student achievement by about 7 percent of a standard deviation within a school-year. In contrast, Busso et. al. (2024) find negative effects of merit-based teacher-hiring on student performance and educational attainment in Colombia. Ome (2013, 2012) also study the Colombian reforms and find no effect of the regulations on student test scores at the high school

[1] See (Aslam, Rawal and Kingdon; 2021) for South Asia, Bedoya (2023), Estrada (2019) for Mexico, Brutti and Torres (2022), and Busso et. al. (2024) for Colombia, Montalvo and Haro (2019) for Ecuador, and Mizala and Schneider, (2019) for Peru and Chile.

[2] Ex ante exams are used primarily for hiring and are usually weighted the highest among other hiring requirements like credentials and interviews. Ex post, teachers are evaluated and scored while teaching.

[3] In Colombia, the exam focuses on three modules: teaching aptitude, subject knowledge and psychometric values. The minimum score needed to pass is 60 percent, and the score system is used to rank applicants, which determines the order in which they are allowed to choose vacancies. Further points are earned for academic credentials such as degrees, attendance of training courses, academic publications, past teaching experience and evaluations, and for holding specific awards. Finally, face-to-face interviews are conducted before final hiring. The exam weight is 65 percent of the final score, the credentials weigh another 20 percent, and the interviews are weighted at 15 percent. Evaluations continue after hiring.

level. This variation in findings not only underscores the importance of this subject, but also the specifics of the stage at which students are evaluated and the evaluation criteria. Estrada (2019) studies the effects of introducing rule-based hiring in Mexico on secondary school students' numeracy and literacy skills and finds positive effects on student outcomes.[4] Araujo and Daniela (2019) study the case of Ecuador and find that test-screened teachers do not perform any better than their peers in general, and yield benefits only when matched with students from disadvantaged backgrounds. Larsen (2013) finds that in the US, the distribution of student test scores increases with stricter occupational licensing, but only in the upper half of the test-score distribution. Also, for most forms of licensing studied in the US, input and output quality improvements due to stricter licensing requirements occured in high-income rather than low-income school districts.[5]

In terms of compositional shifts, Busso et. al. (2024) find that post reform, teacher composition shifted from more experienced contract teachers towards high exam-performing novice teachers in Colombia. The policy sharply increased pre-college test scores of teachers, while decreasing the overall stock of teacher experience. Saavedra et. al (2022) also study Colombia from the labor supply and earnings angle and find that teachers ultimately hired through standardized exams get an earnings premium in the long-run but these exams do not necessarily attract the highest quality teachers. Bedoya et. al. (2023) study the effects of a civil service reform on the skills profile of new primary and lower-secondary school teachers hired in Mexico. They study a nationwide education reform that mandated centrally managed competitive examinations to determine teacher hiring and promotion decisions. They find that teachers hired post-reform have higher cognitive

[4] In Mexico, the exam focused on measuring cognitive skills, knowledge of the teaching subject, mastery of teaching methods, and ethics. One exam is held for each type of teaching position (primary school teacher, mathematics lower secondary school teacher, etc.). Some types of positions were restricted to graduates of teacher training schools or with specific college majors. The number and type of avail-able teaching positions by state and the exam results were widely publicized by media outlets and were available on a dedicated web page. The advertised positions were not associated with specific schools. The applicants are ranked by state and teacher type according to their exam results or a weighted average of the test score and other criteria (often university GPA), in case certain states opted for the latter. In half the states, selected applicants had to pass an additional exam (typically a health exam). Hires with an exam score below a state threshold had to undergo remedial training, defined at the state level. Using a difference-in-difference analysis, Estrada (2019) finds that moving from no rule-based hires in a school to only rule-based hires increased the school's mathematics test score by 0.53 standard deviations and the Spanish score by 0.32 standard deviations. He also notes that rule-based hires had a better academic background than discretionary hires, as measured by average university grade point average (GPA). He does not however find a statistical association between university GPA or other characteristics and teacher performance—a common feature in the literature on teacher effectiveness.

[5] For a review of the literature on teacher hiring practices and impacts in the United States, see Guarino et. al. (2006) and Darling-Hammond, L., & Berry, B. (1999).

skills than teachers hired before the reform, and this change is driven by an improvement in the bottom of the skills distribution of newly hired teachers.[6]

Similarly, Wiswall (2007) provides suggestive evidence of more stringent hiring policies leading to lower teacher quality on average in the United States. They also find evidence of a lower labor supply of teachers and an increase in the average length of their teaching careers. Larsen (2013) also finds that more restrictive licensing laws in the US lead to some first-year teachers of high input quality to opt out of the occupation. However, among teachers who remain in the occupation for multiple years, stricter licensing appears to increase input quality at most quantiles of the teacher quality distribution.

As noted, most of the empirical studies have focused on policy changes in Latin America and the United States and much less has been said about such reforms in the South Asia region. Moreover, the details of the policies in each context in terms of the content of the exam, as well as the scoring of the various components eventually considered for hiring are different from our context. This is the first paper that studies the causal impact of standardized testing in the teaching profession on teacher and student outcomes in the South Asia Region. Moreover, we add to both the aforementioned strands of literature in the context of a lower middle-income country.[7] First, we contribute to the studies on teacher composition by studying the impacts on the educational qualification of teachers hired post-reforms and on the types of degrees they specialize in. Secondly, we use a unique and first-ever collected dataset that helps to compare the impact of teacher hiring policies on student outcomes at the national and subnational level.[8]

[6] See Bedoya et. al. (2023) for details. The ENLACE score of new teachers increased between 2.7 to 3.8 percentile points after the SPD reform.

[7] For a review of the literature on how to staff hard-to-staff schools in low- and middle-income countries, see see Eavns and Amina (2023).

[8] Despite the strength of students being significantly high in public schools (62 percent), comparable and consistent data on teacher and student outcomes is not available across all provinces. The Global Education Policy Dashboard dataset is first unique dataset of the kind which can be used for building a consistent narrative on schooling quality in public schools across the country at the sub-national level. Previous other surveys like SABER SD conducted by the World Bank have been limited to the Punjab province only.

The reform we study here was instituted in 2016 whereby all candidates were mandated to appear in a standardized exam before they could be hired as public-school teachers.[9] Prior to NTS, candidates were required to hold specific educational qualifications, such as the Primary Teaching Certificate (PTC) and/or the Certificate of Teaching (CT). The standardized tests were primarily administered by the National Testing Service (NTS), thus becoming commonly known as NTS reforms. The NTS process tested the candidates' knowledge in multiple subjects like English, Mathematics, Science, Current Affairs and the like. The candidates' final merit was calculated using a weighted scoring system, which considered both their academic qualification and the NTS score.

The NTS test was primarily designed to assess candidates' knowledge in scientific subjects, without evaluating their pedagogical skills or teaching effectiveness. Given prevailing norms in Pakistan that favor candidates with strong science backgrounds, the test aimed to attract high-achieving individuals into the teaching profession. However, this approach overlooked the critical importance of teaching ability, classroom communication skills and pedagogical practices. As a result, the test may have inadvertently prioritized content mastery over the competencies actually required to communicate that content and hence improve student learning outcomes.

The test was also intended to improve transparency and to reduce nepotism in public sector hiring as noted by Aslam et. al. (2021). Moreover, the idea was that the standardized approach would encourage the recruitment of local teachers, where feasible, and may help to promote merit-based hiring practices, which were previously vulnerable to political interference. Although the NTS initially faced logistical challenges, such as managing large numbers of candidates and ensuring accurate scoring, improvements were made overtime to address these issues (source). The implementation of standardized assessments represents a substantial change in the recruitment landscape, with the potential to influence the qualifications and preparedness of teachers entering the public education system.

This paper evaluates the impact of the 2016 NTS reforms by leveraging the first provincially representative data from a nationwide school survey conducted in 2023, the Global Education

[9] The teacher recruitment process in Pakistan came across significant changes between the years 2012 and 2016, with timelines varying across provinces, but since a majority of the changes happened around the year 2016, we use 2016 as the separating year for pre- and post-reform analysis.

Policy Dashboard Survey. The identification strategy assumes that the evaluated reform was the only major change that happened in 2016 and was common to all provinces.[10] After controlling for all other observable teacher and school characteristics, the major difference between teachers hired before and after 2016 is that of being hired under the new criteria of coming in through standardized tests. In terms of teacher outcomes, we evaluate the impact of being hired post-2016 on teachers' educational qualification, pedagogical scores measured using the TEACH Tool, teacher content knowledge and their intrinsic motivation. In terms of student outcomes, we use student performance on verbal and mathematics tests administered to fourth grade students.

We find that teachers, especially females, are significantly more likely to hold post-graduate degrees post-reform. These teachers are also likely to have a STEM background. That said, their higher educational qualifications and STEM backgrounds do not translate into better teacher or student level outcomes. We find no improvements to teacher pedagogical practices or student test scores post-2016. What we do find however, is that female teachers hired post-2016 possess higher content knowledge of mathematics and lower content knowledge of reading. This means that there may have been changes in the composition of teachers hired ex-post, and they may be majoring more in science versus arts fields.[11] We find that this is indeed the case.

The next section presents the methodological framework, Section III discusses the details of the NTS reform, Section IV presents the data and the estimation methodology, section VI presents the results and Section V concludes.

II. NTS Reform:

Before the reform, teacher hiring process was largely position-based and ad hoc. It relied on a combination of factors such as candidates' background, education level, and availability to serve in specific induction areas. The process lacked consistency and was often influenced by the prevailing political regime, which created a widespread perception of political interference in hiring decisions. This not only undermined the credibility of the recruitment process but also allowed significant room for favoritism and non-merit-based appointments.

[10] Although the timeline of the mandated reforms varied between provinces, but the implementation is mostly understood to have taken place by 2016.

[11] Mathematics is considered a science subject.

The reform was introduced with two primary goals. First, it aimed to standardize the teacher recruitment process across the board, ensuring that all candidates were selected through a uniform and transparent system, free from political influence. Second, it sought to improve the overall quality of the teaching workforce by attracting and selecting more qualified individuals. To support this, the recruitment process placed greater emphasis on STEM (Science, Technology, Engineering, and Mathematics) subjects, which are generally seen as attracting more academically capable candidates. The idea was that by focusing on these areas, the system would naturally filter in more competent teachers.

To implement this standardized approach, provincial school education departments outsourced the testing process to the National Testing Services (NTS), an independent organization known for conducting standardized exams for university admissions and job applications, including those for teaching positions. The NTS test for teacher recruitment consists of two components: a general test and a subject-specific test. The general test accounts for 30% of the total score and includes sections on English (15%) and Analytical Reasoning (15%). The subject test, which carries a 70% weight, typically focuses on Math, Physics, and Chemistry, regardless of the specific teaching role being applied for. The test often includes multiple choice questions, based on memorization and rote-learning.

To be eligible for teaching positions in the public sector, candidates must meet several prerequisites, including age limits, minimum qualifications specific to the post, a passing score on the NTS test, and completion of any required training. Historically, about one-third of candidates who take the test pass it by scoring above 50%. This pass rate has remained relatively stable over time and across different provinces, indicating a consistent level of difficulty and candidate performance.

III. Framework:

The introduction of testing services as a criterion for hiring teachers can significantly impact the educational profiling of teachers through two distinct channels. The first mechanism, is what we label ***selection filter mechanism***, suggests that the implementation of a testing service could encourage more educated and qualified individuals to pursue teaching careers. This mechanism operates under the assumption that the testing service is rigorous and is thereby attracting candidates with advanced degrees and strong educational backgrounds. If this mechanism dominates, we should observe an increase in the qualifications of teachers, as more individuals with higher degrees and better educational credentials will be motivated to join the teaching profession.

Conversely, the second mechanism, what we refer to as the ***lowering entry barrier mechanism***, posits that the testing service could lead to the entry of less educated individuals into the teaching profession. This could occur if availability of testing as the key hiring determinants is making it easier for individuals with lower qualifications to pass. As a result, the profession may attract candidates who meet the minimum requirements of the test but lack higher educational credentials. If this mechanism dominates, we should observe a decrease in the qualifications of teachers, as the testing service may inadvertently lower the entry barriers for less educated individuals.

Understanding which mechanism dominates is crucial for assessing the overall impact of testing services on teacher qualifications. If the selection filter mechanism prevails, we can expect an enhancement in the qualification of teachers hired. On the other hand, if the lowering entry barrier mechanism dominates, there may be a decline in qualifications of teachers being hired. The results presented in this paper highlight that selection filter mechanism prevails, as a result teachers hired under the new standards are more educated.

IV. Data and Methodology

This paper uses data from the GEPD survey conducted in 2023/2024. GEPD is the first provincially representative cross-sectional survey of public schools in Pakistan. The survey collects detailed school, student and teacher level information from 800 schools in the provinces of Punjab, Sindh, KPK and Balochistan and 100 schools in the capital Islamabad (ICT). A province-wise breakdown of the sample sizes in the Pakistan GEPD survey can be found in the Appendix. Each province conducts an annual school census of public schools which collects information on school facilities and infrastructure, student enrollment, number of teaching and non-teaching staff. In addition, some provinces maintain administrative databases that monitor and track teacher qualifications, background as well as year of hiring.

In order to assess the impacts of NTS on student and teacher outcomes, we run the following two specifications:

$$Y_{ts} = \alpha + \beta_1 \, |Hiring\ year_{ts} - 2016| + \beta_2 \, |Hiring\ year_{ts} - 2016|^2 + \beta_3(Teacher\ hired\ post\ 2016_{ts}) + Teacher\ characteristics_{ts} + School\ characteristics_s + Province\ fixed\ effects_s + \varepsilon_{ts}$$

$$Y_{its} = \alpha + \beta_1 \, |Hiring\ year_{ts} - 2016| + \beta_2 \, |Hiring\ year_{ts} - 2016|^2 + \beta_3(Teacher\ hired\ post\ 2016_{ts}) + Student\ characteristics_{is} + Teacher\ characteristics_{ts} + School\ characteristics_s + Province\ fixed\ effects_s + \varepsilon_{ts}$$

Where Y_{ts} is the outcome variable for teacher t in school s and Y_{its} is the outcome variable for student i with teacher t in school s.We use four outcome variables for teachers; (i) teacher's educational qualification, (ii) teaching practices (pedagogy), (iii) teachers' content knowledge and (iv) teachers' intrinsic motivation. For students we use test scores on verbal and mathematic examinations conducted for fourth-grade students.[12]

V. Identification Strategy:

The identification strategy utilizes a Regression Discontinuity in Time (RDiT) design, capitalizing on the sharp temporal cutoff introduced by the implementation of the policy in 2016. Unlike traditional regression discontinuity designs that exploit a discontinuity in a continuous assignment variable (e.g., test scores), RDiT leverages time itself as the running variable, with the policy change serving as the cutoff point. This approach is particularly useful when treatment assignment is determined by a specific date or year, as is the case here, being year of hiring.

To model the effect of the policy, we include both linear and quadratic specifications of time elapsed since the policy's introduction. This allows us to flexibly control for underlying time trends that might otherwise confound the estimated treatment effect. The key assumption in RDiT is that, in the absence of the policy, outcomes would have evolved smoothly over time. Any discontinuous jump at the cutoff can therefore be interpreted as the causal effect of the policy.

[12] See Appendix Table A1 for details.

Our dataset includes information on the year of entry into the system, enabling us to compare cohorts entering just before and just after the 2016 policy change. This local comparison around the cutoff helps mitigate concerns about unobserved confounders that vary over longer time horizons. We follow the methodological framework outlined by Hausman and Rapson (2017), which emphasizes the importance of bandwidth selection and robustness checks in RDiT settings.

VI. Results:

Does NTS impact the educational attainment and composition of teachers?

Table 1 presents the results on the effects of NTS on teacher's level of education using GEPD data. The dependent variable takes a value of 1 for all teachers who have completed more than a bachelors degree and a value of zero if they have a bachelors degree or lower. We find that teachers hired after the reform were 5 percentage point (pp) more likely to have more than a bachelors degree and this goes up to almost 16-17 pp for female teachers in various specifications. Overall, teachers in urban areas are likely to me more educated but post-2016, it is the rural regions that are getting more educated teachers.

The GEPD data provides limited insight into teachers' educational backgrounds, capturing only whether a teacher holds less than a bachelor's degree, a bachelor's degree, or a master's degree. As a result, it does not allow us to assess the impact of NTS-based hiring reforms on the specific fields of study pursued by teachers or whether they hold additional education-related qualifications such as B.Ed. or M.Ed. To address this gap, we draw on administrative data from the Punjab Education Census. This dataset enables us to examine whether teachers hired after the reforms were more likely to have specialized in science versus arts, whether they were more or less likely to hold education degrees, and whether they were more likely to possess qualifications beyond a bachelor's degree—the latter being comparable to what is captured in the GEPD.

Table 2a shows that in Punjab female teachers hired post-2016 were more likely to have more than a bachelors degree, were more likely to graduate in science versus arts majors and were less likely to have a degree in education. Moreover, teachers hired in urban areas post-2016 were less likely to be more educated or to major in science and/or to have an education related degree. Table 2b shows that in KP female teachers hired post-2016 were more likely to have more than a bachelors degree, were as likely to graduate in science versus arts majors and were more likely to have a degree in education; be it a B. Ed or an M. Ed. Moreover, teachers hired in urban areas post-2016 were less likely to be more educated.

Table 2c disentangles the effects of NTS on educational attainment of teachers in Punjab by splitting the dependent variables into sub-categories of education. Interestingly we find that post-2016, female teachers' likelihood of getting a B. Ed is higher overall, but because their likelihood of getting a bachelors degree in any subject is lower, the overall effect of teachers who come in with a bachelors degree and an additional B.Ed is lower. This makes sense because post NTS, teachers did not need an education specific degree to be hired. It also makes sense that for female teachers hired post-2016, the probability of having an M. Ed or a masters and an M. Ed is both negative. Thus, what we see is more female teachers coming in with a masters degree, which is not necessarily in education and are more like to have a science background.

Table 2d further disentangles the effects of NTS on educational attainment of teachers in KP by splitting the dependent variables into sub-categories of education. In KP we find that post-2016, teachers' likelihood of having a B. Ed is higher and an M. Ed is lower. But for females, the likelihood of getting a degree in education, be it a B. Ed for an M. Ed is higher post 2016. This is true despite the fact that post NTS, teachers did not need an education specific degree to be hired. We also see whether more female teachers are coming in with a bachelors and a B. Ed or a masters and an M. Ed. Here we find that for female teachers in KP, the probability of having both a bachelors and a B. Ed is lower, while having both a masters and an M. Ed. is higher. Thus, what we see in KP is more female teachers coming in with a masters degree in science, a masters degree in education and/or both.

Does NTS affect teacher's pedagogical practices, content knowledge and intrinsic motivation?

Table 3 presents the effects of NTS reforms on teacher's score on the teach tool.[13] Controlling for teacher age, teacher CPD and induction training, school characteristics, class size, regional and

[13] Teach tool evaluates teacher performance on pedagogical practices. The tool has 9 components. These include: Supportive Learning Environment, Positive Behavioral Expectations, Lesson Facilitation, Checks for Understanding, Feedback, Critical Thinking, Autonomy, Perseverance, Social & Collaborative Skills.

provincial dummies, we find no significant effects of the reform on the overall teach score or any of its components.

On content knowledge, we find that although female teachers had higher content knowledge of literacy as compared to male teachers, for those hired post-2016, content knowledge of literacy was lower (Table 4a).[14] For math content knowledge, we find opposite effects. Overall, female teachers have lower content knowledge of math as compared to male teachers, but for those hired post-2016, their math content knowledge is higher. If we find that more female teachers are graduating in science subjects ex-post, this finding may make sense. Table 4b and 4c indeed show that it is the case in both KP and Punjab.

GEPD also collects information on teachers' intrinsic motivation through five ranking statements on why they joined the teaching profession. These include: (i) I always wanted to become a teacher, (ii) I like teaching, (iii) Teaching provides a steady career path, (iv) Teaching allows me to shape child and adolescent values, and (v) Teaching allows me to benefit the socially disadvantaged. Ranking on each question take a value from 0 to 2. To estimate the possible effects of NTS on intrinsic motivation, we create an index using all five statements on intrinsic motivation. First. we create a dummy variable which equals 1 if the ranking on the statement is greater than zero and zero otherwise, and then take a simple average of the five dummy variables—hence giving equal weightage to each component. Below we present results on whether the motivation of teachers changed on the intrinsic motivation index post reform (Table 5).[15] We find that the female teachers post-reform report lower intrinsic motivation. Similarly, more educated teachers post reform have lower intrinsic motivation.

Do teachers hired post 2016 produce better learning outcomes for students?

We find that the introduction of NTS-based teacher hiring reforms has had negligible effects on student literacy and numeracy outcomes. Despite the policy's aim to enhance teacher quality and, by extension, student learning, our analysis does not reveal any statistically significant improvements in student performance attributable to the reform. Specifically, students taught by

[14] On average, fewer female teachers were hired post 2016.
[15] We also present separate effects of the reform on each dimension in the Appendix Table A7.

teachers hired after 2016—when the NTS reforms were implemented—do not perform better in literacy or numeracy assessments compared to those taught by teachers hired prior to the reform, even when the newer teachers hold at least a bachelor's degree.

We also find consistent patterns in student performance across demographic and geographic lines. Students in urban areas outperform their rural peers in both literacy and numeracy, highlighting persistent disparities in educational opportunity and resource allocation. Gender differences are also evident: girls tend to achieve higher scores in literacy but lower scores in numeracy relative to boys.

Moreover, we find no statistically significant relationship between teachers' formal educational qualifications—such as holding a bachelor's or master's degree—and student learning outcomes. This challenges the assumption that academic credentials alone are a reliable proxy for teaching effectiveness. These findings suggest that improving student outcomes may require a broader focus on pedagogical training, classroom practices, and ongoing professional development, rather than relying solely on formal qualifications or standardized entry tests.

VII. Threats to Identification/Alternative Explanations:

To ensure the internal validity of the empirical strategy, several potential threats to identification and alternative explanations for the observed effects are considered and addressed.

First, one might be concerned that the estimated effects are confounded by concurrent policy reforms or structural changes at the provincial level. However, the empirical specification includes province fixed effects, which absorb all time-invariant provincial characteristics, as well as year fixed effects, which control for national-level shocks common across provinces. Given that most education policy in Pakistan is decentralized following the 18th Constitutional Amendment, the inclusion of province fixed effects is particularly important. These controls account for differences in provincial education budgets, administrative capacity, and any other fixed provincial characteristics. A review of policy documents and administrative data confirms that there were no major changes to higher education policy during the study period that could plausibly explain the observed outcomes.

Second, changes in resource availability within schools, such as shifts in teacher remuneration or school funding, could potentially bias the results. However, budgetary data from provincial education departments indicate that no significant changes occurred in education financing or teacher salary structures during the relevant period. Teacher compensation remained stable, and no new incentive schemes or performance-based pay reforms were introduced. Thus, macro-level changes in resource allocation are unlikely to be driving the results.

Third, the study period coincides with the implementation of teacher hiring reforms, which introduced a standardized recruitment process. Under the new system, all prospective teachers were required to pass a standardized test, effectively eliminating informal hiring channels and discretionary appointments. This reform created a uniform and merit-based entry mechanism into the teaching profession, and the analysis explicitly focuses on the period following this institutional change. As such, the observed effects are plausibly attributable to the standardized hiring process rather than to unobserved variation in recruitment practices.

Fourth, the potential impact of the COVID-19 pandemic is considered. While the pandemic began in late 2019, its economic and educational disruptions in Pakistan were most pronounced in 2020 and beyond. The primary period of analysis precedes these disruptions, and any residual effects are likely captured by the year fixed effects. Moreover, Pakistan experienced a relatively moderate economic contraction compared to global trends, and school closures were implemented unevenly across provinces. Given the timing and scope of the pandemic, it is unlikely to be a confounding factor in the analysis.

Finally, natural disasters, such as floods, which have affected various regions of Pakistan in recent years, are also considered. These events tend to be localized and are therefore captured by the province fixed effects. To the extent that such shocks are time-varying and province-specific, their influence is further mitigated by the inclusion of province-year interactions in robustness checks (if applicable). Consequently, the likelihood that natural disasters systematically bias the estimated treatment effects is minimal.

Taken together, these considerations support the credibility of the identification strategy and suggest that the estimated effects are not likely to be driven by confounding policy changes, macroeconomic shocks, or unobserved heterogeneity across provinces.

VIII. Conclusion

The introduction of standardized recruitment through the National Testing Service (NTS) in Pakistan marked a significant policy shift aimed at improving the quality of public sector teachers through merit-based hiring. This reform, implemented in 2016, was grounded in the belief that raising entry standards—particularly through subject-matter testing—would attract more qualified candidates and ultimately enhance student learning outcomes. Our analysis, drawing on quasi-experimental evidence from the 2023 Global Education Policy Dashboard and administrative census data, provides a nuanced picture of the reform's impact.

We find that the NTS reform did succeed in altering the composition of the teaching workforce. Teachers hired after the reform were more likely to hold postgraduate degrees and to have specialized in science-related fields. This shift was especially pronounced among female teachers, who were significantly more likely to possess advanced degrees and science majors, regardless of the level or subject they were assigned to teach. These changes suggest that the reform functioned as a selection filter, attracting more academically accomplished individuals into the profession.

However, these gains in formal qualifications did not translate into improved classroom practices or student learning outcomes. Teachers hired under the new system performed similarly to their predecessors in both pedagogical assessments and in terms of the literacy and numeracy outcomes of their students. This disconnect between teacher credentials and classroom effectiveness raises important questions about the design and focus of recruitment standards.

One key limitation of the NTS approach is its narrow emphasis on subject-matter knowledge—particularly in STEM fields—without adequately assessing candidates' pedagogical skills, communication abilities, or motivation to teach. While content knowledge is essential, it is not sufficient on its own to ensure effective teaching. The findings underscore that the substance of recruitment standards matters: a test that prioritizes content over teaching aptitude may inadvertently exclude candidates with strong instructional potential while favoring those with academic credentials but limited classroom effectiveness. Recruitment reforms must be complemented by robust initial teacher education (ITE) programs and ongoing professional development that equip teachers with the practical skills needed to translate knowledge into learning.

Going forward, teacher selection standards should be reoriented to emphasize a more holistic profile of teaching competence. This includes not only deep subject-matter expertise (ideally beyond the level they will teach) but also pedagogical knowledge, classroom management skills, and the ability to engage and motivate students. Recruitment processes should incorporate assessments of these competencies—through structured interviews, teaching demonstrations, or pedagogical tests—alongside content-based evaluations.

In parallel, ITE programs must be strengthened to ensure that incoming teachers are not only knowledgeable but also well-prepared to teach. This means integrating theory with practice, offering sustained classroom exposure, and fostering a strong professional identity. Without such alignment between recruitment, preparation, and practice, reforms risk becoming procedural rather than transformative.

References

Araujo P., Maria Daniela (2019) : Measuring the effect of competitive teacher recruitment on student achievement: Evidence from Ecuador, BERG Working Paper Series, No. 150, ISBN 978-3-943153-71-2, Bamberg University, Bamberg Economic Research Group (BERG), Bamberg

Aslam, M., Rawal, S., Kingdon, G. (2021). The Political Economy of Teachers in South Asia. In: Sarangapani, P.M., Pappu, R. (eds) Handbook of Education Systems in South Asia. Global Education Systems. Springer, Singapore. https://doi.org/10.1007/978-981-15-0032-9_20

Bari, F., Raza, R., Khan, B. H., Aslam, M., & Maqsood, N. (2015). An Investigation into Teacher Recruitment and Retention in Punjab. Institute of Development and Economic Alternatives (IDEAS).

Bedoya, Juan; de Hoyos, Rafael; Estrada, Ricardo. 2023. Rule-Based Civil Service: Evidence from a Nationwide Teacher Reform in Mexico. Policy Research Working Papers; 10540. © World Bank.

Brutti, Z., & Torres, F. S. (2022). Turning around teacher quality in Latin America: Renewed confidence and lessons from Colombia. Economic Analysis and Policy, 73, 62-93.

Busso, Matias, Montaño, Sebastián, Munoz-Morales, Juan and Pope, Nolan G., (2024), The Unintended Consequences of Merit-Based Teacher Selection: Evidence from a Large-Scale Reform in Colombia, No 17294, IZA Discussion Papers, Institute of Labor Economics (IZA)

Darling-Hammond, Linda, and Barnett Berry. “Recruiting Teachers for the 21st Century: The Foundation for Educational Equity.” *The Journal of Negro Education*, vol. 68, no. 3, 1999, pp. 254–79

Estrada, Ricardo. "Rules versus discretion in public service: Teacher hiring in Mexico." *Journal of Labor Economics* 37.2 (2019): 545-579.

Estrada, Ricardo, and María Lombardi. 2023. “Skills and Selection into Teaching: Evidence from Latin America.” Economía LACEA Journal 22(1): 218–239

Evans, David K., and Amina Mendez Acosta. "How to recruit teachers for hard-to-staff schools: A systematic review of evidence from low-and middle-income countries." Economics of Education Review 95 (2023): 102430.

Guarino, C. M., Santibañez, L., & Daley, G. A. (2006). Teacher Recruitment and Retention: A Review of the Recent Empirical Literature. Review of Educational Research, 76(2), 173-208.

Larsen, B. (2013). Occupational licensing and quality: Distributional and heterogeneous effects in the teaching profession. Available at SSRN 2387096.

Ricardo Estrada (2019). Rules versus Discretion in Public Service: Teacher Hiring in Mexico, CAF–Development Bank of Latin America

Wiswall, M. (2007). Licensing and occupational sorting in the market for teachers. Unpublished manuscript, Department of Economics, New York University.

Tables

Table 1: The impact of NTS on teacher education

Dependent variable: Teacher has more than a bachelors degree = 1	(1)	(2)	(3)
|Hiring year - 2016|	-0.036**	-0.032**	-0.033**
	(0.016)	(0.016)	(0.016)
|Hiring year - 2016| squared	0.002	0.001	0.001
	(0.001)	(0.001)	(0.001)
Teacher hired after 2016	0.013	-0.027	0.000
	(0.045)	(0.062)	(0.063)
Teacher female		-0.074	-0.080
		(0.056)	(0.057)
Teacher hired after 2016 * teacher female		0.147**	0.158**
		(0.074)	(0.075)
Urban	0.108***	0.110***	0.162***
	(0.040)	(0.040)	(0.055)
Urban*Teacher hired after 2016			-0.106
			(0.080)
Constant	0.925***	0.821***	0.797***
	(0.051)	(0.144)	(0.143)
Region fixed effects	Y	Y	Y
Province fixed effects	Y	Y	Y
Observations	533	533	533
R-squared	0.222	0.230	0.232

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. All regressions include province fixed effects (omitted = Punjab) and control for teachers' age. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table 2a: The impact of NTS on teacher composition

	(1)	(2)	(3)
Dependent Variable	Has more than a bachelors degree	Has a degree in science versus arts	Has a degree in education
|Hiring year - 2016|	0.002	-0.020***	-0.002**
	(0.002)	(0.005)	(0.001)
|Hiring year - 2016| squared	-0.001***	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Teacher hired after 2016	-0.054***	-0.154***	0.003*
	(0.006)	(0.019)	(0.001)
Teacher female	0.049***	-0.148***	0.009***
	(0.003)	(0.006)	(0.001)
Teacher hired after 2016 * teacher female	0.043***	0.090***	-0.009***
	(0.008)	(0.011)	(0.001)
Urban	0.021***	0.080***	0.014***
	(0.003)	(0.006)	(0.001)
Urban*Teacher hired after 2016	-0.046***	-0.103***	-0.015***
	(0.004)	(0.016)	(0.001)
Constant	0.571***	1.665***	1.083***
	(0.028)	(0.034)	(0.008)
District FE	Y	Y	Y
Observations	235,038	224,567	230,681
R-squared statistic	0.030	0.230	0.036

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using Punjab Census Data 2022.

Table 2b: The impact of NTS on teacher composition

	(1)	(2)	(3)	(4)
Dependent Variable	Has a B.A	Has a bachelors degree and a B. Ed	Has an M.A	Has a masters degree and an M. Ed
|Hiring year - 2016|	-0.000	-0.000	0.002	0.044***
	(0.002)	(0.002)	(0.002)	(0.003)
|Hiring year - 2016| squared	0.001***	0.001***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Teacher hired after 2016	0.055***	0.057***	-0.054***	-0.099***
	(0.006)	(0.006)	(0.006)	(0.008)
Teacher female	-0.048***	-0.044***	0.049***	0.133***
	(0.003)	(0.003)	(0.003)	(0.006)
Teacher hired after 2016 * teacher female	-0.044***	-0.038***	0.043***	-0.029***
	(0.008)	(0.008)	(0.008)	(0.008)
Urban	-0.021***	-0.015***	0.021***	0.041***
	(0.003)	(0.002)	(0.003)	(0.006)
Urban*Teacher hired after 2016	0.046***	0.034***	-0.046***	-0.028***
	(0.004)	(0.004)	(0.004)	(0.007)
Constant	0.430***	0.362***	0.571***	0.044
	(0.028)	(0.022)	(0.029)	(0.027)
District FE	Y	Y	Y	Y
Observations	235,043	230,673	235,043	230,673
R-squared statistic	0.030	0.032	0.030	0.105

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using Punjab Census Data 2022

Table 2c: The impact of NTS on teacher composition

	(1)	(2)	(3)
Dependent Variable	Has more than a bachelors degree	Has a degree in science versus arts	Has a degree in education
|Hiring year - 2016|	-0.040***	0.011***	-0.013***
	(0.002)	(0.003)	(0.003)
|Hiring year - 2016| squared	0.001***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)
Teacher hired after 2016	-0.165***	0.115***	-0.137***
	(0.014)	(0.011)	(0.012)
Teacher female	-0.030**	-0.095***	-0.046***
	(0.014)	(0.009)	(0.016)
Teacher hired after 2016 * teacher female	0.144***	-0.010	0.077***
	(0.015)	(0.010)	(0.011)
Urban	0.054***	0.056***	0.070***
	(0.011)	(0.009)	(0.010)
Urban*Teacher hired after 2016	-0.078***	0.012	-0.010
	(0.013)	(0.010)	(0.012)
Constant	0.550***	0.813***	0.880***
	(0.040)	(0.027)	(0.040)
District FE	Y	Y	Y
Observations	137,177	135,585	116,843
R-squared statistic	0.112	0.141	0.062

Notes: The sample is restricted to all teachers hired between the year 2000 to 2024. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using KP Census Data 2024.

Table 2d: The impact of NTS on teacher composition

	(1)	(2)	(3)	(4)
Dependent Variable	Has a B.A	Has a bachelors degree and a B. Ed	Has an M.A	Has a masters degree and an M. Ed
|Hiring year - 2016|	0.044***	0.014***	-0.040***	-0.032***
	(0.002)	(0.001)	(0.002)	(0.002)
|Hiring year - 2016| squared	-0.001***	-0.000***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)
Teacher hired after 2016	0.156***	0.052***	-0.165***	-0.220***
	(0.012)	(0.006)	(0.014)	(0.013)
Teacher female	-0.003	-0.008**	-0.031**	0.029*
	(0.009)	(0.003)	(0.014)	(0.016)
Teacher hired after 2016 * teacher female	-0.116***	-0.035***	0.145***	0.060***
	(0.011)	(0.005)	(0.015)	(0.011)
Urban	-0.025***	-0.002	0.054***	0.086***
	(0.009)	(0.004)	(0.011)	(0.011)
Urban*Teacher hired after 2016	0.049***	0.018**	-0.078***	-0.050***
	(0.011)	(0.007)	(0.013)	(0.012)
Constant	0.539***	0.187***	0.545***	0.427***
	(0.022)	(0.010)	(0.040)	(0.027)
District FE	Y	Y	Y	Y
Observations	137,607	116,843	137,607	116,843
R-squared statistic	0.122	0.033	0.112	0.070

Notes: The sample is restricted to all teachers hired between the year 2000 to 2024. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using KP Census Data 2024.

Table 3: The impact of NTS on standardized teacher scores on the Teach Tool

	(1)	(2)	(3)	(4)
|Hiring year - 2016|	-0.002	0.002	0.003	0.003
	(0.033)	(0.033)	(0.034)	(0.034)
|Hiring year - 2016| squared	-0.000	-0.000	-0.000	-0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Teacher hired after 2016	-0.071	-0.065	-0.082	-0.135
	(0.148)	(0.147)	(0.175)	(0.180)
Teacher is female	-0.102	-0.092	-0.076	-0.075
	(0.162)	(0.162)	(0.139)	(0.140)
Hired post 2016*female	0.056	0.029		
	(0.163)	(0.163)		
Has more than a bachelors degree		0.162	0.141	0.149
		(0.111)	(0.145)	(0.146)
Has more than a bachelors*Hired post 2016			0.043	0.035
			(0.193)	(0.193)
Urban	0.154	0.146	0.147	0.052
	(0.100)	(0.100)	(0.101)	(0.132)
Urban*Teacher hired after 2016				0.191
				(0.187)
Constant	-0.427	-0.524	-0.514	-0.479
	(0.359)	(0.361)	(0.359)	(0.358)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	530	530	530	530
R-squared statistic	0.144	0.148	0.148	0.150

Notes: The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table 4a: The impact of NTS on the Literacy content knowledge (standardized) of teachers

	(1)	(2)	(3)	(4)
|Hiring year - 2016|	-0.032	-0.028	-0.029	-0.028
	(0.038)	(0.038)	(0.039)	(0.039)
|Hiring year - 2016| squared	0.001	0.001	0.001	0.001
	(0.003)	(0.003)	(0.003)	(0.003)
Teacher hired after 2016	-0.125	-0.105	-0.195	-0.238
	(0.160)	(0.161)	(0.207)	(0.220)
Teacher is female	0.342**	0.359**	0.134	0.136
	(0.174)	(0.173)	(0.142)	(0.143)
Hired post 2016*female	-0.414**	-0.461**		
	(0.190)	(0.191)		
Has more than a bachelors degree		0.173	0.235	0.236
		(0.125)	(0.165)	(0.165)
Has more than a bachelors*Hired post 2016			-0.210	-0.208
			(0.220)	(0.220)
Urban	0.169	0.162	0.163	0.099
	(0.107)	(0.108)	(0.108)	(0.144)
Urban*Teacher hired after 2016				0.135
				(0.205)
Constant	0.724*	0.612	0.659	0.684*
	(0.378)	(0.386)	(0.406)	(0.405)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	347	347	347	347
R-squared statistic	0.255	0.261	0.251	0.252

Notes: The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table 4b: The impact of NTS on the Math content knowledge (standardized) of teachers
+G1:K26G28G1:K27G1:K27G1:K28G28G1:K27

	(1)	(2)	(3)	(4)
|Hiring year - 2016|	-0.075	-0.072	-0.075	-0.073
	(0.055)	(0.054)	(0.056)	(0.055)
|Hiring year - 2016| squared	0.003	0.003	0.003	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Teacher hired after 2016	-0.299	-0.325	-0.548	-0.480
	(0.269)	(0.265)	(0.347)	(0.358)
Teacher is female	-0.713**	-0.681**	-0.394*	-0.390*
	(0.275)	(0.273)	(0.220)	(0.218)
Hired post 2016*female	0.401	0.385		
	(0.249)	(0.246)		
Has more than a bachelors degree		0.343**	0.047	0.019
		(0.141)	(0.235)	(0.240)
Has more than a bachelors*Hired post 2016			0.514	0.552
			(0.336)	(0.337)
Urban	0.111	0.096	0.109	0.262
	(0.143)	(0.140)	(0.144)	(0.210)
Urban*Teacher hired after 2016				-0.282
				(0.266)
Constant	1.580**	1.442**	1.598**	1.543**
	(0.617)	(0.620)	(0.646)	(0.642)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	170	170	170	170
R-squared statistic	0.503	0.517	0.519	0.522

Notes: The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table 4c: The impact of NTS on teachers' majoring in science fields

	(1)	(2)	(3)	(4)
|Hiring year - 2016|	-0.020***	-0.019***	-0.020***	-0.020***
	(0.005)	(0.005)	(0.005)	(0.005)
|Hiring year - 2016| squared	-0.000	-0.001**	-0.001*	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)
Hired 2016 or after	-0.169***	-0.178***	-0.159***	-0.136***
	(0.021)	(0.021)	(0.020)	(0.018)
Female	-0.145***	-0.138***	-0.138***	-0.101***
	(0.007)	(0.006)	(0.006)	(0.011)
Hired 2016 or after * female	0.084***	0.091***	0.092***	0.047***
	(0.011)	(0.011)	(0.011)	(0.015)
Has more than a bachelors degree		-0.154***	-0.142***	-0.120***
		(0.006)	(0.008)	(0.007)
Female*Has more than a bachelors degree				-0.043***
				(0.010)
Has more than a bachelors*Hired 2016 or after			-0.025**	-0.052***
			(0.010)	(0.013)
Female*Has more than a bachelors*Hired 2016 or after				0.053***
				(0.015)
Urban	0.050***	0.051***	0.051***	0.051***
	(0.003)	(0.003)	(0.003)	(0.003)
Constant	1.667***	1.755***	1.742***	1.722***
	(0.034)	(0.036)	(0.034)	(0.035)
District FE	Y	Y	Y	Y
Observations	224,567	224,562	224,562	224,562
R-squared statistic	0.229	0.241	0.241	0.241

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using Punjab Census Data 2022.

Table 4d: The impact of NTS on teachers' majoring in science fields

	(1)	(2)	(3)	(4)
|Hiring year - 2016|	0.011***	0.007**	0.004	0.003
	(0.003)	(0.002)	(0.002)	(0.002)
|Hiring year - 2016| squared	-0.001***	-0.001***	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)
Hired 2016 or after	0.116***	0.099***	0.232***	0.251***
	(0.010)	(0.010)	(0.013)	(0.013)
Female	-0.096***	-0.098***	-0.090***	-0.111***
	(0.009)	(0.010)	(0.009)	(0.010)
Hired 2016 or after * female	-0.009	0.005	0.005	-0.088***
	(0.010)	(0.009)	(0.010)	(0.011)
Has more than a bachelors degree		-0.102***	-0.002	-0.015
		(0.009)	(0.008)	(0.011)
Female*Has more than a bachelors degree				0.027**
				(0.011)
Has more than a bachelors*Hired 2016 or after			-0.158***	-0.192***
			(0.013)	(0.015)
Female*Has more than a bachelors*Hired 2016 or after				0.128***
				(0.017)
Urban	0.062***	0.063***	0.059***	0.059***
	(0.008)	(0.009)	(0.009)	(0.008)
Constant	0.812***	0.869***	0.744***	0.756***
	(0.027)	(0.027)	(0.028)	(0.027)
District FE	Y	Y	Y	Y
Observations	135,585	135,585	135,585	135,585
R-squared statistic	0.141	0.150	0.155	0.158

Notes: The sample is restricted to all teachers hired between the year 2000 to 2024. The dependent variable takes a value of 1 if teachers major in science subjects and zero if they major in arts subjects. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using KP Census Data 2024

Table 5: Impact of NTS on teachers' intrinsic motivation index

	(1)	(2)	(3)	(4)
|Hiring year - 2016|	-0.005	-0.005	-0.006*	-0.006*
	(0.003)	(0.003)	(0.003)	(0.003)
|Hiring year - 2016| squared	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Teacher hired after 2016	0.003	0.004	0.021	0.018
	(0.013)	(0.013)	(0.017)	(0.018)
Teacher is female	-0.002	-0.001	-0.017	-0.017
	(0.016)	(0.016)	(0.013)	(0.013)
Hired post 2016*female	-0.025	-0.030*		
	(0.017)	(0.017)		
Has more than a bachelors degree		0.029***	0.050***	0.050***
		(0.010)	(0.014)	(0.014)
Has more than a bachelors*Hired post 2016			-0.045**	-0.045**
			(0.018)	(0.018)
Urban	0.003	0.001	-0.000	-0.005
	(0.010)	(0.010)	(0.010)	(0.013)
Urban*Teacher hired after 2016				0.010
				(0.018)
Constant	0.382***	0.365***	0.354***	0.356***
	(0.036)	(0.036)	(0.037)	(0.037)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	530	530	530	530
R-squared statistic	0.075	0.090	0.095	0.096

Notes: The dependent variable is the motivation index. Each motivation statement is used to create a dummy variable that takes a value of 1 if the ranking on it >=1, and a zero otherwise. All five dummy variables are averaged to make the index. The motivation statements include "I always wanted to become a teacher, I like teaching, Teaching provides a steady career path, Teaching allows me to shape child and adolescent values, Teaching allows me to benefit the socially disadvantaged". The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. We also run the specification with urban and post2016 interaction and the full model with all interactions (post2016 and urban, post 2016 and teacher female, and post 2016 and whether the teacher has more than a bachelors degree), and the effects are similar to the ones presented in the table. We therefore skip them for brevity. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table 6a: The effects of NTS on Student Literacy Scores

	(1)	(2)	(3)	(4)
	Standardized literacy Score			
Teacher characteristics	-	-	-	-
|Hiring year - 2016|	0.031	0.032	0.028	0.028
	(0.025)	(0.025)	(0.025)	(0.025)
|Hiring year - 2016| squared	-0.002	-0.002	-0.002	-0.002
-	(0.002)	(0.002)	(0.002)	(0.002)
Teacher hired after 2016	-0.061	-0.057	0.056	0.037
	(0.106)	(0.106)	(0.152)	(0.157)
Teacher is female	0.215*	0.219*	0.217*	0.221*
	(0.125)	(0.125)	(0.125)	(0.126)
Hired post 2016*female	0.054	0.045	0.057	0.051
	(0.128)	(0.128)	(0.128)	(0.128)
Has more than a bachelors degree		0.050	0.122	0.123
		(0.086)	(0.105)	(0.105)
Has more than a bachelors*Hired post 2016			-0.153	-0.152
			(0.151)	(0.151)
Urban	0.253***	0.249***	0.245***	0.217**
	(0.066)	(0.066)	(0.067)	(0.092)
Urban*Teacher hired after 2016				0.060
				(0.130)
Student characteristics				
Student age	-0.003	-0.003	-0.003	-0.003
	(0.011)	(0.011)	(0.011)	(0.011)
Student gender (female=1)	0.120**	0.119**	0.123**	0.122**
	(0.051)	(0.051)	(0.051)	(0.051)
Constant	-0.194	-0.233	-0.302	-0.285
	(0.315)	(0.309)	(0.315)	(0.318)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
Observations	9,060	9,060	9,060	9,060
R-squared	0.197	0.197	0.198	0.199

Notes: The regression includes a sample of Grade 4 students tested for literacy. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** $p<0.01$, ** $p<0.05$, * $p<0.1$). Source: Authors calculations based on GEPD 2023-24.

Table 6b: The effects of NTS on Student Numeracy Scoes

	(1)	(2)	(3)	(4)
	Standardized Numeracy Score			
Teacher characteristics	-	-	-	
|Hiring year - 2016|	0.032	0.035	0.032	0.032
	(0.025)	(0.025)	(0.025)	(0.025)
|Hiring year - 2016| squared	-0.002	-0.002	-0.002	-0.002
-	(0.002)	(0.002)	(0.002)	(0.002)
Teacher hired after 2016	0.041	0.052	0.133	0.139
	(0.116)	(0.116)	(0.156)	(0.164)
Teacher is female	0.089	0.100	0.098	0.097
	(0.127)	(0.125)	(0.126)	(0.124)
Hired post 2016*female	-0.029	-0.051	-0.043	-0.041
	(0.127)	(0.126)	(0.127)	(0.126)
Has more than a bachelors degree		0.127	0.178	0.178
		(0.083)	(0.117)	(0.117)
Has more than a bachelors*Hired post 2016			-0.109	-0.110
			(0.152)	(0.152)
Urban	0.298***	0.289***	0.286***	0.296***
	(0.062)	(0.062)	(0.062)	(0.089)
Urban*Teacher hired after 2016				-0.019
				(0.131)
Student characteristics				
Student age	-0.001	-0.001	-0.001	-0.001
	(0.013)	(0.013)	(0.013)	(0.013)
Student gender (female=1)	-0.168***	-0.170***	-0.167***	-0.167***
	(0.054)	(0.054)	(0.054)	(0.054)
Constant	-0.256	-0.355	-0.404	-0.410
	(0.351)	(0.350)	(0.349)	(0.355)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
Observations	9,060	9,060	9,060	9,060
R-squared	0.090	0.092	0.093	0.093

Notes: The regression includes a sample of Grade 4 students tested for literacy and numeracy. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Appendix

(i) Teachers Education

The first outcome is a dummy variable that takes on a value of 1 if teacher t in school s has more than a bachelors degree and a value of zero for teachers with a bachelors degree or lower. Overall, 10 percent of the teachers have less than a bachelors degree, 32 percent have a bachelors degree and the remaining 58 percent have more than a bachelors degree. The proportion of teachers with more than a bachelors degree varies significantly by whether they were hired pre- or post-2016. Pre-2016, 1 in 2 teachers was likely to have more than a bachelors degree and this increased to 7 in 10 teachers post 2016.

(ii) Teaching Practices

Here we use the 'teach score' as the outcome variable, estimated using the teach tool, which is a classroom observation tool to test teacher pedagogy and teaching practices.[16] Teach score is a continuous variable which takes a value from 1 to 5 and is the average of three dimensions. The first dimension is *Classroom culture*, which is the average score on (i) supportive learning environment and (ii) positive behavioral expectations. The second dimension is *Instruction*, which is the average score on (i) lesson facilitation, (ii) checks for understanding, (iii) feedback, and (iv) critical thinking. The third dimension is *Socioemotional skills*, which is the average score on (i) autonomy, (ii) perseverance, and (iii) social and collaborative skills.

The average teach score among all GEPD teachers is 2.7 and this does not vary much by teacher hiring year. In terms of provinces, the highest score is in the Capital of Islamabad (3.1) and lowest is in Punjab (2.6). It is important to note that the teach score does not vary by whether the teacher was hired before or after 2016. We also use the three main components of the teach score as outcome variables, namely: (i) classroom culture, (ii) instruction, and (iii) socio-emotional skills, and the results are consistent (available on request).

[16] See https://www.worldbank.org/en/topic/education/brief/teach-helping-countries-track-and-improve-teaching-quality for details.

(iii) Teacher Content Knowledge

This variable measures the teacher's knowledge of verbal and math content. In GEPD a subset of teachers was tested on either numeracy or literacy. The content knowledge variable is a continuous variable, and teachers can score between 0 and 100. Anyone attaining 80 percent or more is considered to have mastered the content. Verbal content knowledge of GEPD teachers is 60 percent and math content knowledge are 48 percent on average. Given this, only 21 percent of the tested teachers have mastered literacy content and 13 percent have mastered math content. Overall, teachers hired post-2016 have higher math and lower verbal content knowledge, but these differences are not statistically significant.

(iv) Teacher's intrinsic motivation

The intrinsic motivation is an index based on responses to five 'motivation statements'. Every teacher responds to each statement with a ranking. These responses are used to create a dummy variable that takes a value of 1 if the ranking on it >=1, and a zero otherwise. All five dummy variables are averaged to arrive at the motivation index. The motivation statements read: (i) I always wanted to become a teacher, (ii) I like teaching, (iii) Teaching provides a steady career path, (iv) Teaching allows me to shape child and adolescent values, and (v) Teaching allows me to benefit the socially disadvantaged.

Student Outcomes

GEPD tested 4th-grade students on their mathematics and verbal skills. The assessment module is based on the Service Delivery Indicators (SDI) Survey Instruments with some adaptations to the model and questions. First, the assessments were designed to be conducted in a group setting rather than one-on-one, to simplify data collection for the enumerators and reduce costs. Second, 11 new literacy items were added to align the instrument better with the Global Proficiency Framework (GPF) developed by a multiagency partnership for the purpose of monitoring progress toward Sustainable Development Goal 4.1. These items were selected from among the publicly released PIRLS items for 4th grade. In total, the assessment included the following items for literacy and numeracy.

i. 24 Literacy Items: • Letter Identification (3 items) • Word Recognition (7 items) • Reading Comprehension Story (3 items) • Reading Comprehension Story II (11 items).
ii. 15 Numeracy Items: Number Sense (4 items) • Arithmetic (11 items) • Word Problem (1 item) • Sequences (1 item).

School Infrastructure Index

The school infrastructure index is based on the average of the following four components:

i. *Drinking water*, which is based on the main source of drinking water available at the school. This is binary variable that takes on the value 1 if the source is piped water, protected well/spring, packaged bottled water, tanker-truck, or cart and 0 otherwise.
ii. *Functioning toilet*, which is a binary indicator that takes on a value of 1 if the following conditions are met: separate toilet exists for boys and girls, toilet is clean, toilet is private, usable and handwashing is available. It takes on the value 0 if otherwise.
iii. *Disability accessibility*, which the average of the following seven binary variables: accessible road, a wheelchair ramp at the entrance, an entrance wide enough to fit a ramp, class entrance accessible disabled students, accessible toilet, and disability screening at school.

Internet, a binary variable indicating whether the PCs and laptops at schools have internet connectivity.

Appendix Tables

Table A1: The impact of NTS on teacher education

Dependent variable: Teacher has more than a bachelors degree = 1	(1)	(2)	(3)
Teacher hired after 2016	0.051	-0.001	0.023
	(0.040)	(0.061)	(0.062)
Teacher female		-0.076	-0.081
		(0.056)	(0.057)
Teacher hired after 2016 * teacher female		0.162**	0.172**
		(0.074)	(0.076)
Urban	0.097**	0.098**	0.142**
	(0.040)	(0.040)	(0.056)
Urban*Teacher hired after 2016			-0.091
			(0.081)
Constant	0.815***	0.759***	0.741***
	(0.037)	(0.138)	(0.137)
Region fixed effects	Y	Y	Y
Province fixed effects	Y	Y	Y
Observations	533	533	533
R-squared	0.210	0.219	0.220

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. All regressions include province fixed effects (omitted = Punjab) and control for teachers' age. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table A2a: The impact of NTS on teacher composition

	(1)	(2)	(3)
Dependent Variable	Has more than a bachelors degree	Has a degree in science versus arts	Has a degree in education
Teacher hired after 2016	-0.049***	-0.112***	0.006***
	(0.006)	(0.016)	(0.001)
Teacher female	0.050***	-0.148***	0.009***
	(0.004)	(0.006)	(0.001)
Teacher hired after 2016 * teacher female	0.042***	0.091***	-0.009***
	(0.009)	(0.011)	(0.001)
Urban	0.021***	0.080***	0.014***
	(0.003)	(0.007)	(0.001)
Urban*Teacher hired after 2016	-0.048***	-0.104***	-0.015***
	(0.005)	(0.018)	(0.001)
Constant	0.709***	1.991***	1.097***
	(0.024)	(0.030)	(0.008)
District FE	Y	Y	Y
Observations	235,038	224,567	230,681
R-squared statistic	0.024	0.213	0.035

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using Punjab Census Data 2022.

Table A2b: The impact of NTS on teacher composition

	(1)	(2)	(3)	(4)
Dependent Variable	Has a B.A	Has a bachelors degree and a B. Ed	Has an M.A	Has a masters degree and an M. Ed
Teacher hired after 2016	0.048***	0.052***	-0.048***	-0.161***
	(0.006)	(0.006)	(0.006)	(0.007)
Teacher female	-0.049***	-0.044***	0.050***	0.135***
	(0.004)	(0.003)	(0.004)	(0.007)
Teacher hired after 2016 * teacher female	-0.043***	-0.037***	0.042***	-0.034***
	(0.009)	(0.008)	(0.009)	(0.008)
Urban	-0.021***	-0.015***	0.021***	0.043***
	(0.003)	(0.002)	(0.003)	(0.006)
Urban*Teacher hired after 2016	0.047***	0.035***	-0.048***	-0.032***
	(0.005)	(0.004)	(0.005)	(0.008)
Constant	0.300***	0.260***	0.709***	-0.019
	(0.024)	(0.019)	(0.024)	(0.024)
District FE	Y	Y	Y	Y
Observations	235,043	230,673	235,043	230,673
R-squared statistic	0.025	0.028	0.024	0.098

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using Punjab Census Data 2022

Table A2c: The impact of NTS on teacher composition

	(1)	(2)	(3)
Dependent Variable	Has more than a bachelors degree	Has a degree in science versus arts	Has a degree in education
Teacher hired after 2016	-0.068***	0.102***	-0.096***
	(0.011)	(0.010)	(0.010)
Teacher female	-0.052***	-0.095***	-0.060***
	(0.014)	(0.010)	(0.017)
Teacher hired after 2016 * teacher female	0.161***	-0.009	0.084***
	(0.014)	(0.010)	(0.011)
Urban	0.052***	0.057***	0.070***
	(0.013)	(0.009)	(0.010)
Urban*Teacher hired after 2016	-0.062***	0.008	-0.004
	(0.015)	(0.011)	(0.012)
Constant	0.404***	0.868***	0.895***
	(0.047)	(0.030)	(0.041)
District FE	Y	Y	Y
Observations	137,177	135,585	116,843
R-squared statistic	0.064	0.140	0.052

Notes: The sample is restricted to all teachers hired between the year 2000 to 2024. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using KP Census Data 2024.

Table A2d: The impact of NTS on teacher composition

	(1)	(2)	(3)	(4)
Dependent Variable	Has a B.A	Has a bachelors degree and a B. Ed	Has an M.A	Has a masters degree and an M. Ed
Teacher hired after 2016	0.065***	0.022***	-0.068***	-0.146***
	(0.008)	(0.004)	(0.011)	(0.010)
Teacher female	0.014	-0.001	-0.052***	0.010
	(0.009)	(0.003)	(0.014)	(0.016)
Teacher hired after 2016 * teacher female	-0.129***	-0.037***	0.162***	0.066***
	(0.011)	(0.005)	(0.014)	(0.011)
Urban	-0.023**	-0.001	0.052***	0.085***
	(0.010)	(0.005)	(0.013)	(0.012)
Urban*Teacher hired after 2016	0.031**	0.013	-0.062***	-0.038***
	(0.014)	(0.008)	(0.015)	(0.013)
Constant	0.718***	0.212***	0.400***	0.388***
	(0.029)	(0.013)	(0.047)	(0.024)
District FE	Y	Y	Y	Y
Observations	137,607	116,843	137,607	116,843
R-squared statistic	0.079	0.020	0.065	0.049

Notes: The sample is restricted to all teachers hired between the year 2000 to 2024. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using KP Census Data 2024.

Table A3: The impact of NTS on standardized teacher scores on the Teach Tool

	(1)	(2)	(3)	(4)
Teacher hired after 2016	-0.064	-0.061	-0.078	-0.134
	(0.143)	(0.143)	(0.173)	(0.179)
Teacher is female	-0.100	-0.090	-0.074	-0.074
	(0.161)	(0.161)	(0.138)	(0.138)
Hired post 2016*female	0.059	0.030		
	(0.163)	(0.162)		
Has more than a bachelors degree		0.163	0.143	0.151
		(0.111)	(0.145)	(0.145)
Has more than a bachelors*Hired post 2016			0.043	0.034
			(0.186)	(0.186)
Urban	0.151	0.144	0.146	0.051
	(0.099)	(0.099)	(0.100)	(0.131)
Urban*Teacher hired after 2016				0.193
				(0.186)
Constant	-0.417	-0.508	-0.495	-0.463
	(0.341)	(0.340)	(0.341)	(0.339)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	530	530	530	530
R-squared statistic	0.144	0.148	0.148	0.150

Notes: The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table A4a: The impact of NTS on the Literacy content knowledge (standardized) of teachers

	(1)	(2)	(3)	(4)
Teacher hired after 2016	-0.084	-0.068	-0.171	-0.225
	(0.158)	(0.159)	(0.207)	(0.220)
Teacher is female	0.353**	0.370**	0.149	0.151
	(0.176)	(0.175)	(0.143)	(0.144)
Hired post 2016*female	-0.392**	-0.449**		
	(0.192)	(0.192)		
Has more than a bachelors degree		0.196	0.244	0.244
		(0.124)	(0.164)	(0.164)
Has more than a bachelors*Hired post 2016			-0.181	-0.180
			(0.220)	(0.221)
Urban	0.149	0.145	0.146	0.070
	(0.107)	(0.108)	(0.108)	(0.144)
Urban*Teacher hired after 2016				0.163
				(0.204)
Constant	0.677*	0.556	0.607	0.640
	(0.363)	(0.370)	(0.394)	(0.392)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	347	347	347	347
R-squared statistic	0.249	0.256	0.246	0.248

Notes: The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table A4b: The impact of NTS on the Math content knowledge (standardized) of teachers

	(1)	(2)	(3)	(4)
Teacher hired after 2016	-0.207	-0.231	-0.479	-0.417
	(0.261)	(0.256)	(0.335)	(0.349)
Teacher is female	-0.708**	-0.674**	-0.380*	-0.374*
	(0.278)	(0.277)	(0.224)	(0.221)
Hired post 2016*female	0.415*	0.399		
	(0.249)	(0.247)		
Has more than a bachelors degree		0.341**	0.026	-0.004
		(0.137)	(0.234)	(0.239)
Has more than a bachelors*Hired post 2016			0.549*	0.589*
			(0.326)	(0.328)
Urban	0.085	0.069	0.086	0.229
	(0.140)	(0.137)	(0.141)	(0.206)
Urban*Teacher hired after 2016				-0.267
				(0.264)
Constant	1.354**	1.237**	1.374**	1.344**
	(0.624)	(0.618)	(0.630)	(0.631)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	170	170	170	170
R-squared statistic	0.493	0.507	0.509	0.512

Notes: The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table A4c: The impact of NTS on teachers' majoring in science fields

	(1)	(2)	(3)	(4)
Hired 2016 or after	-0.127***	-0.135***	-0.146***	-0.121***
	(0.018)	(0.018)	(0.019)	(0.018)
Female	-0.145***	-0.138***	-0.138***	-0.107***
	(0.007)	(0.006)	(0.006)	(0.011)
Hired 2016 or after * female	0.085***	0.091***	0.090***	0.042***
	(0.011)	(0.011)	(0.011)	(0.015)
Has more than a bachelors degree		-0.140***	-0.146***	-0.128***
		(0.007)	(0.009)	(0.008)
Female*Has more than a bachelors degree				-0.036***
				(0.010)
Has more than a bachelors*Hired 2016 or after			0.013	-0.018
			(0.011)	(0.012)
Female*Has more than a bachelors*Hired 2016 or after				0.058***
				(0.015)
Urban	0.049***	0.050***	0.051***	0.050***
	(0.004)	(0.003)	(0.003)	(0.003)
Constant	1.994***	2.094***	2.100***	2.084***
	(0.031)	(0.032)	(0.032)	(0.032)
District FE	Y	Y	Y	Y
Observations	224,567	224,562	224,562	224,562
R-squared statistic	0.212	0.221	0.221	0.222

Notes: The sample is restricted to all teachers hired between the year 2000 to 2022. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using Punjab Census Data 2022.

Table A4d: The impact of NTS on teachers' majoring in science fields

	(1)	(2)	(3)	(4)
Hired 2016 or after	0.103***	0.096***	0.233***	0.253***
	(0.009)	(0.009)	(0.013)	(0.012)
Female	-0.096***	-0.101***	-0.091***	-0.113***
	(0.010)	(0.010)	(0.009)	(0.009)
Hired 2016 or after * female	-0.009	0.007	0.006	-0.086***
	(0.010)	(0.009)	(0.010)	(0.010)
Has more than a bachelors degree		-0.102***	-0.000	-0.013
		(0.009)	(0.009)	(0.011)
Female*Has more than a bachelors degree				0.028**
				(0.011)
Has more than a bachelors*Hired 2016 or after			-0.161***	-0.194***
			(0.013)	(0.015)
Female*Has more than a bachelors*Hired 2016 or after				0.128***
				(0.016)
Urban	0.061***	0.063***	0.059***	0.059***
	(0.008)	(0.009)	(0.008)	(0.008)
Constant	0.867***	0.909***	0.763***	0.774***
	(0.030)	(0.029)	(0.027)	(0.026)
District FE	Y	Y	Y	Y
Observations	135,585	135,585	135,585	135,585
R-squared statistic	0.140	0.150	0.155	0.158

Notes: The sample is restricted to all teachers hired between the year 2000 to 2024. The dependent variable takes a value of 1 if teachers major in science subjects and zero if they major in arts subjects. Teacher characteristics include teacher age. Robust standard errors are clustered at the level of the district (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations using KP Census Data 2024

Table A5: Impact of NTS on teachers' intrinsic motivation index

	(1)	(2)	(3)	(4)
Teacher hired after 2016	0.007	0.007	0.022	0.018
	(0.012)	(0.012)	(0.017)	(0.018)
Teacher is female	-0.002	-0.001	-0.015	-0.015
	(0.016)	(0.016)	(0.013)	(0.013)
Hired post 2016*female	-0.023	-0.028*		
	(0.017)	(0.017)		
Has more than a bachelors degree		0.031***	0.049***	0.050***
		(0.010)	(0.014)	(0.014)
Has more than a bachelors*Hired post 2016			-0.039**	-0.040**
			(0.018)	(0.018)
Urban	0.001	-0.001	-0.002	-0.008
	(0.010)	(0.010)	(0.010)	(0.013)
Urban*Teacher hired after 2016				0.012
				(0.018)
Constant	0.369***	0.352***	0.341***	0.343***
	(0.033)	(0.034)	(0.035)	(0.035)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
N	530	530	530	530
R-squared statistic	0.069	0.086	0.089	0.090

Notes: The dependent variable is the motivation index. Each motivation statement is used to create a dummy variable that takes a value of 1 if the ranking on it >=1, and a zero otherwise. All five dummy variables are averaged to make the index. The motivation statements include "I always wanted to become a teacher, I like teaching, Teaching provides a steady career path, Teaching allows me to shape child and adolescent values, Teaching allows me to benefit the socially disadvantaged". The sample is restricted to all teachers hired from 2000 to 2022. All regressions include province (omitted province = Punjab) and region (urban=1) fixed effects. All regressions also include teacher characteristics and school characteristics. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. We also run the specification with urban and post2016 interaction and the full model with all interactions (post2016 and urban, post 2016 and teacher female, and post 2016 and whether the teacher has more than a bachelors degree), and the effects are similar to the ones presented in the table. We therefore skip them for brevity. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table A6a: The effects of NTS on Student Literacy Scores

	(1)	(2)	(3)	(4)
	Standardized literacy Score			
Teacher characteristics				
Teacher hired after 2016	-0.073	-0.071	0.062	0.041
	(0.102)	(0.102)	(0.151)	(0.157)
Teacher is female	0.217*	0.221*	0.218*	0.222*
	(0.126)	(0.127)	(0.126)	(0.127)
Hired post 2016*female	0.044	0.036	0.050	0.044
	(0.128)	(0.128)	(0.128)	(0.128)
Has more than a bachelors degree		0.044	0.128	0.129
		(0.087)	(0.104)	(0.105)
Has more than a bachelors*Hired post 2016			-0.177	-0.175
			(0.150)	(0.150)
Urban	0.260***	0.257***	0.251***	0.222**
	(0.065)	(0.066)	(0.066)	(0.091)
Urban*Teacher hired after 2016				0.062
				(0.128)
Student characteristics				
Student age	-0.002	-0.003	-0.002	-0.002
	(0.012)	(0.012)	(0.012)	(0.012)
Student gender (female=1)	0.118**	0.118**	0.123**	0.122**
	(0.051)	(0.051)	(0.051)	(0.051)
Constant	-0.094	-0.126	-0.219	-0.202
	(0.294)	(0.290)	(0.301)	(0.303)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
Observations	9,060	9,060	9,060	9,060
R-squared	0.196	0.196	0.197	0.197

Notes: The regression includes a sample of Grade 4 students tested for literacy. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.

Table A6b: The effects of NTS on Student Numeracy Scoes

	(1)	(2)	(3)	(4)
	Standardized Numeracy Score			
Teacher characteristics	-	-	-	
Teacher hired after 2016	0.024	0.032	0.134	0.142
	(0.112)	(0.112)	(0.155)	(0.164)
Teacher is female	0.089	0.099	0.097	0.096
	(0.128)	(0.126)	(0.126)	(0.125)
Hired post 2016*female	-0.040	-0.062	-0.051	-0.048
	(0.128)	(0.127)	(0.128)	(0.126)
Has more than a bachelors degree		0.119	0.183	0.183
		(0.082)	(0.117)	(0.117)
Has more than a bachelors*Hired post 2016			-0.136	-0.137
			(0.152)	(0.152)
Urban	0.307***	0.300***	0.295***	0.305***
	(0.062)	(0.062)	(0.062)	(0.087)
Urban*Teacher hired after 2016				-0.021
				(0.128)
Student characteristics				
Student age	-0.000	-0.001	-0.000	0.000
	(0.013)	(0.013)	(0.013)	(0.013)
Student gender (female=1)	-0.168***	-0.171***	-0.167***	-0.166***
	(0.054)	(0.054)	(0.054)	(0.054)
Constant	-0.162	-0.249	-0.320	-0.326
	(0.332)	(0.330)	(0.332)	(0.337)
Teacher characteristics	Y	Y	Y	Y
School characteristics	Y	Y	Y	Y
Province fixed effects	Y	Y	Y	Y
Observations	9,060	9,060	9,060	9,060
R-squared	0.088	0.091	0.091	0.091

Notes: The regression includes a sample of Grade 4 students tested for literacy and numeracy. Teacher characteristics include teacher's age, whether the teacher has acquired induction or CPD training, as well as the class-size of teacher. School characteristics include whether the school is a girl's school or a co-education school. The omitted category is boys school. A variable capturing school infrastructure quality (an index of 5 separate variables on toilet facility, drinking water, internet facility, disability access facility and electricity in the school) is also included in all regressions. All regressions are unweighted. Robust standard errors are clustered at the level of the school (*** p<0.01, ** p<0.05, * p<0.1). Source: Authors calculations based on GEPD 2023-24.