Essays on Economics of Education and Agriculture in Developing Countries

ANUSHA GUHA

Thesis submitted to the Department of Economics
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy

UNIVERSITY COLLEGE LONDON

August, 2024

Declaration

I, Anusha Guha, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Anusha Guha
London, United Kingdom
August, 2024

Abstract

This thesis explores two critical aspects of development: education and agriculture.

The first chapter examines teacher effectiveness in Vietnam. Using a panel dataset, we estimate both classroom and teacher effects on academic and non-cognitive skills. Our findings reveal moderate teacher impacts on academic skills, but substantial classroom effects on both cognitive and non-cognitive skills. Through analysis of classroom recordings, we identify key pedagogical practices that contribute to high-quality learning environments. Results highlight significant differences in the specific areas of classroom practice that influence different developmental outcomes.

Given that teacher quality alone explains limited variation in students' attainment, we posit that home inputs and their interaction with school inputs play a crucial role. The second chapter thus investigates the joint role of home and school inputs in the skill formation. It considers parents' time and material investment and child time investment as home inputs, and class value-added as school inputs. Results indicate that school inputs and parental time investment significantly influence cognitive skills, while material investment and child time investment are more important for non-cognitive skills development. We also find that home and school inputs are substitutes in the production of child skills. Notably, parents decrease investment in their children in response to higher classroom quality.

The last chapter shifts focus to agriculture, examining the competition among intermediaries and its impact on agricultural prices by leveraging a novel source of exogenous variation in competition - the introduction of online platforms. Using a unique dataset that captures both inter- and intra-market competition, we demonstrate that trader participation and spatial competition significantly affect farmer prices. We then evaluate a policy initiative which transitioned sales from an offline to an online platform. By investigating these less explored channels, it offers new perspectives on the factors shaping farm-gate prices in developing economies and the potential of online platforms to enhance market outcomes.

Impact Statement

This thesis provides significant insights into two critical areas of development economics: education and agriculture, with the aim of informing policies that can foster economic development in less advanced economies.

The first two chapters address a significant gap in the literature concerning teacher effectiveness and the interplay of home and school inputs in child skill formation in Lower- and Middle-Income Countries (LMICs). While extensive research exists on teacher value-added in high-income countries, evidence from LMICs is scarce. Our study in Vietnam provides novel evidence on teacher impacts across a broad spectrum of skills, including non-cognitive abilities crucial for long-term outcomes. This research is particularly timely given Vietnam's ongoing education reforms aimed at equipping students with skills needed for a modern economy.

The findings on key pedagogical practices, in the first chapter, provide policymakers with crucial insights into the characteristics that distinguish high-quality learning environments. This understanding helps identify specific aspects of the classroom environment to target for improvement. Moreover, the research highlights how various dimensions of educational quality contribute to the development of a wide range of skills essential for children's long-term success, enabling policymakers to strategically enhance performance in areas of policy interest.

The second chapter on the interplay between home and school inputs highlights the importance of considering household responses when estimating education production functions. First, whether parents believe home investments to be complements or substitutes to school investments in child human capital production will play a key role in determining the impact of any government investments in quality of schooling on pupil outcomes. The effect of such policies will depend on the crowding-in or crowding-out of parents' home investment. Second, if households adjust their home inputs in response to changes in school inputs, not taking these responses into account in the estimation of the education production function produces estimates of policy effects rather than production function parameters since they include behavioural responses by the parents (Todd and Wolpin (2003)).

The final chapter examines under-explored channels influencing prices in the agriculture sector, a critical livelihood source in many developing countries. By quantifying the impact of restricted market participation, government regulations, and geographical market concentration, the thesis provides valuable estimates of potential gains from market reforms. The findings on intermediaries competition have implications beyond agriculture, as similar institutions exist in other contexts. The chapter also evaluates a policy shifting sales from offline to online platforms, offering insights into how online platforms can enhance market outcomes by increasing competition and dismantling collusive practices. In the context of India's push towards electronic unified markets, this research provides valuable estimations of potential gains from such reforms.

By bridging gaps in the understanding of teacher effectiveness, home-school input interactions, and agricultural market dynamics, this thesis provides policymakers with evidence-based insights for designing more effective interventions in education and agriculture.

Statement of Conjoint Work

Note on the joint work in Anusha Guha's thesis "Essays on Economics of Education and Agriculture in Developing Countries":

The chapter, "Measuring the impacts of teachers in Vietnam: teacher value-added and student skills" is co-authored work between Anusha Guha, Pedro Carneiro, Paul Glewwe and Sonya Krutikova. Guha and Krutikova took the lead in conducting the analysis and writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript. The chapter, "The Joint Role of Home and School inputs in Children's Learning" is co-authored work between Anusha Guha, Pedro Carneiro, Michele Giannola and Sonya Krutikova. Each author contributed equally. The chapter, "Market Reform and Farmer Prices: The Case of Indian Agriculture" is single authored by Anusha Guha.

Candidate signature: Anusha Guha
Date: November 28, 2024
Senior Author signature: Pedro Carneiro
Date: November 28, 2024

Acknowledgements

This thesis marks the end of a challenging but beautiful roller coaster of learning, excitement, wicked challenges, with a few loops of frustration and stretches of routine along the way.

First and foremost, I want to express my heartfelt gratitude to my advisors, Aureo de Paula and Orazio Attanasio, for their generous support and encouragement, particularly during times when I felt lost. I often approached our meetings with a bit of nervousness, but I am really grateful for all the time and support they have given me right from the start.

For a considerable duration of my time as a PhD student, I have been working as a Research Officer at the Institute for Fiscal Studies (IFS). I cannot think of a more stimulating and supportive place to complement UCL. I am grateful to Sonya Krutikova and all other colleagues at IFS, who broadened my learning to real world applications.
I am also very grateful to Pedro Carneiro. Pedro is a brilliant researcher and an incredibly generous person. Working with him and Sonya has been one of the best experiences of my PhD, and I am very fortunate to be able to learn from them.

I am deeply appreciative of my co-authors, Michele Giannola and Paul Glewwe, with whom I've had the privilege of collaborating over the past few years. I would also like to thank my peers in the research community at UCL. I feel very lucky to have met not only amazing researchers but also made lifelong friends in Dajana, Morgane, Uta, Mia, Anna, Alison, Lorenzo, Alessandro, Moritz, Justus, Thomas, Elena and Vlad. I have immensely benefited from our conversations about life, not just academia.

I must acknowledge the deep gratitude that I owe to my family. Ultimately, this was only possible due to their unwavering support, manifested through daily calls from hours ahead and regular visits from thousands of miles away. I feel the utmost gratitude toward my mother, Sumi, and my father, Alok, whose belief in me has been a constant source of strength. I dedicate this work to my parents, whose encouragement pushed me forward and lifted me up, even in moments of doubt. I am also deeply grateful for my dear siblings, Ayush, Ayesha, and Jiya and my aunt, Rennie; your support has been invaluable. My heartfelt gratitude to my dear friends Surabhi and Atul, who have probably seen me distressed by my work and have ensured that I don't take myself too seriously. Last but not least, I want to thank my incredible partner Suyash for being there during my toughest times and for helping me see this through to a new chapter of our lives together.

Contents

List of Tables

List of Figures

Chapter 1

Measuring the impacts of teachers in Vietnam: teacher value-added and student skills

1.1 Introduction

The impact of teacher quality on student outcomes has been a focal point of education research for decades. There is a growing body of literature showing that access to high quality teachers has not only short-term gains for students at all levels of schooling but can also have lasting positive effects on later-life outcomes (Chetty, Friedman, and Rockoff (2014)). These findings, initially centered on High Income Countries (HICs) such as the United States, have recently expanded to include Lower- and Middle-Income Countries (LMICs) such as Ecuador, India, and Pakistan (Bau and Das (2020); Singh (2015); Araujo et al. (2016)).

However, this literature has two significant limitations. First, it primarily focuses on teacher effectiveness in improving literacy and numeracy skills, neglecting the broader spectrum of skills and capabilities crucial for long-run outcomes (Heckman, Stixrud, and Urzua (2006)). Second, while existing research quantifies the difference that having a better teacher makes for learning outcomes, it offers limited insights on what are the key pedagogical practices that goes into the production of high-quality school learning environments.

These gaps in our understanding are even more pronounced in LMICs, where there is a pressing need to improve overall learning outcomes. The question of how much schools and teachers contribute to the development of non-academic skills remains largely unexplored. Moreover, it is not enough for policy makers to know what test score gains can be achieved through moving a child from a low quality to a high-quality classroom. They also need to know : (1) what distinguishes higher and lower quality school learning environments and, therefore, what areas of classroom environment to target; and (2) how different dimensions of quality map into the development of wide range of skills that children need for long-term success and, therefore, how to boost performance in the dimensions that are of particular policy interest.

Our study addresses these critical gaps by examining teacher effectiveness in Vietnam, an outlier among LMICs due to its exceptionally high levels of learning achievement relative to its income level (Dang et al. (2023)). Vietnam's highly centralized education system with respect to teacher recruitment, training and curriculum provides an interesting setting to examine these questions. It is not clear, a priori, how much variation exists in teacher effectiveness and whether having a higher or lower quality teacher matters as much as has been found in other LMIC contexts.

We approach these questions by using several key methodologies. We first estimate teacher effectiveness as the teacher's contribution to student learning in an academic year. By utilizing test scores from two cohorts taught by the same teacher, we can differentiate teacher effects from classroom effects. This method allows us to isolate the specific impact of individual teachers on student outcomes. We then assess whether teachers who are effective at raising academic cognitive skills are equally adept at fostering broader cognitive and non-cognitive development.

Building on this, we then identify the key ingredients of high-quality classrooms by examining specific features of the classroom environment that are crucial for student learning. This part of our research provides robust empirical evidence in the domain where data remains sparse, contributing to a more comprehensive understanding of what makes an effective learning environment.

Our analysis here is grounded in the education literature, which distinguishes between two broad domains of school quality - *structural* and *process*. Structural quality is comprised of what are generally seen as more distal but easy to regulate determinants of children's outcomes such as teacher qualifications and classroom setting characteristics (e.g. class size and quality of facilities). However, existing evidence suggests that in fact they are poor stand-alone predictors of student achievement and development (Hanushek (2003); Hanushek and Rivkin (2006); Ganimian and Murnane (2016)). Process quality, on the other hand, captures what are likely to be more proximal determinants of children's outcomes such as pedagogical practices, quality of the curriculum and quality of interaction with the teachers. While evidence on the role of process quality is sparse, it suggests that they have much higher explanatory power (Araujo et al. (2016); Hamre et al. (2014)) and that the benefits of classroom quality for children's outcomes are not necessarily uniform across outcomes (Zaslow et al. (2016)).

To this end we leverage unique panel data from 140 primary schools and over 10,000 primary school pupils in Vietnam. The data include extensive classroom observation, one-to-one interviews with teachers, and a variety of pupil assessments measuring core academic skills, as well as a set of higher order cognitive and non-cognitive capabilities among primary school children. These were selected to include capabilities that are considered to be important dimensions of child development in the psychology and education literature and have been identified as priority skills in the new primary school curriculum currently being developed in Vietnam. In order to measure quality of teaching and learning environment we conducted interviews with the head teachers (school principals) at the schools that the children in our sample attend, as well as interviews with their classroom teachers. We also filmed these classroom teachers while they were teaching two core subjects - math and Vietnamese. We then code the quality of teaching observed in the videos using a classroom observation tool known as *Teach*, and construct measures of several key dimensions of classroom process quality (Molina, Pushparatnam, and Wilichowski (2019); Molina, Fatima, et al. (2020)).

In examining teacher effectiveness and classroom quality, we encounter several identification and measurement challenges. A primary concern stems from the non-randomness of assignment of pupils to teachers in many real-world settings. This non-randomness implies that the observed correlations between measures of classroom quality and student development may be confounded by unobserved pupil characteristics correlated with both factors. Identifying causal relationships, therefore, requires study designs with plausible sources of exogenous variation in classroom practices, which can be challenging to find. To address this issue, we leverage the panel nature of our data as well as the inclusion of multiple classrooms per school-grade-year in our sample. By using two classrooms per cohort per school we are able to estimate classroom-specific effects within a school-grade-year. To address concerns about sorting within a school across classrooms, we can rely on teacher and principal reports on how children and teachers are assigned to classes. It is worth noting that, in Vietnam, concerns about sorting within schools are minimal, as streaming is against official government guidance. Nonetheless, we additionally estimate value-added measures conditioning on a rich set of student and parent characteristics that could drive the non-random sorting. This approach is similar to the methodology used by Bau and Das (2020) and Chetty, Friedman, and Rockoff (2014).

The second challenge lies in separating the variance in test scores due to teachers from the variance due to classroom random shocks. To address this, we leverage a unique aspect of our dataset: in a subset of schools, we observe two different cohorts taught by the same teacher in grade 2. This allows us to use the covariance in class effects across the classes taught by the same teacher to estimate the variance of the teacher effect purged of classroom shocks such as random compositional classroom changes over time or noise next to a classroom, following methodologies by Hanushek and Rivkin (2012) and McCaffrey et al. (2009). However, a limitation of our design is that we can estimate teacher value added in this way only for academic skills in grade 2. For academic skills in grade 3, as well as higher order cognitive and non-cognitive skills, our estimates are of the combined effects of teachers and of classroom shocks.

Another significant challenge involves identifying the key ingredients of classroom quality, particularly in measuring elements of process quality. This requires sophisticated measures of teaching practices and of the classroom environment, as well as comprehensive measures of children's skills. Many well-established tools for measuring teaching practices require observation and coding of lessons by highly trained observers, which is both time-consuming and often very costly due to high licensing fees. Additionally, many have been developed for HICs and require significant adaptation and validation for application outside these settings. To address this, we employ the *Teach* tool, which has been designed for broader applicability across different contexts.

It is important to note that our study focuses on within-school teacher and classroom variation and is unable to capture any variation across schools.

Despite these obstacles, our comprehensive approach allows us to contribute valuable insights into teacher effectiveness and classroom quality, particularly in the context of LMICs like Vietnam.

The results from our analysis reveal several key insights. First, we find that classroom quality in Vietnam has a considerable impact on children's learning. A one standard deviation increase in classroom quality for students in grades 2, 3 and 4 increases students' mathematics scores by 0.20 - 0.22 standard deviations. This suggests that if a pupil moves from a class in the bottom 5th percentile of the classroom value added distribution to the top 95th percentile, his/her math test scores would increase by up to 0.72 standard deviation. For all the grades 2, 3 and 4, value-added estimates for Vietnamese are somewhat lower than mathematics. For grades 2 and 3, they are around 0.15 standard deviations, which implies up to a 0.49 of a standard deviation improvement in response to movement from a class in the bottom 5th percentile of the class value added distribution to a classroom in the top 95th percentile. These classroom effects fall between those observed in some other developing countries. Specifically, these effects are slightly lower than those found by Bau and Das (2020) for Pakistan, but somewhat larger than those reported by Araujo et al. (2016) for Ecuador and Buhl-Wiggers et al. (2022) for Uganda.

Separating out the effects of differences in teacher quality from classroom shocks, we estimate grade 2 *teacher* effects of 0.07 standard deviations for math and Vietnamese. To put this into perspective, this impact represents only 18-25 percent of the impact of one year of schooling on median test scores (0.28 to 0.39 of a standard deviation for grade 2 students). In practical terms, moving a student from a teacher in the bottom 5th percentile of quality to the 95th percentile would increase scores on math and Vietnamese by 0.23 standard deviations. This teacher effect of 0.07 standard deviations is only half the size of that found by Bau and Das (2020) for Pakistan: they estimate that a one standard deviation improvement in teacher quality results in a 0.15 standard deviation increase in test scores, which in Pakistan is almost one half of an annual test score gain of an additional year of schooling. In contrast, it is only slightly lower than the 0.09 standard deviation teacher effect reported by Araujo et al. (2016) in the context of Ecuador. Using comparative *Teach* data from other countries, we provide a plausible explanation for why teachers appear to explain less variation in student outcomes. This comparison highlights the contextual differences that may account for the lower teacher effect sizes observed in our study.

Second, we estimate class effects for a set of additional higher order cognitive skills - executive functions (EF) - which capture children's ability to maintain their attention, focus on the task at hand, and maintain relevant information in mind. These skills are related to two sets of competencies targeted in the Vietnam New Curriculum - Self-study and Self-managed Learning, as well as Problem Solving and Creativity. EF consists of three domains: inhibitory control, working memory and cognitive flexibility. We use validated direct assessment tasks that have been used internationally to capture each of these domains. We find that, in fact, classroom value added is somewhat higher for these skills than for Vietnamese skills: a one standard deviation increase in classroom quality results in a 0.21 standard deviation increase in executive functioning. In contrast, this is higher than the 0.07 standard deviation classroom effect estimated for Ecuador.

We also collected measures of "non-cognitive" skills, including self-perception across several domains (scholastic, social and physical) as well as propensity for independent mastery of school subjects rather than reliance on teachers. Existing studies suggest that more independent students with higher perceived competence attain better success at school. In addition, these competencies link to the Vietnam New Curriculum focus on academic, physical and social competence, as well as competence in self-study and self-managed learning. We find classroom effects for these skills in the range of 0.15-0.16 standard deviations.

Third, we find that our classroom effects for cognitive and non-cognitive skills are not highly correlated with each other. The correlation between the mathematics and Vietnamese class effects is 0.40, while the correlations between these two cognitive skills and the self-perception and mastery scores range from -0.09 to 0.06. This suggests that the classrooms, and likely the teachers, that are relatively effective at developing students' cognitive skills are not the same classrooms, and likely not the same teachers, that are most effective at developing students' non-cognitive skills.

Our final analysis draws on classroom recordings of teachers' teaching evaluated using the *Teach* tool which provides insights into the specific strengths and areas for improvement among primary school teachers in Vietnam. The results reveal that teachers in Vietnam excel in creating supportive learning environments and setting positive behavioral expectations for students. They also demonstrate proficiency in using questions and prompts to check for understanding and in monitoring students during independent and group work. However, the analysis also identifies areas where teachers could improve. On average, they are less adept at promoting critical thinking, fostering student autonomy, and developing social and collaborative skills within the classroom.

Building on this, we find evidence to suggest that there are important differences in the specific areas of classroom practice that matter the most for different developmental outcomes. Practices categorised as “Instruction”, which includes using questions and prompts to check for understanding and in monitoring students during independent and group work, are important for the acquisition of academic skills - math and Vietnamese - but less so for Executive function and non-cognitive skills. In contrast, practices within the “Socioemotional” domain, such as fostering autonomy, perseverance and social and collaborative skills appear to be more relevant for the development of some non-cognitive skills.

These results resonate with evidence in the handful of other studies that look at differences in the role that different dimensions of classroom practices play in the formation of different types of skills. Though there are very few, the ones that exist find that instructional quality is strongly associated with children’s academic skills, whereas measures of emotional support are more predictive of children’s social, emotional, and behavioral skills (Burchinal, Kainz, and Cai (2011); Curby, Rimm-Kaufman, and Ponitz (2009); Mashburn et al. (2008)).

Finally, we explore the role that measures of structural quality, such as teacher-student ratios, teacher qualifications and experience, and quality of the classroom infrastructure. As found in other studies, we find little to suggest that these dimensions of quality are strong predictors of children’s skills.

The rest of this paper is structured as follows. We start with a detailed description Vietnam’s education system in Section 1.2. We then describe the key measures that we use in the analysis in Section 1.3 and the methods used to construct them in Section 1.4. Section 1.4 also sets out our main empirical strategy. Main results are presented in Section 1.5, and Section 1.6 concludes.

1.2 Education in Vietnam

Vietnam's education system is structured into primary school (grades 1-5, beginning at age 6), lower secondary school (grades 6-9), and upper secondary school (grades 10-12). Additionally, Vietnam offers pre-primary education (for ages 3-5), secondary vocational training schools, and various post-secondary institutions. As of 2020, Vietnam had more than 12,000 primary schools, 8,000 lower-secondary schools, and 2,300 upper secondary schools.[1]

Nearly all primary schools in Vietnam are state-managed public schools. As of 2016, about three-quarters of primary schools were providing "full day" (6 hours) instruction; the others received only "half day" (3.5 hours) instruction, with schools usually operating two shifts. The government aims to extend full-day schooling to poorer areas, but progress has been slow. Despite the relatively short school days and school years compared to global standards, Vietnamese students performance on international assessments is much higher than the performance of students from other LMICs (Dang et al., 2023).This maybe due in part to many Vietnamese children attending additional "extra study" classes, though the extent of this varies by province.

Administratively, the Ministry of Education and Training (MoET) in Hanoi retains formal authority over the entire education system. MoET works with other line ministries to determine investments in education, and plays the leading role in education planning and in determining the content of curricula (London (2011)).

Vietnam's education system differs from those in other lower- and middle-income countries. One key difference is Vietnam's high degree of centralization. MoET retains control over teacher recruitment, training, and the development and implementation of the curriculum. This centralization contrasts with the more decentralized systems, where regional and local authorities have more influence over educational policies and practices. Furthermore, Vietnam's focus on additional "extra study" classes is unusual compared to many developing countries. In most countries, extra tuition is less formalized and more dependent on family resources and regional availability.

[1] `https://www.gso.gov.vn/en/homepage`

A notable aspect of Vietnam's education system is its ability to achieve high attainment levels while maintaining relatively equitable outcomes. An analysis of Programme for International Student Assessment (PISA) data reveals that Vietnam exhibits one of the lowest variations in reading test scores among both high-income and low-and-middle-income countries. Additionally, in mathematics, its achievement is comparable to that of high-income countries, while also demonstrating one of the lowest variations among these high-income nations (see Figure 1.A6). This low variation suggests that students in Vietnam demonstrate a more uniform level of academic achievement, indicating a more equitable education system with a smaller gap between high- and low-performing students. The consistency in test scores may also be attributed to a uniform level of teacher quality across the country, which is a result of Vietnam's highly centralized teacher recruitment, training and assignment processes.

This is an especially policy relevant time for this study as the Vietnamese government is keen to reform the current education system in order to equip students with the skills needed for a modern economy. In particular, the government has issued various guidelines that encourage students to develop their non-cognitive skills, including social skills, creativity, and self-learning ability, and that discourage teachers from using ineffective teaching practices such as passive learning and rigid memorization (Government of Vietnam (2014)). In 2018, the Vietnamese government announced a major education reform – the "Fundamental and Comprehensive Reform of Education" – that has started to implement major revisions to the official curriculum, pedagogical methodology, and teacher professional development in order to provide Vietnamese children with the skills they need to be effective participants in Vietnam's economy and society. Our study is able to shed light on the role that schools and teachers currently play in shaping this wider set of skills.

1.3 Data and Measurement

The data used in this paper were collected from 140 primary schools that are approximately nationally representative of all of Vietnam's primary schools.[2] For each of the 140 schools, data were collected for two adjacent

[2]To validate this, we compare our sample of students to the students in the Vietnam Household Living Standards Survey (VHLSS). Appendix Table 1.A4 shows this com-

cohorts of students, those who started grade 2 in the 2017-18 school year (henceforth cohort 1), and those who started grade 2 in the 2018-19 school year (cohort 2). Data were collected from cohort 1 for the 2017-18, 2018-19 and 2019-20 school years, when they were in grades 2, 3 and 4, and data were collected from cohort 2 for the 2018-19, 2019-20 and 2020-21 schools years, when they were in grades 2, 3 and 4. Table 1.1 shows when the data were collected from the two cohorts.

Table 1.1: Dates of Data Collection for Two Cohorts of Students

Date	Nov 2017	April 2018	Nov 2018	April 2019	April 2020	April 2021
Cohort 1	Grade 2	Grade 2		Grade 3	Grade 4	
Cohort 2			Grade 2	Grade 2	Grade 3	Grade 4

For both cohorts, more than 5,000 students were tested in mathematics and Vietnamese in their classroom when they were in grades 2, 3 and 4. The data were collected from a random sample of 20 students per class, and each test took one hour. Appendix Table 1.A5 provides the attrition rate over different rounds. We collected more detailed data on a random subsample of slightly more than 1600 students in each cohort. These pupils, and their parents, were interviewed. The pupils also completed assessments for an extended set of skills, including Executive Functioning (EF) and a battery of non-cognitive skills. Table 1.2 provides some basic information on both cohorts when they were in grade 2. As can be seen, the two adjacent cohorts are quite similar; none of the differences is significant at the 5% level, although the differences for birth order and gender are significant at the 10% level.

A key issue regarding estimation is whether the teachers instructing cohort 1 students in grade 2 were the same teachers teaching grade 2 when cohort 2 progressed to that level in the subsequent academic year. This is crucial for accurately estimating the variance of teacher fixed effects, as it requires the same teachers to instruct students in the same grade over two consecutive years. This is the case for 160 teachers, who are spread across 100 of the 140 schools. Yet in only 60 of these schools are *both* grade 2 teachers the

parison, examining several key variables available in both datasets, and finds that our data closely match the VHLSS figures.

Table 1.2: Student-Level Summary Statistics

	Cohort 1		Cohort 2	
	Mean	S.D.	Mean	S.D.
Age (in years)	7.416	0.558	7.386	0.564
Ethnic minority	0.232	0.422	0.249	0.432
Male	0.539	0.499	0.506	0.500
Birth order	1.985	1.130	1.898	1.116
Mother's highest grade	8.554	3.361	8.577	3.470
Father's highest grade	8.688	3.266	8.496	3.247
Father's age (in years)	36.977	6.508	36.812	5.955
Mother's age (in years)	34.007	5.859	34.031	5.557
Students tested (Academic Skills)	5070		5209	
Students interviewed (including Non-cog skills)	1654		1673	
Number of Schools	140		140	
Number of Classrooms	276		279	

Notes: This table reports means and standard deviations of the characteristics of students measured at the beginning of Grade 2.

same for both cohorts. While this reduces the sample of schools by more than half, this is a sufficient sample to carry out the estimation procedure described in Section 1.4. We also provide a descriptive table comparing this subset of schools with the full sample. Appendix Table 1.A7 presents means and standard deviations of some characteristics of students, teachers, and parents for both groups. Our analysis reveals that the subset of schools used for the estimation of teacher value-added is highly similar to the full sample across all dimensions.

The more discouraging situation, however, is that in grade 3, only 23 of the 140 schools had both of the grade 3 teachers teaching cohort 1 continue to teach grade 3 in following school year. This attrition is rather severe and so we do not estimate the variance of teacher effects for grade 3 (although we do estimate class fixed effects, which do not require following the same teacher over time).

Data were also collected from about 280 teachers in each grade using a teacher questionnaire. Table 1.3 presents some basic characteristics of teachers, separately for grades 2, 3 and 4. The typical teacher for these grades is about 42 years old, and the vast majority (90% in grade 2, and 78% in grades 3 and 4) are women. About one out of eight is a member of an ethnic minority group, and about 90% are married. About two thirds have a university degree, and well over 90% have had pedagogical training at the university level. Almost all (92% to 94%) have permanent positions.

Table 1.3: Teacher-Level Summary Statistics

	Grade 2		Grade 3		Grade 4	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age (in years)	41.179	8.736	42.873	8.262	41.717	8.140
Married	0.909	0.288	0.884	0.320	0.903	0.296
Male	0.099	0.299	0.216	0.413	0.221	0.416
Have university education or higher	0.620	0.486	0.634	0.483	0.689	0.464
Have university pedagogical training	0.916	0.278	0.933	0.251	0.962	0.192
Years of experience	20.226	9.380	21.705	8.829	20.498	8.895
Have a permanent position	0.916	0.278	0.922	0.269	0.941	0.236
Belongs to ethnic minority	0.139	0.346	0.142	0.349	0.107	0.310
Number of teachers	276		268		289	

Notes: This table reports means and standard deviations of the characteristics of teachers measured at the end of each academic year.

1.3.1 Academic Skills Assessments

Information on the cognitive and non-cognitive tests administered is provided in Appendix Table 1.A6 . In all math and Vietnamese tests administered at different grades, common (linking) items were included in order to put students' performance on these tests on a common scale, using item response theory (IRT).

The items were developed for each grade by the Vietnam Institute of Educational Sciences. All tests administered assess domains relevant to the Vietnamese curriculum. The Math test assesses three skills: arithmetic, measurement and quantities, and geometry and the Vietnamese test assesses four language skills: vocabulary, rhetoric, grammar, and reading comprehension. They both assess the following cognitive domains: knowledge, understanding and application. We conducted two pilot tests, the first was done in May of 2017, before the first round of data collection, and the second was in January of 2019, before collecting data from grade 4. For both pilot tests, about 300 students and their teachers participated for each grade.

To facilitate comparisons of student performance across different grades and between the two cohorts, we consolidated all test results into unified datasets, one for mathematics and another for Vietnamese language skills. We then employed Item Response Theory (IRT) analysis to construct latent measures of mathematical and Vietnamese language proficiency for each student. This approach allowed us to identify and select the best performing test items (questions). For ease of interpretation, we normalized these latent scores for both mathematics and Vietnamese language. Within each grade level, the scores were standardized to have a mean of zero and a standard deviation of one. For additional details on test construction, distribution of test scores (to examine floor and ceiling effects), and validity of the tests, please refer to Appendix 1.A.1.

1.3.2 Executive Functioning Measures

In addition to these academic skills we also administered measures intended to capture high-level cognitive processes that are known as "executive functions", which enable individuals to concentrate (P. W. Burgess and Simons (2005), Espy (2004)). Students use executive functioning skills to maintain concentration in class, consolidate taught material and apply it to complex problems. Executive functioning was measured using two different tests: "hearts and flowers" and "backward digit span". The former is a computerized task where a stimulus appears on the right or left of the screen. The rules are: (a) For one stimulus, press on the same side as the stimulus (called the congruent condition); (b) For the other stimulus, press on the side opposite the stimulus, which requires inhibiting the natural tendency to activate the hand on the same side as the stimulus (called the incongruent condition). This task requires both working memory and inhibitory control. The backward digit span tests only working memory; respondents are required to repeat a series of numbers back to the assessor in reverse order, and the final score is equal to the longest sequence of numbers that a student was able to say backwards. As with scoring of the Math and Vietnamese tests, we use IRT model to construct a combined score of these higher order cognitive skills. This measure was part of an extended set of measures administered to a sub-sample of the children who had been tested in Math and Vietnamese - around 7 children per class. Due to logistical challenges, we were only able to administer Executive Functioning tasks to Cohort 1 in grade 3 and not at all to Cohort 2.

1.3.3 Non-Cognitive Skill Measures

The first measure of non-cognitive skills is the self-perception profile. It asks students to compare themselves to hypothetical children with high and low skills or characteristics of various types (e.g. academic performance, social skills, and physical appearance). This measures students' self-perceptions of these skills. The second measures intrinsic vs. extrinsic motivation skills, of which only one scale was administered: independent mastery vs. dependence on the teacher. This measures the extent to which the student likes to solve problems independently or prefers to depend on the teacher for help. We applied Item Response Theory (IRT) modeling to construct scores for the non-cognitive scales, mirroring the method used before.

A comprehensive overview of our non-cognitive and executive functioning assessments is provided in Appendix 1.A.2. This includes detailed descriptions of the domains measured, analyses of test score distributions, and evidence supporting the validity and consistency of these instruments.

1.3.4 Class and Teacher Quality

We measure the quality of what goes on in the classroom through teacher interviews and direct observation (filming lessons). The key measures captured through teacher interviews for the two cohorts are presented in Table 1.3.

In order to construct measures of classroom quality from the videos using the *Teach* tool, we hired a team of coders who were trained to rate teachers' practices as per the guidelines set out in the *Teach* manual.[3] The coders were trained to rate teachers' practices on a low-medium-high scale, with a focus on 28 *behaviours* using four 15-minute segments of the filmed lessons. Examples of these behaviours include "Teacher asks open-ended questions" or "Teacher encourages goal setting"; the full set is listed in Table 1.A8 . These direct observations are unavailable for grade 4, as our recordings were limited to teachers in grades 2 and 3.

According to the *Teach* framework, the 28 *behaviours* map into nine *elements* of classroom environment which can be aggregated into three broad *areas*: Classroom Culture, Instruction, and Socioemotional Skills. Guided by the scores for the 28 behaviours, the coders were trained to assign a score on a 5-point scale to each of the nine elements, with 1 indicating lowest level. [4] We average the scores for each of the nine elements across the four segments of filmed lessons coded for each teacher within each cohort. To

[3] `https://documents1.worldbank.org/curated/en/872291641201520569/pdf/Teach-Primary-Observer-Manual.pdf`

[4] Following the scoring of individual behaviors, evaluators determine overall element scores on a scale of 1 to 5 as per the *Teach* manual. This process involves a thorough review of behavior-level descriptions and careful consideration of the observed classroom dynamics.

obtain raw scores for the three areas and a Global Teach score, we average the scores of the relevant elements. Table 1.A8 illustrates the mapping of behaviors into elements and elements into areas, while Table 1.8 presents descriptive statistics and score distributions by areas and elements.

While the *Teach* framework was initially developed with three hypothesized areas - Classroom Culture, Instruction, and Socioemotional Skills - we opt for a more nuanced approach. Instead of simply averaging element scores to create area measures, we employ factor analysis to derive three indices representing distinct domains of teaching practices/classroom quality. This method allows us to explore various combinations of the nine elements and to assess whether the data truly reflect these hypothesized aspects of teaching quality. Our factor analysis results largely align with the original three areas (detailed item loading can be found in Appendix Table 2.A5). Furthermore, we extract a single common factor, termed the Global Teach factor, which incorporates information from all nine of the elements. Each element demonstrates a factor loading of at least 0.3 on this Global factor, except for Social and Collaborative Skills element (which had a factor loading of 0.21).

1.4 Methodology

Having explained how our main measures are constructed, we now discuss the main model for our empirical analyses. The overall objective of this paper is to understand the characteristics and behavior of teachers in Vietnam that make some of them more productive than others. The first step is to estimate teacher productivity, and the second step is to examine which observable characteristics of those teachers appear to make them more productive.

1.4.1 Class and Teacher Effects

Here we present the methodology used to estimate school and teacher value added, and to estimate the impact of teacher characteristics and behavior on student learning.

In order to estimate teacher productivity, which is defined as the value added of a teacher on students' learning, the starting point is to estimate a classroom fixed effect for a given time period. This can be done by estimating the following regression equation:

$$Y_{icst}^{k} = \alpha_{cs}^{k} + \beta Y_{icst-1}^{k} + \epsilon_{icst}^{k} \tag{1.1}$$

where Y_{icst}^{k} is the end-of-school-year score of student i in classroom c in school s at time t on a test for subject k (mathematics, Vietnamese), α_{cs} are classroom indicators, Y_{icst-1}^{k} is the beginning-of-school-year score on subject k and ϵ_{icst}^{k} is an i.i.d. error term. The classroom fixed effects (α_{cs}) are estimates of classroom value added under the assumption that, conditional on controls (primarily the test score at the beginning of the school year or at the end of the previous grade), students are randomly assigned to classrooms. It is plausible that students are not randomly assigned to schools, even conditional on past learning, but it is generally accepted that the assumption of conditional random assignment within school is reasonable (e.g., Chetty, Friedman, and Rockoff (2014)).[5] This requires, however, calculating classroom value added relative to the school mean, which ignores any cross-school variation in school quality.

Therefore, since we have two classrooms per school, as in most papers in this literature, we redefine each classroom effect relative to the school average to address the issue of sorting of teachers and/or students into schools. This will estimate classroom effects using only variation across classrooms within schools. The demeaned classroom effect, denoted by λ_{cs}^{k}, is:

$$\lambda_{cs}^{k} = \alpha_{cs}^{k} - \frac{\sum_{c=1}^{C_s} N_{cs}\alpha_{cs}^{k}}{\sum_{c=1}^{C_s} N_{cs}} \tag{1.2}$$

where C_s is the number of classrooms in a school and N_{cs} is the number of students in the classroom c in school s.

To measure the overall contribution of (variation in) classroom quality to (variation in) student learning, it is useful to estimate the variance of the classroom effect, which can be denoted by $(V(\lambda_{cs}^{k}))$. In order to avoid overestimating the true variance of λ_{cs}^{k}, due to sampling error in the estimates of λ_{cs}^{k}, we apply a shrinkage procedure, following, for example, Chetty, Friedman, Hilger, et al. (2011). This correction can be expressed

[5] In our setting, within-school sorting of students across classes is unlikely to be a significant concern due to legal restrictions. However, to address any potential non-random assignment of students to classrooms within schools, we employ two additional strategies, discussed later.

as follows:

$$V(\lambda_{cs}^{k}) = V(\hat{\lambda}_{cs}^{k}) - E\left\{\frac{(\sum_{d=1}^{C_s} N_{ds}) - N_{cs}}{N_{cs}(\sum_{d=1}^{C_s} N_{ds})}\sigma^2\right\} \tag{1.3}$$

where σ^2 is the within-classroom variance of residual student learning (the variance of ϵ_{icst}^{k} in equation (1.1)).

The (within-school) variance in class effects in equation (1.3) includes both variation in teacher quality across classrooms as well as random "classroom shocks" that reflect random differences in (average) pupil characteristics over time (over cohorts) and random events that happen on the day of the test. To separate out variation in teacher quality (variation in teacher fixed effects) from "classroom shocks", one can estimate the covariance over time of the classroom effect for classrooms that have the same teacher in the two time periods. [6] That is, one can estimate:

$$Cov(\lambda_{cs}^{kt}, \lambda_{cs}^{k,t+1})$$

If students are randomly assigned to teachers within schools, then the square root of this covariance is an estimate of the standard deviation of the teacher effects alone; the classroom shocks are uncorrelated over time and so drop out of this covariance term.

1.4.2 Classroom quality and teacher practices

The second objective is to study how different dimensions of classroom quality and teacher practices affect student learning. A key empirical challenge in identifying this relationship is that one cannot measure all relevant school inputs or control perfectly for child characteristics such as ability. We address this issue in two ways. First, we exploit the panel dimension of our data. Following Todd and Wolpin (2007), Fiorini and Keane (2014), Del Bono et al. (2016) and Keane, Krutikova, and Neal (2022b) among others, we proxy for these omitted inputs and ability by a lagged test score alongside some additional controls for time-invariant characteristics of the child. Second, we conduct the analysis at the within school level, exploiting the feature that our sample includes two classrooms within each year-grade-cohort, which accounts for any sorting of pupils to schools. Our main empirical specification for this analysis is as follows:

[6] Our methodology here lies on the assumption that teacher quality is fixed over time or teachers do not change their teaching methods in response to class characteristics.

$$Y^k_{icmsg} = \alpha^k_{msg} + X^{\intercal}_{cmsg}\beta + \gamma Y^k_{icmsg-1} + Z^{\intercal}_{cmsg}\theta + \nu^k_{icmsg} \tag{1.4}$$

where Y^k_{icmsg} is the end-of-school-year score of student i belonging to cohort m in classroom c in school s in grade g on a test for subject k (mathematics, Vietnamese or Non-cognitive); X_{cmsg} is a vector of our measures of classroom quality and teacher practices (*Teach* scores); $Y^k_{icmsg-1}$ is the lagged measure of the outcome; Z is a vector of child background characteristics including age, sex and ethnicity; and ν^k_{icmsg} is an i.i.d. error term. In the regression analysis we pool the two cohorts and grades together, so α_{msg} is the school-cohort-grade fixed effect.

Another concern in identifying this relationship is that reverse causality may bias the coefficients on our contemporaneous *Teach* scores. Consider a scenario where a teacher is randomly assigned to a group of particularly disruptive or less engaged students. This teacher might receive lower *Teach* score as a result. However, these same students could also be slow learners, even after accounting for their observable characteristics. This scenario could lead to an underestimation of the teacher's true effectiveness. To mitigate this potential source of bias, we use teacher's *Teach* score from the other cohort, or lagged score, rather than their current score in most of our estimates. This approach helps to isolate the teacher's impact from the confounding effects of their current student composition. Therefore, we restrict our sample to students who have teachers that taught two cohorts.

Our analysis encompasses cognitive skill attainment for both grades 2 and 3 (for which we film classes), but focuses solely on grade 3 for non-cognitive skill attainment. This approach is necessary because, as previously explained, we control for lagged measures of the outcome to address potential endogeneity from omitting key inputs. For grade 2, we have the requisite cognitive skill data to estimate our main model. However, we lack lagged measures of executive functioning and non-cognitive skills from the beginning of grade 2. In contrast, for grade 3, we can utilize all assessment results from the end of grade 2 as our lagged measures, allowing for a comprehensive analysis of both cognitive and non-cognitive outcomes.

1.5 Results

1.5.1 Classroom and Teacher Value Added Estimates

Academic Skills

We start by estimating the proportion of the overall variance of student test scores in math and Vietnamese that is accounted for by variance in classroom fixed effects, for both cohorts for grades 2, 3 ad 4.

Table 1.4 presents estimates of the within school standard deviations of classroom and teacher effects, adjusted for sampling error, for mathematics and Vietnamese scores, combining both cohorts. Classroom effects on mathematics for grade 2 pupils suggest that a one standard deviation increase in classroom quality results in around a 0.19 standard deviation improvement in mathematics scores. Estimates for Vietnamese are somewhat lower, around 0.16. These estimates are based on almost all of the schools for both cohorts. We find very similar size class effects for grades 3 and 4.

Table 1.4: Within-School Standard Deviations of Classroom and Teacher Effects

	Class Effects			Teacher Effect
	Grade 2	Grade 3	Grade 4	Grade 2
Mathematics	0.195	0.217	0.218	0.069
	(0.018)	(0.010)	(0.014)	(0.020)
Vietnamese	0.157	0.145	0.204	0.070
	(0.012)	(0.008)	(0.009)	(0.014)
Students	9829	9317	8777	4613
Schools	139	139	138	60
Classes	533	520	507	120

Notes: In this table, we report the within-school standard deviations of classroom and teacher effects adjusted for sampling error. The first three columns give us classroom effects using a pooled sample of the two cohorts. The last column reports teacher effects and uses only those schools in which we have the same grade 2 teachers teaching the two cohorts. Bootstrapped standard errors in parenthesis.

For grade 2, we are able to separate the effects of classroom shocks from the effects of teacher quality, using a sub-sample of 60 primary schools for which we have the same grade 2 teachers teaching the two cohorts (so that we observe each teacher teaching two different classes). As expected, teacher effects are smaller than class effects, yet this reduction is more pronounced for math than for Vietnamese (a reduction of almost one half).

These teacher effects of 0.07 for both math and Vietnamese are a little less than the effects of 0.09 for both math and language found by Araujo et al. (2016) in Ecuador.[7] Furthermore, these teacher effects can be used to estimate the standard deviation of the shocks, which are 0.18 for math and 0.14 for reading[8], suggesting that these effects can be quite large, as has been found in other studies. In fact, for both math and Vietnamese the standard deviation of the classroom shock effect is larger than that of the teacher effect. These large classroom shock effects may indicate the presence of significant peer effects.

Value Added: Robustness

We also report some robustness tests of the validity of our classroom and teacher value added estimates. In Vietnam, within-school sorting of students across classes is unlikely to be a significant concern as it is not allowed by the law. Despite these rules, there are occasional reports of schools allocating classrooms to students, not based on ability, but based on their location.

Thus, to address any potential non-random assignment of students to classrooms within schools, we employ two additional strategies. First, we estimate the class and teacher effects using a sub-sample of schools where in we include only those schools where the principal reported to allocating classrooms randomly to students. The results from this sample are presented in Appendix Table 1.A10 . Our findings show that even with this restricted sample, our value-added measures remain largely unchanged. Second, we estimated our measures while controlling for a rich set of student and parental characteristics that could potentially influence non-random sorting into classes. These characteristics include mother's age, parental education levels, household income, student's gender and other relevant factors. This approach limited our sample to only those students with available parental data, which reduced our sample size to one-third of the original. Our analysis suggests that inclusion of these additional controls had minimal impact on our estimates, as demonstrated by the high cor-

[7] We compare Vietnam's *Teach* score to other countries utilizing this tool for measuring teacher practices and classroom quality in Table 1.A12 . We observe that Vietnam exhibits the lowest variation in the Global *Teach* score among all countries assessed. This suggests a higher degree of uniformity in teaching practices across Vietnam. The consistency in teaching practices could be attributed to Vietnam's highly centralized system. As a result of this uniformity, it is possible that teachers explain less of the variation in student outcomes as compared to other countries.

[8] These are calculated as $(\text{class effect}^2 - \text{teacher effect}^2)^{0.5}$

relations between estimates derived from models with and without these controls in this sub-sample (see Table 1.A11).

Executive Functioning

Table 1.5 presents our estimates of the within-school standard deviation of classroom effects on EF. Class effects on executive functioning can be estimated only for Cohort 1 grade 3, as executive functioning tasks were administered only at the end of grades 2 and 3 and only for Cohort 1. Furthermore, they were administered to a sub-sample of around seven children per class. Due to some compositional changes in classes between grades 2 and 3, however, our sample contains some classes that have even fewer than seven children in grade 3 who completed the executive functioning tasks. We restrict our analysis sample to classes that had at least 3 children who completed these tasks, leaving us with a sample of 132 schools, 272 classes and 1,373 children. The EF score has been standardised to have a mean zero and standard deviation of 1 within the full sample of children who completed this assessment.

Our estimates suggest that classroom effects on EF are comparable to those on mathematics but higher than on Vietnamese. A one standard deviation increase in classroom quality leads to an improvement in executive functioning of 0.21 standard deviations. If we assume that the classroom shocks for executive function are the average of classroom shock effects that we imputed for mathematics and Vietnamese, then the teacher effect on the composite measure of EF would be about 0.14 standard deviations.

Non-cognitive skills

The final set of classroom effect estimates is also presented in Table 1.5. It shows effects on our measures of non-cognitive skills, including self-perception which aggregates the scholastic, social and physical domains, and a mastery score (see Section 1.3 for further details). These scales were administered to the same subset of children who completed the EF assessment in Cohort 1, as well as to a sub-sample of children in Cohort 2 (also about seven per class). In order to estimate the classroom effects on non-cognitive skills, we pool the two cohorts (to maximise sample size) but (as before) we exclude classes in which less than three children completed the non-cognitive scales. The scores are standardised to have a mean zero and a standard deviation of one within the grade.

Table 1.5: Within-School Standard Deviations of Classroom Effects on EF and Non-Cognitive Skills

Assessment score	Class Effects
EF: HF and BDS	0.214
Students	1,373
Schools	132
Classes	272
Self-Perception	0.147
Independent Mastery	0.161
Students	3110
Classes	520
Schools	139

Notes: This table reports the within-school standard deviations of classroom effects adjusted for sampling error. The classroom effect on EF uses data from only one cohort; while classroom effects on self-perception and mastery uses data from both the cohorts. We can only estimate classroom effects for grade 3.

These estimates are smaller in magnitude compared to those observed for EF: a one standard deviation improvement in classroom quality results in a 0.15 standard deviation improvement in self-perception and 0.16 standard deviation increase in independent mastery. We know of no comparable estimates for either self-perception or independent mastery. Therefore, providing these credible estimates of classroom effects on non-cognitive skills represents a major contribution of our study.

1.5.2 Classroom Effect Correlations

Having estimated classroom effects on math, Vietnamese, EF and non-cognitive skills, we can now ask whether good teachers are good for students' acquisition of competencies in all domains or whether some are good at, for example, teaching academic skills while others are good at fostering non-cognitive skills. To do this, we simply look at the correlations between the classroom effects estimated for each of our measures of cognitive and non-cognitive skills. Table 1.6 shows these estimates for grade 3 (as this is the grade for which we have the most complete data).

The first results to point out in this table is the high positive correlation between math and Vietnamese classroom effects (0.40), which indicates that classrooms, and presumably the teachers in them, that are successful in raising students' math skills are also successful in increasing their reading skills. A second result of interest is the moderate positive corre-

Table 1.6: Pairwise Correlation of Classroom Effects

Variables	EF	SP	MAS	Viet	Math
EF	1.000				
SP	-0.074	1.000			
	(0.239)				
MAS	0.211***	0.123***	1.000		
	(0.001)	(0.005)			
Vietnamese	0.139**	-0.066	0.001	1.000	
	(0.027)	(0.133)	(0.979)		
Math	0.143**	-0.089**	0.060	0.404***	1.000
	(0.023)	(0.044)	(0.173)	(0.000)	

*Notes: This table reports the pairwise correlation coefficient of classroom value added estimates for 520 teachers, except column 1 which corresponds to 253 teachers as EF measures were collected only from one cohort. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.*

lation between EF and math and Vietnamese skills; this also makes sense because EF is hypothesised to support students' learning skills. Finally, it is interesting that there is almost no, or even negative, correlation between cognitive and non-cognitive class effects. This suggests that teachers, or, at least, classroom environments, that are effective for students' acquisition of cognitive skills are different from those that are effective for their acquisition of non-cognitive skills. Specifically, we observe that teachers who excel at fostering mathematics skills seem less effective at cultivating students' self-perception.

Table 1.7: Pairwise Correlation of Student's Test Scores

Variables	EF	SP	MAS	Viet	Math
EF	1.000				
SP	0.016	1.000			
	(0.561)				
MAS	0.062**	0.130***	1.000		
	(0.021)	(0.000)			
Vietnamese	0.255***	-0.008	0.090***	1.000	
	(0.000)	(0.663)	(0.000)		
Math	0.295***	0.015	0.101***	0.691***	1.000
	(0.000)	(0.422)	(0.000)	(0.000)	

*Notes: This table reports the pairwise correlation coefficient of students test scores from both the cohorts in Grade 3, except column 1 which corresponds to only one cohort. Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.*

To ensure that this pattern is not driven by students' test scores, we examined the correlation between individual students' cognitive skills and their non-cognitive skills (Table 1.7). Notably, we find no negative correlation between the self-perception skills and cognitive skills at the student level. This lack of correlation at the student level strengthens our interpretation of the classroom effect findings, indicating that the observed patterns are likely attributable to differences in teaching strategies or classroom environments.

1.5.3 Teacher behaviour and children learning

In this section we relate differences across classrooms in learning outcomes to the practices and behaviors of teachers. We start by exploring what the *Teach* measures tell us about classroom practices in Vietnam. We then turn to assessing the relationship between these processes and the cognitive and non-cognitive skills of primary school students. In the final part of the analysis, we estimate models which include additional measures of teacher and classroom characteristics to also capture the *structural* dimension of classroom quality.

Classroom Practices

Table 1.8 shows the distribution of *Teach* scores by areas and elements. The last column shows the mean score for each area/element, which is also shown graphically in Figure 1.1. First, these scores are similar to what has been found in other countries but with slightly better performance of teachers in the Socioemotional and Instruction areas (Bassi, Medina Pedreira, and Nhampossa (2019), Molina, Fatima, et al. (2020)).

As is evident from Figure 1.1, primary teachers in Vietnam are more skilled in creating a supportive learning environment, setting positive behavioral expectations, facilitating lessons and checking for understanding, than promoting critical thinking, perseverance and social/collaborative skills. For example, over 60% of teachers in both cohorts got a score of 4 or 5 (5 being highest) on the Supportive Learning Environment element, which, as we show in Table 1.A8 , incorporates treating students respectfully, using positive language, responding to students' needs and challenging stereotypes. In contrast, less than 10% of teachers got an equally high score on the Social and Collaborative Skills element, which includes promotion of student collaboration with peers and interpersonal skills.

Table 1.8: *Teach* Score Distribution by Areas and Elements

	Distribution of scores					
	1-2	2-3	3-4	4-5	Mean score	
Areas						
Class culture	0.34	5.83	62.17	31.66	3.48	
Instruction	2.97	53.94	41.26	1.83	2.78	
Socioemotional	40.46	52.34	7.20	0.00	2.06	
Global teach score	1.49	70.29	28.23	0.00	2.77	
	1	2	3	4	5	Mean Score
Elements						
Supportive Learning Environment	0.34	4.91	34.17	60.46	0.11	3.55
Positive Behavioral Expectations	0.57	5.14	50.63	40.46	3.20	3.40
Lesson Facilitation	0.34	35.20	44.80	18.06	1.60	2.85
Checks for Understanding	1.03	9.26	33.03	46.40	10.29	3.55
Feedback	19.09	22.51	35.66	16.91	5.83	2.67
Critical thinking	19.09	63.31	12.00	4.80	0.80	2.05
Autonomy	5.60	60.00	24.69	9.03	0.69	2.40
Perseverance	4.91	87.31	7.43	0.34	0.00	2.03
Social and Collaborative Skills	54.40	22.29	15.43	7.66	0.23	1.77

Notes: In this table, we present distribution of the overall Teach score, Teach's areas, and Teach's elements for Grade 3. We find similar distribution for Grade 2 teachers. The top panel presents the distribution of the scores under 1-point intervals (as a result of aggregation across the elements). This was computed by averaging the score across multiple segments per teacher

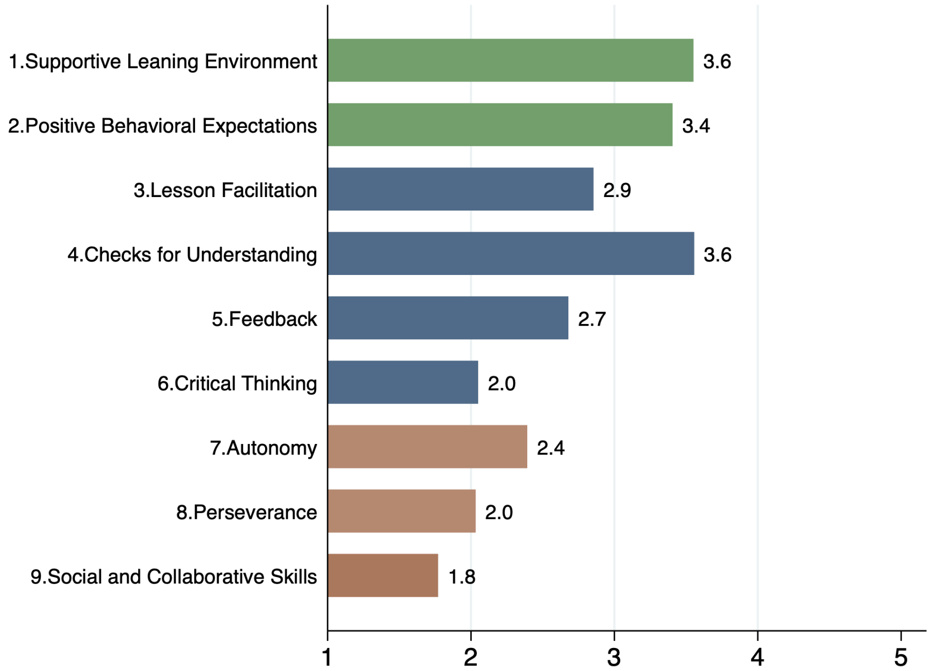


Figure 1.1: Distribution of Teach Scores by Element

Table 1.9 further shows that while teachers who are good at promoting a strong "Class Culture" (encompassing creating a supportive learning environment and setting positive behavioral expectations) also perform well in the "Instruction" area (including on lesson facilitation, checking for understanding and providing feedback) and "Socioemotional Skills" area. However, doing well in the "Instruction" area is not correlated with doing well in the "Socioemotional Skills" area.

Table 1.9: Correlation between different *Teach* Areas

	Class Culture	Instruction	Socioemotional
Class Culture	1.0000		
Instruction	0.448***	1.0000	
	0.0000		
Socioemotional	0.335***	0.010	1.0000
	0.0000	0.0000	

*Notes: This table reports the pairwise correlations between different Teach areas. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.*

We now turn to the core part of this analysis, which seeks to explore the relationship between the classroom practices we have just discussed and students' acquisition of skills. We start by looking at the impact of classroom practices summarised by (1) Global *Teach* factor score; and (2) factor scores corresponding to the three areas - "Classroom Culture", "Instruction" and "Socioemotional" on cognitive and non-cognitive skills of students. The skill measures include: (1) academic skills (Vietnamese and math); (2) higher order cognitive skills (EF); (3) non-cognitive skills (global self-perception score, as well as degree of independence in mastery). To this end we follow the estimation strategy set out in Section 1.4.2 and estimate the model in Eq(1.4) for the range of skills that we measure. The results are presented in Tables 1.10 and 1.11.

Table 1.10 suggests that classroom practices, as measured by the global *Teach* score are most relevant in explaining variation in academic skills across students in our sample: higher global *Teach* scores are associated with higher scores in math and Vietnamese. To mitigate potential reverse causality, we also use teacher's *Teach* score from the other cohort taught by the same teacher (henceforth lagged score) in addition to contemporaneous scores. For Vietnamese, estimates remain largely consistent regardless of which score we use. However, for math, we observe slightly lower estimates when using lagged scores. A one s.d. increase in the global *Teach* score is associated with a 0.08 s.d. increase in math skills and about a 0.04 s.d. increase in Vietnamese skills.

Breaking the global score down into the three "areas", we find that the positive relationship between the *Teach* score and academic skills, both math and Vietnamese, is driven in particular by the importance of "Instruction"

Table 1.10: *Teach* and Child Learning (Academic Skills)

	Math	Math	Vietnamese	Vietnamese
Lagged Teach				
Global *Teach*	0.076***		0.035**	
	(0.016)		(0.018)	
Class Culture		0.044*		0.012
		(0.023)		(0.024)
Instruction		0.046**		0.033
		(0.023)		(0.021)
Socioemotional		0.018		0.001
		(0.016)		(0.017)
Current Teach				
Global Teach	0.097***		0.032*	
	(0.017)		(0.018)	
Class Culture		0.060**		0.001
		(0.025)		(0.021)
Instruction		0.032*		0.032*
		(0.017)		(0.019)
Socioemotional		0.058***		0.007
		(0.014)		(0.016)
Students	7221	7221	7221	7221
Classroom	325	325	325	325
Schools	100	100	100	100

*Notes: The table reports estimates from regressions of test scores on Teach scores. All regressions are limited to children in classes where the teachers taught both the cohorts. All regressions include baseline student test scores and school-cohort-grade fixed effects. Standard errors (in parentheses) are clustered at teacher level.*** p<0.01, ** p<0.05, * p<0.1.*

for learning. This construct relates to the degree to teachers checks for understanding, gives feedback on student's work etc.

Examining the impact of classroom practices on non-cognitive skills (Table 1.11), we find no significant relationship between the aggregate measure of classroom practice quality and either Executive Function or other non-cognitive skills. It is important to note a key methodological difference in our approach to analyzing academic versus non-cognitive skills: while for academic skills, we were able to assess the relationship using both contemporaneous and lagged *Teach* scores, for non-cognitive skills, we are restricted to using only contemporaneous *Teach* scores. This limitation arises because: (a) we lack lagged measures of non-cognitive skills for grade 2; and (b) in grade 3, where we do have lagged measures of these skills, we face a scarcity of common teachers between the two cohorts. This prevents us from using lagged *Teach* scores. If we were to restrict our sample to only those schools where teachers have taught two cohorts and also have lagged non-cognitive skill measures, our sample size would be drastically reduced.

Table 1.11: *Teach* and Child Learning (Non-Cognitive Skills)

	SP	SP	Mastery	Mastery	EF	EF
Current Teach						
Global *Teach*	0.001		0.036		-0.011	
	(0.031)		(0.031)		(0.033)	
Class Culture		-0.055		0.016		-0.056
		(0.035)		(0.039)		(0.036)
Instruction		0.026		-0.019		0.033
		(0.038)		(0.040)		(0.049)
Socioemotional		0.083*		0.078*		0.040
		(0.043)		(0.046)		(0.038)
Students	2604	2604	2604	2604	1162	1162
Classrooms	461	461	461	461	250	250
Schools	139	139	139	139	125	125

*Notes: The table reports estimates from regressions of non-cogntiive test scores on Teach scores. All regressions include baseline student test scores and school-cohort fixed effects. The last two columns corresponds to EF which uses data only from cohort 1. Standard errors (in parentheses) clustered at the teacher level.*** p<0.01, ** p<0.05, * p<0.1.*

Our analysis of the impact of specific *Teach* areas on student outcomes reveals a nuanced picture of the relationship between classroom practices and skill development. While we found no significant overall relationship between the *Teach* global score and either Executive Function or non-cognitive skills, a closer look at specific areas shows that the “Socioemotional” area of teaching practices shows a statistically significant positive impact on students’ non-cognitive skills (self perception and independent mastery). Interestingly, teachers in our sample received their lowest scores in this area (as shown in Table 1.8).

Another interesting finding from our analysis is that the “Instruction” area, which is pivotal for cognitive skills, shows no significant impact on non-cognitive skills. This aligns with the low correlation we observed between instructional and socioemotional practices in our *Teach* analysis (see Table 1.9), suggesting that proficiency in instructional techniques does not necessarily translate to equal competence in socioemotional practices. This observation reinforces our initial insight: teachers who excel in fostering students’ cognitive development may not be effective in nurturing non-cognitive skills. The clear distinction between these two teaching domains helps explain the low correlation we initially found between cognitive and non-cognitive classroom effects. This underscores the importance of a balanced approach in teacher training and evaluation, one that addresses both cognitive and non-cognitive skill development to ensure comprehensive student growth.

Structural Quality and Teacher Characteristics

Before drawing out key conclusions, we conduct the final step of our analysis - estimation of a model which combines the *Teach* measures of process quality that we have been analysing with measures of teacher characteristics and structural quality. These measures include teacher characteristics (sex, ethnicity, marital status, education and experience) and conditions at work (including contractual arrangement, quality of the classroom infrastructure and student-teacher ratio). As before, the regressions also include controls for basic child characteristics and lagged measure of the outcome variable.

Table 1.12: Teacher Characteristics, *Teach* Score and Child Outcomes

	Math	Vietnamese	SP	Mastery	EF
Lagged *Teach* Score	0.077***	0.039**			
	(0.018)	(0.019)			
Current *Teach* Score			0.031	0.034	-0.015
			(0.032)	(0.033)	(0.035)
Permanent position	0.050	-0.026	-0.083	0.031	0.235
	(0.068)	(0.079)	(0.178)	(0.143)	(0.146)
Classroom needs repair	0.015	-0.102***	-0.175	-0.208**	-0.046
	(0.052)	(0.035)	(0.107)	(0.099)	(0.098)
Has 4 years of experience	-0.103	-0.031	0.058	0.132	0.063
or less	(0.101)	(0.086)	(0.221)	(0.177)	(0.048)
University education	-0.016	-0.041	-0.087	0.257***	0.203**
or higher	(0.032)	(0.031)	(0.085)	(0.070)	(0.092)
Student-Teacher ratio	-0.018**	-0.039***	-0.100	0.092	0.031
	(0.009)	(0.007)	(0.074)	(0.074)	(0.020)
Age	-0.003	-0.001	-0.001	0.005	-0.001
	(0.002)	(0.002)	(0.005)	(0.005)	(0.006)
Male dummy	0.002	0.054	0.063	-0.005	-0.202**
	(0.026)	(0.048)	(0.089)	(0.094)	(0.088)
Marriage dummy	-0.102***	-0.096***	-0.367***	-0.073	-0.162
	(0.035)	(0.035)	(0.122)	(0.128)	(0.107)
Minority dummy	0.065	-0.054	-0.350***	0.024	0.000
	(0.060)	(0.117)	(0.111)	(0.118)	(0.097)
Students	7221	7221	2604	2604	1162
Schools	100	100	461	461	250
Classrooms	325	325	139	139	125
R-squared	0.624	0.648	0.177	0.216	0.346
F-test (p-value)	0.00	0.00	0.00	0.01	0.00

Note: All regressions control for school FE, cohort dummy, student's ethnicity, age and gender. The regressions from first two columns are limited to children in schools in which we have teachers that have taught the two cohorts. While the last three columns include all the observations for which we have EF and non-cognitive skills. For these regressions, we also only use Current Teach scores. Standard errors in parentheses clustered at teacher level . * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.12 presents the results. On the whole, while the global *Teach* score continues to have a statistically significant positive effect on math and Vietnamese, with a few exception all the additional variables do not appear to have any impact on the measured skills. Among the few consistent trends is the finding of a negative effects of having married teachers on cognitive skills. Having a classroom which is in good state and does not need repairing seems to be especially beneficial for attainment in Vietnamese skills and mastery. We also find that having more students per teacher within a school has a negative impact on student's academic outcomes. None of the other teacher characteristics, including contract type or experience seems to be associated with student learning. However, having university education seems to matter for EF and mastery.

1.6 Conclusion

High-quality teachers are known to increase student learning at all levels of schooling and have long-lasting positive impacts on their students' later-life outcomes. However, most evidence for this comes from the U.S. and is largely limited to effects on students' numeracy and literacy skills. Crucially, evidence is lacking on what distinguishes more effective teachers from less effective ones, especially in lower and middle-income countries.

This paper addresses these questions for Vietnam, using a rich longitudinal data set. We present the first estimates of teacher effects on academic skills for Vietnam, a lower-middle-income country. Our findings reveal that the magnitude of teacher effects in Vietnam is lower than estimates from other countries, suggesting that variation in teacher quality may play a smaller role in driving inequalities in primary school attainment in Vietnam compared to other settings. This is particularly noteworthy given Vietnam's high performance in mathematics and verbal skills compared to much wealthier countries, as seen on its performance on the PISA. This suggests that Vietnam manages to achieve not only a high standard of teaching but also a higher degree of uniformity in the quality of teaching that primary school children receive than do other countries.

Extending beyond core academic subjects, we estimate classroom effects on a wider range of competencies, making this one of very few studies to do so. We find that classroom quality explains as much, if not more, variation in these broader skills. Our analysis of classroom practices using the *Teach* observation tool provides crucial insights into how different dimensions of quality map into development in the wide range of skills that children need for long-term success. This part of our research provides robust empirical evidence in this domain where data remain sparse, contributing to a more comprehensive understanding of what makes an effective learning environment and how it impacts various aspects of student development.

Interestingly, we find that better overall classroom practices result in improved performance in math and Vietnamese, but not necessarily in non-cognitive skills. The "Instruction" domain appears most crucial for academic skills but does not affect non-cognitive skills, while the "Socioemotional" domain is more important for non-cognitive skill development. This aligns with the low correlation we observed between instructional and socioemotional practices in our analysis, suggesting that proficiency in instructional techniques doesn't necessarily translate to equal competence in socioemotional practices that are effective in nurturing non-cognitive skills. We also explored the role of structural quality measures, such as teacher qualifications and experience, and quality of the classroom infrastructure. As found in other studies, we find little to suggest that these dimensions of teacher quality are strong predictors of children's skills.

The overall findings in this paper have significant implications for education policy in Vietnam, particularly in light of the country's priority to foster non-cognitive skills in students over the past decade. Our results suggest a substantial need for teachers to develop stronger skills in fostering autonomy, perseverance and social/collaborative skills among their pupils, which are especially important for non-cognitive skill development.

While our study demonstrates the value of teacher effect estimates and classroom observations in understanding educational quality, it also highlights the complexity of measuring teacher effectiveness. Value-added measures provide useful insights but are not the sole indicator of teaching quality. Our use of the *Teach* observation tool complements these measures by highlighting specific classroom practices that contribute to student outcomes. This approach, combining teacher effect estimates, classroom observations, and assessments of various student outcomes, provides a more comprehensive understanding of educational quality.

In conclusion, this research makes a significant contribution to our understanding of teacher effectiveness and classroom quality in a lower-middle-income country context. It offers valuable insights for policymakers and educators seeking to improve educational outcomes across a broad spectrum of skills crucial for long-term success, particularly in settings similar to Vietnam. By providing a more nuanced picture of what constitutes effective teaching and high-quality learning environments, our study paves the way for more targeted and effective educational interventions.

1.A Chapter 1 Appendix

Appendix 1.A.1 provides details on academic skills construction and distribution, while Appendix 1.A.2 gives details on non-cognitive skills tests.

1.A.1 Cognitive Assessments

Test Scoring

The cognitive tests are the primary outcome variables which will be used to measure student learning. Below we show the distribution of the cognitive test scores and an analysis of the quality of the test data using Item Response Theory.

All items administered under Maths and Literature tests were multiple choice questions, responses to which were coded as "1" for correct and "0" for incorrect. Table 1.A1 shows the raw cognitive test scores, and Figure 1.A1 their distributions. We see that the raw test scores exhibit adequate variation. The distributions of the raw Maths test scores are approximately normal for all grades in both the cohorts. The high density to the right in first cohort's Vietnamese grade 2 test indicates that the test was too easy. The distribution of the second cohort's grade 3 Vietnamese test scores have a similar shape, but it is less pronounced.

Linking test scores

The raw test scores however are not comparable across grades and cohorts as different questions were administered in each test. There were however some common (linking) items that were included in order to put students' performance on these tests on a common scale, using item response theory (IRT). This approach involved combining all test observations into a single data set (separately for Math and Vietnamese) and deriving IRT scores through a unified analysis. Our methodology treated all question responses as binary (correct or incorrect) and we opted for a 2-parameter logistic model in our analysis.

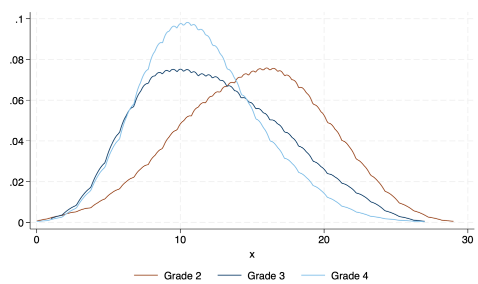
To validate our approach, we also examined the empirical fit of the estimated Item Characteristic Curves (ICCs) for each assessment round. Our analysis showed that the test questions effectively differentiate between students of varying achievement levels. Importantly, we selected the best performing items and also found no evidence of differential item function-

Table 1.A1 : Raw Cognitive Test Scores

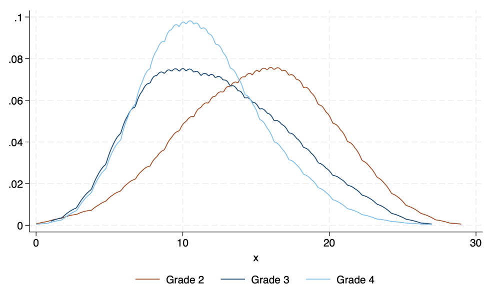
	Mean	S.D.	Min	Max	N
Cohort 1					
Math G2 (Baseline)	10.83	4.40	0	24	5008
Math G2	15.01	5.00	0	29	5041
Math G3	12.22	4.85	1	27	4836
Math G4	11.38	4.07	0	27	4689
Vietnamese G2 (Baseline)	12.52	4.86	0	25	5031
Vietnamese G2	18.75	6.64	0	30	5012
Vietnamese G3	16.39	6.55	0	30	4846
Vietnamese G4	17.90	5.99	0	30	4709
Cohort 2					
Math G2 (Baseline)	10.86	4.59	0	24	5159
Math G2	13.28	5.48	0	29	5148
Math G3	13.07	5.18	0	28	4961
Math G4	11.90	4.32	0	29	4695
Vietnamese G2 (Baseline)	14.08	6.21	0	30	5140
Vietnamese G2	14.80	6.75	0	30	5149
Vietnamese G3	16.54	6.60	0	30	4970
Vietnamese G4	17.74	5.98	1	30	4715

Notes: This table provides the mean, standard deviation, minimum and maximum of student's total raw scores for each grade and subject. The students were tested at the end of each academic year. For Grade 2, we also tested students in the beginning of the year.

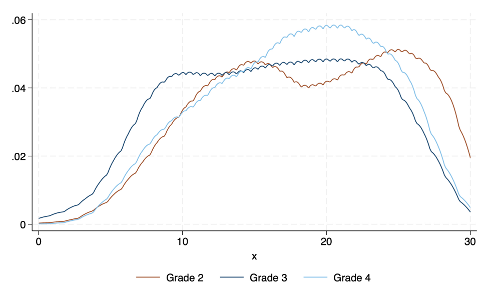
Figure 1.A1: Distribution of Raw Percentage Correct Scores



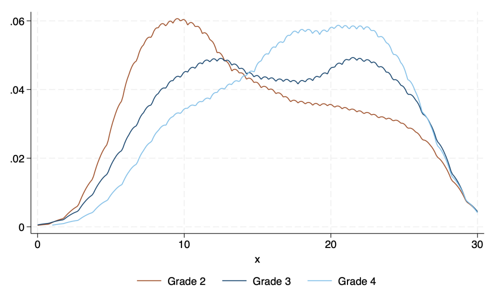
mathematics score, Cohort 1



mathematics score, Cohort 2



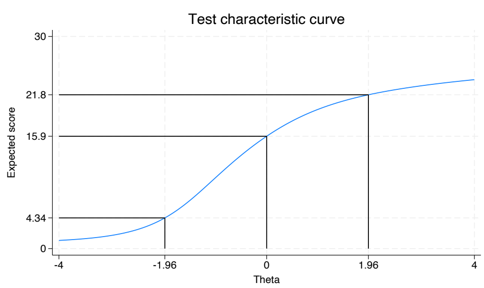
Vietnamese score, Cohort 1



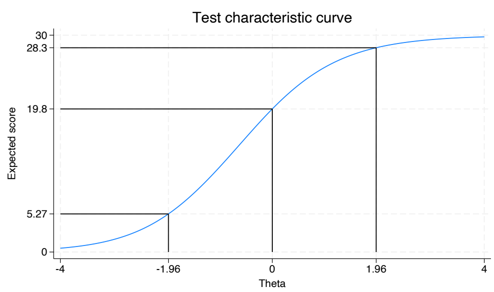
Vietnamese score, Cohort 2

ing across the different assessment rounds, ensuring the consistency and comparability of our measurements over time[9]. This IRT-based approach thus allows us to create a unified scale for student achievement, facilitating meaningful comparisons across different groups and assessment periods in our study.

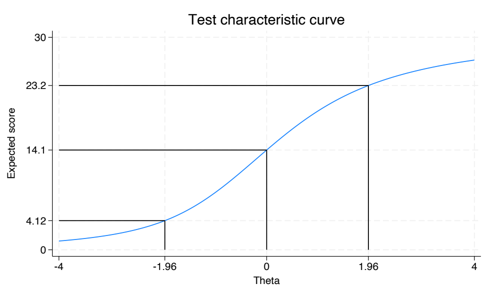
In Figure 1.A2 we also present the Test Characteristic Curve (TCC) for both cohorts for grade 2. The TCC's support the overall validity of the tests as higher ability students (higher theta) are expected to attain a higher score, and we do not see evidence of test ceiling or floor effects. The Test Information Functions (TIF), presented in Figure 1.A3 indicates information the test is able to give for different parts of the distribution of the latent trait of ability (theta). The more information the test provides for a given value of theta, the more accurately theta is estimated at that level. The figures show that for both cohorts most information is attained for students in the middle of the ability distribution. We also see that noise in the measures is more of an issue for Cohort 1 than Cohort 2. This makes sense since we were able to adjust the assessments for Cohort 2 following analysis of Cohort 1 data, eliminating questions that did not work well.



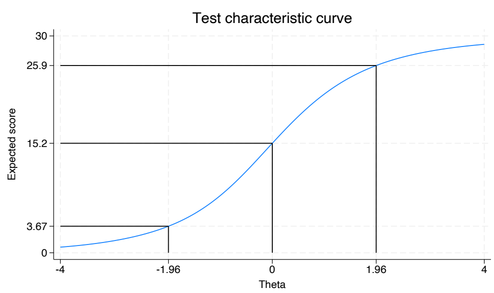
Math score, Cohort 1



Vietnamese score, Cohort 1



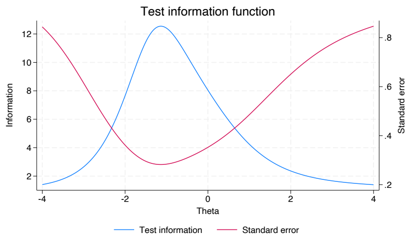
Math score, Cohort 2



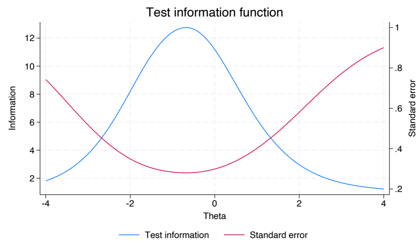
Vietnamese score, Cohort 2

Figure 1.A2: Test Characteristic Curves

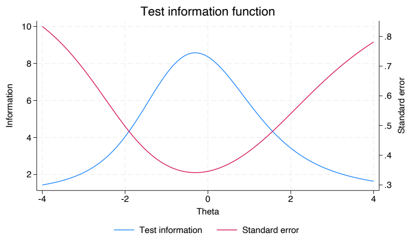
[9]A good performing item is the one that generally lie in the -2.5 to 2.5 difficulty range and that which helps discriminate between low and high ability students.



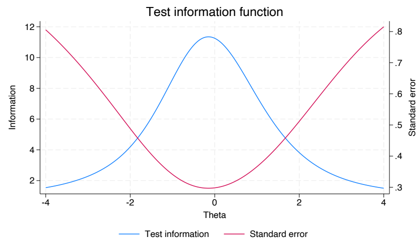
Math score, Cohort 1



Vietnamese score, Cohort 1



Math score, Cohort 2



Vietnamese score, Cohort 2

Figure 1.A3: Test Information Function

1.A.2 Non-Cognitive Assessments

Selection of Non-cognitive measures

Selected domains of non-cognitive skills were chosen to address key questions such as: 'What are the effect of non-cognitive skills on the acquisition of cognitive skills?' and 'What are the effects of individual teachers on the acquisition of non-cognitive skills relative to the prior ability of students and parental investments?'. Therefore, domains were chosen based on their importance to learning in the classroom environment and the potential for these domains to be influenced by teachers.

Executive functioning skills are predictors of long-term outcomes such as schooling achievement (Borella, Carretti, and Pelegrina (2010)), health (Will Crescioni et al. (2011)) and labour market success (Bailey (2007)). In the context of educational attainment executive functioning is essential as children with low levels of executive functioning will be unable to focus and can disrupt the learning environment of their peers. This project aims to quantify the impact of different attributes of individual teachers on the executive functioning of their students in order to investigate which specific teacher attributes contribute to development of executive functioning skills.

A key non-cognitive variable predictive of long term life outcomes including academic attainment, lifetime earnings, employment and the likelihood of engaging in risky behaviours is self-esteem and hence the self perception profile measure was selected. This measure can be used to assess teacher effectiveness in terms of non-cognitive outcomes and can also be a control variable in terms of cognitive outcomes. Self-characterization has been proved to change with age and with the area of a child's life. Differentiation starts when children approach middle childhood in areas such as scholastic competence, social competence, physical appearance and behavioural conduct.

Finally, the Intrinsic and Extrinsic Motivation instrument was selected, as an individual's motivation is a determinant of cognitive outcomes since it is related to the effort children exert and their ability to conduct independent study. In addition, the teacher's non-cognitive value-added effect on student's motivation can be computed to investigate the influence of teaching on students' motivation.

Distribution and Validity of assessments

Executive functioning is made up of three domains- inhibitory control, working memory, and cognitive flexibility. Together these higher order cognitive skills are required for reasoning, problem solving, and planning. As detailed in Section 1.3 we measure EF using the HF and BDS tasks. Figure 1.A4 shows the distributions of the raw scores for each of these tasks for Cohort 1[10]. These show that most students get a high score on the HF task (Panel (a)), on average getting about 21 out of 30 trials correct, and that most of them are able to repeat at least three numbers in reverse order. This corresponds to Level 2 in BDS task (Panel (b)). Cronbach's alpha measures the reliability of these instruments, which is the extent to which individual items are measuring the same construct. Table 1.A2 then shows the Cronbach alpha measure of internal consistency (Cronbach (1949)) which estimates the proportion of variance in the HF and BDS scores which can be attributed to true score variance. While it has been shown that the application of a simple "rule-of-thumb" threshold for what is considered an adequate level of internal consistency is not appropriate (e.g. Taber (2018)), we do see that it is reassuringly high for both sets of tasks at 77% and 91% respectively.

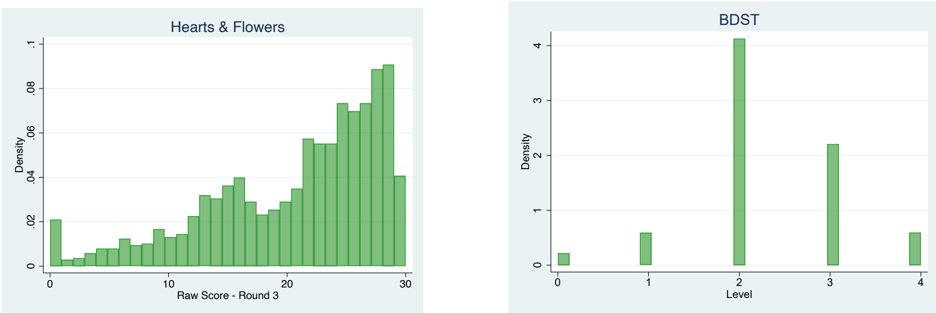
Table 1.A2 : Executive Functioning Cronbach's Alpha

	Cronbach's Alpha
BDS	0.77
Hearts and Flowers	0.91

Self-esteem is related to competences such as self-managed learning, self-assert, career self-orientation, self-study, social acceptance, communication and cooperation. In the tests, the specific domains of scholastic competence, social acceptance and physical appearance were assessed. The other non cognitive scale in our assessment is independent mastery vs. dependence on teacher. It assesses children's attitudes toward learning and mastery in classroom.

Thus, our non-cognitive measures include a global scale of students' self perception in the scholastic, social and physical appearance domains, as well

[10]We could only administer these tasks for Cohort 1



(a) Hearts and Flowers

(b) Backward Digit span Task

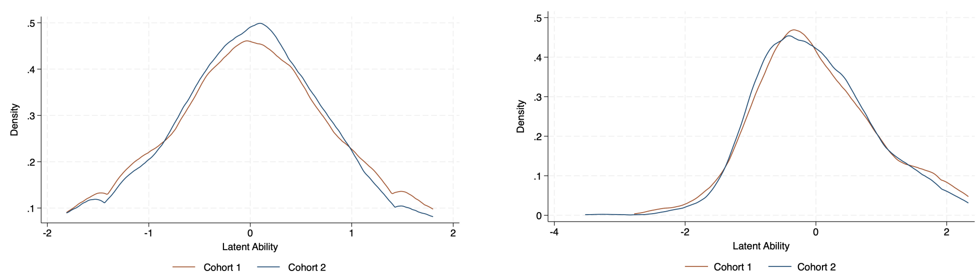
Figure 1.A4: Density Graph for Executive Functioning Measures

as a scale capturing the degree to which students are intrinsically versus extrinsically motivated in application to mastery of school material. Table 1.A3 shows the alpha coefficients for the global Self-Perception score, and the Intrinsic vs. Extrinsic Motivation scale for the two cohorts. These are similar across cohorts and suggest that for all of the scales the majority of variance in each of the scores can be attributed to true score variance.

Table 1.A3 : Self Perception Profile and Independent Mastery's Cronbach's Alpha

	Cronbach's Alpha	
	Cohort 1	Cohort 2
Total Self perception	0.78	0.78
Independent Mastery	0.74	0.73

As with scoring of the Math and Vietnamese tests, we use IRT model to construct the scores of these non-cognitive scales. Instead of using a logit model, however, we use an ordered logit model to map responses to unobserved latent factors since the responses to the non-cognitive scales are on a 5 point likert scale. We plot the distribution of the IRT scores of these scales in Figure 1.A5, and we find that all scales have adequate variation. The Independent Mastery scale has a distribution which is approximately normal which indicates that students are normally distributed across a spectrum of being internally or externally motivated. The self-perception scale is slightly skewed to the right hand side with a higher density in the right tail than in the left. Therefore, a high proportion of students rate themselves as having a high competency in school work, social relationships and physical tasks.



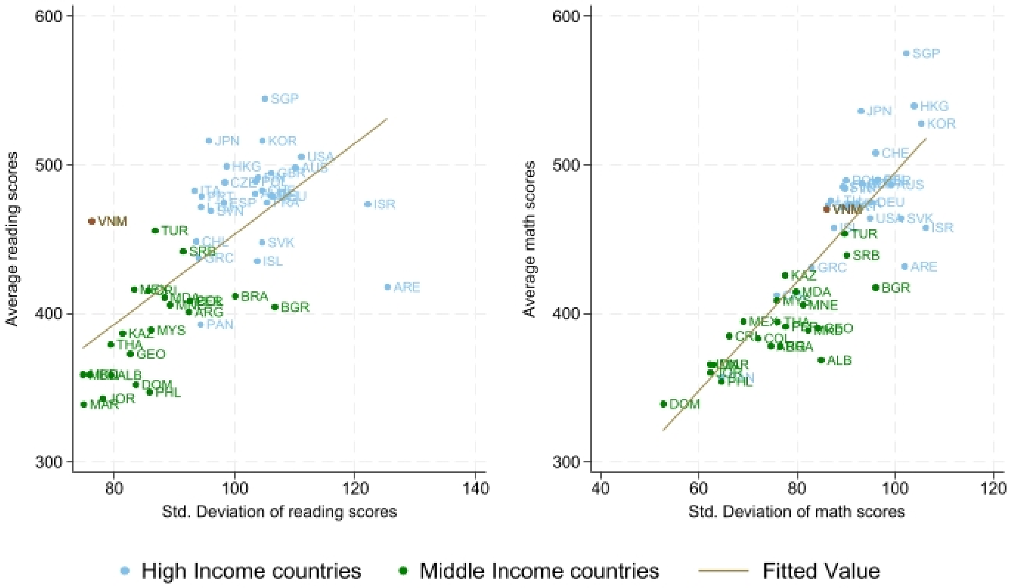
(a) Independent Mastery

(b) Self Perception

Figure 1.A5: Distribution of Self Perception and Intrinsic vs. Extrinsic Profile Subscales

1.A.3 Additional Figures

Figure 1.A6: Mean test scores in 2022 PISA, by standard deviations



1.A.4 Additional Tables

Table 1.A4 : Sample Representation

	Sample	VHLSS
Father's highest grade	8.6	7.5
Mother's highest grade	8.5	7.1
Household size	4.72	4.75
Father's age	36.9	37.8
Mother's age	34	34. 9
Percentage of Minority Population	23.2	24.8
Percentage of public school	100	99.2

Notes: This table compares our sample of students to the students in the Vietnam Household Living Standards Survey (VHLSS).

Table 1.A5 : Sample Size and Missing Responses

Instrument	Max. Sample	Sample	% Missing
Cohort 1			
Cognitive Tests G2 B	5,186	5,008	3.4%
Cognitive Tests G2 E	5,186	5,012	3.4%
Cognitive Tests G3	5,186	4,836	8.0%
Cognitive Tests G4	5,186	4,716	9.0%
Student Interview G2	1,654	1,654	0%
Student Interview G4	1,654	1,546	6.5%
Student Non-Cognitive Tests G2	1,654	1,654	0%
Student Non-Cognitive Tests G3	1,654	1,591	3.8%
Student Non-Cognitive Tests G3	1,654	1,541	6.8%
Parent Interview G2	1,654	1,628	3.87%
Parent Interview G4	1,654	1,512	8.5%
Teacher Interview G2	277	274	1%
Teacher Interview G3	277	268	3.2%
Teacher Interview G4	277	268	3.2%
Cohort 2			
Cognitive Tests G2 B	5,257	5,140	2.2%
Cognitive Tests G2 E	5,257	5,169	1.6%
Cognitive Tests G3	5,257	4,990	5.0%
Cognitive Tests G4	5257	4789	9.0%
Student Interview G2	1674	1,673	0%
Student Interview G4	1674	1,524	9.0%
Student Non-Cognitive Tests G2	1,674	1,665	0.5%
Student Non-Cognitive Tests G3	1,674	1594	4.7%
Student Non-Cognitive Tests G4	1,674	1524	9.0%
Parent Interview G2	1,674	1,638	2.1%
Parent Interview G4	1,674	1,500	10.0%
Teacher Interview G2	277	275	0.7%
Teacher Interview G3	278	273	2.0%
Teacher Interview G4	278	273	2.0%

Table 1.A6 : Cognitive and Non-Cognitive Tests Administered

Instrument	**Description**
Cognitive Test (C1, start of G2)	Maths Test with 25 items and Vietnamese test with 25 items
Cognitive Test (C1, end of G2)	Maths- 30 items and Vietnamese-30 items (M: 11 common items ; V: 12 common items from start of G2)
Cognitive Test (C2, start of G2)	Maths-25 items and Vietnamese-30 items (M: 16 common items; V: 12 common items from Cohort 1's G2 test
Cognitive Test (C2, end of G2)	Maths-30 items and Vietnamese-30 items (M: 12 common items ; V: 15 common items from start of G2)
Cognitive Test (G3 and G4)	Maths-30 items and Vietnamese-30 items (M: 11-12 common items; V: 13-14 common items from previous tests)
Non-Cognitive Test (C1, G2 and G3)	Executive Functioning, Self Perception Profile, Intrinsic vs.Extrinsic Motivation Scales
Non-Cognitive Test (C2, G2 and G3)	Executive Functioning, Self Perception Profile, Intrinsic vs.Extrinsic Motivation Scales

Notes: C1 corresponds to Cohort 1; C2 corresponds to Cohort 2 and G2,G3 and G4 correspond to Grades 2, 3 and 4 respectively.

Table 1.A7 : Descriptive Statistics

	Full Sample		Sub-sample	
	Mean	S.D.	Mean	S.D.
Student Characteristics				
Mathematics score in G2	14.189	5.312	14.348	5.285
Vietnamese score in G2	16.802	6.985	17.151	6.957
Male	0.521	0.500	0.520	0.500
Age (in years)	7.365	0.537	7.347	0.522
Parents Characteristics				
Father's age	37.019	6.205	37.475	6.225
Mother's age	34.105	5.677	34.323	5.420
Father's highest grade	8.649	3.245	8.852	3.164
Mother's highest grade	8.603	3.402	8.812	3.274
Teacher Characteristics				
Age (in years)	41.322	9.020	42.252	8.446
Married	0.916	0.278	0.894	0.309
Male	0.112	0.315	0.098	0.298
Highest education	0.616	0.487	0.585	0.495
Years of experience	19.888	9.471	21.089	9.161
Permanent contract	0.907	0.290	0.911	0.287
Number of Schools	140		60	

Notes: This table presents the means and standard deviations of students, parents and teachers characteristics. It compares the full sample with a subset of 60 schools used for estimating teacher value-added.

Table 1.A8 : *Teach* Framework

	Areas	Elements	Behaviors
Time on Task		Time on Learning	
Quality of Teaching Practices	Classroom Culture	Supportive Learning Environment	1.1 The teacher treats all students respectfully 1.2 The teacher uses positive language with students 1.3 The teacher responds to students' needs 1.4 The teacher does not exhibit bias and challenges stereotypes
		Positive Behavioral Expectations	2.1 The teacher sets clear behavioral expectations for classroom activities 2.2 The teacher acknowledges positive student behavior 2.3 The teacher redirects misbehavior and focuses on the expected behavior, rather than the undesired behavior
	Instruction	Lesson Facilitation	3.1 The teacher explicitly articulates the objectives of lesson and relates classroom activities to the objectives 3.2 The teacher explains content using multiple forms of representation 3.3 The teacher makes connections in the lesson that relate to other content knowledge or students' daily lives 3.4 The teacher models by enacting or thinking aloud
		Checks for Understanding	4.1 The teacher uses questions, prompts or other strategies to determine students' level of understanding 4.2 The teacher monitors most students during independent/group work 4.3 The teacher adjusts teaching to the level of students
		Feedback	5.1 The teacher provides specific comments or prompts that help clarify students' misunderstandings 5.2 The teacher provides specific comments or prompts that help identify students' successes
		Critical Thinking	6.1 The teacher asks open-ended questions 6.2 The teacher provides thinking tasks 6.3 The students ask open-ended questions or perform thinking tasks
	Socioemotional Skills	Autonomy	7.1 The teacher provides students with choices 7.2 The teacher provides students with opportunities to take on roles in the classroom 7.3 The students volunteer to participate in the classroom
		Perseverance	8.1 The teacher acknowledges students' efforts 8.2 The teacher has a positive attitude towards students' challenges 8.3 The teacher encourages goal setting
		Social and Collaborative Skills	9.1 The teacher promotes students' collaboration through peer interaction 9.2 The teacher promotes students' interpersonal skills 9.3 Students collaborate with one another through peer interaction

Table 1.A9 : *Teach* Areas: Factor Loading

Elements	Class Culture	Instruction	Socioemotioal	Global
Supportive Learning Environment	*0.478*	0.178	0.135	*0.496*
Positive Behavioral Expectations	*0.410*	0.219	0.140	*0.460*
Lesson Facilitation	0.158	*0.387*	0.032	*0.343*
Checks for Understanding	0.149	*0.364*	0.097	*0.355*
Feedback	0.154	*0.461*	-0.106	*0.335*
Critical thinking	0.138	0.239	0.199	*0.316*
Autonomy	0.238	0.036	*0.452*	*0.373*
Perseverance	*0.408*	0.067	0.071	*0.357*
Social and Collaborative Skills	0.036	-0.054	*0.485*	0.210

Notes: The table reports the factor loading of all the nine elements. In Teach original framework, the first two elements map to Classroom Culture; the next four elements, i.e, Lesson Facilitation, Checks for Understanding, Feedback and Critical thinking make up Instruction and the last three elements map to Sociomotional areas of teacher practices. Loading larger than 0.3 in absolute value are in italics.

Table 1.A10 : Within-School Standard Deviations of Classroom and Teacher Effects - Sub -sample

	Class Effects			Teacher Effect
	Grade 2	Grade 3	Grade 4	Grade 2
Mathematics	0.205	0.205	0.199	0.074
	(0.022)	(0.009)	(0.016)	(0.022)
Vietnamese	0.165	0.135	0.194	0.078
	(0.013)	(0.008)	(0.013)	(0.011)
Students (M)	7298	6875	6434	3936
Students (V)	7283	6883	6481	4590
Schools	101	101	101	50
Classes	394	384	374	100

Notes: In this table, we report the within school standard deviation of classroom and teacher effects adjusted for sampling error for a sub-sample of schools where according to principal interview, children were randomly allocated across classes. Bootstrapped standard errors in parenthesis.

Table 1.A11 : Within-School Standard Deviations of Classroom and Teacher effects

	Class Effects			Teacher Effect
	Grade 2	Grade 3	Grade 4	Grade 2
Mathematics				
No controls	0.224	0.240	0.220	0.073
Additional controls	0.221	0.236	0.219	0.069
Correlation	**0.989**	**0.990**	**0.990**	
Vietnamese				
No controls	0.164	0.179	0.184	0.070
Additional controls	0.163	0.175	0.183	0.068
Correlation	**0.981**	**0.982**	**0.980**	
Students	2353	2243	2100	1000
Schools	139	139	138	60
Classes	533	520	507	120

Notes: In this table, we report the within school standard deviation of classroom and teacher effects with and without additional controls for a sub-sample of students for whom we have parental data. We also report the correlation between the estimates derived from models with and without the controls.

Table 1.A12 : *Teach* Global Score Variation within Countries

Country	S.D.
VNM	0.30
ECA1	0.30
SSA6	0.31
SSA7	0.35
SSA3	0.44
SSA8	0.44
ECA2	0.46
SAR3	0.47
LAC2	0.48
LAC3	0.50
EAP2	0.51
SSA5	0.52
EAP3	0.52
SSA4	0.53
SAR4	0.53
SSA2	0.53
MNA2	0.54
SAR1	0.55
EAP1	0.57
SSA1	0.59
SAR2	0.59
LAC1	0.69
MNA1	0.85

Notes: The table reports the standard deviation of the Global Teach score across all countries where the Teach tool has been used. Given that these scores are measured consistently across countries, we believe they should be comparable between different countries. However, different number of video segments were used to create these scores in some countries. Here SSA corresponds to Sub-Saharan Africa; MNA corresponds to Middle East and North Africa; LAC is Latin America; SAR is South Asia Region; EAP is East Asia Pacific; ECA is Europe and Central Asia and VNM is Vietnam.

Chapter 2

The Joint Role of Home and School Inputs in Children's Learning

2.1 Introduction

It is well established that development during childhood is critical for lifetime human capital accumulation. Large literature exist on the roles played by key actors in this process – parents and schools. However, to date the great majority of papers treat these in isolation – the Child Development literature has focused on the role of parents, while the school Value Added literature on the role of schools. Clearly though children's learning and development outcomes are a product of the combination of these inputs which, furthermore, may interact and be inter-dependent. An important next step in understanding how skills are formed, and what drives formation of inequalities between children is to bring these two literature together.

This is what we do in this paper: we study the joint role played by parents and schools in the formation of skills of primary school-age children in Vietnam. Vietnam is an outlier among Lower-Middle Income countries with respect to the very high levels of learning achieved for its income level (Dang et al. (2023)). Understanding how this is achieved is a policy and academically relevant question. In the first chapter, we have presented findings which show that variation in teacher quality alone explains less variation in primary school pupil attainment than has been found in other contexts. This result suggests that differences in home inputs or the ways in which home inputs respond to, and interact with school inputs may play an especially important role in this context. This motivates the aim of this paper, which assesses the combined role of home and school inputs, as well as how a better characterisation of the skill production function can help us anticipate how parents might respond to policies that aim to improve school quality.

We utilise a rich longitudinal data set from 140 primary schools and over 10,000 pupils in second grade of primary school in Vietnam (age 7). We estimate separate skill production functions for cognitive skills - Math and Vietnamese as well as non-cognitive skills. These include measures of school and home investments as inputs into these production functions. Our primary measure of school inputs is the class value added (VA) estimate which subsumes classroom-level shocks as well as the teacher effect. In order to capture key features of the home environment and parent's material and time investment in the sampled child, we conducted in-depth interviews with the parents of the children in our sample. We also consider the role of time-investment into learning at home by the child, which might be particularly important in a high performing setting like Vietnam.

A key identification challenge in our analysis is the potential sorting of students and teachers which would bias our estimates of class effects, as well as of the parental response to classroom quality. We sample two classes per cohort per school in order to be able to estimate class effects within a school-grade-year, thus addressing concerns about non-random selection of pupils into schools. Concerns about sorting within schools are minimal in Vietnam, where streaming is against official government guidance. This is supported by reports from teachers and Head Teachers in our sample about how students and teachers are assigned to classes, as well as the role parents play in influencing the class that their child is in. Thus, the non-random sorting of students to classes is limited.

Even with robust measures of class quality we face the risk that omitted home inputs, unobserved shocks and child lagged ability may bias our estimates of the production function parameters. We address these concerns by exploiting the panel aspect of our data in order to estimate Value-added production functions where we assume that lagged test scores subsumes all the lagged inputs and ability (as in Todd and Wolpin (2007); Fiorini and Keane (2014); Keane, Krutikova, and Neal (2022a)). Several recent studies find value-added models are a reliable way to control for latent ability.[1] We also estimate our production function under an alternate specification using contemporaneous inputs only and compare the results with our benchmark specification which includes contemporaneous inputs as well as lagged test

[1]These studies rely on simulation (Guarino et al. (2015)) or comparison of experimental and VA estimates (Angrist, Pathak, and Walters (2013); Deming et al. (2014); Muralidharan and Sundararaman (2013)).

scores.

Having assessed the joint role of home and school inputs, we then focus on the relationship between the two. In particular, we examine whether and how parents respond to variation in the quality of the class that their child is in. This will depend on what they believe about the relationship between home and school inputs in the child skill production function, on their perception of the quality of schooling their child receives, as well as on their preferences over consumption and their children's human capital. This is a critical question for at least two reasons. First, clearly whether parents believe home investments to be complements or substitutes to school investments in the child human capital production will play a key role in determining the impact of any government investments in quality of schooling on pupil outcomes. In particular the effect of such policies will depend on the *crowding-in* or *crowding-out* of parents' home investment, which depends on parents' preferences and beliefs. Second, if households adjust their home inputs in response to changes in school inputs, not taking these responses into account in the estimation of the education production function produces estimates of policy effects rather than production function parameters since they include behavioural responses by the parents (Todd and Wolpin (2003); Das, Dercon, et al. (2013)).

Our results show that both school and home environments play crucial roles in the development of academic and non-cognitive skills during the early years of primary education. The analysis reveals that, when controlling for the quality of school inputs, within the home environment, parental time investment stands out as the variable most strongly and positively associated with children's academic skills. Interestingly, the benefits of parental time investment appear consistent across children of varying baseline academic abilities[2].

However, we observe significant variations in the effects of other inputs based on a child's initial academic ability. For children with higher baseline academic ability, their own time investments and material resources contribute substantially to their learning progress. Furthermore, while parental time investment emerges as the most critical factor in fostering academic skills, our data shows that material investments and child time investment correlate most strongly with the development of non-cognitive skills. This highlights the diverse impacts of different types of home inputs

[2]These abilities are measured at the beginning of the school year.

on various aspects of child development. We also find that, on the whole, home inputs generate higher returns for children who are in lower quality classrooms - suggesting that home and school inputs are substitutes in academic skill production function.

We then analyse whether parents respond to differences in classroom quality by changing investments in their children. Analysing parental behaviour shows that parents response aligns with the substitutability we find in the production function between school and home inputs. Specifically, when classroom quality improves, parents tend to *decrease* their material investments at home. This reduced investment in response to increased classroom quality suggests that, for average household, school inputs and home inputs are perceived as substitutes in the child's skill production function. This finding aligns with previous research by Greaves et al. (2023), Das, Dercon, et al. (2013) and Pop-Eleches and Urquiola (2013). Interestingly, we also find a contrasting response observed in children's time allocation. As classroom quality improves, we find a significant increase in the time children devote to homework. We hypothesize that this could be due to higher-quality classrooms/teachers assigning more homework, necessitating greater time investment from students.

Finally, we compared parental reports about the quality of their child's teacher with our 'objective' class value-added (VA) measure. This analysis helps to assess whether parents can distinguish between good and bad teachers. The results suggest that parents may not be well-informed about teacher quality or may lack critical information as they do not necessarily assign higher scores to objectively better teachers. As in Almås, Attanasio, and Jervis (2023), we therefore highlight the important role of measurement for development of context appropriate behavioural models which allow us to understand and anticipate parental behaviour (see also Caplin (2021)).

This paper makes several contributions to the literature. We are aware of only one other recent study which combines school and home inputs in the estimation of the child skill production function, but does so for a different context (US) and younger (pre-school age) children (Agostinelli, Saharkhiz, and Wiswall (2019)). There is also only a handful of studies that study the question of parental response to school inputs. Findings in this literature are mixed. While Das, Dercon, et al. (2013) find that in the context of India and Zambia parental response to unanticipated increases in public education spending are consistent with a situation in which household believe public and household educational spending to be substitutes in the production function, using directly elicited parental beliefs data, Attanasio, Boneva, and Rauh (2019) find the reverse to be the case in the UK. Using data on actual parental behavior also in the UK context, however, Greaves et al. (2023) show that parents actually reduce time investment in children in response to unanticipated school quality improvements.

Finally, we study a new context - Vietnam – which is a particularly interesting case since there is evidence that parents here are more vested in their children's education than parents in other contexts. For example, the survey conducted on behalf of the by Varkey Foundation by Ipsos MORI between 2017 and 2018, shows that Vietnam is the second among the countries in the sample in terms of parents' time spent helping their children with homework.[3]

The remainder of the paper proceeds as follows: Section 2.2 provides some details about the study context. We discuss the data in Section 2.3 and present the empirical strategy in Section 2.4. Our main findings are described in Section 2.5 and their implications are discussed in Section 2.6. Section 2.7 concludes.

2.2 Education in Vietnam

Vietnam's primary and secondary education system is divided into primary school (grades 1-5, starting at age 6), lower secondary school (grades 6-9), and upper-secondary school (grades 10-12). Vietnam also has pre-primary education (for ages 3-5), secondary vocational training schools, and many different post-secondary institutions. As of 2020, Vietnam had more than 12,000 primary schools, 8,000 lower-secondary schools, and 2,300 upper secondary schools.[4]

Virtually all primary schools in Vietnam are state-managed and thus are public schools. Primary education is also compulsory and tuition free. As

[3]The survey includes 27,380 parents' interviews across 29 countries (Argentina, India, Singapore, Australia, Indonesia, South Africa, Brazil, Italy, South Korea, Canada, Japan, Spain, China, Kenya, Turkey, Colombia, Malaysia, Uganda, Estonia, Mexico, United Kingdom, Finland, Peru, United States, France, Poland, Vietnam, Germany and Russia). All interviews were conducted online between the 8th December 2017 and 15th January 2018. The report can be found at `https://www.varkeyfoundation.org/media/4340/vf-parents-survey-18-single-pages-for-flipbook.pdf` (accessed June 2022).

[4]https://www.gso.gov.vn/en/homepage/

of 2016, about three-quarters of primary schools were providing "full day" (6 hours) instruction; the others received only "half day" (3.5 hours) instruction, with schools usually operating two shifts. An explicit goal of the government is to extend full-day schooling to poorer localities, but this has proceeded slowly. Although the aim was to achieve full-day schooling by 2015, review of progress in 2012 flagged that the goal was not going to be reached (Harris (2012)) and a 2018 Ministry of Education Action Plan still identified universal full-day schooling as a goal (MOET (2018)). Ho Chi Minh City's latest educational objective aims to implement full-day schooling in all primary schools and 70% of secondary schools. Despite Vietnam's notably brief school days and academic years by global standards, the practice of "extra study" classes is widespread. Parents across the country, including in rural areas, enroll their primary school-aged children in these additional sessions. However, the time and resources allocated to these additional classes differ among provinces.

Administratively, the Ministry of Education and Training (MoET) in Hanoi retains formal authority over the entire education system. MoET works with other line ministries to determine investments in education, and plays the leading role in education planning and in determining the content of curriculum (London (2011)).

2.3 Data and Measurement

2.3.1 Study and Sample

The data used in this paper were collected from 140 primary schools that is nationally representative of all of Vietnam's primary schools. These include both "half-day" and "full-day" instruction schools [5]. For each of the 140 schools, data were collected for two adjacent cohorts of students: those who started Grade 2 in the 2017-18 school year (henceforth Cohort 1), and those who started Grade 2 in the 2018-19 school year (Cohort 2). Data were collected for Cohort 1 in the 2017-18, 2018-19 and 2019-20 school years, when they were in Grades 2, 3 and 4, and for Cohort 2 for the 2018-19, 2019-20 and 2020-21 school years, when they were in Grades 2, 3 and 4.

For both cohorts, more than 5,000 students were tested in Math and Vietnamese in their classroom when they were in Grade 2, 3 and 4. The data

[5] In our sample, 130 out of 140 schools offer "full-day" schooling and 80% of our sample students attend "full-day" schooling.

were collected from a random sample of 20 students per class and each test lasted an hour. To gain deeper insights, we also collected more detailed data on a random sub-sample of slightly more than 1,600 students in each cohort. This extended data collection included comprehensive assessments of these students' non-cognitive skills and in-depth interviews with their primary caregivers.

In this paper we use data collected for Grade 2 only. Parental data, used in this analysis, were only collected when the children in the sample were in Grades 2 and 4. However, we focus only on Grade 2 because this is when we observe a substantial number of teachers teaching students across the two cohorts (a feature we exploit in the analysis)[6]. Table 2.1 provides descriptive statistics for the two cohorts when they were in Grade 2; these show that they are similar on basic observable characteristics such as age, ethnicity and parental age and education.

The pupils are spread across 140 schools and, in most cases, 2 classes per school. A feature that we exploit in this paper is that in 100 of these schools we have teachers that *did not* move with the class, which means that we observed the same teachers teaching classes from two neighbouring cohorts of children. We exploit this feature to increase the robustness of our analysis (as we discuss below). Therefore, our final analysis sample consists of children who were in Grade 2 either in 2017/18 or 2018/19, for whom we conducted Math and Vietnamese assessments at the beginning and end of Grade 2, and who were in schools where teachers *did not* move with the class. This gives us a sample of around 6000 children, spread across 100 schools. For a third of this sample, parents were interviewed at the end of Grade 2.[7]

2.3.2 Key Measures

2.3.2.1 Academic Skill Assessments

The Math and Vietnamese tests for each grade were developed by the Vietnam Institute of Educational Sciences. All tests administered as part of this study assess domains relevant to the school curriculum in Vietnam. The Math test assesses three skills: arithmetic, measurement and quan-

[6]In Grade 4, we only have 20 teachers that are common across the two cohorts

[7]Comparison between students with and without parental data is shown in Appendix Table 2.A1 . Sub-samples exhibit similar characteristics, with no statistically significant differences at the 5% level, except for minority status.

Table 2.1: Student-Level Summary Statistics by Cohort

	Cohort 1	Cohort 2
Outcomes		
Baseline Mathematics score (Raw)	11.17	11.14
	(4.424)	(4.563)
Baseline Vietnamese score (Raw)	12.99	14.36
	(4.849)	(6.056)
Controls		
Student's age at interview (in years)	7.420	7.366
	(0.547)	(0.538)
Male	0.538	0.495
	(0.499)	(0.500)
Ethnic minority	0.203	0.219
	(0.402)	(0.413)
Mother's age at interview (in years)	33.81	33.75
	(5.703)	(5.309)
Father's age at interview (in years)	36.91	36.78
	(6.401)	(5.796)
Father's highest grade	8.789	8.590
	(3.242)	(3.206)
Birth Order of the student	1.937	1.893
	(1.050)	(1.084)
Mother's education		
No education	0.133	0.128
	(0.341)	(0.335)
Completed primary education	0.220	0.211
	(0.427)	(0.421)
Completed secondary education	0.524	0.525
	(0.500)	(0.499)
Completed secondary education	0.125	0.135
	(0.342)	(0.353)
Students tested (Math,Vietnamese)	5070	5209
Students interviewed (including non-cognitive)	1654	1673
Schools	140	140
Classrooms	276	279

Notes: This table reports means and standard deviations of the characteristics of students and their parents across the two cohorts in Grade 2.

tities, and geometry. The Vietnamese test assesses four language skills: vocabulary, rhetoric, grammar and reading comprehension. Both tests assess three cognitive domains - knowledge, understanding and application. We piloted the tests twice: in May, 2017, before the first round of data collection, and January, 2019, before Grade 4 data collection. As part of each of the pilots we administered the tests to about 300 students per grade and collected comments on the tests from their teachers. The best performing items (test questions) were selected for the final versions of the tests based on analysis of their psychometric properties using factor analysis. The initial Math and Vietnamese tests given at the beginning of Grade 2 both had 25 items; tests administered at the end of Grade 2 had 30 items.

In all math and Vietnamese tests administered, common (linking) items were included in order to put students' performance on these tests on a common scale. To facilitate comparisons of student performance across different grades and between the two cohorts, we consolidated all test results into unified datasets, one for mathematics and another for Vietnamese language skills. We then use a structural measurement model to construct estimates of underlying latent factors capturing ability in Math and Vietnamese from students' responses on the tests. These techniques combine the information contained in the available measures efficiently. Building on a long history of application in psychometric (e.g. Van Der Linden and Hambleton (1997)) we use a measurement model based on Item Response Theory (IRT) which uses a logit model to map indicators of responses to the test questions onto unobserved latent factors. We estimate the measurement models by maximum likelihood using an Expectation-Maximization (EM) algorithm and using the parameters of the model we then estimate children's latent ability at different point in time. For ease of interpretation, these math and Vietnamese latent scores were standardized to have a mean of zero and a unit standard deviation within each grade. See Das and Zajonc (2010) and Singh (2020a) for recent application of these methods in economics.

2.3.2.2 Non-Cognitive Skills Measure

Self-Perception Profile

Our first measure of non-cognitive skills is designed to evaluate how students view themselves in various domains. It employs a comparative approach, asking students to relate themselves to fictional children who exhibit either high or low levels of specific skills or traits. The areas assessed include Academic performance (how students perceive their scholastic abilities), Social skills (how students perceive their interpersonal capabilities) and Physical appearance (how students perceive their own looks and physical attributes). We create a global scale of students' self perception which encompasses all three domains: scholastic, social and physical appearance.

Intrinsic vs. Extrinsic Motivation Skills

The second measure focuses on students' motivational orientations. In the context provided, only one scale was administered: independent mastery vs. dependence on the teacher. This scale assesses the extent to which a student enjoys solving problems on their own, indicating intrinsic motivation vs the degree to which a student relies on or prefers teacher assistance, suggesting more extrinsic motivation.

For both our non-cognitive measures, we applied Item Response Theory (IRT) model to construct latent non-cognitive ability, allowing for a more robust assessment of these skills.

2.3.2.3 School Inputs

In line with the education literature, we capture classroom quality as latent fixed effects, leveraging the fact that we have multiple students per class and beginning and end of grade test-scores. We estimate the following regression equation:

$$Y^k_{icst} = \alpha^k_{cs} + \tau Y^k_{icst-1} + \epsilon^k_{icst} \tag{2.1}$$

where Y^k_{icst} is end-of-school-year score of student i in classroom c in school s at time t on a test for subject k (mathematics, Vietnamese), α_{cs} are classroom indicators, Y^k_{icst-1} is beginning-of-school-year score on subject k and ϵ^k_{icst} is an i.i.d. error term. The classroom fixed effects (α_{cs}) are estimates of class VA under the assumption that, conditional on controls (test score at the beginning of the school year) students are randomly assigned to classrooms. Class VA is the effect that being in a given class, including being taught by a given teacher, has on a student's learning. This is the measure we use to capture the quality of schooling that children in the sample receive.

There are two main ways in which the assumption of random assignment of students to classes may not be satisfied in a real-life setting, such as ours. First, certain types of students may choose certain types of schools. Second, within schools there may be some systematic ways in which students are allocated to classes; for example they may be streamed by ability.

Sampling two classes per cohort per school allows us to address concerns about sorting across schools by conducting the analysis at the within school level. In order to do this we follow a common approach adopted in related literature (e.g. Chetty, Friedman, Hilger, et al. (2011); Araujo et al. (2016)) and redefine each classroom effect relative to the school average.

The demeaned classroom effect, used in the analysis, denoted by λ^k_{cs}, is:

$$\lambda_{cs}^{k} = \alpha_{cs}^{k} - \frac{\sum_{c=1}^{C_s} N_{cs}\alpha_{cs}^{k}}{\sum_{c=1}^{C_s} N_{cs}} \tag{2.2}$$

where C_s is the number of classrooms in a school and N_{cs} is the number of students in the classroom c in school s.

Further, in this setting, within school sorting across classes in a year-group is unlikely to be a big issue. First, this is not allowed by law: Resolution No. 02-NQ/HNTW of the Central Committee of the Party stipulates "Do not organize selected classes at all levels, do not organize specialized schools at primary and secondary school levels, except schools for the gifted in the arts and sports". Official Dispatch No. 2449 dated May 27, 2016 of the Ministry of Education and Training also states: "It is strictly forbidden to organize specialized schools and selected classes at the preschool, primary and secondary school levels in any form".

Despite these rules there are reports that, on occasion, schools have a "select" class for high ability children with the remaining children allocated randomly across the "normal" classes. During sampling we asked schools whether they had such a class, excluding these from the sampling frame if schools reported having one. However, there were still some schools where head-teachers reported to allocating classrooms to students, not based on ability, but based on their location. As a robustness check, we show results from the estimates of the production function (for both cognitive and non-cognitive skills) wherein we restrict our sample to only those schools where the principals reported random classroom assignment. In our companion paper (Carneiro et al. (2023)), we also show that our value-added estimates remains largely unchanged even when restricting to this sub-sample of schools.

2.3.2.4 Home Inputs

We collected a rich set of measures of the children's home environments and parental investment through interviews with the primary caregivers of the children. These include the time the child spends studying at home, how much help the child gets with home study from adult members of the household, availability of key study materials to the child at home - access to study space and key study materials including notebook, textbook, exercise kit, pencils etc. We also have measures of financial investments in child learning captured through educational expenditure on the child in the last 12 months on a range of inputs including study books, materials, travel to and from school, exam fees, extra-curricular and educational activities such as extra classes and tutoring.

We use this detailed data to construct three distinct measures of home inputs: parental material investment, parental time investment and child time investment. Child's time investment captures the amount of time that the child spends on homework on a school day of a typical school week and is measured in hours. This variable ranges from 0 to 5 hours, and, on average, children in our sample spend 1.5 hours studying at home. We standardize this variable to have a mean zero and standard deviation of one in the whole sample. Parental time investment combines the number of activities parents do with the child including reading, telling stories, singing, doing arts, playing sports, studying, talking, watching TV. We also use an alternate measure of the parental time investment that incorporates information on both the types of activities and their weekly frequency, using a factor model.[8] We interpret the retained factor as an index of parental time investment. The final input we consider in the production function is parents' material investment, which is constructed using yearly educational expenditure (education expenditure includes expenditure on the following items: books, stationary and other study materials, private tuition, and extra classes).[9] All inputs, including parental material investment, are standardized to have a mean of zero and a standard deviation of one in our sample. This standardization allows for easier comparison and interpretation of effects across different input measures in our analysis of human capital production.

2.4 Empirical Strategy

We start by estimating a child skill production function which includes school and home inputs. We use the three measures of home inputs described in the previous section: parental material investment, parental time investment and child time investment. We capture school inputs through the demeaned classroom fixed effect (class VA) estimated following the methodology described in Section 2.3. These classroom effects in-

[8]Following standard procedure, we retain one factor with an eigenvalue greater than one. Appendix Table 2.A5 shows that all items load positively on this factor

[9]The largest expenditure categories are study materials (44%), books (34%) and private tuition (19%). Only 25% of our sample students attend extra classes and 20% attend private tutoring.

clude classroom shocks as well as teacher effects. Since in this analysis we utilise this classroom fixed effect in individual level regressions with pupil test score as the dependent variable, we use the classroom value-added of the teacher estimated using test scores of the other cohort taught by that teacher.

A challenge in the estimation of child skill production functions is that one cannot measure all relevant inputs or control perfectly for child latent ability. The literature discusses various strategies to deal with this issue, as outlined by researchers such as Fiorini and Keane (2014), Del Bono et al. (2016), and Todd and Wolpin (2007). Our approach utilizes a value-added model, which combines contemporaneous inputs with lagged test scores and some parental background characteristics. Thus, we address this issue by exploiting the panel dimension of our data where the lagged test score (using assessments administered at the beginning of Grade 2) subsumes all the past investments and lagged ability. This specification assumes that the effect of inputs declines with age at a rate determined by the coefficient on the lagged test score. It's worth noting that other specifications dealing with the endogeneity of omitted lagged inputs, such as the cumulative model or first differences, require more extensive panel data. These models typically include lagged inputs and lagged parental background characteristics as proxies for unobserved inputs and lagged ability. However, given our data limitations, these alternative approaches were not feasible for our study.[10]

Our empirical specification is as follows:

$$Y^k_{icmst} = \beta_1 \lambda^k_{cmst} + x^{\intercal}_{icmst}\beta_2 + \beta_3 Y^k_{icms,t-1} + h^{\intercal}_{icmst}\gamma + \theta^k_{mst} + \nu_{icmst} \quad (2.3)$$

where Y^k_{icmst} is the end of year (end of Grade 2) test score in subject k - Mathematics, Vietnamese, and non-cognitive skills (Self perception and Independent Mastery) - of child i from class c, cohort m in school s at time t (end of Grade 2). $Y^k_{icms,t-1}$ is the child's score in subject k at time $t-1$ (beginning of Grade 2) [11]; $h^{\intercal}_{icmst}$ is a vector of home inputs (described

[10] While cognitive skills were measured both at the beginning and end of the academic year, we interviewed parents only at the end of Grade 2. The absence of parental data from the beginning of Grade 2 means we lack lagged measures of home inputs. This limitation prevents us from testing alternative specifications which control for lagged home inputs.

[11] We do not assess students' non-cognitive skills at the beginning of Grade 2. Owing to this data limitation, we use combined test scores (average of Mathematics and Vietnamese) from beginning of Grade 2 as lagged score in the non-cognitive skills production

above); λ^k_{cmst} is the class fixed effect capturing school inputs; $x^{\intercal}_{icmst}$ are additional time-invariant controls for the home environment of the child including maternal education, maternal age and household wealth, and child characteristics including gender, age and ethnicity. ν^k_{icmsg} is an i.i.d. error term. In the regression analysis we pool the two cohorts together, so θ_{mst} is the school-cohort fixed effect.

Lastly, in our study, we encountered a common challenge in empirical research: missing data. While we had access to a larger sample of student test scores and classroom value-added measures, our information on parental inputs was limited to a random subsample of approximately seven students per classroom. Instead of excluding observations with missing parental inputs and analyzing only the complete cases, we adopt the approach proposed by Abrevaya and Donald (2017), incorporating a linear imputation model for the missing parental inputs into our set of moment conditions and estimating the main model using the Generalized Method of Moments (GMM) framework. This method has been shown to yield efficiency gains compared to complete data estimators when dealing with partially missing covariates. Our strategy allowed us to leverage different sample sizes for different aspects of our analysis: using the smaller subsample of 1,700 students with complete data to estimate the relationship between test scores, value-added, and parental inputs, and the larger sample of 6,000 students for estimating classroom value-added and the relationship between test scores and value-added. This approach, grounded in the literature on missing data problems (Little and Rubin (2019); Muris (2010)), enabled us to maximize the use of available information across different parts of our analysis, aiming to increase the precision of our estimates and the robustness of our findings. A brief explanation of this missing data methodology is provided in the Appendix 2.A.1.

After estimating the child skill production function, we turn to the question of whether parental and child inputs at home are sensitive to the quality of inputs at school. To do this we estimate the following model:

$$H_{icmst} = \delta_1 \lambda^k_{mcst} + x^{\intercal}_{icmst}\delta_2 + \theta_{mst} + \upsilon_{icmst} \tag{2.4}$$

This model captures the association between H_{icmst} - each of the measures of parental and child inputs in vector $h^{\intercal}_{icst}$ in Eq 2.3 - and class quality measures by λ^k_{cmst} - the class fixed effect.

function.

2.5 Results

2.5.1 Child Skill Production Functions

We start by presenting estimates of the child skill production functions in Tables 2.2 and 2.3. Table 2.2 shows the production function estimates separately for Math and Vietnamese. All regressions control for child baseline score and school-cohort fixed effects; they also include controls for household characteristics (maternal education, household wealth, mother's age) and child controls (age, gender, and ethnicity).

In all specifications school input, captured by the class VA estimate, is highly statistically significant. Our class VA estimates are derived using test scores from other cohorts, not the current one. Importantly, we find significant positive associations between parents' time and child attainment across all measures of academic skills. This suggest that one standard deviation increase in parental time investment is associated with an increase of 0.06 standard deviations in Mathematics skills and 0.05 standard deviations in Vietnamese skills. [12] These investments encompass activities such as reading to the child, playing sports, and other interactive engagements. Parental time investment appears to have a stronger effect on mathematics skills compared to Vietnamese skills. As a robustness check, we also employed an alternative specification using a factor model for parental time investment. The results of this analysis, presented in Appendix Table 2.A4 , closely aligns with our primary specification.

Looking at the non-cognitive skill production, Table 2.3 shows that while parental time investment matters for academic skills, child's time and material investment appear to have a greater impact on increasing non-cognitive skills. This pattern may be explained by the nature of non-cognitive skill development. Increased time spent by children studying at home is likely to enhance their scholastic self-perception. The increased engagement of children with academic material can lead to improved understanding and performance, which in turn reinforces their positive self-perception in scholastic domain. Meanwhile, the positive association between material investment and independent mastery, can be attributed to the resources these

[12] We also compare the results of our specification with a model using only contemporaneous inputs in Appendix Table 2.A2 . Our findings indicate that inputs significant in our specification remain significant in the alternative model, albeit with slightly larger magnitudes.

investments provide. When parents invest in educational materials such as books, study guides, or private tuition, they equip their children with tools to engage in independent learning and problem-solving. This increased access to learning resources may foster a sense of autonomy and competence in children.

Table 2.2: Production Function (Math and Vietnamese)

	(1) Math	(2) Math	(3) Vietnamese	(4) Vietnamese
Child's time	0.012	0.011	0.005	0.007
	(0.018)	(0.018)	(0.018)	(0.018)
Material Investment	-0.000	-0.002	0.013	0.012
	(0.023)	(0.023)	(0.020)	(0.020)
Parent's time	0.061***	0.062***	0.050***	0.049***
	(0.019)	(0.019)	(0.018)	(0.018)
Classroom VA		0.188**		0.276***
		(0.074)		(0.056)
Baseline score	0.631***	0.629***	0.740***	0.740***
	(0.013)	(0.013)	(0.013)	(0.013)
Child controls	Yes	Yes	Yes	Yes
Observations	5929	5929	5929	5929

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's VA. Standard errors are clustered at classroom level. *** p<0.01, ** p<0.05, * p<0.1.*

A limitation in our methodology when assessing non-cognitive production function is that we do not have self-perception and mastery measures at the beginning of Grade 2. To address this limitation, we use lagged cognitive scores as proxies for lagged non-cognitive scores. This approach is based on the assumption that omitted inputs affect current parental investment primarily through lagged cognitive scores.

We also report the results from a sub-sample comprising of schools where principals reported random student assignment to classrooms as a robustness check in Appendix table 2.A3 . Even in this sample, inputs significant in our primary specification remain significant, though with slightly smaller magnitudes.

Next, we take a closer look at the role that home and school inputs play in shaping children's skills by investigating whether (1) returns to a given input differ depending on levels of other inputs - for example, is parental time investment more important for children going to lower or higher quality classrooms?; and (2) whether the effects of inputs differ for children with lower and higher levels of ability - for example, do higher ability children

Table 2.3: Production Function (Non-Cognitive)

	(1) SPP	(2) SPP	(3) Mastery	(4) Mastery
Child's time	0.073**	0.073**	0.041	0.041
	(0.031)	(0.031)	(0.030)	(0.030)
Material Investment	0.055	0.055	0.073**	0.072**
	(0.035)	(0.035)	(0.033)	(0.033)
Parent's time	-0.034	-0.034	-0.003	-0.003
	(0.033)	(0.033)	(0.032)	(0.032)
Classroom VA		0.042		-0.088
		(0.188)		(0.210)
Baseline Score	0.003	0.003	0.050	0.051
	(0.033)	(0.033)	(0.035)	(0.035)
Child controls	Yes	Yes	Yes	Yes
Observations	1436	1436	1436	1436

*Notes: The baseline score here is the lagged academic score (average of Math and Vietnamese). All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's cognitive VA. Standard errors are clustered at classroom level.*** p<0.01, ** p<0.05, * p<0.1.*

benefit more or less from receiving high quality teaching than lower ability children?

Insights into these interactions are important for several reasons. First, they allow us to advance understanding of the process of skill formation in childhood and the role that different inputs play individually and in combination with each other for different children. Second, they are important for thinking about potential impacts of education policy. For example, they can help us learn which children stand to benefit most from policies which improve quality of schools and, thus, whether such policies would increase or reduce attainment gaps between different groups of children. We can also get some insights into how parents might respond to such policies. For example, if we find that home and school inputs are substitutes in the skill production function then, given plausible assumptions about parental preferences, optimal parental response to an improvement in school quality could be to reduce home investment into children's skills. As the discussion in the next section makes clear, complete answers to these questions are beyond the scope of this paper; nevertheless, we are able to make some progress relative to the existing literature, motivating directions for future work on this topic.

To investigate these questions we add interactions between all of the inputs, as well as between inputs and baseline scores to the production functions estimated above. As before, we show subject specific measures - Math (Table 2.4) and Vietnamese (Table 2.5) . The first column of Tables 2.4 and 2.5 correspond to columns 2 and 4 of Table 2.2 respectively, and they do not include any interaction terms. Column 2 hones in on variation in returns to inputs by child baseline academic ability, showing interactions of child baseline score with each of the home and school inputs. In Columns 3-5 we study whether and how returns to inputs vary depending on levels of other inputs: Column 3 shows interactions between the school input (class VA) and each of the home inputs, followed by Column 4 which shows interactions between home inputs and school inputs along with baseline score. Finally, in Column 5 we include the full set of interactions.

This analysis yields several interesting findings. Most importantly, in both Tables 2.4 and 2.5, we observe that children with higher baseline ability benefit more significantly from higher quality teachers/classrooms.[13] This is evidenced by the statistically significant coefficient on the interaction term between class value-added (VA) and baseline test scores. This finding holds for both Mathematics and Vietnamese. The estimates reveal that higher quality teachers are more effective for children with higher baseline skills. These results are consistent with the notion of complementarity highlighted in the literature (Cunha and Heckman (2007)) and with the evidence from several developing countries showing that despite being enrolled in school, many children are not learning because of previously accumulated learning delays (Muralidharan, Singh, and Ganimian (2019)).

This insight highlights the critical importance of early childhood development interventions in preparing children for successful learning experiences in primary school and beyond. Recent research by Andrew et al. (2024) demonstrates that early interventions not only has short term gain but can yield significant returns in the medium run too by enhancing children's ability to learn. Specifically, they find that children who received an early intervention exhibited higher school readiness four years later.

Next, we consider whether the effects of inputs vary depending on the levels of other inputs. To do this we allow for interactions between all of the home and school inputs in Column (5). Here, we see negative and statistically significant coefficients on the interaction terms between class VA and parental time investment. The key input for children appears to be parental time, as indicated by the positive and statistically significant

[13]In this paper, since the school input captured by class VA subsumes classroom shocks as well as teacher effects, we use the terms classroom effect, classroom quality and teacher quality interchangeably.

coefficient on the uninteracted parental time investment term in Column 5 of Table 2.4. The negative coefficient on the interaction term between school input and parental time input, which is statistically significant in both Math and Vietnamese specification, suggests children going to lower quality classrooms benefit more from spending time with parents than those going to higher quality classrooms. A one s.d. increase in parental time investment increases test scores of students going to lower quality classroom (lower by one s.d.) by 0.16 s.d. more than those going to higher quality classroom. These are consistent with substitutability between home and school inputs in the child skill production function.

Table 2.4: Production Function with Interactions (Mathematics)

	(1)	(2)	(3)	(4)	(5)
Child's Time	0.011	0.015	0.010	0.013	0.012
	(0.018)	(0.019)	(0.018)	(0.019)	(0.019)
Material Investment	-0.002	-0.005	-0.002	-0.005	-0.013
	(0.023)	(0.025)	(0.024)	(0.025)	(0.025)
Parent's Time	0.062***	0.061***	0.065***	0.064***	0.062***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Classroom VA	0.188**	0.195***	0.167**	0.170**	0.165**
	(0.074)	(0.067)	(0.066)	(0.067)	(0.066)
Baseline score	0.629***	0.631***	0.629***	0.630***	0.631***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
VA x Child			0.070	0.056	0.058
			(0.101)	(0.100)	(0.100)
VA x Material			-0.001	-0.003	-0.042
			(0.126)	(0.129)	(0.137)
VA x Parent			-0.238***	-0.234***	-0.221***
			(0.087)	(0.086)	(0.085)
VA x BL		0.114*	0.154**	0.145**	0.149**
		(0.069)	(0.071)	(0.071)	(0.071)
Bl x Child		0.033*		0.032*	0.031*
		(0.017)		(0.017)	(0.017)
BL X Material		0.008		0.008	0.008
		(0.019)		(0.018)	(0.017)
BL X Parent		-0.016		-0.014	-0.018
		(0.017)		(0.017)	(0.018)
Parent x Child					-0.003
					(0.018)
Parent x Material					0.025
					(0.016)
Material x Child					0.008
					(0.020)
Child controls	Yes	Yes	Yes	Yes	Yes
Observations	5929	5929	5929	5929	5929

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's VA. BL refers to Baseline scores administered at the beginning of Grade 2. Standard errors are clustered at classroom level. *** p<0.01, ** p<0.05, * p<0.1.*

Our analysis also reveals differential impacts of home inputs on children's

Table 2.5: Production Function with Interactions (Vietnamese)

	(1)	(2)	(3)	(4)	(5)
Child's Time	0.007	0.014	0.007	0.012	0.016
	(0.018)	(0.017)	(0.018)	(0.017)	(0.018)
Material Investment	0.012	-0.004	0.012	-0.005	0.012
	(0.020)	(0.021)	(0.020)	(0.021)	(0.021)
Parent's Time	0.049**	0.047***	0.047***	0.047***	0.050***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Classroom VA	0.276***	0.294***	0.302***	0.303***	0.301***
	-0.056	(0.057)	(0.065)	(0.065)	(0.065)
Baseline score	0.740***	0.736***	0.739***	0.736***	0.734***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
VA x Child			-0.084	-0.076	-0.090
			(0.115)	(0.112)	(0.113)
VA x Material			-0.113	-0.150	-0.178
			(0.140)	(0.152)	(0.144)
VA x Parent			-0.196	-0.204*	-0.210*
			(0.122)	(0.122)	(0.122)
VA x BL		0.142*	0.203**	0.200**	0.194**
		(0.084)	(0.093)	(0.091)	(0.091)
BL x Child		0.044**		0.044**	0.045**
		(0.018)		(0.018)	(0.019)
BL X Material		0.029**		0.032**	0.028*
		(0.013)		(0.014)	(0.015)
BL x Parent		-0.004		-0.005	0.003
		(0.018)		(0.018)	(0.018)
Parent x Child					-0.010
					(0.017)
Parent x Material					-0.044***
					(0.015)
Material x Child					0.006
					(0.017)
Child controls	Yes	Yes	Yes	Yes	Yes
Observations	5929	5929	5929	5929	5929

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's VA. BL refers to Baseline scores administered at the beginning of Grade 2. Standard errors are clustered at classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.*

learning outcomes based on their baseline academic ability levels. This heterogeneity in effects provides further insight into the dynamics of skill formation across the ability spectrum.

For children with higher baseline academic ability, we find that their own time investment plays a crucial role in educational development. This is evidenced by positive and statistically significant coefficients on the interaction term between child time investment and child ability for both mathematics and Vietnamese language skills.

When it comes to material educational investments made by parents, we observe a more nuanced pattern. For Vietnamese language skills, higher-ability children appear to derive greater benefits from these investments.

The interaction term between material inputs and child ability is positive and significant for Vietnamese. This indicates that material investments are more productive for learning outcomes of higher-ability children, with less apparent impact on lower-ability children in this subject area. These findings suggest that children with higher baseline abilities are better positioned to leverage both their own time investments and the material resources provided by parents, creating a compounding effect on their skill development.

Focusing on these results by subject in Tables 2.4 and 2.5 show that heterogeneity in productivity of own time and material between higher and lower ability children is most evident for Vietnamese.

In the remainder of the chapter we take stock of our findings to consider what they imply for thinking about impacts of education policies which improve school quality.

2.6 Discussion

If we consider the model proposed by Currie and Almond (2011) with a Cobb-Douglas utility function defined over parents' consumption and child human capital, and a constant elasticity of substitution for the child's human capital production function, then parents' responses to an increase in school quality depend only on the technology parameter (see also Greaves et al. (2023)). Within this framework, in a world where, as we find, school and home inputs are substitutes, educational policy which improves the quality of school should result in a reduction in parental investment.

To check whether this is the case we estimate the model described in Equation 2.4, where the outcome variables are our measures of home investment (parental material and time investment, and child time investment) and the explanatory variable is our measure of school quality - the class VA. We look at response of home investment to class VA calculated for Math and Vietnamese separately. As before, we include controls for home and child characteristics as well as school-cohort fixed effects.

The findings presented in Table 2.6 show that both parents and children do indeed adjust their investment behaviour to the quality of classroom, though only in some dimensions. Examining the child's time investment measure first, we observe positive associations across specifications, though statistical significance is only achieved for mathematics. This suggests that as classroom quality improves, children tend to increase their own input. To illustrate, the estimates in the first row of column 5 indicate that a one standard deviation improvement in the quality of mathematics instruction corresponds to a 0.35 standard deviation increase in a child's time investment.

Shifting our focus to the material response in relation to class quality, the estimates reveal negative associations between class value-added (VA) and parental material investments. This pattern suggests that parents tend to reduce their material investment when there's an increase in the quality of teaching, specifically for Vietnamese.

Table 2.6: Parental Response to School Quality

	Expenditure		Time - parents		Time - child	
Class VA	(1)	(2)	(3)	(4)	(5)	(6)
Math	-0.237		-0.070		0.350***	
	(0.190)		(0.109)		(0.119)	
Vietnamese		-0.357*		-0.116		0.112
		(0.216)		(0.123)		(0.153)
Child Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1511	1511	1511	1511	1511	1511
R^2	0.464	0.464	0.345	0.352	0.331	0.333

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth, average classroom characteristics and school-cohort fixed effects. Class VA is the other cohort's VA. Standard error (in parenthesis) are clustered at classroom level. *** p<0.01, ** p<0.05, * p<0.1.*

A response by parents to improvements in teacher quality aligns with the predictions of theoretical frameworks in the presence of substitutability between home and school inputs in the skill production function. To understand this, we first note that parents respond to what they *perceive* teacher quality to be. In order to interpret their responses we, therefore, need to understand how parents' perceptions map to the objective measures of quality (class VA) that we use in our analysis as well as the measure based on teacher practices (*Teach*). To look at this we utilise information collected from parents on how good they think their child's teacher is as a measure of parental perceptions of teacher quality and look at the relationship between this measure and the objective measure we use in the analysis.

Parent's rating of their child's teacher quality is measured on a five point scale ranging from "very low" to "very high". Our data reveals a generally positive perception of teachers among parents. Notably, only 0.1 percent of teachers receive a "very low" quality rating, while approximately 75 percent are categorized as having "high" or "very high" quality. We employ multiple measures of teacher quality to ensure a comprehensive evaluation - Math and Vietnamese class VA estimates as well as *Teach* global score[14].

We look at the relationship of parental rating for each teacher by regressing this rating on different measures of teacher quality. Table 2.7 report results from simple OLS as well as ordered probit. The results reveal that parents do not necessarily have accurate assessments of teachers' Value-Added (VA) estimates. This suggests that parents may not be accurately identifying high-performing teachers based on these quantitative measures of student achievement growth. However, we found a strong alignment between parental ratings and teacher quality as measured by the Current *Teach* observational instrument. This indicates that while parents may not be aware of VA measures, they seem to have a some right perception of teacher quality when it comes to observable classroom practices in the current year. This discrepancy raises important questions about what aspects of teaching quality are most visible or important to parents.

Table 2.7: Parental Perception and Measures of Teacher Quality

	(1) OLS	(2) Ordered probit
Math VA	0.067	0.162
	(0.094)	(0.201)
Vietnamese VA	0.147	0.150
	(0.110)	(0.247)
Lagged *Teach*	-0.019	-0.026
	(0.029)	(0.073)
Current *Teach*	0.038**	0.081**
	(0.018)	(0.038)
Child controls	Yes	Yes
Observations	1520	1520

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's VA. Each cell in the table corresponds to a separate regression. Standard errors are clustered at classroom level. *** p<0.01, ** p<0.05, * p<0.1.*

In sum, this analysis suggests that parents likely have limited or incomplete knowledge about teacher quality. Their response to improved classroom

[14] *Teach* global score has been associated with improved cognitive skills in students. For a more detailed exploration of the *Teach* analysis and its impact on learning outcomes, we refer readers to our companion paper (Carneiro et al. (2023))

quality, however, are consistent with the predictions of standard models used in the literature. One possible explanation could be that parents are responding to observable changes in their child's behavior and observable classroom practices. Our earlier findings showed that children increase their time investment in response to improvements in school quality. This visible shift in the child's engagement with schoolwork could also be a factor influencing parents' beliefs about teacher effectiveness.

The above observation connects to a growing body of literature exploring how the investment decisions of parents in the process of child development are shaped by their perception or beliefs about the production of child development, how mis-perceptions can lead to a sub-optimal level and mix of investments, and how targeted interventions can improve child development by correcting these mis-perceptions (e.g. Boneva and Rauh (2018); Attanasio, Cunha, and Jervis (2019b); Cunha, Elo, and Culhane (2020)). In our application, the observed parental response to variation in teacher quality could be rationalised by the belief that home and school inputs are substitutes in the production function. However, we lack direct measures of these beliefs in our data, preventing us from testing this hypothesis conclusively. Attanasio, Boneva, and Rauh (2019) study parental beliefs in the UK context and find that parents believe home and school inputs to be complements. In fact, whether and how beliefs can be measured, as well as how they can be used in analysis, is a subject of a growing new literature (e.g. Almås, Attanasio, and Jervis (2023)) and is a promising area for future research.

2.7 Conclusion

In this paper, we study the joint role played by home and school inputs in the formation of academic skills of primary school-age children in Vietnam. We utilise a rich longitudinal data set from 100 primary schools and around 6,000 primary school pupils to estimate production functions for Math, Vietnamese and non-cognitive skills. We consider the role of school inputs (measured by class VA), parents' time and material investments, as well as children's own time investment. We allows for interactions between these different inputs, as well as between inputs and child ability.

Our results show that, in addition to the school inputs, parental time investment is relevant in the academic skill production function and is positively associated with child attainment. For non-cognitive skills, however, we find material and child time investment to be more important. We observe that a child's own time investment is more productive for those with higher baseline achievement in both math and Vietnamese skills. Interestingly, while material investments at home also yield higher returns for higher-ability children in Vietnamese, this effect is not observed for math skills. At the same time, our results suggest that school and home inputs are substitutes in child skill production, especially parental time; kids going to to lower quality classrooms benefit more from home inputs than those going to higher quality classrooms.

We highlight that an in-depth understanding of the process of skill formation which takes into account all of the key inputs is critical for designing education policies which achieve desired goals. An important factor in this is whether parents are sensitive to what schools do: do they respond to changes in school quality? and, if so, do these changes crowd-in or crowd-out parents' efforts at home? We find that, as has been found elsewhere, parents do indeed respond to changes in quality of school inputs. We also show, however, that, in order to anticipate these responses accurately we need both an accurate characterisation of the skill production function, as well as context appropriate behavioural models.

The finding of substitutability between home and school inputs in our production function analysis, seen through the lens of models of parental behaviour used in past studies, would suggest that policies which improve quality of schooling in Vietnam would crowd-out parental investment at home. We provide evidence supporting this hypothesis in our sample as we find that higher classroom quality is correlated with *lower* investment at home. However, we cannot say that they are able to accurately differentiate between higher and lower quality teachers as the parental assessments of teacher quality does not align well with objective measures. Our findings underscore the importance of further investigating parental decision-making processes in future research. Innovations in measurement techniques could enable the development of context-appropriate theoretical models. Such models would enhance our understanding and ability to accurately predict parental behavioral responses to education policies, ultimately 'opening the black box' of parental decision-making.[15]

[15]For recent examples see Attanasio, Cunha, and Jervis (2019a), Delavande and Zafar (2019), Adams-Prassl and Andrew (2020), Ameriks, Briggs, Caplin, Lee, et al. (2020), Ameriks, Briggs, Caplin, Shapiro, et al. (2020), and Giannola (2020).

2.A Chapter 2 Appendix

2.A.1 GMM with Missing Data

The main regression of interest is the equation of test scores on home inputs, baseline score and other cohort's class value added.

$$Y^k_{icmst} = \beta_1 \lambda^k_{cmst} + x^{\intercal}_{icmst}\beta_2 + \beta_3 Y^k_{icms,t-1} + h^{\intercal}_{icmst}\gamma + \theta^k_{mst} + \nu_{icmst} \quad (2.5)$$

To simplify our notation, we can rewrite this equation as:

$$Y_i = \alpha_1 H_{i1} + \alpha_2 H_{i2} + \alpha_3 H_{i3} + Z_i'\beta + \theta + \nu_i = W_i'\gamma + \nu_i \quad (2.6)$$

where H_1, H_2, H_3 are home input variables, Z is a vector of all covariates that are never missing (including an intercept term), and ν_i is the residual which is assumed to be orthogonal to the regressors. A significant challenge in our analysis is the incomplete observation of home inputs for the entire sample and hence we cannot estimate the above equation using all the students in our sample. However, in the spirit of the paper by Abrevaya and Donald, 2017, we assume that their exists linear projections of all home inputs on the consistently observed variables (Z) :

$$H_{i1} = Z_i'\delta_1 + \epsilon_{i1}$$

$$H_{i2} = Z_i'\delta_2 + \epsilon_{i2}$$

$$H_{i3} = Z_i'\delta_3 + \epsilon_{i3}$$

Combining these equations with our main regression yields:

$$Y_i = Z_i'(\delta_1\alpha_1 + \delta_2\alpha_2 + \delta_3\alpha_3 + \beta) + \nu_i + \epsilon_{i1}\alpha_1 + \epsilon_{i2}\alpha_2 + \epsilon_{i3}\alpha_3 = Z_i'(\delta_1\alpha_1 + \delta_2\alpha_2 + \delta_3\alpha_3 + \beta) + \eta_i \quad (2.7)$$

To handle missing data, we introduce an indicator variable m, which equals 1 when home inputs are missing for an observation. Given that our sub-sample of students with parent interviews was randomly selected, we can assume our data is Missing Completely at Random (MCAR). This crucial assumption implies that m is statistically independent of both regressors and error terms, forming the basis for our Generalized Method of Moments (GMM) approach.

Using all the above equations, we construct a vector of moment functions:

$$\begin{pmatrix} (1-m_i)W_i(Y_i - \alpha_1 H_{i1} - \alpha_2 H_{i2} - \alpha_3 H_{i3} - Z_i'\beta) \\ m_i Z_i(Y_i - Z_i'(\delta_1\alpha_1 + \delta_2\alpha_2 + \delta_3\alpha_3 + \beta)) \\ (1-m_i)Z_i(H_{i1} - Z_i'\delta_1) \\ (1-m_i)Z_i(H_{i2} - Z_i'\delta_2) \\ (1-m_i)Z_i(H_{i3} - Z_i'\delta_3) \end{pmatrix} = \begin{pmatrix} g_{1i}(\alpha_1,\alpha_2,\alpha_3,\beta,\delta_1,\delta_2,\delta_3) \\ g_{2i}(\alpha_1,\alpha_2,\alpha_3,\beta,\delta_1,\delta_2,\delta_3) \\ g_{3i}(\alpha_1,\alpha_2,\alpha_3,\beta,\delta_1,\delta_2,\delta_3) \\ g_{4i}(\alpha_1,\alpha_2,\alpha_3,\beta,\delta_1,\delta_2,\delta_3) \\ g_{5i}(\alpha_1,\alpha_2,\alpha_3,\beta,\delta_1,\delta_2,\delta_3) \end{pmatrix}$$

Under the MCAR assumption, we expect $E(g_i(\alpha_1,\alpha_2,\alpha_3,\beta,\delta_1,\delta_2,\delta_3)) = 0$, which provides us with more moment conditions than parameters to be estimated. By leveraging these additional moment conditions, we can reduce the variance of our estimates and improve their precision.

For models incorporating interaction terms, we employ a similar methodology. It's worth noting that the GMM approach offers efficiency gains over alternative methodologies such as imputation. For a comprehensive understanding of these advantages, readers are directed to the work of Abrevaya and Donald, 2017.

2.A.2 Robustness Checks

Table 2.A1 : Student Summary: Parental Data Sub-sample

	No Parental Data	Parental Data
Mathematics score in G2	14.147	14.120
	(5.346)	(5.265)
Vietnamese score in G2	16.776	16.687
	(6.964)	(7.039)
Student's age in G2 (in years)	7.351	7.400
	(0.524)	(0.561)
Male	0.521	0.520
	(0.499)	(0.499)
Ethnic minority	0.201	0.242
	(0.398)	(0.428)
Students	6905	3326
Schools	140	140

Notes: This table reports means and standard deviations (in paranthesis) of the characteristics of students with and without parental data in Grade 2.

Table 2.A2 : Production function by Model Specification

	(1) Math CT	(2) Math VA	(3) Vietnamese CT	(4) Vietnamese VA
Child's time	0.024	0.010	0.005	0.007
	(0.023)	(0.018)	(0.024)	(0.018)
Material Investment	0.025	-0.002	0.031	0.012
	(0.027)	(0.023)	(0.028)	(0.020)
Parent's time	0.082***	0.061***	0.077***	0.049***
	(0.025)	(0.019)	(0.025)	(0.018)
Classroom VA	0.373***	0.190**	0.607***	0.276***
	(0.100)	(0.074)	(0.085)	(0.056)
Baseline score		0.629***		0.740***
		(0.013)		(0.013)
Child controls	Yes	Yes	Yes	Yes
Observations	5929	5929	5929	5929

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's VA. CT corresponds to the model which uses only contemporaneous inputs. VA corresponds to the specification which uses lagged test score as well as contemporaneous inputs. Standard errors are clustered at classroom level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.*

Table 2.A3 : Production Function for a Sub-Sample (Cognitive and Non-Cognitive)

	(1) Math	(2) Vietnamese	(3) SPP	(4) Mastery
Child's time	0.014	0.014	0.075**	0.036
	(0.020)	(0.019)	(0.035)	(0.034)
Material Investment	-0.005	0.008	0.060	0.063*
	(0.025)	(0.023)	(0.040)	(0.038)
Parent's time	0.052**	0.041**	-0.028	0.013
	(0.021)	(0.020)	(0.036)	(0.036)
Classroom VA	0.203**	0.291***	-0.014	-0.126
	(0.080)	(0.065)	(0.202)	(0.215)
Baseline Score	0.634***	0.744***	0.010	0.041
	(0.014)	(0.014)	(0.033)	(0.037)
Child controls	Yes	Yes	Yes	Yes
Observations	4875	4875	1164	1164

*Notes: This table reports the production function estimates for the sub-sample of schools where in the principal reports that students are randomly assigned to classrooms. The baseline score here is the lagged Math and Vietnamese academic score for each subject and the average of Math and Vietnamese score for non-cognitive regressions. All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's cognitive VA. Standard errors are clustered at classroom level.*** p<0.01, ** p<0.05, * p<0.1.*

Table 2.A4 : Production Function using Parental Time Factor (Math and Vietnamese)

	(1) Math	(2) Math	(3) Vietnamese	(4) Vietnamese
Child's time	0.016	0.015	0.012	0.013
	(0.019)	(0.019)	(0.019)	(0.019)
Material Investment	0.002	0.000	0.003	0.004
	(0.025)	(0.025)	(0.023)	(0.023)
Parent's time	0.057***	0.056***	0.052***	0.052***
	(0.019)	(0.019)	(0.020)	(0.020)
Classroom VA		0.183**		0.276***
		(0.074)		(0.056)
Baseline score	0.632***	0.630***	0.741***	0.742***
	(0.013)	(0.013)	(0.013)	(0.013)
Child controls	Yes	Yes	Yes	Yes
Observations	5929	5929	5929	5929

*Notes: All regressions control for student's age, gender, ethnicity, maternal education, mother's age, household wealth and school-cohort fixed effects. Class VA is the other cohort's VA. Standard errors are clustered at classroom level. *** p<0.01, ** p<0.05, * p<0.1.*

Table 2.A5 : Parental Time Inputs: Factor Loadings

Items	Loading
Reads a book with child	*0.6145*
Tells a story to the child	*0.6436*
Sings to, or with your child	*0.5876*
Do activities like arts and crafts	*0.6176*
Play board games or do puzzles	*0.7008*
Play sports or exercise together	*0.6674*
Talk about family history or ethnic heritage	*0.4091*
Working on a hobby	*0.7029*
Attending concerts or movies outside of school	*0.4140*
Spending time just talking	*0.3697*
Helps child with Vietnamese	*0.6462*
Helps child with Mathematics	*0.6254*

Notes: This table reports the factor loading of all the items of parental time investment. All items have factor loading larger than 0.30 in absolute value.

Chapter 3

Market Reform and Farmer Prices: The Case of Indian Agriculture

3.1 Introduction

Agriculture plays a pivotal role in India's economy, serving as the primary source of livelihood for a significant portion of the population. The agriculture sector employs over 42% of the total workforce and more than 70% of the rural population. However, despite its immense scale and importance, this sector continues to face widespread market distortions and challenges that profoundly impact the lives of tens of millions of farmers. The farmers remain poor, facing considerable variations in the prices they receive for the same produce across different regions and seasons. This paradox – where a sector so vital remains so vulnerable – underscores the need for a deeper understanding of the factors driving these market inefficiencies and price variations.

This paper aims to explore the factors influencing farmer prices, with a focus on the role of government regulations and competition among intermediaries. While extensive research has examined the reasons behind poorly integrated agricultural markets, leading to significant price variations across regions - such as information friction, trade costs, and lack of infrastructure - a significant gap remains in our understanding of how competition among intermediaries and government regulations impact inter-regional trade patterns within the country. Moreover, the persistence of substantial price disparities, despite technological advancements, improved infrastructure, and the proliferation of farmer-support apps, suggests that these factors alone do not provide a complete picture. This research, therefore, seeks to fill this gap by examining how the interplay between intermediaries, market competition, and regulatory frameworks shapes the prices farmers receive for their produce.

India's agricultural sector operates within a regulatory framework, with the Agricultural Produce Market Committee (APMC) Act playing a vital role. This act, which mandates that farmers sell their produce exclusively in government-regulated markets though open-outcry auctions, was introduced to protect farmers by establishing regulated marketplaces for fair auctions and standardized practices. However, over time, the Act's implementation has led to unintended consequences, such as limited competition due to new market entrants facing obstacles in obtaining licenses, potentially suppressing prices for farmers.

Within this framework, this paper examines the dynamics of India's agricultural markets in two sections, each addressing critical aspects of agricultural market efficiency in developing economies. The first section studies the impact of trader competition on farm-gate prices. We utilize a novel dataset which allows us to capture both inter- and intra-market competition, providing a comprehensive view of market dynamics.

The distribution of markets across space creates significant variation in farmers' market access. Some regions exhibit high market density, while others are characterized by sparse, geographically dispersed markets. This spatial variation, combined with substantial transportation costs, effectively limits farmers' outside options. Furthermore, there is limited competition among traders within a market, as studies have shown that traders tend to collude within a market and also actively work to prevent the entry of new competitors (Meenakshi and Banerji (2005)). This creates another constraint for farmers: not only are they unable to access other markets, but they also face restricted competition among buyers within their local market. Consequently, farmers may accept lower prices due to a lack of alternatives.

Thus, the geographical distribution of markets and limited within-market competition, exacerbated by government regulations, contribute significantly to traders' ability to exert influence over price. In such a context, to effectively enhance the welfare of farmers, policies need to specifically target increasing competition among intermediaries. Our findings suggest that both inter- and intra-market competition significantly impact farm-gate prices. Specifically, we find that a one standard deviation increase in our measure of spatial competition leads to a 1.4-2.4% increase in crop prices, while a similar increase in within-market competition measure results in a 2% price increase.

Building on these results, the second part of our study focuses on a new policy initiative: the introduction of the electronic National Agriculture Market (eNAM). This national unified market aims to address the issues of limited competition identified in our the first section by breaking down barriers in agricultural markets. eNAM provides a common electronic platform enabling farmers to sell their crops to buyers in these markets. By liberalizing trader licensing and facilitating remote trading online, eNAM seeks to enhance competition among traders both spatially and within individual markets. We study the impact of the eNAM policy, examining how effective this initiative has been in increasing competition and, consequently, improving prices for farmers. Our results demonstrate that eNAM implementation led to a 3.1-3.4% increase in farmers' prices. Furthermore, we find evidence of increased trader participation following the introduction of eNAM, suggesting that the policy has been effective in reducing entry barriers and enhancing market competition.

This paper makes two major contributions to the understanding of agricultural markets in India. First, it offers a detailed analysis of how agricultural markets work, examining both market location and local competition as key factors in price determination. Second, we evaluate the impact of a significant nationwide policy reform: the electronic National Agriculture Market (eNAM). Our analysis uses high-frequency price data collected from one of the states of India, covering periods both before and after the eNAM implementation. To our knowledge, this presents one of the first studies to utilize micro-level data in assessing the economic impacts of the eNAM initiative.

The work in our paper connects to different strands of literature: : the role of intermediaries in trade, market competitiveness in developing countries, and the impact of market integration and infrastructure on agricultural market performance.

Firstly, this paper contributes to the growing body of literature examining the role of intermediaries in trade (Dhingra and Tenreyro (2017); Bardhan, Mookherjee, and Tsumagari (2013); Grant and Startz (2022)). The existing research presents a nuanced view of how intermediaries affect market efficiency and producer welfare. One strand of research highlights the positive role of intermediaries as trade facilitators. For instance, Antras and Costinot (2011) demonstrate how intermediaries can improve trade outcomes by reducing search frictions and connecting producers with markets. Similarly, Ahn, Khandelwal, and Wei (2011) emphasize the importance of intermediaries in enabling trade, particularly for smaller firms or in markets with significant barriers to entry. However, another significant body of work focuses on the potential negative impacts of intermediaries, particularly in rural and developing economies. These studies suggest that intermediaries can sometimes exploit their position to extract excessive profits at the expense of producers. For instance, Besley and Burgess (2000) and Goyal (2010) have explored this phenomenon in the Indian context, attempting to quantify the efficiency gains that could be achieved by eliminating exploitative intermediaries from the market chain. Recent work by Bartkus et al. (2022) adds to this perspective. Their study of a cooperative intervention that eliminates middlemen, allowing fishermen to bring goods directly to market, finds a substantial improvement in the prices received by producers, thus underscoring the potential benefits of direct market access for primary producers. Our research aligns more closely with this second perspective, providing new empirical evidence on the impact of intermediaries' competition on farmer prices.

This study also adds to our understanding of market competitiveness in developing countries. A common concern in developing markets is lack of competitiveness, potentially allowing traders or intermediaries to earn economic rents or charge significant markups. Extensive research has examined various characteristics of market imperfections in these contexts.

While a set of paper studies spatially varying markups arising from high transportation costs (Miller and Osborne (2014); Atkin and Donaldson (2015); Donaldson (2015); Asturias, Garcia-Santana, and Ramos (2019); Chatterjee (2023)), another set of papers have explored collusive practices among traders in various contexts, such as vegetable markets, grain markets etc. (Banerjee et al. (2022); Bergquist and Dinerstein (2020); Mitra et al. (2018); Banerji and Meenakshi (2004)). Additionally, research has focused on how entry barriers or market size can limit competition. For instance, Leone, Macchiavello, and Reed (2024)'s work on the cement industry shows that limited competition stems from small market size. Others have examined the effects of increasing competition by introducing new competitors, either large retailers or small firms (Busso and Galiani (2019); Dhingra and Tenreyro (2017)).

Our paper adds to this literature by leveraging novel micro-data to investigate how both the geographic distribution of intermediaries and the level of competition within a market are crucial determinants of farm-gate prices. By providing a more holistic picture of what shapes competitive dynamics in agricultural markets, we aim to inform effective policy interventions that enhance market efficiency and improve outcomes in developing country agricultural markets. Understanding the nature of competition is vital for effective policy design.

Finally, the paper also relates to the broader literature on market integration and the role of technological infrastructure in enhancing market performance. Donaldson (2018) shows that improvements in transport infrastructure, particularly railroads, led to substantial welfare gains in the Indian economy by facilitating market integration. In the agricultural sector, Tomar (2016) examined the benefits of state-specific market integration through an online platform, revealing significant welfare gains for farmers. Similarly, Aggarwal (2018) demonstrated that market integration, facilitated by a rural road development program, increased the variety of goods available in local markets. A related strand of literature has focused specifically on addressing inefficiencies in agricultural markets through information technology. For instance, Goyal (2010)'s research on the soybean market emphasizes the critical role of information provision in improving market outcomes. Jensen (2007)'s work in South India also shows how mobile phone reduced information asymmetries, leading to reduction in price dispersion. These studies underscore the potential of technological advancements to mitigate market inefficiencies. Our paper contributes to these two areas by examining the National Agricultural Market (eNAM) policy in India, a nationwide initiative aimed at integrating agricultural markets online. Using micro-level data from before and after the policy's implementation, we explore its effects on prices, providing novel insights into how online integration can influence market dynamics. Furthermore, the eNAM policy addresses another source of inefficiency: the reduced trader participation in offline markets. By enabling traders to participate without being physically present, the online platform has the potential to increase market participation. By focusing on this national initiative, our research contributes to the ongoing discourse on market integration and agricultural welfare providing valuable insights into the effects of such a reform.

The rest of the paper is organized as follows. In section 3.2, I provide a background on the agricultural markets in India and the policy reform. In the following section 3.3, I outline my data sources and key measures. Section 3.4 presents my empirical analysis with results and finally in section 3.5, I provide concluding remarks.

3.2 Agriculture Market in India and Policy Reform

Agriculture is a state subject in India and the regulation of agricultural markets are governed by state specific Agricultural Produce Market Committee (APMC) Acts. The Act mandates that the sale and purchase of agricultural commodities must occur in designated market yards or *mandis*. Farmers within a state are required to sell their produce only in these regulated markets, although they have the freedom to choose any of these markets within their state boundaries. On the other side of the transaction, traders who wish to procure goods must obtain licenses to operate in these markets, typically long-term licenses ranging from 5 to 10 years, but are restricted to operating in only one market in their state. These regulated markets are spatially dispersed across the country, as illustrated in Figure 3.1. This distribution leads to varying degrees of market concentration in different areas.

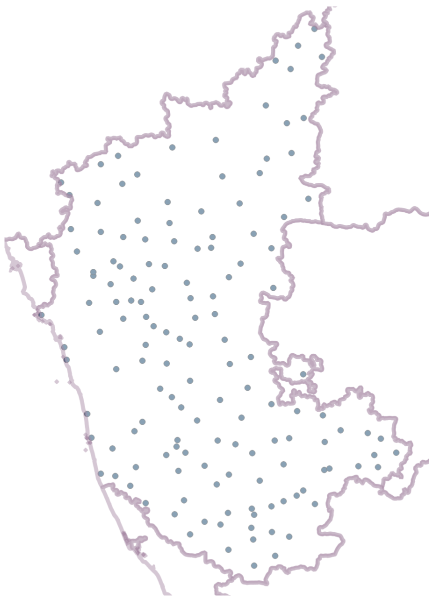


Figure 3.1: Spatial Distribution of Agricultural Markets in Karnataka, India.

The price in these *mandis* is determined through an open outcry auction system (second-price bid auction). Farmers display their produce in the market yard, and traders gather around the lot and bid on each lot. The trader with the highest bid wins, and the process repeats for each lot. This system aims to create a stable and regulated trading environment in the regulated markets. Figure 3.2 shows the timeline of events.

To understand the significance of these regulated markets, it's crucial to examine the context that led to their creation. They were established with the intention of empowering farmers to secure fair prices for their produce by shielding them from exploitation by middlemen and fostering a competitive pricing environment. As Chand (2012) and Acharya (2004) note, this system was designed to eliminate malpractices and market imperfections, creating transparent marketing conditions and liberating farmers from trader exploitation, a significant concern in pre-APMC agricultural markets.

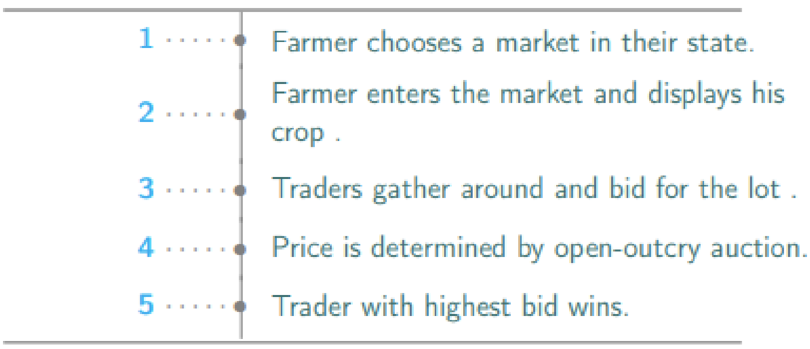


Figure 3.2: Timeline of trade

During the 1960s and 1970s, around the spread of the Green Revolution, most states enacted and enforced Agricultural Produce Marketing Regulation acts, which helped make the country self-reliant in food grain. As agricultural production boomed, new market yards were constructed, and marketing committees were constituted to frame and enforce rules. A map in Appendix 3.A2 illustrates how areas with higher agricultural yields saw a greater concentration of APMC markets, reflecting the system's responsiveness to increased production.

However, as the country achieved food self-sufficiency, its marketing system became inefficient. Farmers reliance on intermediaries (traders) for services strengthened their position. These developments have led to the current state of regulated markets, which face several challenges such as trader associations effectively barring new entrants, leading to widespread cartelization and stifling competition. Moreover, the marketing monopoly power provided to the state has discouraged private investment in market infrastructure. As a result, the number of markets has not increased proportionally with production growth, leading to overcrowded markets that put sellers at a significant disadvantage. The stagnation in market development is evident in the fact that hardly any new markets have been created since 2000, as shown in the Figure 3.A1 .

This combination of factors – restricted market access, lack of infrastructure, dominance of entrenched trading associations, and limited selling options – has created a system that amplifies the economic challenges faced by farmers.

In response to these challenges, the Government of India introduced the electronic National Agriculture Market (eNAM) in 2016. This platform aims to create a unified national market for agricultural commodities through an online trading system, addressing key issues by promoting transparency, reducing information asymmetry, and enhancing market access for farmers. The eNAM system promotes the liberal licensing of traders, buyers, and commission agents without the precondition of physical presence or shop ownership, potentially breaking the cartelization of trader associations and allowing new entrants into the market. It also requires states to issue a single unified license to traders, enabling them to operate across all regulated mandis within the state, thereby increasing competition.

eNAM is being implemented in phases, with the initial phase networking 585 APMC mandis across 18 states. The rollout began in April 2016 with a pilot launch in 21 mandis, expanding to approximately 400 mandis between November 2016 and March 2017. The remaining 185 mandis joined by March 2018. The second phase, initiated afterward, further expanded the network by integrating an additional 415 mandis. Not all the markets in the state have been integrated, which allows us to compare integrated and non-integrated markets across states to analyze any effect on the crop prices.

3.3 Data and Key Measures

We use a rich micro-level dataset, combined from different sources, on prices of crops sold in regulated markets, their geo-locations, market characteristics, integration details, etc. This comprehensive data allows us to understand price variation across a state and study the effect of market reform and competition on these prices. This section provides information on the data sets we have gathered for this research. Table 3.1 presents summary statistics of the data used.

Data on Grain prices

The grain price data utilized in this study is gathered from the Ministry of Agriculture's *Agmarknet* platform. This comprehensive dataset provides weekly information on three price movements— maximum, minimum, and modal prices—for all major crops traded in each of the markets across all states where markets are regulated under APMC act. Our analysis focuses specifically on modal prices[1] of rice in 17 states over the period 2008-2018. We exclude states with negligible rice production and consequently minimal trade in the markets. The data also includes information on rice quality and variety, allowing us to categorize rice into three broad grades: coarse, medium, and fine. Our study primarily concentrates on medium and coarse rice, which together constitute around 94% of the sample. Fine grade rice is typically not traded in these markets as it is often procured directly by the government.

Data on measures of competition

Within-market **Competition**

We employ two measures to assess intra-market competition in agricultural markets in India. *Within-market* competition refers to the level of competition among traders operating within a single market:

Registered Traders: We scrap data from the agriculture website to determine the number of traders registered in each market. These registrations represent traders licensed to operate in the markets, indicating the number of potential bidders. It's important to note that these licenses are typically long-term (5 to 10 years), resulting in primarily cross-sectional variation

[1]The modal price is the price at which the maximum quantity of a crop is sold in the market in a given period

rather than temporal changes[2]

Trader Participation: Recognizing the potential limitations of relying on registered trader data as an indicator of actual market participation, we gathered detailed information on traders participation across all major markets in a single Indian state for the years 2016 to 2018. In this context, trader participation is defined as the total number of traders that were present in the market and took part in rice trading on a given day. This rich dataset provides information on trader engagement on a daily basis, offering both cross-sectional and temporal variation. Importantly, it allows us to identify the true number of traders specifically engaged in rice trading.

Our analysis of this detailed participation data yields an insightful finding: on average, only about 20% of registered traders actively participate in rice trading on any given day. This substantial gap between registration and actual participation can be attributed to two factors: (i) not all registered traders attend the market on a daily basis, and (ii) some registered traders may engage in trading goods other than rice. These findings highlight the critical distinction between potential and actual competition in these markets, emphasizing the need for a more nuanced understanding of market dynamics beyond the number of licensed traders.

The use of both measures provides a more comprehensive view of market competition. While the registered trader data offers a broad perspective across all Indian markets, the participation data from one state allows for a more granular analysis of daily market participation, albeit for a geographically limited sample.

***Between-market* Competition**

Geo-spatial data on market locations is crucial in analyzing competition between traders across space. The majority of these markets were established between 1960 and 1985, during the peak of India's Green Revolution when grain production increased exponentially. To obtain precise location data, we geo-code the markets using Google Maps API. This geo-coding process allowed us to compute distances between markets using Google's Distance Matrix API. Additionally, we calculated the time taken to travel from each market to its respective district headquarters as this provides

[2]This measure however has limitations. While new licenses may be issued annually, our data doesn't capture these incremental changes. However, given the high entry barriers in these markets (discussed above) this measure can be viewed as fixed number of potential bidders.

further context on market accessibility.

To measure competition across space, we draw from the trade literature, where trade cost is commonly used as a standard measure of competition. In this study, we proxy trade cost using the distance between markets. Specifically, I adopt an approach similar to that used by Macchiavello and Morjaria (2015) and Chatterjee (2023), calculating a weighted sum of all markets within the state, with weights inversely proportional to distance. This method assigns larger weight to closer markets, reflecting the reality that nearby markets are likely to exert more competitive pressure.
Competition index:

$$b\_competition_w^m = \sum_{n \in M \setminus m} \frac{1}{distance_{mn}}$$

where M is the set of all markets in the state.

We also consider an alternative measure: the total number of markets within a 100-kilometer radius of each market. This provides a different perspective on market density and potential competition.

It's important to note that during the time-frame of this analysis, very few new markets were created. Consequently, the variation in our competition measure is primarily cross-sectional, derived from the geographical placement of markets. For ease of interpretation, this measure is standardized to have a mean of zero and a unit standard deviation within each state.

Data on Integration

The electronic National Agricultural market (eNAM) portal provides information on the roll out of the policy reform in different states from where we gather information on the time of the integration of different markets across states. Although we cannot determine the exact dates of implementation, we know the months when specific markets were integrated into the system. We also collect information on the characteristics of the market where eNAM was introduced and the ones where it wasn't. These are reported in Table 3.A1 .

Data on Other Controls

Market Characteristics: We extract information from the *Agmarknet* website on market characteristics like storage availability, year of establishment of the market, annual income, number of villages served by the markets. This information however is not well reported and we lack this information for almost 10% of the markets in my sample.

Rainfall and Yields: We obtained district-wise estimates of crop yields from the Ministry of Agriculture website, providing a granular view of agricultural productivity across India. Complementing this, we collected rainfall data at the monthly level for each district from the Indian Meteorological Department.

Table 3.1: Summary Statistics

Prices	Mean	S.D.
Medium Grade		
Model Price	2.55	0.11
Maximum Price	2.58	0.12
Minimum Price	2.52	0.12
Coarse Grade		
Model Price	2.32	0.11
Maximum Price	2.36	0.10
Minimum Price	2.29	0.12
Fine Grade		
Model Price	3.06	0.19
Maximum Price	3.13	0.21
Minimum Price	2.91	0.22
Independent variables		
Rice yield (tonnes/hectare)	2.33	0.85
Annual rainfall (millimeters)	1082	646
Storage facility availability	0.63	0.48
Time to district headquarter (minutes)	60	45

Notes: This table reports the summary statistics for the key variables. The top panel reports prices which are measured in rupees per kilogram. The bottom panel reports other controls. The within and between market competition measures are standardised at the state level. Details of the source of each variable can be found in the data section.

State-level data

In addition to the aggregated data collected for all of India, we gathered detailed transaction-level data from 14 major agricultural markets in the state of Chhattisgarh, India. This data, originally recorded in handwritten ledgers, captures every auction conducted on a given day, including comprehensive information on farmers and traders, their villages of origin, and their trading partners. This granular dataset provides details on which trader purchased what quantity from which farmer, offering insights into potential network effects - a promising area for future research. To analyze the impact of the eNAM initiative, we collected and digitized these records for one year before and one year after the policy implementation. While our primary analysis focuses on prices and trader participation in six of these markets, this comprehensive dataset serves as a rich foundation for broader research[3].

3.4 Empirical Strategy and Results

In this section, we present our empirical strategy and analyze the impact of market competition and policy reform on farmers' prices. Our analysis unfolds in two main parts. First, we examine how competition among traders, both within and across markets, affects farmers' income. For this analysis, we utilize data from 2008 to 2016, deliberately focusing on the period before the introduction of eNAM. Second, we evaluate the effects of the online market reform (eNAM) on farmers' prices, concentrating on the years 2013 to 2018. This approach allows us to capture the impact both before and after eNAM's implementation.

Market competition and prices

This subsection delves into the relationship between market competition and the prices farmers receive for their crops, focusing on rice. Our analysis focuses on two key forms of competition among traders: *between-market* and *within-market*. We investigate how these influence prices across different markets using a regression model that accounts for various other factors. The model takes the following form for each grade (quality) of rice:

$$\begin{aligned} log\ p_{mdst} = \alpha_0 + \alpha_1\ b\_competition_m + \alpha_2\ w\_competition_m \\ + \alpha_3\ \ storage_{mdst} + \mathbf{Z}'_{dst}\alpha_4 + \delta_s + \delta_t + \nu_{mdst} \end{aligned} \tag{3.1}$$

where p_{mdst} represents the price received by farmers for a particular rice grade in market m in district d of state s at time t. The key independent variables of interest are *between-market* competition denoted by $b\_competition_m$, which measures competition across markets in space, and

[3]This extensive data collection and digitization process was conducted over 2022-23, with the final digitized data becoming available in June 2024. The use of only six markets in this paper is due to the late availability of data for the remaining markets.

within-market competition presented by $w\_competition_m$, which measures competition within a specific market. We include several control variables ($\mathbf{Z}$) such as crop yield and rainfall which account for local production levels, and population density which serves as a proxy for local demand at the district level. We also control for the time taken to go to the district headquarter, from market m, which is a large retail market where the intermediaries could potentially sell the goods[4]. *storage* is a binary variable indicating the availability of storage facilities in a market. Having storage infrastructure available may increase farmers' ability to hold inventory, which could in turn affect prices. To account for this factor, we include local storage capacity as a control variable in our analysis. In the specification, we also control for time-invariant state characteristics (e.g., state level polices, local taste, geographic conditions) by including state fixed effects δ_s, as well as changes over time that affect all states in the same way by adding month-year fixed effects, δ_t. These help us control for cost inflation and seasonal fluctuations in prices. All prices are at monthly level and standard errors are clustered at the district level to account for potential within-district correlation.

For our measure of *within-market* competition, we firstly use the number of traders registered in each market. Although this measure doesn't precisely quantify the number of traders active in these markets, it serves as a credible proxy for the within-market competition. *Between-market* competition is measured as distances to the nearest markets or the number of markets within a certain radius.

A key challenge in our analysis is addressing potential endogeneity concerns in our competition measures. One might worry that prices could influence the establishment of new markets as well as number of licensed traders in a market, thereby biasing our estimates. However, our research setting offers several appealing features that facilitate identification.

First, the markets in our study were primarily established during the Green Revolution. Our data on the years of establishment of these markets, as illustrated in Figure 3.A1 , shows that less than 0.7% of the markets have been created since 2008, which marks the beginning of our analysis period. This historical context helps mitigate concerns about contemporaneous market creation in response to price changes. Second, the variation in

[4]The intermediaries have freedom to sell their goods anywhere and have no legal constraints.

the number of traders (or potential bidders) is also exogenous, as there are limitations on the number of trader licenses issued within a market, and traders are licensed to operate in a market for extended periods. Moreover, obtaining new licenses is notably difficult due to existing trader associations effectively barring new entrants. Consequently, the variation in the number of traders in a market is cross-sectional and does not vary over time.

To address remaining concerns about geographically varying factors that could confound our analysis, we incorporate a comprehensive set of cross-sectional control variables. These include crop yield and rainfall, which account for local production levels and hence could affect prices. We also control for population density, which serves as a proxy for local demand. Additionally, to further isolate the effects of competition on prices, we run separate models for different rice qualities.

The regression results are reported in Table 3.2. All six columns correspond to different regressions for various grades of rice. We employ two measures of between-market competition: the weighted sum of all markets within the state (Columns 1-3) and the number of markets within a 100km radius (Columns 4-6). In interpreting our results, we focus primarily on the coefficients α_1 and α_2 in equation (3.1).

Our analysis reveals a significant positive relationship between spatial market density and farmers' prices, as indicated by a positive and significant α_1. Specifically, when using the weighted sum of all markets within the state as our competition measure, we find that a one standard deviation increase in this measure corresponds to a 1.4% rise in prices for medium-grade rice and a 2.4% increase for coarse-grade rice. When we employ our alternative measure of spatial competition, the results remain statistically significant, albeit with smaller magnitudes.

Turning to *within-market* competition, our findings indicate that a greater number of potential traders within a market benefits farmers through higher prices, as evidenced by a positive α_2. The coefficient on the potential number of traders is significant across all grades of rice, with a one standard deviation increase leading to a 1.1% to 1.4% rise in prices. This suggests that if a farmer moves from a market in the bottom 5th percentile of the trader distribution within a state to the top 95th percentile, his price could increase by approximately 3.6% for the medium quality of rice. Our findings emphasize the critical role of competition among intermediaries in shaping farmers' outcomes, both across and within markets.

Table 3.2: Competition and Farm Price

	(1)	(2)	(3)	(4)	(5)	(6)
	Log price					
	Medium	Coarse	Fine	Medium	Coarse	Fine
Within-market competition	0.011***	0.013***	0.014***	0.011***	0.013***	0.013***
	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)
Between-market (weighted avg)	0.014***	0.024***	0.005			
	(0.005)	(0.006)	(0.012)			
Between-market (100km radius)				0.008*	0.018***	0.002
				(0.005)	(0.006)	(0.008)
Storage availability	0.008*	0.020***	0.010	0.008*	0.020***	0.009
	(0.004)	(0.005)	(0.010)	(0.004)	(0.005)	(0.010)
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49807	17559	5594	49807	17559	5594
R-squared	0.757	0.606	0.379	0.756	0.606	0.379

*Notes: All regressions control for district-year crop yield, population density. distance to district headquarter as well as lagged monthly rainfall. Robust standard errors (in parenthesis) are clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.*

Robustness checks: To ensure the reliability of our findings, we conduct some robustness checks, which are presented in the Appendix. We explored alternative price measures, re-running our analyses using maximum and minimum prices instead of modal prices. This approach allows us to examine whether our findings hold across different points of the price distribution, providing insights into the consistency of our results. The results of these robustness checks, detailed in the Appendix (Tables 3.A2 and 3.A3), largely support our main findings. The consistency across these alternative specifications strengthens confidence in the validity of our estimates and overall conclusions regarding the impact of market competition on farmers' prices. A key limitation of our study is the lack of data on trader types. Consequently, we cannot distinguish between the number of large traders and small traders, preventing us from accounting for potential asymmetries between big and small bidders in our analysis.

While our previous results provide insights into how *within-market* competition, measured by number of potential traders, impacts prices, we acknowledge that this measure doesn't measure temporal trader participation in rice-auctions and only captures cross-sectional variation. There maybe variations in the number of traders that operate in a market across time. To address this limitation, we conduct a more granular analysis focusing on major markets in one specific state of India. For these markets, we have detailed daily information on trader participation from year 2016 (before eNAM), allowing us to capture temporal variation in *within-market* competition. We focus only on medium and coarse grades of rice as there is very minimum sale of fine quality rice in these markets.[5]

Table 3.3 presents our regression results from this state-level analysis, encompassing both rice grades combined and separate analyses for medium and coarse grades.[6] The results demonstrate that *within-market* competition, measured through actual trader participation, consistently exhibits a significant positive impact on farm prices across all specifications. The coefficient on this competition measure is positive and statistically significant in most models, ranging from 0.016 to 0.020. This suggests that a one s.d. increase in trader participation is associated with a 1.6% to 2.0% increase in farm prices, holding other factors constant[7]. The consistently significant effect of *within-market* competition highlights that farmers benefit from having multiple traders competing for their produce within the same market.

The impact of *between-market* competition varies depending on the measure used. Using the weighted average measure, we find a positive effect of between-market competition, particularly for medium-grade rice. The coefficient for medium-grade rice indicates that a one standard deviation increase in this competition measure is associated with a 4.2% increase in farm prices. When analyzing all grades collectively with grade fixed effects, the effect remains positive and significant, showing a 3.0% increase. In contrast, our alternative competition measure, which considers the number of nearby markets, doesn't yield significant effects.

This section emphasizes the critical importance of intermediary competition, both across and within markets, in determining farmers' outcomes. By conducting analyses at both the national and state levels, we provide

[5]Chhattisgarh, the state from which our data is drawn, ranks as one of the highest in terms of rice procurement at the minimum support price among Indian states. As a result, we observe very little trade in fine grade rice, which is primarily procured by the government. `https://pib.gov.in/PressReleasePage.aspx?PRID=1951287`

[6]The calculation of standard errors in this analysis, which involves six clusters, utilizes asymptotic formula.

[7]There could be a potential endogeneity concern related to trader participation, for example, due to sale characteristics unobserved in our data. To address this, we employ an instrumental variable (IV) approach, instrumenting actual trader participation with monthly participation rates from other years. Our identifying assumption is that this participation rate is likely correlated with current participation due to seasonal patterns, but unlikely to be directly related to current unobserved market conditions. Results (Appendix Table 3.A4) show marginally higher coefficients than OLS estimates, suggesting limited endogeneity concerns.

Table 3.3: Competition and Farm Price for Chattisgarh

	(1)	(2)	(3)	(4)	(5)	(6)
	Log price					
	All	Medium	Coarse	All	Medium	Coarse
Within-market competition	0.020***	0.020**	0.019*	0.017*	0.017**	0.016
	(0.004)	(0.005)	(0.010)	(0.007)	(0.006)	(0.010)
Between-market (weighted avg)	0.030*	0.042**	0.019			
	(0.014)	(0.014)	(0.027)			
Between-market (100km radius)				0.009	0.006	0.009
				(0.016)	(0.020)	(0.024)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Grade FE	Yes			Yes		
Observations	140846	127108	13738	140846	127108	13738
R-squared	0.674	0.692	0.504	0.672	0.687	0.511

*Notes: All regressions control for district-year crop yield, population density as well as lagged monthly rainfall. In Columns 1 and 4 , we report results together for all grades of rice with grade dummy. Robust standard errors (in parenthesis) are clustered at market level. *** p<0.01, ** p<0.05, * p<0.1.*

a comprehensive understanding of agricultural markets. The national-level analysis offers broad insights across the country, while the state-level analysis allows us to examine more nuanced, day-to-day competitive dynamics.

We compare our estimates of the effects of increasing inter and intra-market competition to findings from other studies on Indian agricultural markets that examine different channels for increasing farmers' outcome. For instance, Goyal (2010) found that the introduction of kiosks providing wholesale price information led to a 1-3 percent increase in soybean prices. Similarly, Tomar (2016)'s work on Karnataka, which investigated a different online trading platform, reported a 4% increase in groundnut prices with the introduction of online trading platform. These comparisons suggest that enhancing market competition along the dimensions discussed in this paper can yield price benefits for farmers that are on par with other interventions. Initiatives like eNAM, which have the potential to boost competition both between and within markets, could hence substantially improve farmers' prices.

The impact of eNAM on prices

This subsection aims to identify the causal impact of the market reform, eNAM, on farmers' prices in India. As detailed in Section 3.2, eNAM was designed to increase competition among traders through several mechanisms: easing entry for intermediaries with liberal licensing, allowing re-

mote trading through an online platform, and issuing a single unified license to traders which enables them to operate across all integrated mandis. However, the last aspect of inter-market trade is yet to materialize. Consequently, our analysis will focus on how eNAM affects *within-market* competition. For this analysis, we exploit the variation in eNAM adoption across different agricultural markets in India. Under the assumptions described below, the variation generated by the eNAM roll out allows us to estimate the causal impact of online trading on farmers' prices using a difference-in-differences (DD) strategy. This approach compares the before-after difference in outcomes between markets where eNAM was introduced and markets where eNAM had not yet been implemented during our sample period. The current analysis assumes that there are no spillovers from the eNAM treated markets to the markets in the control group.
As a baseline specification, we estimate the following two-way fixed effect (TWFE) model:

$$log\ p_{mdst} = \beta_0 + \theta enam_{mdst} + \gamma_m + \gamma_t + \mathbf{X}_{dt}\beta + \epsilon_{mdst} \qquad (3.2)$$

where p_{mdt} is farmers' price in market m in district d of state s at time t, γ_m represents market fixed effects and γ_t represents time fixed effects. The variable *enam* is a dummy variable that indicates whether market has implemented eNAM in time t. We use the monthly price for a specific grade of rice averaged over four months as our outcome variable.[8] The market fixed effects control for time-invariant market characteristics (e.g., spatial competition, storage facilities), while time fixed effects account for changes that affect all markets similarly. Standard errors are clustered at the market level.

The key identifying assumptions for our DD strategy are parallel trends and homogeneity of treatment effects. The parallel trends assumption posits that in the absence of eNAM implementation, the prices in treated and control markets would have evolved along parallel trends. The homogeneity assumption requires that the average treatment effects are relatively consistent across treated markets and over time. Under these assumptions, the coefficient of interest θ identifies the Average Treatment Effect on the Treated (ATT) of eNAM policy on farmers' prices.

[8]We use four-month averages instead of monthly prices due to uncertainty in eNAM implementation timing. The initial rollout occurred between November 2016 and March 2017, with varying dates across markets. Lacking precise implementation dates for each market, we analyze at the trimester level to avoid potential errors arising from this timing uncertainty.

Our analysis covers the period from 2013 to 2018. We focus on two groups of markets: those integrated with eNAM before March 2017 (treated group) and those not treated at all during our study period (control group). Markets that implemented eNAM between March 2017 and March 2018 are excluded from our sample to ensure a clear comparison between early-treated and never-treated markets.

Table 3.4 presents the estimated impact of the policy reform on rice prices for two grades: medium and coarse. Panel A looks at the modal price as the outcome variable and Panel B presents results from the maximum price. In the odd columns when we only have market and time fixed effects, we find that eNAM significantly increases farmers' prices for medium-grade rice by 3.3% and coarse-grade rice by 3.1%. However, these estimates may capture effects from time-varying factors such as local crop yield. To address this concern, we include time-varying controls in the even columns. The results remain robust, with only marginal changes in the magnitude of the estimates. However, the coefficient for coarse-grade rice loses statistical significance in this specification. Overall, our findings indicate that eNAM implementation leads to an increase in farmers' prices of approximately 3.1% to 3.4% of the mean for medium and coarse grades of rice, respectively.

Examining the effect on maximum prices in Panel B, we observe similar results for medium-grade rice. However, we find no significant effect on coarse-grade rice prices. One potential explanation for this discrepancy is that coarse-grade rice, being of lower quality, may deter traders from purchasing online without physical inspection. Our estimates are slightly smaller than those reported in a previous study (Tomar (2016)) examining a similar policy in one state of India for a different crop. They find that the introduction of a unified online system led to 4% increase in groundnut prices.

To probe the plausibility of the parallel trends assumption, and study the dynamics of treatment effects, we estimate an event-study version of the TWFE model with indicators for distance to/from the introduction of eNAM. Specifically, we estimate the following specification:

$$log\ \ p_{mdst} = \delta_\tau \sum_{\tau=-8}^{5} D_{\tau(mdst)} + \alpha_m + \gamma_t + \mathbf{X}_{dt}\beta + \epsilon_{mdst}$$

where $D_{\tau(mdst)}$ is set of indicator variables that take value one if for the market m in time t, the introduction of eNAM was τ periods away. The

Table 3.4: The Effect of eNAM on Prices

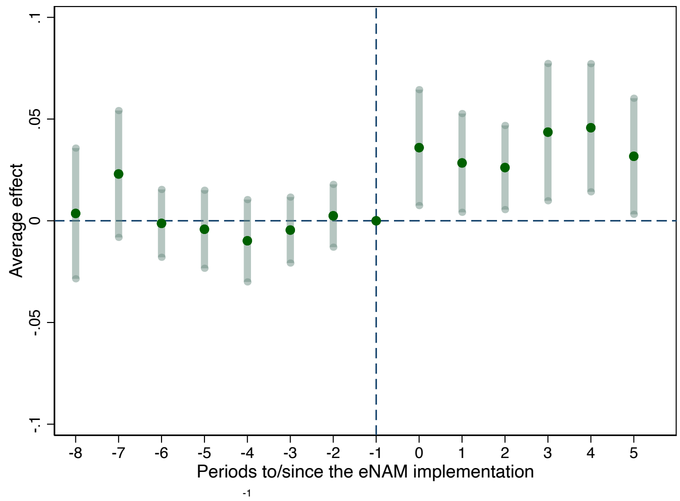
	Medium Grade		Coarse Grade	
	(1)	(2)	(3)	(4)
Panel A: Model Price				
eNAM implemented	0.033***	0.034***	0.031*	0.032
	(0.008)	(0.008)	(0.018)	(0.019)
Market FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Controls		Yes		Yes
Observations	6705	6705	2610	2610
R-squared	0.738	0.744	0.694	0.705
Panel B: Maximum Price				
eNAM implemented	0.033***	0.035***	0.023	0.024
	(0.008)	(0.008)	(0.022)	(0.023)
Market FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
Controls		Yes		Yes
Observations	6705	6705	2610	2610
R-squared	0.760	0.762	0.664	0.680

*Notes: This table reports the effect of eNAM policy reform on model and maximum prices of two different grades of rice. Each row represents a separate regression. In panel A, the dependent variable is the model price and in panel B, the dependent variable is the maximum price. The unit of observation is a market and time is running in four-month intervals. The set of controls include district specifying time varying controls like rice yield and rainfall. *** p<0.01, ** p<0.05, * p<0.1.*

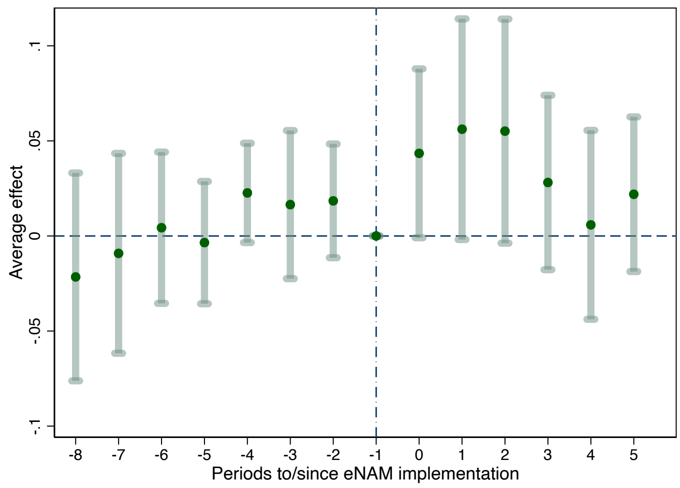
parameter δ_τ is the dynamic treatment effect which captures the treatment effects τ periods before or after the eNAM policy introduction. We normalize the coefficient on the period just prior to the reform to zero, so that the estimates can be interpreted as the difference in outcomes relative to those in the period preceding the reform.

Figures 3.3 and 3.4 present the results of our event study analysis, showing point estimates and 95% confidence intervals for each period relative to the period before eNAM implementation. The coefficients for periods preceding eNAM implementation are close to zero and statistically insignificant. This lack of significant pre-trends supports the parallel trends assumption, strengthening the validity of our difference-in-differences approach. On the other hand, the post-implementation coefficients illustrate the dynamic effects of eNAM on farmers' prices over time. In Panel A of Figure 3.3, for medium grade of rice, we observe immediate and significant positive changes in prices following eNAM implementation. These effects persist throughout the post-implementation period, suggesting a sustained positive impact of the reform on medium-grade rice prices. For coarse grade, while we initially see significant positive effects, these impacts diminish after a few periods post-implementation. We find similar effects on the maximum price of the markets presented in Figure 3.4.

Figure 3.3: The Effect of eNAM on Prices: Modal Price



Panel A: Medium Grade

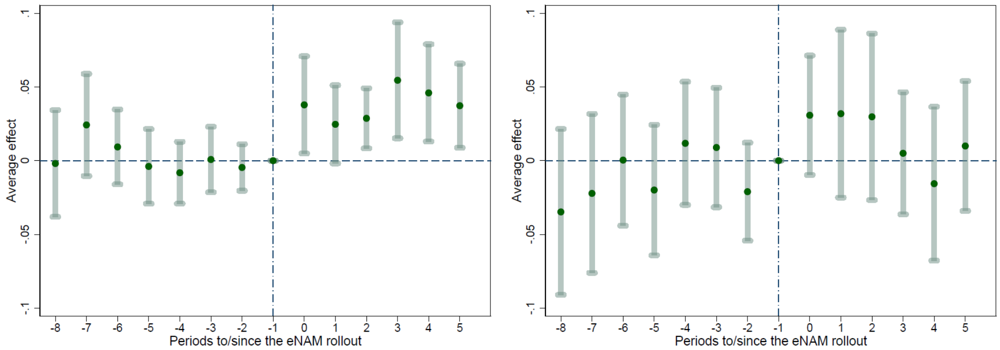


Panel B: Coarse Grade

Notes: These figures show the event-study graphs for the effect of eNAM policy on modal prices. Panel A presents the results for the medium grade of rice and panel B presents the results for the coarse grade. The reference group is the period preceding the eNAM implementation. The circles give the estimated coefficients normalized by the mean of the period prior to the policy reform, and the vertical lines indicate the 95 percent confidence interval.

Having established that eNAM policy reform led to an increase in rice prices, we now turn our attention to understanding the potential mechanisms driving this effect. Three primary channels emerge as potential explanations: First, reduced entry costs through remote trading, potentially leading to increased trader participation and enhanced outside options for farmers within the market. Second, the transition from offline to

Figure 3.4: The Effect of eNAM on Prices: Maximum Price



Panel A: Medium Grade

Panel B: Coarse Grade

Notes: These figures show the event-study graphs for the effect of eNAM policy on the maximum price in a market. Panel A presents the results for the medium grade of rice and panel B presents the results for the coarse grade. The reference group is the period preceding the eNAM implementation. The circles give the estimated coefficients normalized by the mean of the period prior to the policy reform, and the vertical lines indicate the 95 percent confidence interval.

online platforms potentially disrupting existing collusive behaviors among traders. Third, an increase in potential bidders due to liberal licensing of traders. While we lack information on the number of new licenses issued, we can investigate whether trader participation increases following eNAM implementation.

Our analysis from the previous section suggests that a one standard deviation increase in trader participation results in a 2% increase in prices. To study the effect of eNAM on trader participation, we again analyze our state-level data collected from Chhattisgarh for a period of 12 months surrounding the policy reform. Specifically, given that eNAM was first introduced in this state in November, we analyze the data from March 2016 to March 2017. This granular dataset allows us to examine changes in trader participation more directly and to assess whether the introduction of eNAM had a significant impact on trader participation in the market.

Our dataset comprises of daily observations from six major markets in Chhattisgarh. Importantly, these markets were integrated into the eNAM system in two phases: three markets were enrolled initially, while the remaining three were integrated four months later. This variation in implementation timing allows us to employ a difference-in-differences approach, comparing the outcomes of the first treated markets with the later treated markets. We restrict our analysis period to the eight months before and four months after the initial eNAM implementation, ending just before the later group of markets were enrolled.

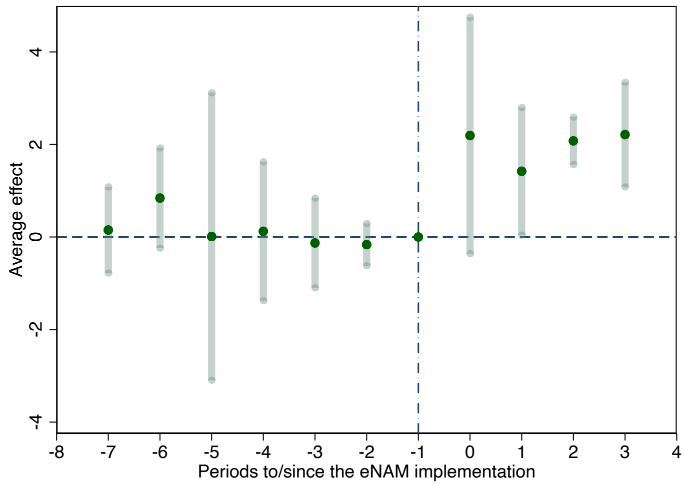
We employ the event-study approach as before, comparing trader participation in the early eNAM-implemented markets in Chhattisgarh to those markets that are not yet treated within our sample period. To provide a comprehensive picture of eNAM's impact, we also analyze its effect on prices within these six markets. This allows us to verify that the price increases through eNAM implementation, observed in our broader dataset, are also there in this smaller sample.

The results of these analyses are presented in Figure 3.5, which shows event-study graphs for both trader participation and prices. The figures reveal a clear jump in the number of trader participants following eNAM implementation. Concurrently, we observe a significant increase in prices, consistent with our findings from the broader dataset. These results provide strong evidence that eNAM increases prices, potentially by increasing trader participation.

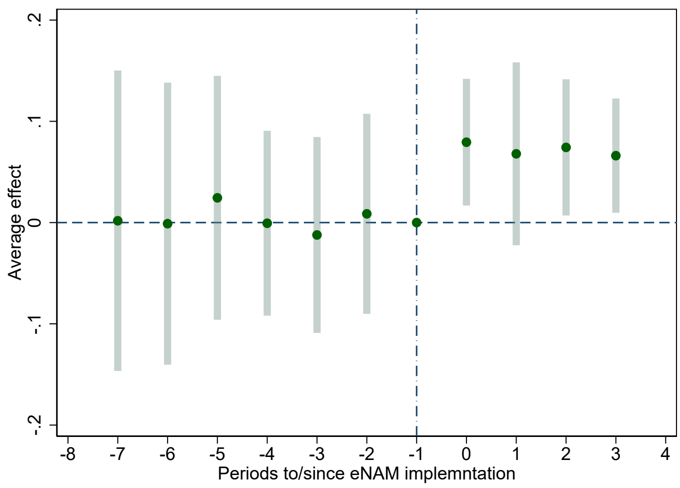
This analysis offers crucial insights into the mechanisms driving the observed price increases post-eNAM implementation. The simultaneous increase in trader participation and prices suggests that reduced entry barriers and increased competition play a substantial role in the observed price increases. However, we cannot rule out the possibility that other factors also contribute to these price changes.

One such factor warranting consideration is the potential disruption of collusive behaviors. The transition from offline to online platforms may have significant implications for market dynamics. There is a prevailing view that open outcry auctions are more susceptible to bidder collusion due to face-to-face interactions and the ability to react to opponent behavior in real-time. Thus, eNAM's implementation could have increased prices by disrupting potential collusion among traders. However, detecting and quantifying collusion is challenging, particularly in the context of agricultural markets with limited data availability. Disentangling the effects of increased trader participation from those of dismantling of collusive practices is beyond the scope of this paper due to such data limitations. Nevertheless, to lay the groundwork for future research in this area, we present in Appendix 3.A.3 a model of second-price auctions with endogenous entry. This model demonstrates how reduced entry costs could lead to increased trader participation and, consequently, higher prices. The model provides a framework for detecting collusion in markets transitioning from open-outcry to online auctions. With access to more granular data from be-

Figure 3.5: The Effect of eNAM on Trader Participation and Prices



Panel A: Trader participation



Panel B: Prices (Medium Grade)

Notes: These figures show the event-study graphs for the effect of eNAM policy on trader participation and prices. The sample uses data from one state of India for six major markets. The reference group is the period preceding the eNAM implementation. The circles give the estimated coefficients normalized by the mean of the period prior to the policy reform, and the vertical lines indicate the 95 percent confidence interval.

fore and after the policy change, future research could employ this model to identify potential collusive behavior and differentiate between various mechanisms driving price changes. Such analysis would offer valuable insights into the full impact of digital market platforms on agricultural trade dynamics.

3.5 Conclusion

This paper provides new insights into the determinants of agricultural prices in India, focusing on the role of intermediaries, competition, and policy reform. Our analysis reveals that both between markets and within-market competition among traders significantly influence the prices farmers receive for their produce.

Our research also evaluates the impact of the electronic National Agriculture Market (eNAM) policy, demonstrating that this reform led to a 3.1-3.4% increase in farmers' prices for rice. Notably, this study is one of the first to empirically assess the effects of eNAM on both prices and trader participation. The policy aims to increase competition through various mechanisms, including easier market licensing and the issuance of single unified licenses to traders, enabling them to operate across all regulated *mandis* within a state. While the policy's broader goal of facilitating inter-mandi and inter-state trade has not yet fully materialized, our findings suggest that even the partial implementation has yielded positive results. We show that the implementation of eNAM is associated with increased trader participation which is a key driver of price increase. This finding highlights the potential of online platforms to enhance market efficiency and improve farmer welfare. Given the importance of spatial competition revealed in our study, the full implementation of cross-state trading under eNAM could potentially lead to even greater price increases for farmers.

Two key insights emerge from our analysis. First, it is crucial to understand the form of competition among traders in agricultural markets, as this guides policymakers in identifying specific aspects to target. Our findings suggest that addressing information gaps or enhancing infrastructure alone may not be sufficient to improve farmers' well-being in regions where there is imperfect competition. Second, online trading platforms can yield multiple benefits, primarily through boosting trader participation. While not directly tested in our study, these platforms may also have the potential to dismantle collusive practices existing in open ascending auctions. In the context of the Indian government's push towards fully implementing electronic unified markets, our research offers valuable estimations of the potential gains from such reforms.

To conclude, this paper contributes to the literature by illuminating previously unexplored channels that influence farm-gate prices. We demonstrate that geographical variation in market concentration and restricted market participation due to government regulations and trader cartelization have sizeable impacts on agricultural prices. Furthermore, we show how the implementation of an online platform enhances these prices by reducing barriers to entry. These findings have important implications for policy design in agricultural markets and potentially in other sectors with similar market structures in developing economies.

3.A Chapter 3 Appendix

3.A.1 Additional Tables

Table 3.A1 : eNAM Summary Statistics

	Not-Integrated		Integrated	
	Mean	S.D.	Mean	S.D.
Storage Facility	0.62	0.48	0.64	0.47
Spatial Competition (weighted avg)	0.71	0.40	0.71	0.38
Number of licensed traders	230.77	131.06	246.76	160.18
Population Density (per sq km.)	523	658	520	290
Area Served (Villages)	170	191	199	215
Share of crop				
Coarse Grade	21.72	25.9	24.18	28.7
Medium Grade	72.76	28.9	69.39	30.1
Fine Grade	5.52	19.1	6.36	25.0

Notes: This table reports the characteristics of the markets that implemented eNAM and those that did not. Note that some variables are missing across certain markets.

Table 3.A2 : Competition and Maximum Price

	(1)	(2)	(3)
		Log price	
	Medium	Coarse	Fine
Within-market competition	0.014***	0.021***	0.019***
	(0.002)	(0.005)	(0.005)
Between-market (weighted avg)	0.012**	0.017***	0.003
	(0.005)	(0.006)	(0.013)
Storage availability	0.003	0.015***	0.014
	(0.004)	(0.005)	(0.012)
State FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	49807	17559	5594
R-squared	0.731	0.560	0.350

*Notes: All regressions control for district-year crop yield, population density. distance to district headquarter as well as lagged monthly rainfall. Standard errors (in parenthesis) are clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.*

Table 3.A3 : Competition and Minimum Price

	(1)	(2)	(3)
		Log price	
	Medium	Coarse	Fine
Within-market competition	0.010***	0.013***	0.014***
	(0.003)	(0.005)	(0.004)
Between-market (weighted avg)	0.013**	0.017***	0.005
	(0.006)	(0.007)	(0.013)
Storage availability	0.005	0.024***	0.004
	(0.008)	(0.005)	(0.011)
State FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes
Observations	49807	17559	5594
R-squared	0.731	0.560	0.350

*Notes: All regressions control for district-year crop yield, population density. distance to district headquarter as well as lagged monthly rainfall. Standard errors (in parenthesis) are clustered at district level. *** p<0.01, ** p<0.05, * p<0.1.*

Table 3.A4 : Competition and Farm price for Chattisgarh: IV

	(1)	(2)	(3)
		Log price	
	All	Medium	Coarse
Within-market Competition	0.024**	0.024**	0.021*
	(0.011)	(0.012)	(0.011)
Between-market (weighted avg)	0.009	0.038	0.029
	(0.037)	(0.039)	(0.036)
Time FE	Yes	Yes	Yes
Grade FE	Yes		
Observations	140846	127108	13738
R-squared	0.341	0.326	0.593

*Notes: All regressions control for district-year crop yield, population density as well as lagged monthly rainfall. In Column 1 , we report results together for all grades of rice with grade dummy. Standard errors (in parenthesis) are clustered at market level. The coefficient on the instrument in the first-stage regression is 0.028 (0.003) for all the regressions. *** p<0.01, ** p<0.05, * p<0.1.*

3.A.2 Additional Figures

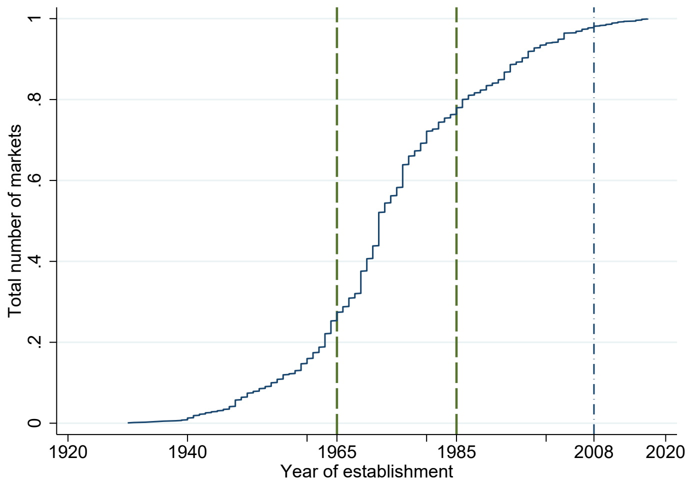


Figure 3.A1 : Growth in Number of Markets Over the Years

Notes: This figure shows the growth in the number of markets created over the years. The dotted line marks the beginning of our analysis period, while the period between the green lines presents the peak of the Green Revolution when most markets were established. interval.

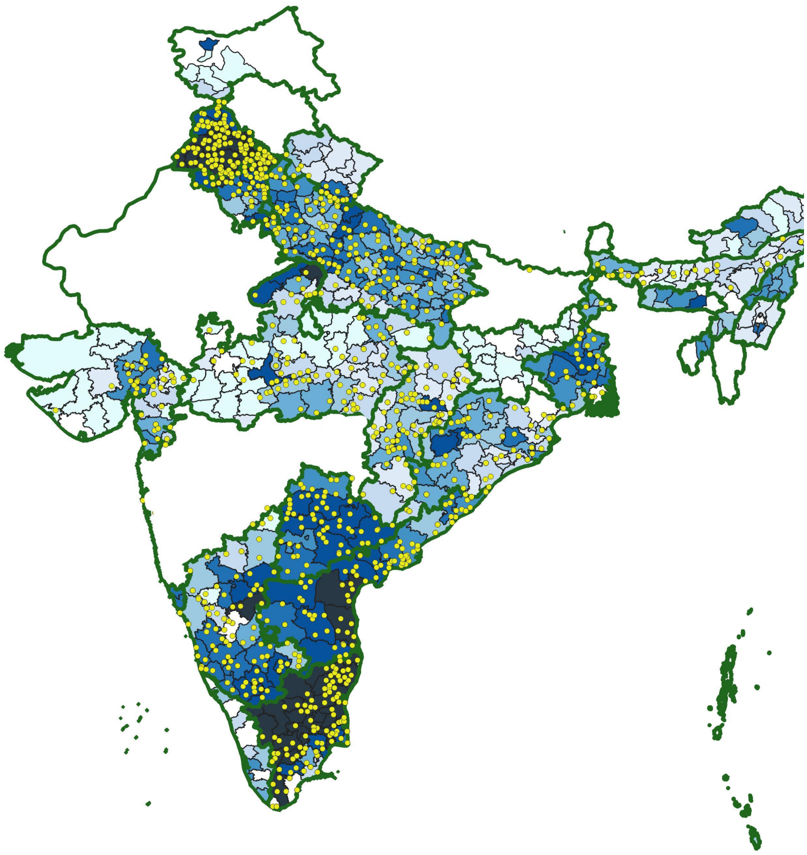


Figure 3.A2 : Market Placement and Average Crop Yield (1960-1980)
Notes: The map illustrates the distribution of agricultural markets and crop yields averaged over the period 1960-1980. Dots represent regulated market (mandi) locations. Shades of blue indicate average crop yields per area, with darker shades representing higher yields. White regions reflect lack of yield data for certain states.

3.A.3 Model

The model examines a spatially dispersed agricultural economy, focusing on the interactions between farmers, regulated crop markets or *mandis*, and traders. The economy consists of geographically dispersed farmers who produce crops at various locations. Regulated markets are exogenously placed throughout the region. Farmers sell their harvest at the nearest local market. Traders are licensed to operate in specific markets and cannot freely move between them. Their primary decision is whether to enter the market on a given day.

The model employs a ascending auction format for trading, reflecting the practice common in these agricultural markets. There are L market locations in the state, and at each location l, there is a set N_l of traders. To participate in the market, each trader must incur an entry cost d_i drawn from a distribution K_d. Upon paying this cost, trader i enters the market and learns their private value v_{ij} for each unit j . The private value is drawn from a common distribution F_v with density f_v. Valuations v_{ij} are independent across bidders i and units j. Each trader's payoff is given by the sum of all values for the units they end up buying, minus the total price they have to pay.

$$v_{ij} \overset{iid}{\backsim} F_v$$

The auction follows a standard ascending auction format. The price rises from a minimum bidding price, and the auction concludes and the price is finalized when all but one participating bidder has dropped out. A trader's strategy consists of two components: (i) Bidding strategy $b_i(\cdot; n)$ which specifies trader i's bid as a function of their valuation and the set of participating bidders and (ii) Entry strategy p_i which determines the probability of the trader entering the auction. The model focuses on an equilibrium characterized by a pair of bidding and entry strategies such that each bidder's strategy maximizes their profits, conditional on entering the market and traders enter the market if and only if their expected profit from entry exceeds the entry cost.

Equilibrium Bidding

Let's consider a participating bidder i with a private value v_i, bidding in a market location where the set of participants is n_l. Given the ascending auction format, all participants have a dominant strategy of bidding up to their true valuation. In the equilibrium with n_ℓ bidders, where all bidders follow this dominant strategy, the winning bidder pays the second highest valuation. The expected payoff of bidder i, from bidding on a unit, conditional on entering and bidding up to their valuation v_i, can thus be expressed as follows:

$$\pi(v_i; n) = G^2_{v,n_l}(v_i) \int (v_i - x) g^2_{v,n_l} dx$$

where G^2_{v,n_l} and g^2_{v,n_l}, respectively, denote the cdf and pdf of the second highest of n_l valuations. If all the players play their equilibrium bid strategies then the expected profit of the trader before knowing their valuation would be:

$$\pi_i(n) = \frac{1}{n_l} \mathbb{E}(v^{(1)}_{n_l} - v^{(2)}_{n_{tl}})$$

$$= \frac{1}{n_l} \Big[\int x \cdot g^{(1)}_{v,n_{tl}}(x) dx - \int x \cdot g^{(2)}_{v,n_l}(x) dx \Big]$$

where $g^{(1)}(.)$ and $g^{(2)}(.)$ are first and second highest valuation density function.

From the distribution of the transaction prices $G(\cdot)_{p,n_l}$, which is the second highest valuation, we can identify F_v, that is, the value distribution following Athey and Haile (2002).

$$G_{p,n_l}(x) = F^{(n-1:n)}(x) = \Big(n_l F_v(x)^{n_l - 1} - (n_l - 1) F_v(x)^{n_l} \Big)$$

Using the transaction prices observed in a given day, we can thus estimate our value distribution. These would then provide an estimate of $\pi(Z, N, n)$, where Z is a vector of observed sale characteristics.

Equilibrium Entry

I now characterize equilibrium entry. In this stage, traders compare the ex ante expected profit to their entry cost d_i. Let $\pi(n)$ denote the expected profit for a trader given the set of participants n. At location l the set of possible bidders is N_l. Bidders simultaneously decide whether to enter the auction. If they find it optimal to enter an auction for a single unit, they find it optimal to enter all auctions, and vice verse. All traders are symmetric thus ex-ante expected profit for a trader from participating is :

$$\Pi(p) = \frac{Q_l}{n_l} \sum_{n \subset N} \pi(n) Pr(n | N, i \in n)$$

where p is the entry probability and $Pr(n|N, i \in n)$ is the probability that n bidders enter given i enters. Q_l is the total number of units sold in the market l. It reflects that at the entry stage, traders are uncertain about how many other traders will enter. We study the symmetric equilbira of entry game which involves mixed strategies played by traders. Here each potential trader enters with probability p. The number of traders then follow a binomial distribution

$$Pr[n|N, i \in n] = \binom{N-1}{n-1} p^{n-1}(1-p)^{N-n}$$

Entering is optimal if the expected profit $\Pi(p)$ exceeds the entry cost. The entry cost threshold is defined as $D(p) = \Pi(p)$. Thus only those traders with entry cost below th entry threshold, $D(p)$, will enter the market. The equilibrium entry probability, p, would then solve $p = K[D(p)]$.

Bringing the model and data together can help us assess the relationship between our empirical findings and the theory we proposed. Our model provides a way to detect potential collusion in agricultural markets, particularly in open auction settings. By leveraging entry and bidding data from the post-eNAM reform period, we can estimate the model parameters assuming competitive behavior in online trading. The calibrated model then could allow us to predict equilibrium outcomes for each sale in our sample. The important application of this model lies in its ability to generate out-of-sample predictions for open auctions, which are not used in the initial parameter estimation. By comparing these predictions to actual outcomes in open auctions, we can assess whether observed prices align with competitive behaviour. If observed prices consistently fall below predicted competitive prices, with two or more traders, it may indicate collusive behavior. This approach provides a quantitative tool to evaluate the competitiveness of open auctions and potentially identify instances where traders may be engaging in collusive practices. Given the concerns about collusion in these agricultural markets (Meenakshi and Banerji (2005)), this model could thus offer a valuable mechanism for analysing competitiveness in agricultural trade.

References

Abrevaya, Jason and Donald, Stephen G (2017). "A GMM approach for dealing with missing data on regressors". *Review of Economics and Statistics* 99.4, pp. 657–662.

Acharya, Shabd Swaroop (2004). *Agricultural marketing in India.* Oxford and IBH publishing.

Adams-Prassl, Abi and Andrew, Alison (2020). "Preferences and beliefs in the marriage market for young brides".

Aggarwal (2018). "Do rural roads create pathways out of poverty? Evidence from India". *Journal of Development Economics* 133, pp. 375–395.

Aggarwal, Nidhi, Jain, Sargam, and Narayanan, Sudha (2017). "The long road to transformation of agricultural markets in India: Lessons from Karnataka". *Economic and Political Weekly*, pp. 47–55.

Aggarwal, Shilpa, Giera, Brian, Jeong, Dahyeon, Robinson, Jonathan, and Spearot, Alan (2022). "Market access, trade costs, and technology adoption: Evidence from northern Tanzania". *Review of Economics and Statistics*, pp. 1–45.

Agostinelli, Francesco, Saharkhiz, Morteza, and Wiswall, Matthew J. (2019). "Home and Schools in the Development of Children". *NBER Working Paper Series* 26037.

Ahn, JaeBin, Khandelwal, Amit K, and Wei, Shang-Jin (2011). "The role of intermediaries in facilitating trade". *Journal of International Economics* 84.1, pp. 73–85.

Allen, Treb and Atkin, David (2022). "Volatility and the Gains from Trade". *Econometrica* 90.5, pp. 2053–2092.

Almås, Ingvild, Attanasio, Orazio, and Jervis, Pamela (2023). *Economics and measurement: New measures to model decision making.* Tech. rep. National Bureau of Economic Research.

Ameriks, John, Briggs, Joseph, Caplin, Andrew, Lee, Minjoon, Shapiro, Matthew D, and Tonetti, Christopher (2020). "Older Americans would work longer if jobs were flexible”. *American Economic Journal: Macroeconomics* 12.1, pp. 174–209.

Ameriks, John, Briggs, Joseph, Caplin, Andrew, Shapiro, Matthew D, and Tonetti, Christopher (2020). “Long-Term-Care Utility and Late-in-Life Saving”. *Journal of Political Economy* 128.6, pp. 2375–2451.

Andrew, Alison, Attanasio, Orazio, Augsburg, Britta, Cardona-Sosa, Lina, Day, Monimalika, Giannola, Michele, Grantham-McGregor, Sally, Jervis, Pamela, Meghir, Costas, and Rubio-Codina, Marta (2024). *Early Childhood Intervention for the Poor: Long Term Outcomes.* Tech. rep. National Bureau of Economic Research.

Angrist, Joshua D, Pathak, Parag A, and Walters, Christopher R (2013). “Explaining charter school effectiveness”. *American Economic Journal: Applied Economics* 5.4, pp. 1–27.

Antras, Pol and Costinot, Arnaud (2011). “Intermediated trade”. *The Quarterly Journal of Economics* 126.3, pp. 1319–1374.

Araujo, M Caridad, Carneiro, Pedro, Cruz-Aguayo, Yyannú, and Schady, Norbert (2016). “Teacher quality and learning outcomes in kindergarten”. *The Quarterly Journal of Economics* 131.3, pp. 1415–1453.

Asturias, Jose, Garcia-Santana, Manuel, and Ramos, Roberto (2019). “Competition and the welfare gains from transportation infrastructure: Evidence from the Golden Quadrilateral of India”. *Journal of the European Economic Association* 17.6, pp. 1881–1940.

Athey, Susan and Haile, Philip A (2002). “Identification of standard auction models”. *Econometrica* 70.6, pp. 2107–2140.

Athey, Susan, Levin, Jonathan, and Seira, Enrique (2011). “Comparing open and sealed bid auctions: Evidence from timber auctions”. *The Quarterly Journal of Economics* 126.1, pp. 207–257.

Atkin, David and Donaldson, Dave (2015). *Who’s Getting Globalized?: The Size and Implications of Intra-national Trade Costs.* w21439. National Bureau of Economic Research Cambridge, MA.

Attanasio, Orazio, Boneva, Teodora, and Rauh, Christopher (2019). *Parental beliefs about returns to different types of investments in school children.* Tech. rep. National Bureau of Economic Research.

Attanasio, Orazio, Cunha, Flavio, and Jervis, Pamela (2019a). *Subjective Parental Beliefs. Their Measurement and Role.* Tech. rep. National Bureau of Economic Research.

Attanasio, Orazio, Cunha, Flavio, and Jervis, Pamela (2019b). "Subjective Parental Beliefs. Their Measurement and Role". *NBER Working Paper*, No. 26516. ISSN: 1098-6596. DOI: 10.1017/CBO9781107415324.004.

Attanasio, Orazio, Meghir, Costas, and Nix, Emily (2020). "Human capital development and parental investment in India". *The Review of Economic Studies* 87.6, pp. 2511–2541.

Bailey, Charles E (2007). "Cognitive accuracy and intelligent executive function in the brain and in business". *Annals of the New York Academy of Sciences* 1118.1, pp. 122–141.

Banerjee, Abhijit, Fischer, Greg, Karlan, Dean, Lowe, Matt, and Roth, Benjamin N (2022). *Does the Invisible Hand Efficiently Guide Entry and Exit? Evidence from a Vegetable Market Experiment in India*. Tech. rep. National Bureau of Economic Research.

Banerji, A and Meenakshi, JV (2004). "Buyer collusion and efficiency of government intervention in wheat markets in northern India: An asymmetric structural auctions analysis". *American Journal of Agricultural Economics*, pp. 236–253.

Bardhan, Pranab, Mookherjee, Dilip, and Tsumagari, Masatoshi (2013). "Middlemen margins and globalization". *American Economic Journal: Microeconomics* 5.4, pp. 81–119.

Bartkus, Viva Ona, Brooks, Wyatt, Kaboski, Joseph P, and Pelnik, Carolyn (2022). "Big fish in thin markets: Competing with the middlemen to increase market access in the Amazon". *Journal of Development Economics* 155, p. 102757.

Bassi, Marina, Medina Pedreira, Octavio, and Nhampossa, Lucia Jose (2019). "Education Service Delivery in Mozambique: A Second Round of the Service Delivery Indicators Survey". *Report. Washington, DC: World Bank Group.*

Bau, Natalie and Das, Jishnu (2020). "Teacher value added in a low-income country". *American Economic Journal: Economic Policy* 12.1, pp. 62–96.

Behrman, Jere R, Pollak, Robert A, and Taubman, Paul (1982). "Parental preferences and provision for progeny". *Journal of political economy* 90.1, pp. 52–73.

Bergquist, Lauren Falcao and Dinerstein, Michael (2020). "Competition and entry in agricultural markets: Experimental evidence from Kenya". *American Economic Review* 110.12, pp. 3705–3747.

Besley, Timothy and Burgess (2000). “Land reform, poverty reduction, and growth: Evidence from India”. *The Quarterly Journal of Economics* 115.2, pp. 389–430.

Blankenship, Tashauna L, O’Neill, Meagan, Ross, Alleyne, and Bell, Martha Ann (2015). “Working memory and recollection contribute to academic achievement”. *Learning and individual differences* 43, pp. 164–169.

Boneva, Teodora and Rauh, Christopher (2018). “Parental Beliefs about Returns to Educational Investments—The Later the Better?” *Journal of the European Economic Association* 16.6, pp. 1669–1711.

Borella, Erika, Carretti, Barbara, and Pelegrina, Santiago (2010). “The specific role of inhibition in reading comprehension in good and poor comprehenders”. *Journal of Learning Disabilities* 43.6, pp. 541–552.

Buhl-Wiggers, Julie, Kerwin, Jason T, Smith, Jeffrey, and Thornton, Rebecca (2022). “Learning More about Teachers: Estimating Teacher Value-Added and Treatment Effects on Teacher Value-Added in Northern Uganda”.

Burchinal, Margaret, Kainz, Kirsten, and Cai, Yaping (2011). “How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings.”

Burgess, Paul W and Simons, Jon S (2005). “18 Theories of frontal lobe executive function: clinical applications”. *The effectiveness of rehabilitation for cognitive deficits* 211.

Busso, Matias and Galiani, Sebastian (2019). “The causal effect of competition on prices and quality: Evidence from a field experiment”. *American Economic Journal: Applied Economics* 11.1, pp. 33–56.

Callaway, Brantly and Sant’Anna, Pedro HC (2021). “Difference-in-differences with multiple time periods”. *Journal of Econometrics* 225.2, pp. 200–230.

Caplin, Andrew (2021). *Economic Data Engineering*. Tech. rep. National Bureau of Economic Research.

Carneiro, Pedro, Glewwe, Paul, Guha, Anusha, and Krutikova, Sonya (2023). “Measuring the impacts of teachers in Vietnam: Teacher value-added and student skills”.

Chand, Ramesh (2012). “Development policies and agricultural markets”. *Economic and Political Weekly*, pp. 53–63.

Chatterjee, Shoumitro (2023). “Market power and spatial competition in rural india”. *The Quarterly Journal of Economics* 138.3, pp. 1649–1711.

Chatterjee, Shoumitro and Kapur, Devesh (2017). “Six puzzles in Indian agriculture”. *India Policy Forum*. Vol. 13. 1. National Council of Applied Economic Research, pp. 185–229.

Chetty, Raj, Friedman, John N, Hilger, Nathaniel, Saez, Emmanuel, Schanzenbach, Diane Whitmore, and Yagan, Danny (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project STAR”. *The Quarterly Journal of Economics* 126.4, pp. 1593–1660.

Chetty, Raj, Friedman, John N, and Rockoff, Jonah E (2011). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. Tech. rep. National Bureau of Economic Research.

— (2014). “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood”. *American Economic Review* 104.9, pp. 2633–79.

Cronbach, Lee J (1949). “Essentials of psychological testing.”

Cunha, Flavio, Elo, Irma, and Culhane, Jennifer (Sept. 2020). “Maternal subjective expectations about the technology of skill formation predict investments in children one year later”. *Journal of Econometrics*. ISSN: 03044076. DOI: 10.1016/j.jeconom.2020.07.044.

Cunha, Flavio and Heckman, James (2007). “The technology of skill formation”. *American Economic Review* 97.2, pp. 31–47.

Curby, Timothy W, Rimm-Kaufman, Sara E, and Ponitz, Claire Cameron (2009). “Teacher–child interactions and children’s achievement trajectories across kindergarten and first grade.” *Journal of educational psychology* 101.4, p. 912.

Currie, Janet and Almond, Douglas (2011). “Human capital development before age five”. *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1315–1486.

Dang, Hai-Anh, Glewwe, Paul, Lee, Jongwook, and Vu, Khoa (2023). “What Explains Vietnam’s Exceptional Performance in Education Relative to Other Countries? Analysis of the 2012, 2015, and 2018 PISA Data”. *Economics of Education Review* 96, p. 102434.

Das, Jishnu, Dercon, Stefan, Habyarimana, James, Krishnan, Pramila, Muralidharan, Karthik, and Sundararaman, Venkatesh (2013). “School inputs, household substitution, and test scores”. *American Economic Journal: Applied Economics* 5.2, pp. 29–57.

Das, Jishnu and Zajonc, Tristan (2010). “India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement”. *Journal of Development Economics* 92.2, pp. 175–187. ISSN: 03043878. DOI: 10.1016/j.jdeveco.2009.03.004.

Davidson, Matthew C, Amso, Dima, Anderson, Loren Cruess, and Diamond, Adele (2006). “Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching”. *Neuropsychologia* 44.11, pp. 2037–2078.

Del Bono, Emilia, Francesconi, Marco, Kelly, Yvonne, and Sacker, Amanda (2016). “Early maternal time investment and early child outcomes”. *The Economic Journal* 126.596, F96–F135.

Delavande, Adeline and Zafar, Basit (2019). “University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints”. *Journal of Political Economy* 127.5, pp. 2343–2393.

Deming, David J, Hastings, Justine S, Kane, Thomas J, and Staiger, Douglas O (2014). “School choice, school quality, and postsecondary attainment”. *American Economic Review* 104.3, pp. 991–1013.

Dhingra, Swati and Tenreyro, Silvana (2017). *Piggy-back exporting, intermediation, and the distributional gains from trade in agricultural markets.*

Diamond, Adele and Wright, Andy (2014). “An effect of inhibitory load in children while keeping working memory load constant”. *Frontiers in psychology* 5, p. 213.

Dillon, Brian and Dambro, Chelsey (2017). “How competitive are crop markets in sub-Saharan Africa?” *American Journal of Agricultural Economics* 99.5, pp. 1344–1361.

Donaldson, Dave (2015). “The gains from market integration”. *Annual Review of Economics* 7.1, pp. 619–647.

— (2018). “Railroads of the Raj: Estimating the impact of transportation infrastructure”. *American Economic Review* 108.4-5, pp. 899–934.

Donaldson, Dave and Hornbeck, Richard (2016). “Railroads and American economic growth: A “market access” approach”. *The Quarterly Journal of Economics* 131.2, pp. 799–858.

Espy, Kimberly Andrews (2004). “Using developmental, cognitive, and neuroscience approaches to understand executive control in young children”. *Developmental Neuropsychology* 26.1, pp. 379–384.

Feenstra, Robert C and Weinstein, David E (2017). “Globalization, markups, and US welfare”. *Journal of Political Economy* 125.4, pp. 1040–1074.

Fiorini, Mario and Keane, Michael (2014). “How the allocation of children’s time affects cognitive and noncognitive development”. *Journal of Labor Economics* 32.4, pp. 787–836.

Ganimian, Alejandro J and Murnane, Richard J (2016). “Improving education in developing countries: Lessons from rigorous impact evaluations”. *Review of Educational Research* 86.3, pp. 719–755.

Gao, Pei and Lei, Yu-Hsiang (2021). “Communication infrastructure and stabilizing food prices: evidence from the telegraph network in China”. *American Economic Journal: Applied Economics* 13.3, pp. 65–101.

Giannola, Michele (2020). *Parental Investment, and Intra-household Inequality in Child Human Capital: Evidence from a Lab-in-the-field Experiment*. Tech. rep. Mimeo, University College London.

Government of Vietnam (2014). *Resolution No. 44/NQ-CP dated June 09, 2014*.

Goyal, Aparajita (2010). “Information, direct access to farmers, and rural market performance in central India”. *American Economic Journal: Applied Economics* 2.3, pp. 22–45.

Grant, Matthew and Startz, Meredith (2022). *Cutting out the middleman: The structure of chains of intermediation*. Tech. rep. National Bureau of Economic Research.

Greaves, Ellen, Hussain, Iftikhar, Rabe, Birgitta, and Rasul, Imran (2023). “Parental responses to information about school quality: Evidence from linked survey and administrative data”. *The Economic Journal* 133.654, pp. 2334–2402.

Guarino, Cassandra, Reckase, Mark, Stacy, Brian, and Wooldridge, Jeffrey (2015). “A comparison of student growth percentile and value-added models of teacher performance”. *Statistics and Public Policy* 2.1, pp. 1–11.

Hamre, Bridget, Hatfield, Bridget, Pianta, Robert, and Jamil, Faiza (2014). “Evidence for general and domain-specific elements of teacher–child interactions: Associations with preschool children’s development”. *Child development* 85.3, pp. 1257–1274.

Hanushek, Eric A (2003). “The failure of input-based schooling policies”. *The Economic Journal* 113.485, F64–F98.

Hanushek, Eric A and Rivkin, Steven G (2006). “Teacher quality”. *Handbook of the Economics of Education* 2, pp. 1051–1078.

— (2012). “The distribution of teacher quality and implications for policy”. *Annu. Rev. Econ.* 4.1, pp. 131–157.

Harris, R (2012). “Vietnam EFA action plan 2003–2015”. *Review and update.*

Harter, S (1980). “A scale of intrinsic versus extrinsic orientation in the classroom (Available from Susan Harter, Department of Psychology, University of Denver, Denver, CO 80208)”.

Harter, Susan (1981). “A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components.” *Developmental psychology* 17.3, p. 300.

— (2012). “Self-perception profile for adolescents: Manual and questionnaires”. *Denver, CO: Univeristy of Denver, Department of Psychology.*

Heckman, James, Stixrud, Jora, and Urzua, Sergio (2006). “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior”. *Journal of Labor Economics* 24.3, pp. 411–482.

Jacob, Brian A, Lefgren, Lars, and Sims, David P (2010). “The persistence of teacher-induced learning”. *Journal of Human resources* 45.4, pp. 915–943.

Jensen, Robert (2007). “The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector”. *The Quarterly Journal of Economics* 122.3, pp. 879–924.

Jochmans, Koen (2020). “Testing Random Assignment to Peer Groups”.

Keane, Michael, Krutikova, Sonya, and Neal, Timothy (2022a). “Child Work and Cognitive Development: Results From Four Low To Middle Income Countries”. *Quantitative Economics* 13.2, pp. 425–465.

— (2022b). “Child work and cognitive development: Results from four low to middle income countries”. *Quantitative Economics* 13.2, pp. 425–465.

Koedel, Cory and Rockoff, Jonah E (2015). “Value-added modeling: A review”. *Economics of Education Review* 47, pp. 180–195.

Krasnokutskaya, Elena and Seim, Katja (2011). “Bid preference programs and participation in highway procurement auctions”. *American Economic Review* 101.6, pp. 2653–2686.

Lemos, Marina S and Veríssimo, Lurdes (2014). “The relationships between intrinsic motivation, extrinsic motivation, and achievement, along elementary school”. *Procedia-Social and Behavioral Sciences* 112, pp. 930–938.

Leone, Fabrizio, Macchiavello, Rocco, and Reed, Tristan (2024). “The High and Falling Price of Cement in Africa”. *American Economic Journal: Applied Economics.*

Little, Roderick JA and Rubin, Donald B (2019). *Statistical analysis with missing data*. Vol. 793. John Wiley & Sons.

London, Jonathan D (2011). *Education in Vietnam*. Institute of Southeast Asian Studies.

Macchiavello, Rocco and Morjaria, Ameet (2015). "The value of relationships: evidence from a supply shock to Kenyan rose exports". *American Economic Review* 105.9, pp. 2911–45.

Maitra, Pushkar, Mitra, Sandip, Mookherjee, Dilip, Motta, Alberto, and Visaria, Sujata (2017). "Financing smallholder agriculture: An experiment with agent-intermediated microloans in India". *Journal of Development Economics* 127, pp. 306–337.

Marchant, Gregory J (1991). "A profile of motivation, self-perception, and achievement in Black urban elementary students". *The Urban Review* 23.2, pp. 83–99.

Mashburn, Andrew J, Pianta, Robert C, Hamre, Bridget K, Downer, Jason T, Barbarin, Oscar A, Bryant, Donna, Burchinal, Margaret, Early, Diane M, and Howes, Carollee (2008). "Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills". *Child development* 79.3, pp. 732–749.

McCaffrey, Daniel F, Sass, Tim R, Lockwood, JR, and Mihaly, Kata (2009). "The intertemporal variability of teacher effect estimates". *Education Finance and Policy* 4.4, pp. 572–606.

Meenakshi, JV and Banerji, A (2005). "The unsupportable support price: an analysis of collusion and government intervention in paddy auction markets in North India". *Journal of Development Economics* 76.2, pp. 377–403.

Melitz, Marc J and Ottaviano, Gianmarco IP (2008). "Market size, trade, and productivity". *The Review of Economic Studies* 75.1, pp. 295–316.

Miller, Nathan H and Osborne, Matthew (2014). "Spatial differentiation and price discrimination in the cement industry: evidence from a structural model". *The RAND Journal of Economics* 45.2, pp. 221–247.

Mitra, Sandip, Mookherjee, Dilip, Torero, Maximo, and Visaria, Sujata (2018). "Asymmetric information and middleman margins: An experiment with Indian potato farmers". *Review of Economics and Statistics* 100.1, pp. 1–13.

MOET (2018). *Circular No. 03/2018/TT-BGDDT*.

Molina, Ezequiel, Fatima, Syeda Farwa, Ho, Andrew Dean, Melo, Carolina, Wilichowski, Tracy Marie, and Pushparatnam, Adelle (2020). "Measuring the quality of teaching practices in primary schools: Assessing the validity of the Teach observation tool in Punjab, Pakistan". *Teaching and Teacher Education* 96, p. 103171.

Molina, Ezequiel, Pushparatnam, Adelle, and Wilichowski, Tracy (Jan. 2019). *Teach: Tackling the learning crisis, one classroom at a time.*

Muralidharan, Karthik, Singh, Abhijeet, and Ganimian, Alejandro J (2019). "Disrupting education? Experimental evidence on technology-aided instruction in India". *American Economic Review* 109.4, pp. 1426–1460.

Muralidharan, Karthik and Sundararaman, Venkatesh (2013). *Contract teachers: Experimental evidence from India.* Tech. rep. National Bureau of Economic Research.

Muris, Chris (2010). *Efficient GMM estimation with a general missing data pattern.* Tech. rep. Citeseer.

Oberle, Eva and Schonert-Reichl, Kimberly A (2013). "Relations among peer acceptance, inhibitory control, and math achievement in early adolescence". *Journal of Applied Developmental Psychology* 34.1, pp. 45–51.

Obradović, Jelena, Sulik, Michael J, Finch, Jenna E, and Tirado-Strayer, Nicole (2018). "Assessing students' executive functions in the classroom: Validating a scalable group-based procedure". *Journal of Applied Developmental Psychology* 55, pp. 4–13.

Pop-Eleches, Cristian and Urquiola, Miguel (2013). "Going to a better school: Effects and behavioral responses". *American Economic Review* 103.4, pp. 1289–1324.

Rosenthal, Eve N, Riccio, Cynthia A, Gsanger, Kristen M, and Jarratt, Kelly Pizzitola (2006). "Digit Span components as predictors of attention problems and executive functioning in children". *Archives of Clinical Neuropsychology* 21.2, pp. 131–139.

Singh, Abhijeet (2015). "Private school effects in urban and rural India: Panel estimates at primary and secondary school ages". *Journal of Development Economics* 113, pp. 16–32.

— (2020a). "Learning More with Every Year: School Year Productivity and International Learning Divergence". *Journal of the European Economic Association* 18.4, pp. 1770–1813. ISSN: 15424774. DOI: `10.1093/jeea/jvz033`.

— (2020b). "Learning more with every year: School year productivity and international learning divergence". *Journal of the European Economic Association* 18.4, pp. 1770–1813.

Sun, Liyang and Abraham, Sarah (2021). “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects”. *Journal of Econometrics* 225.2, pp. 175–199.

Taber, Keith S (2018). “The use of Cronbach’s alpha when developing and reporting research instruments in science education”. *Research in science education* 48, pp. 1273–1296.

Teach Primary: Observer Manual (n.d.).

Todd, Petra E and Wolpin, Kenneth I (2003). “On the specification and estimation of the production function for cognitive achievement”. *The Economic Journal* 113.485, F3–F33.

— (2007). “The production of cognitive achievement in children: Home, school, and racial test score gaps”. *Journal of Human capital* 1.1, pp. 91–136.

Tomar, Shekhar (2016). “Gains from Agricultural Market Reform: Role and Size of Intermediaries”. *Indian School of Business*.

Van Der Linden, Wim J and Hambleton, Ronald K (1997). *Handbook of Modern Item Response Theory*. Ed. by Wim J. van der Linden and Ronald K. Hambleton. New York, NY: Springer New York, pp. 1–28. ISBN: 978-1-4419-2849-8. DOI: 10.1007/978-1-4757-2691-6.

Will Crescioni, A, Ehrlinger, Joyce, Alquist, Jessica L, Conlon, Kyle E, Baumeister, Roy F, Schatschneider, Christopher, and Dutton, Gareth R (2011). “High trait self-control predicts positive health behaviors and success in weight loss”. *Journal of Health Psychology* 16.5, pp. 750–759.

Zaslow, Martha, Anderson, Rachel, Redd, Zakia, Wessel, Julia, Daneri, Paula, Green, Katherine, Cavadel, Elizabeth W, Tarullo, Louisa, Burchinal, Margaret, and Martinez-Beck, Ivelisse (2016). “I. Quality thresholds, features, and dosage in early care and education: introduction and literature review”. *Monographs of the Society for Research in Child Development* 81.2, pp. 7–26.