8-2022

Evaluating Impacts of Development Programs on Female Education in Afghanistan

Ahmad Shah Mobariz

University of Arkansas, Fayetteville

Evaluating Impacts of Development Programs on Female Education in Afghanistan

A dissertation submitted in partial fulfillment
of the the requirements of the degree of
Doctor of Philosophy in Economics

by

Ahmad Shah Mobariz
Bangalore University
Bachelor of Business Management, 2012
South Asian University
Master of Arts in Development Economics, 2014

August 2022
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

____
Arya B. Gaduh, Ph.D.
Dissertation Chair

____
Gary D. Ferrier, Ph.D.
Committee Member

____
Gema Zamarro Rodriguez
Committee Member

Abstract

This dissertation evaluates the effects of three development interventions on female education in Afghanistan: 1) effects of foreign military withdrawal on females' demand for higher education; 2) impacts of PEZAK, a community-driven university entrance preparation, on student enrollment and performance in tertiary education; and 3) long-term effects of National Solidarity Program (NSP), that established gender-balanced local development councils, on female enrollment. Foreign military withdrawal increased female participation in higher education by 0.3 percentage points from a base value of 0.05 percent participation per capita. The PEZAK increased test scores by 0.17 standard deviations and had a positive effect on enrollment in top programs. Female students exposed to the PEZAK had a lower likelihood of enrollment in low-rank universities as compared to treated male students. In areas with favorable attitudes toward women, the NSP increased female enrollment in higher grades. In culturally conservative places, the NSP was counterproductive. Findings in this dissertation inform development policies related to women's empowerment in conservative and fragile state settings.

Dedication

I dedicate this work to my father, Amir Mohammad Rahimi, whose vision for my and other children's future brought the first school to our village. I also dedicate this work to my first teacher, Mu'alim Baqir, "the teacher", who installed the seeds of passion for formal education in early days.

Acknowledgments

I thank my Ph.D. Committee, Dr. Gary D. Ferrier, Dr. Arya B. Gaduh, and Dr. Gema Zamarro Rodriguez. I am grateful for their guidance and immense support. My main supervisor, Dr. Arya Gaduh, provided thoughtful advice, challenged my research rigor, and trained me to become an independent researcher. Dr. Gary Ferrier was also my program advisor in connection with the Fulbright Programme. He guided me in the early stages of the Ph.D. program and offered constructive comments on my research ideas. Dr. Gema Zamarro was an incredible mentor who read my draft papers with patience and provided illuminating feedback. I am grateful for having the three of them as my mentors and thank their thoughtful training and genuine support.

I appreciate the mentoring I received from the Faculty of Economics at Sam M. Walton College of Business. Dr. Peter James McGee always listened with passion and rendered direction. Dr. Raja Kali valued talking to students and hearing our problems. They both accommodated pathways to address challenges that I faced in the aftermath of the fall of Afghanistan to the Taliban in August 2021. I would like to thank the leadership of the college and staff for creating an enabling research environment.

My Ph.D. program was sponsored by the Fulbright Programme and the University of Arkansas. I would like to thank the Fulbright Programme for her generous financial support and for facilitating an enriching cultural exchange journey. I thank the Sponsored Students and Special Programs office at the University of Arkansas for their tremendous institutional support.

Many people inspired me in this journey. Baqir Fateh encouraged me to join the Ph.D. program at the University of Arkansas. My family's patience with my frequent disconnections due to the tremendous demanding nature of my Ph.D. program was instrumental. My dream team, Nathaniel Burke, Logan Miller, and James Willbanks were the bedrock in building an in-house copping mechanism to mitigate the stressful Ph.D. life with laughter and friendly exchanges in addition to helping one another academically. My roommates, Arsene Agossadou, Ameldine Dachiroudine, Lemine M'Bareck, and Ahmed Sadek Yousuf were valuable companions. Khalid Ahmadzai, Ikramudin Bahram, Mahdi Faizi, Daniel Gilmour, Abdullah Ibrahimzada, Harun Karim, and their families were my family in Fayetteville. Lina Anaya, Saloni Dasgupta, Catherine Michaud Leclerc, Simon Meulen, Tucker Smith, Pere Taberner, and Cora Wigger, my co-organizers at Graduate Economics of Education Zoom Seminars (GEEZ) and GEEZ attendees were a source of motivation to tackle challenges during the Covid-19 pandemic. The Fulbright scholars at the University of Arkansas had an immense role in fulfilling my cultural exchange responsibilities as part of the Fulbright program. I appreciate all of them.

Chapter 1: Dr. Gary Ferrier, Dr. Arya Gaduh, and Dr. Gema Zamarro Rodriguez provided guidance. Dr. Austin L. Wright at the University of Chicago provided constructive feedback and responded to my questions related to the methodology that I apply in this study. Dr. Travers Barclay Child from China Europe International Business School (CEIBS) facilitated data access from NATO.

Chapter 2: Dr. Arya Gaduh secured funding for data cleaning and provided thoughtful guidance. Ibrahim Qasimi and his office in Kabul graciously furnished available information about the PEZAK program. Esmatullah Soroush facilitated access to student test scores data. Mahdi Alizada, Alikhan Behzad, Ali Sina Doosti, Zahra Hussainzada, Fazila Mirzapoor, Mohammad Reza Sharifi, Mohammad Sharif Sultani, Hanifa Yari, and Latifa Yari provided research assistance in data cleaning. Porsesh Research and Studies Organization (PRSO) in Kabul facilitated data cleaning.

Chapter 3: Dr. Andrew Beath from the World Bank furnished the NSP data and provided valuable institutional information. Hamidullah Gharibzada helped furnish school-village pairs' data from Afghanistan's Ministry of Education. Khodadad Bisharat, Nabi Nazari, and Samad Rezayee facilitated access to Afghanistan's education administrative data.

Contents

Introduction

The gender gap in education outcomes in low-and middle-income countries poses a problem for development. Between 0.4-0.9 percentage points difference in per capita growth rate between East Asia and Sub-Saharan Africa, and South Asia and the Middle East could be attributed to gender differences in educational attainment (Klasen, 2002). Measuring the effects of development programs on female education is important for development programming. This dissertation studies three programs that affect the enrollment and performance of women and students from disadvantaged communities in education with a focus on higher education.

The first chapter studies the effects of foreign military withdrawal on females' demand for higher education in Afghanistan. In the aftermath of the 9/11 attacks, the U.S. launched the War on Terror campaign that led to the collapse of the Taliban regime and the installation of a democratic system in Afghanistan. After a decade of military occupation, between 2011-2015, the United States implemented a conditions-based gradual transfer of security responsibilities to local security forces and evacuation of NATO troops. I exploit the geographic and topographic barriers to the shipment of military equipment from Afghanistan's districts to regional military airbases as a source of variation for physical repatriations of troops. I use the least-cost travel distance between districts and the nearest logistic hubs as an Instrumental Variable (IV) for the departure of forces. I find that withdrawal of foreign forces resulted in a 0.3 percentage point increase in female university participation from a base value of 0.05%, while male participation did not change. I show that declining economic opportunities and massive male emigration explain this difference.

The second chapter evaluates a gender-inclusive education program in Afghanistan that intends to increase the enrollment of students from rural communities and reduce gender disparity in higher education. The program recruited high-performing undergraduate students from top universities in Afghanistan to provide tutoring in math and science to high school students in rural areas. It helped high school students prepare for the university entrance examination. Using a complete enumeration of university entrance exam scores between 2003 and 2019, we estimate the impacts of the program on exam scores, enrollment by university ranking, and passing rate. We use a heterogeneity-robust difference-in-differences approach to estimate the treatment effects. We find that the program increased the university entrance exam scores and enrollment in top universities. The program also increased the overall passing rate. Female students exposed to the program had a lower likelihood of enrollment in low-rank universities as compared to treated male students.

The third chapter examines the long-term impacts of gender-balanced local development councils on female education. We combine a randomized experiment that was conducted to evaluate the effects of Afghanistan's National Solidarity Program (NSP) during 2008-2012 with current administrative education enrollment data for the years 2011-2018. The NSP was a flagship rural development program in Afghanistan during 2002-2017. The NSP evaluation experiment randomly allocated 250 villages to receive NSP and another 250 villages to control group that was intended to receive the program at a later time. The NSP villages established gender-balanced village councils and received block grants to implement projects prioritized by the councils. We matched the NSP villages and schools in the enrollment database using an intermediary village-school pair accessed from Afghanistan's ministry of education. Additionally, we employed name matching between the NSP villages and villages in school administrative data and also geographic coordinate references in both datasets to complete the school-village matching. We find that in areas with favorable attitudes toward women, the program increased female enrollment in higher school grades. On the other hand, in culturally conservative places the NSP produced negative impacts on female enrollment. Our study sheds light on the sustainability of development interventions.

Overall, this dissertation informs three aspects of development programming in fragile-state contexts. It provides a rigorous analysis of unintended gender consequences of a foreign military withdrawal. Women's demand for higher education as a result of the NATO military drawdown from Afghanistan's districts increased because of cultural restrictions on their mobility. Understanding the effects of military withdrawals on development measures such as female education is important for gauging the potential threats of withdrawal from occupation to vulnerable groups such as women in the host communities. The PEZAK program is an example of a community-driven program without external support. It had strong immediate and long-term effects on student test scores and enrollment. On the other hand, the NSP is an example of a large development program that established gender-balanced village councils with the objective of inclusive development for men and women. Our results do not find strong evidence to support the sustainability of this program. Our findings also point to important cultural differences that might result in varying absorption capacities. The NSP improved gender outcomes in areas with favorable attitudes towards women while it worsened in culturally conservative districts. Each chapter in this dissertation contributes to the literature on the economics of education and women empowerment.

Chapter 1. Foreign Military Withdrawal, Male Migration, and Female Education

1.1 Introduction

Foreign military withdrawal, after a long period of occupation, is an economic and political shock to the occupied nation. For example, Fortuna et al. (2021) show that downsizing the U.S. Military Air Base at Lajes in Portugal had negative effects on local income[1]. Moreover, departure of foreign troops after several years of occupation also affects security and political stability of communities in the host nations (Assi, 2016). Evacuation of foreign military establishments could have broader consequences due to termination of economic activities associated with foreign military facilities. While there are some research on the macroeconomic effects of base closures, micro-level studies on impacts of foreign military withdrawals for local communities, especially in fragile states, is scant.

In this chapter, I study the effect of a major foreign military withdrawal on females' demand for higher education in context of the 2011-2015 NATO troop drawdown from Afghanistan's districts. I employ an Instrumental Variable (IV) approach to identify the effects of foreign military evacuation on demand for college enrollment. For this purpose, I utilize logistical bottlenecks imposed by travel distance from the nearest regional military airbase, as an instrument, which provide an exogenous source of variation for military withdrawal planning. Females' demand for university enrollment, especially in conservative societies, is important, because it may translate into actual enrollment.

The outcome of interest is demand for higher education, measured by annual participation per capita in Afghanistan's national university entrance examination. I combine a student exam data accessed from administrative databases in a data scarce context, with other surveys such as the Survey of the Afghan People, ACLED conflict data, and NATO's

[1] Military base closures, in general, have negative effects on employment but limited or no effects on per capita income (Hooker and Knetter, 2001; Andersson et al., 2007; Paloyo et al., 2010; Lee, 2018; Asteris et al., 2018)

quarterly security perception surveys. I construct a panel dataset at the district level that covers a range of variables including university exam participation, income, emigration, pessimism, and district-level violence incidents.

NATO troop drawdown from Afghanistan provides a natural experiment to examine the effect of foreign military withdrawal on human capital decisions. The United States intervened in Afghanistan in 2001 as part of the War on Terror campaign (U.S. Department of State, 2001). Downsizing the U.S. and NATO Military personnel in Afghanistan started in 2011 (NATO, 2010). Withdrawal was preceded by a security transition in which district security responsibilities were transferred to local allies. Between 2011-2015, the number of foreign troops reduced to nearly **10%** of 140,000 combat personnel deployed to Afghanistan.

I exploit geographic and topographic impediments to repatriation of military forces to predict military withdrawal. The pullout of troops was constrained by logistical challenges such as carrying sensitive and heavy military equipment to large military hubs from where they would be shipped outside Afghanistan (Loven, 2013). Following Fetzer et al. (2021), I use the least-cost travel distance between districts and the nearest regional military airbases as an IV for sequential arrangement of military evacuation. Travel cost is the least-cost and the shortest path distance between district centroid and the nearest NATO military headquarters in Afghanistan, which is calculated using Dijkstra's algorithm. Military withdrawal is measured by closure of the last military base at a certain date after the transition announcement. Withdrawal is highly positively correlated with distance from nearest military hub i.e., the "the instrument".

Conditional on district security conditions and district fixed effects that absorbs time-invariant district characteristics, my IV is as good as random. The locations of regional military hubs were determined by their proximity to supply centers outside Afghanistan (NATO, 2006). The Karachi port in Pakistan (the Southern Distribution Network), and contracted airports in Central Asia (the Northern Distribution Network) served as supply entry nodes. The logistic sites inside Afghanistan were constructed in places that were accessible to the supply lines outside of the country. Moreover, withdrawal of troops was not a concern during the construction of these facilities because in the initial years of war, the Taliban resurgence that would lead to the deployment of thousands of troops, was not expected. The United States and NATO did not have a withdrawal schedule. Discussions about the withdrawal became more salient after Osama bin Laden was killed in May 2011. Therefore, the locations of major military sites is fairly unrelated to factors that determine withdrawal and the dependent variables.

The least-cost travel distance from the nearest military hubs is a relevant IV for gradual evacuation of forces. The probability of withdrawal is strongly correlated with the IV. I also examine distance from provincial capitals that serves as economic hubs and distance from national capital where the largest U.S. Military base is located, as potential alternative IVs. There is no correlation between the plausible candidate IVs and the probability of withdrawal.

I address concerns over exclusion restriction in my estimation. I use district fixed effects, year fixed effects, and district-specific linear trends. Time-invariant unobserved characteristics that are correlated with the outcome and maybe correlated with the IV as well, are captured in district fixed effects. District linear trends absorb other district-specific unobserved features in each year. I also include district violence level that might be correlated with distance from military hubs and the outcome.

I find that military withdrawal had a significant positive effect on female participation but did not change male participation in university entrance exam. Female participation increased approximately by 0.3 percentage points from a base value of 0.05% participation per capita. The coefficient for male participation, on the other hand, is positive in magnitude but statistically not significant. This result also holds after controlling for number of violence incidents.

Deterioration of economic opportunities and migration are plausible mechanisms for my findings. Evacuation of foreign military resulted to loss of jobs among locals who were employed by the army in various capacities such as translators, cooks, maintenance, and logistic suppliers. It negatively effected the NGOs operating under the shelter of foreign troops. The Provincial Reconstruction Teams (PRTs), a security program that worked side-by-side with the combat missions as part of the "winning hearts and minds" campaign, also closed as military bases were dismantled or handed over to ANSF (Andrew et al., 2013; Petřík, 2016). Using self-reported income data, I show that evacuation of foreign troops reduced income by 5%. Employing a mediation framework introduced by Dippel et al. (2019), I separate the direct effects of military withdrawal and indirect effects mediated by changes in income. Nearly 88% of the increase in female participation in higher education was driven by changes in district income in the aftermath of foreign military withdrawal.

Shrinking income and unemployment causes distress, displacement, and migration. Migration affects enrollment and explain the difference between male and female participation in the university entrance exam. Using intention to migrate as a measure of migration, I run the 2SLS regressions with migration as an outcome. Withdrawal increased migration tendencies among district residents in general. But male migration was higher than female migration. Female migration was limited because of cultural restrictions on movement of women and hardships in illegal migration routes for women and children. Hence, as a result of the drawdown of foreign troops from Afghanistan's districts, income fell and unemployment increased. Consequently, a large number of men migrated. Because young women lost economic opportunities and did not have alternative options such as migration, they pursued the higher education path.

Furthermore, I examine changes in capacity constraint that could potentially explain my results. In the period 2012-2015, the level of participation in university entrance exam increased substantially. For example, in 2012 participation increased by approximately 60%

compared to 2011. In response to a sharp rise in demand, the Government of Afghanistan expanded enrollment capacity (The Daily Afghanistan-e Ma, 2012; Da Azadi Radio, 2013). Changes in capacity was in response to demand rise not vice verse. Moreover, these changes were applied to the entire country not specific locations. Therefore, shifts in supply of seats cannot explain the differential effects of military withdrawal on male and female demand for higher education.

I also explore male recruitment by insurgent groups as another possible driver of my results. It may be argued that with the departure of foreign troops from Afghanistan's districts, male residents who would otherwise participate in higher education, joined the insurgents. I examine this argument by inspecting violence incidents as measured by the annual number of armed conflict events including shooting, explosions, and suicide attacks. If larger number of youths were recruited among local residents, in the fight between insurgents and NATO-backed local security forces, the insurgent casualties should increase because the ANSF possessed better artillery and air support. I do not find any effect of withdrawal on insurgent fatalities. Even if this was the case for some male members in the district, it does not provide an answer for the increase in female participation.

I examine the robustness of my results in two ways. First, to verify if the results were driven by a particular year, I run separate regressions for each transition year. The effect of withdrawal on female participation is positive in all years but statistically insignificant in the third year, which could be due to smaller sample size. The treatment effect is larger in the second transition year as compared to the first year. Although there are some differences in magnitude, the direction of effect is comparable in all years. The first and second transition years constitute nearly 78% of the sample size.

Second, I employ a difference-in-difference strategy as an alternative identification. It may be argued that areas with greater distance from the nearest military hubs might have had different education participation trajectories before military withdrawal. I use the non-parametric estimation strategy of Callaway and Sant'Anna (2021) and estimate the treatment effects in all possible treatment-control combinations. After matching based on the least-cost distance from the nearest military airbase, the treatment effects are in the same direction as in the IV approach. Moreover, there is no difference in trend between early and late evacuees.

I contribute to three strands of literature. First, my results explain gender differences in human capital accumulation as a result of shocks in conservative settings. Micro-level analyses of this nature in data-scarce and fragile state countries are extremely rare. Some studies have shown that the manpower mobilization in the United States during the World War II increased female participation in labor force but decreased their education attainment (Bound and Turner, 2002; Acemoglu et al., 2004; Jaworski, 2014; Rose, 2018). This indicates a trade-off between employment opportunities and education. With the departure of men to war, more employment opportunities were opened for women that led to more women quitting school. I present an opposite scenario. In the aftermath of foreign troops withdrawal a mix of security deterioration and declining economic opportunities led to an exodus of men. Women on the other hand, could not migrate but chose to apply for higher education. My paper demonstrates a unique case in the context of a fragile country during a security transition.

I also contribute to the burgeoning literature on security transitions. Fetzer et al. (2021) show that handover of security responsibilities to the local allies led to a reduction in violence but physical evacuation of foreign troops resulted to a surge in insurgent assaults and armed conflict fatalities. The existing research has mainly focused on political stability and security side of transitions (Cortright, 2015; Rupesinghe, 1989; Sokov, 2005; Gribincea, 2009; Anderson, 1977; Seng, 2012). There are no studies that explore the development effects of withdrawals. I provide an empirical evidence on the effect of security transitions and foreign military disengagements on development measures such as female education.

Finally, my paper is also relevant to the brain drain literature. Beine et al. (2008) show that prospects of skilled migration has a positive effect on gross human capital formation. But the direction of the effect reverses if migration of skilled professionals is above 20% or the proportion of the tertiary-educated population is below 5% of the population. In my study, military withdrawal increased female participation in higher education, which counts as a brain gain. But it did not change male participation despite changing economic conditions for both men and women. Instead, male migration increased, which is an indication of male brain drain.

My findings have an important policy implication. Drawdown of troops to the extent that the political order and establishments built under foreign occupation, are maintained, can sustain the gains of occupation such as female education in the host communities. The NATO troop drawdown in Afghanistan by approximately 90% transferred the entire security responsibilities to local forces. However, NATO's continued presence with significantly fewer troops in regional bases towards training and advice to local security forces helped maintain trust in the sustainability of the government. Despite some surge in violence after the drawdown, social progress such as female education continued at a higher pace. Male migration also faded away as more restrictions were imposed by European countries and in response the male participation in higher education also picked up after a period of decline.

A complete withdrawal that does not ensure continuity of the political order can lead to further fragility. In February 2020 the U.S. signed a deal with the Taliban based on which the U.S. troops would withdraw completely and unconditionally. This arrangement weakened the position of Islamic Republic of Afghanistan against the Taliban which eventually led to the collapse of the government that was established under NATO's presence, to an extremist regime that imposes severe restrictions on female education (Special Inspector General for Afghanistan Reconstruction, 2022). Since seizing power in 2021, the Taliban has restored revere restrictions on female education and presence of minorities in political decision-making.

The rest of this chapter is organized as follows. Section 1.2 provides an overview of the research context. Section 1.3 introduces my empirical strategy. Section 1.4 describes the data sources. Section 1.5 discusses my results. Section 1.6 illustrates mechanisms for my findings. Section 1.7 concludes the chapter.

1.2 Background

1.2.1 The U.S. Invasion in Afghanistan

Since 1978, Afghanistan has experienced a long period of political instability that has manifested in the forms of coup d'etat, foreign invasion, civil war, terrorism, and insurgency (Chaffetz, 1980). Table 1.1 summarizes major events in Afghanistan since the Soviet invasion.

Table 1.1: Key events in Afghanistan's conflict

1979	The Soviet invasion
1989	The Soviet withdrawal
1990-1996	Civil wars
1996-2001	The Taliban regime
9/11/2001	The Al Qaeda attacks
Nov 2001	U.S intervention
2011-2014	Security transition from NATO to local allies
Feb 2020	The US-Taliban Peace Accord
Aug 2021	The Taliban takeover

The radical reform agenda undertaken by the communist People's Democratic Party faced oppositions. As a result, the Soviet Union invaded Afghanistan in 1979 in support of the communist party. The West supported Afghanistan resistance (Jihad) against the Soviet. Several anti-Soviet Jihadi fractions were formed, trained, and armed in Pakistan among the refugees that had escaped war and were deployed to fight against the Russians. Eventually, the Soviet troops withdrew from Afghanistan in 1989.

The conflict in Afghanistan entered another phase after the withdrawal of the USSR troops. A vicious civil war on power broke out among the Jihadi fractions (Khalilzad, 1995; Rubin, 1989). Between 1989 and 1996, hundreds of thousands of Afghans were killed, wounded, and displaced. Millions sought refuge in the neighboring countries, mainly Pakistan and Iran. The Taliban, a radical ethno-religious group emerged in 1996 (Rashid, 1999). The Taliban launched extreme restrictions on civilian lives such as banning arts and recreational activities, and depriving women from education (Azzi, 1999). Eventually, the Taliban regime who harbored the leader of Al Qaeda, Usama bin Laden, collapsed with the U.S. intervention in Afghanistan as part of the U.S. War on Terror campaign.

After the fall of the Taliban, a new era started in Afghanistan. Under the United Nations mandate, the Afghan Interim Authority (AIA), a coalition of previously war fractions, was formed (Ponzio, 2007). In addition to the few thousands of U.S. military, an International Security Assistance Force (ISAF) mandated by the United Nations was deployed to Afghanistan to support state-building. In the following years, Afghanistan enacted a democratic constitution. In January 2004 the country had its first-ever presidential elections.

In the early years after the Taliban, Afghanistan experienced a period of peace and flow of international aid. Between 2001-2020, the United States alone allocated approximately $143 billion for the reconstruction and security of Afghanistan (Congressional Research Service, 2021). As a result, several development programs were initiated and millions of boys and girls enrolled in school. Three million refugees returned from the neighboring countries. Based on a UNHCR report, by December 2002 approximately 1.8 million refugees returned to Afghanistan (UNHCR, 2021).

In response to the Taliban insurgency, in 2009, the Obama administration increased the number of troops in Afghanistan. As presented in Figure 1.2,the U.S. troops level in Afghanistan reached 100,000 soldiers in 2011. It was during this period that the U.S. endured the largest number of casualties in the war against terrorism in Afghanistan. By the end of 2012, the U.S. had lost 2,169 soldiers most of which were after the new deployment (Brookings, 2015).

The War on Terror campaign concluded in 2011. On May 2, 2011, the U.S. military killed Osama bin Laden in Abbottabad, Pakistan (BBC, 2011). In early 2011 NATO and the Government of Afghanistan decided to gradually transfer the security responsibilities to Afghanistan National Security Forces (ANSF) while continuing training and financial support. President Obama announced his plan to end the U.S. and NATO's combat mission by 2014 (The Wall Street Journal, 2014).

1.2.2 NATO Military Withdrawal

Withdrawal of NATO and ISAF troops started after Bin Laden was killed. The Government of Afghanistan and NATO agreed on a gradual timetable for transfer of security responsibilities from foreign troops to Afghanistan National Security Forces (ANSF). This process is known as the transition process (Inteqal in Farsi). The Joint Afghan-NATO Inteqal Board (JANIB) developed a plan in which districts would be handed over to ANSF in five tranches between 2011 and 2014. A formal ceremony for the conclusion of transition was conducted in December 2014 (NATO, 2010).

The security transition and military withdrawal are two different processes but interconnected. The transition is a formal process for handover of security responsibilities to local allies. On the other hand, military withdrawal which followed the transition, refers to physical evacuation of foreign military personnel and equipment. Military withdrawal started after transition tranche announcement but in most cases it was not complete at the conclusion of security handover. Withdrawal time relative to tranche announcement in months is presented in Panel A of Figure 1.3. The average time gap between transition announcement and physical departure of troops was 13 months.

Variation in time of military withdrawal relative to transition announcement is due to logistical challenges for shipment of military equipment and personnel. The topographic characteristics of Afghanistan and poor road network in that country on one hand, and enormous amount of sensitive heavy military equipment stretched across Afghanistan, on the other hand, posed serious challenges to evacuation efforts. Sequencing of base evacuation became more challenging when Pakistan closed two main land logistic passes to NATO in 2011 (Reuters, 2011). Panel B of Figure 1.3 plots the least-cost travel distance from regional hubs. There is an inverse relationship between withdrawal time relative to security handover and the least-cost travel distance between district centroid and nearest military airbase. Distance between small district-level military posts and regional hubs created a variation in closure date relative to the transition. Remote bases were evacuated earlier.

One concern in determining actual withdrawal was the level of violence. Panel (a) in Figure 1.4 demonstrates transition tranches and Panel (b) plots per capita annual fatalities from violence in 2010, the year before transition. Areas with greater level of violence at the planning stage seem to be assigned to later transition tranches. This is expected because the level of security and the ability of the ANSF to maintain security were the two main conditions for handover of security responsibilities. However, within each tranche, it was the logistical barriers that determined withdrawal.

Remoteness created logistical challenges for sequencing of transition and troop withdrawal. Foreign troops operated in various capacities such as training camps, major outposts and security check-points. As the transition started, bases were handed over to the ANSF. To minimize the risk of military facilities falling in the hand of Taliban insurgents, some bases were completely demolished and their sensitive equipment were shipped to larger military bases where NATO continued to train the ANSF under the new Resolute Support Mission (RSM) after the combat mission was over. As stated above, the existing level of violence played some role in planning of transition phases but not so much in withdrawal decisions.

Remoteness mainly depends on geographic location and topographic condition of districts. In the case of Afghanistan's districts, there is a positive correlation between remoteness from logistic hubs and elevation. This suggests that distance from the nearest military logistical hubs created an exogenous variation in the timing of base closure and physical withdrawal. I utilize this source of variation as a natural experiment to predict the probability of withdrawal conditional on transition tranche announcement. I use closure of the last base documented in Fetzer et al. (2021) as a proxy for complete physical withdrawal from districts.

The withdrawal of ISAF and NATO troops from districts had several consequences. The Taliban insurgency increased, which in turn led to an increase in civilian fatalities. Figure 1.10 plots conflict-induced fatalities for Afghanistan, Iraq, and Syria. As we can see, between 2006-2013 conflict-induced fatalities remained below 10,000 annually but started to increase steadily since 2014. Fetzer et al. (2021) show that escalation of violence was a directly caused by foreign military withdrawal.

Reduction in the number of foreign troops was also accompanied by a reduction in foreign aid and spending in Afghanistan (see Figure 1.11 and Figure 1.12). An aid-dependent country, the economy of Afghanistan also started to shrink from average annual growth of 9% to 1.8% in 2015. The real GDP trend for Afghanistan is presented in Figure 1.13. Along with military withdrawal, the Provincial Reconstruction Teams (PRTs) that were conceived based on the theory of "hearts and minds" to gain support of locals, closed down. The PRTs implemented development projects such as construction of administrative buildings, schools and health facilities. Several NGOs that were either implementing partners with the PRTs or conducted other donor-funded humanitarian and development projects would also close down due to draw-down in funding and potential security risks. Consequently, local people that were recruited in those programs lost jobs.

1.2.3 Education Under the U.S. Military Presence

After collapse of the Taliban in 2001, the Government of Afghanistan and donor countries placed significant attention on education. Expenditure on education constitutes approximately 4% of the country's GDP and 15.7% of the government budget. In this period, Afghanistan made remarkable progress in the education sector. In 2018, literacy rate among individuals aged 15 years or more and 15-24 years was 43% and 65.4%, respectively. Similarly, gross enrollment rate in primary, secondary, and tertiary education was 100%, 55.3%, and 9.7%, respectively. This was a substantial improvement over 21% male enrollment and zero female primary enrollment in 2001.

However, despite the progress, gender gap remains a challenge. In 2018, the difference between male and female gross enrollment in primary, secondary, and tertiary education was 41.38, 30, and 9.3 percentage points, respectively. Literacy rate among males aged 15 years or more was 29.7 percentage points higher than females in the same age category. Effective transition rate from primary to lower secondary education was 94% for males and 85.5% among female students. School life expectancy (1-8 years) was 12.56% for males and 7.72% for females (UNESCO, 2018).

Public education up to six years undergraduate program was free in Afghanistan. Free education includes tuition, textbooks, and stationery. Students enrolled in universities different from their province of residence also receive free dormitory and food or equivalent monetary compensation. Higher education includes four-year degree programs, vocational institutes, higher education institutions for religious education, and two-years higher education programs.

University seats are allocated based on test scores in university entrance examination. Students take the test after graduation from high school. Students select their major and preferred higher education institutions on the test. Available seats for each university-major pair are known to the students, but the minimum requirement is determined by competition. Seats are allocated to the students with the highest scores that selected the institution-major pair on the test. Usually, universities in Kabul, the capital, and major cities admit students with the highest test scores. Among university majors, medicine and engineering programs have the highest minimum test score for admission.

Social support for female education and employment in the post-Taliban period improved over time. But Afghan society has strict discretion over the type of occupation for women. In 2016, 74% of Afghans agreed for women to work outside home. Surveys show support for female employment in education, health, and other services sectors such as government, NGOs, and private sector (The Asia Foundation, 2016). In addition to other factors, social preference over the type of occupation for women can affect the type of university majors selected by female students (Kahn and Ginther, 2017).

Military withdrawal could affect education in several ways. First, between 2001-2012 over 140,000 troops from 43 countries were present in Afghanistan. Thousands of Afghans, primarily men, worked as translators and local facilitators. Similarly, several thousands others, including men and women, worked with the PRTs and development organizations. Withdrawal implies that most locals employed with the military and their derivative organizations would be unemployed. Those without a higher education but still in college age may adopt alternative paths such as applying for higher education. Second, the escalation of violence in the aftermath of military withdrawal could hamper education, particularly for women. Such implications could produce different trajectories for male and female education.

The initiation of withdrawal process seems to have affected demand for higher education. A description of total exam participation by transition tranche is presented in Figure 1.14.

As we can see, the average total participation of the first tranche sharply increased in 2011 after the announcement of the transition. Similarly, the average of districts in the second tranche experienced a similar increase in 2012. Tranches 3-5 had slight increase in participation in 2013. However, what is common among all tranches is that participation fell considerably in 2015 and beyond. The transition was concluded in 2013 and physical withdrawal was completed in 2015. I study whether the difference in participation is due to military withdrawal.

1.3 Empirical Strategy

The goal of this chapter is to estimate the effect of foreign military withdrawal on demand for higher education. I use participation in university entrance examination as a measure of demand for higher education. Participation is the number of high school graduates that took Afghanistan's annual national university entry exam which serves as enrollment mechanism to public university programs. The outcome is also segregated by gender. Furthermore, I divide participation by 2016 district population estimates and therefore construct a measure of participation for every 100 people.

To estimate the causal effect of military withdrawal on demand for higher education, I employ an Instrumental Variable (IV) approach. I use the least-cost travel distance from the nearest regional military airbases that created exogenous variation for sequencing of withdrawal as an instrumental variable (see Section 1.2.2). Regional military airbases served as logistical hubs for shipment of equipment and heavy armories stretched across Afghanistan's approximately 400 districts. Throughout the Operation Enduring Freedom, heavy artillery were airlifted to Afghanistan from the U.S. military bases in Qatar and Kuwait mainly through Pakistan in the South and East, and Central Asian states in the North. Supplies were first deployed to the regional hubs within Afghanistan and then distributed to district-level camps, forward operating bases, and combat outposts.

Between 2001-2011 ten regional military facilities were constructed in Afghanistan. The locations of these mega military infrastructure were initially decided based on their proximity to the supply lines outside Afghanistan. In addition to logistical issues, balancing Afghanistan's regional geopolitics and intelligence were also involved in determining these sites. However, withdrawal of troops was not a consideration in choosing the location of these facilities because in the initial years of war when the bases were constructed, an insurgency that would lead to the deployment of thousands of foreign troops to Afghanistan's remote districts was not expected. Furthermore, risk of equipment falling in the hands of enemies was also not predicted since 42 countries joined hands with a spirit of winning the War on Terror.

Regional airbases became more prominent during withdrawal planning. In the aftermath of Salala attack in November 2011 that a U.S. military operation in border area between Afghanistan and Pakistan killed 25 Pakistani military personnel, Pakistan blocked Khyber Pass and Chaman Pass (Reuters, 2011). The two gates were NATO's main land supply lines that connected Kabul, Afghanistan's capital, and Kandahar, another major city in Afghanistan, to Pakistan's Karachi port on the Arabian Sea. With closure of those crucial gateways, regional airbases were the only options for evacuation of facilities. Therefore, travel distance from nearest hub which I use as an IV to predict withdrawal, is as good as random.

The IV was effective only after transition process started. To this end, I interact the least-cost travel distance from the nearest military hub with ***Post*** that switches on to 1 in 2011. The transition served as an official inauguration of physical withdrawal. Hence, my instrument is an interaction of distance and the binary variable ***Post***. Therefore, in the first stage, I estimate the following linear probability specification:

$$Withdrawal_{i,t} = \beta_0 + \lambda Transition_{i,t} + \gamma(Distance_i \times Post_t) + \alpha_i + \eta_t + \alpha_i \times t + \theta_j X_{i,t} + \varepsilon_{i,t} \quad (1.1)$$

where, $Withdrawal_{i,t}$ is a binary outcome variable for military withdrawal that is equal to 1 when the last military base was closed in district i at time t. $Transition_{i,t}$ is a binary variable that is equal to 1 when the district security was transferred to local allies. There were five transition tranches that are on average six months apart. Since I am interested in the annual participation in university entrance exam, tranches that fall in one particular year are grouped together. $Distance_i$ is the least-cost travel path between district i and nearest regional military airbase. $Post_t$ is an indicator that switches to 1 in 2011 when the transition process started. The interaction term $(Distance_i \times Post_t)$ is the instrument of interest that predicts withdrawal once the transition started. α_i and η_t are district fixed effects and year fixed effects, respectively. $\alpha_i \times t$ captures district-specific linear trends. $X_{i,t}$ represents time-varying control variables such as violence incidents. $\varepsilon_{i,t}$ is the error term that captures unobserved time-varying district characteristics.

In the second stage, I use the predicted withdrawal from Equation 1.1 to estimate the effect of withdrawal on demand for higher education. The following equation specifies the second stage regression:

$$Participation_{i,t} = \beta_0 + \lambda Transition_{i,t} + \omega(\widehat{Withdrawal}_{i,t}) + \alpha_i + \eta_t + \alpha_i \times t + \theta_j X_{i,t} + \varepsilon_{i,t} \quad (1.2)$$

where, $Participation_{i,t}$ is university exam participation for district i at year t. Total participation, male participation, and female participation are estimated in separate regressions. $\widehat{Withdrawal}_{i,t}$ is the predicted withdrawal for district i at time t from Equation 1.1. Rest of the variables are the same as in Equation 1.1.

The underlying assumption in my identification strategy is that the least-cost travel path between districts and nearest military airbase affects participation in higher education only through foreign military withdrawal. Some unobserved time-invariant and time-varying district characteristics that may be correlated with remoteness and potentially affect the outcome, are captured in district fixed effects and district-specific linear trends. It is worth mentioning that distance from military hubs does not necessarily imply remoteness from economic hubs. Later, I will show that distance from economic centers does not predict withdrawal. Under such conditions the exclusion restriction holds in my approach.

The least-cost travel distance from nearest military airbase is a valid instrument. Table 1.3 presents results from OLS regressions for Equation 1.1. Conditional on security transition, military withdrawal is highly correlated with the instrument. One standard deviation increase in least-cost travel distance from nearest hub increases the probability of withdrawal by 8%, conditional on transition. Including violence incidents does not change the results (see column 2 in Table 1.3).

I examine plausible candidate IVs. One may expect that distance from the economic hubs is potentially correlated with the placement of military airbases and could be a better predictor of withdrawal. I explore this possibility by using two alternative distance measures in Equation 1.1. First, I use distance from provincial capital as a plausible instrument. Province and districts are layer 2 and 3 administrative divisions in Afghanistan such that several districts form a province. Each province has a provincial capital that serves as economic and administrative center for the province. Table 1.4 provides results for this alternative instrument. It shows no correlation between probability of withdrawal and the new instrument. Another potential candidate could be distance from the national capital, Kabul, because the largest U.S. Military base was located north of Kabul. Table 1.5 presents the results for this substitute instrument. Conditional on security transition, there is no correlation between probability of military departure from the districts and distance from Kabul City during the transition period. Therefore, the least-cost travel distance from nearest military airbases is the only strong predictor of withdrawal.

The interaction term in Equation 1.1 is an essential component of my IV. The interaction term implies that the IV is only effective once transition process started. The role of distance in predicting withdrawal before the transition was zero. However, one may argue that any arbitrary years after the transition could be considered as interaction term. To investigate this, I interact the distance measure with the binary variable *post* such that $post = 1$ for years greater than 2013. I find that the new IV does not predict withdrawal. Hence, $Post > 2010$ is an integral part of the IV and not an arbitrary interaction term.

1.4 Data

I combine several data sources for this study. Student participation in the university entrance examination comes from Afghanistan university entrance examination data. It contains student test scores, university major admitted, student gender, and university location for the universe of students during 2003-2019. I merge this data with an administrative school database that contains information on school characteristics and precise location of schools from which the students graduated. During the period 2003-2019, in total $2,263,962$ students took the university entrance examination of which 70.5% were male and 29.5% were female. 56% of the total test takers were admitted to a public higher education institution of which approximately 60% were admitted in a four-year university degree program and the rest were admitted in two-year higher education and other lower-ranked programs.

I focus on demand for higher education measured by participation in university entrance examination. I aggregate the data at the district level to construct a measure of district-level demand for higher education. Participation is also segregated by gender. Moreover, district exam participation aggregates are divided by district population, fixed at 2016 population estimates.

Information about security transition is accessed from NATO. Their website provides detailed information about transition timing and districts in each tranche. I overlay maps with tranche data on Afghanistan district shape files to merge security handover information with other district data. I extract information about physical withdrawal and travel cost path between districts and nearest military logistic hubs from Fetzer et al. (2021). They have identified 10 regional hubs and calculated least-cost travel distance between districts' centroid and nearest NATO military airbase using Dijkstra's algorithm. I use this distance measure to construct the instrument defined in Section 1.3.

I use additional data sources for auxiliary analysis and examining mechanisms. I use the Asia Foundation's Survey of Afghan People (SAP) and Afghanistan Nationwide Quarterly Assessment Research (ANQAR) from NATO to construct measures of security and perception about the direction of the economy. The SAP is an annual perception survey that is representative at the province level. Using district names in the SAP data and other databases explained above, I merge them at the district level. However, district-level variables in this data is only indicative not completely representative because the survey sampling was not conducted at district level. It provides information about a range of variables such as income, migration, political participation, security, and natural disasters since 2006. ANQAR is a quarterly survey conducted by NATO since 2008 that I accessed through a restricted contract. It supplements the SAP data in range of other variables such is ethnic composition and local conflict. The two surveys also have questions on local perception about the presence of foreign troops. I use this information to measure agreement and disagreement with foreign military withdrawal.

Furthermore, I use Eurostat to present a picture of the status of immigration from

Afghanistan. Figure 1.16 shows trend in asylum applications in the EU. Panel (a) compares the total and first-time applications in the EU between 2008 and 2020. As we can see, with the start of the Syrian conflict in 2011, there is a spike in the total asylum applications. It reaches a peak in 2015. Around the same time, as shown in panel (b) asylum seekers from Afghanistan also surged. Based on Eurostat figures, Afghanistan had the second-largest asylum applicants after Syria. Over 75% of the asylum seekers from Afghanistan were below age 35.

1.5 Results

1.5.1 Main Results

Table 1.3 presents the regression results for Equation 1.1. The first-stage regression shows that conditional on security transition, probability of foreign military withdrawal is positively correlated with the least-cost travel distance from nearest military airbase during the transition period. Table 1.4 and Table 1.5 presents first-stage results for alternative plausible instruments. We do not see any correlation between the alternative instruments and the probability of military evacuation. For detailed discussion on the first stage see Section 1.3.

The reduced form regression results are presented in Table 1.6. In reduced form regressions, the instrument is directly used as a regressor. There are three outcomes in this table: total participation, male participation, and female participation. There is a strong correlation between female participation in the university entrance exam and least-cost travel distance from nearest military airbases but a weak correlation with male participation. As discussed in Section 1.3, the effect of the least-cost travel distance only comes through foreign military withdrawal.

The second-stage results are presented in Table 1.7 and Table 1.8. Table 1.7 provides estimates OLS and 2SLS regression results for total participation in university entrance exam.

In OLS regressions where military withdrawal is directly used as an independent variable, the coefficient on military withdrawal is small in magnitude and statistically in significant. However, in these regressions withdrawal is correlated with the error term and the coefficients on withdrawal do not present unbiased estimates. The true causal effects of withdrawal on total participation in university entrance exam are presented in column (2) and column (4) of Table 1.7. The coefficient on withdrawal in the 2SLS regression presented in column (3) is positive and significant at 5% significance level. Military withdrawal from districts increases per capita total participation in university entrance examination by approximately 0.4 percentage points. Including violence events per capita in column (4) does not change the results.

Heterogeneous effect of military withdrawal on male and female demand for higher education is presented in Table 1.8. Panel A provides results for OLS regressions and Panel B shows results from 2SLS regressions. Results for both OLS and 2SLS regressions for male demand for higher education are small in magnitude and statistically insignificant. OLS regression results for female demand for higher education is small but highly significant. The effect of withdrawal on female participation in higher education in the 2SLS regressions is approximately 0.3 percentage points. These results do not change by including violence incidents in the regressions. Therefore, foreign military evacuation had no effect on male demand for higher education and substantially positive effect on female demand for higher education.

The effect of withdrawal on total district demand for higher education is entirely driven by the increase in female participation in university entrance exam. I find that districts from which foreign military evacuates has 0.4 percentage points more participation in the entrance exam. While the effect of withdrawal on male participation is zero, female participation increased by around 0.3 percentage points. This result is not sensitive to inclusion of violence levels as an additional control.

1.5.2 Robustness

Bellow, I explore two measures to examine the robustness of results discussed in Section 1.5.1. One may argue that treatment effects could be driven by a specific transition tranche. To examine this hypothesis, I estimate Equation 1.1 for each tranches. I employ a difference-in-difference design to investigate pre-trends.

Treatment effects in each transition tranche. — In the main results, I estimate treatment effects conditional on security transition. As discussed in Section 1.3 departure of foreign troops started only after handover of security responsibilities to local forces. The transition was completed in five phases that were announced approximately in six-months intervals. Since I use annual data, I consider transition tranches within a year in one group which gives three transition groups. In the first and second transition year, respectively, 150 and 169 districts were handed over to ANSF. In the final transition year 96 remaining districts were submitted to local forces.

To examine heterogeneity across transition years, I estimate treatment effects in each transition subsample. Results for this robustness check is presented in Table 1.9. All regressions include district fixed effects, year fixed effects, district-specific linear trends, and violence events per capia. Results for OLS regressions are provided in Panel A. As in the main analysis, the coefficient on withdrawal in tranche-by-tranche OLS regressions are small in magnitude and statistically insignificant. Results for 2SLS regressions are presented in Panel B. The effect on male demand for higher education is insignificant in all transition years.

The effect of withdrawal on female participation in university entrance exam is consistent with the main results for majority of districts. In the first and second transition years, withdrawal increased female participation by approximately 0.3 and 0.6 percentage points, respectively. The magnitude of the effect in the third year is approximately 0.2 percentage points but statistically not significant which could be due to smaller sample size. Therefore, while there are some heterogeneity across transition years, the sign of the effect for the majority of districts is in the same direction as in the main results.

Difference-in-Difference identification. — I employ a different identification to examine robustness of results presented in Section 1.5.1. The staggered timing of foreign military evacuation from Afghanistan's districts also provides a natural experiment for an event study design. One approach is two-way fixed effects specification. However, in settings with different treatment times, difference in weights from past treatment can contaminate the treatment effects (Goodman-Bacon, 2021; Athey and Imbens, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; De Chaisemartin and d'Haultfoeuille, 2020; Borusyak et al., 2021). To exploit the sequential timing of military withdrawal in studying treatment effects of troops evacuation in an event study design, I apply the non-parametric solution of Callaway and Sant'Anna (2021). In this method, treatment effect is calculated in all treatment-control combinations separately and then aggregated for each event year. I present the aggregate of all event years in a single coefficient.

Results for non-parametric event study approach are presented in Table 1.11. There are two sets of results in this table. Results with simple difference-in-difference do not account controls and estimates with IPW accounts for difference in least-cost travel distance from nearest military airbases by inverse probability weighing proposed by Abadie (2005). The effect of military withdrawal on female participation in university entrance exam is positive and significant at 5% significance level in both estimates and no effects on male participation. However, the magnitude of the effect is relatively small as compared to instrumental variable approach.

Dynamic effects of military withdrawal on university entrance exam participation are presented in Figure 1.5, Figure 1.6, and Figure 1.7. Panel (a) in these Figures show simple difference-in-difference and Panel (b) presents results with inverse probability weighting.

As we can see, differences in outcome before military withdrawal was zero. The dynamic effect of withdrawal is positive for both male and female. However, comparing confidence intervals around treatment effect, the effect on male participation is less precise which is consistent with significant coefficient on female participation and insignificant estimates for male participation in Table 1.11.

Therefore, all robustness checks point to the same direction. Military withdrawal did not affect male demand for higher education but increased female demand. In the following section I explore potential mechanisms for this differential effect.

1.6 Mechanisms

1.6.1 Empirical evidence

What can explain the differential impact of military withdrawal for men and women? Withdrawal increased female demand for higher education but did not change male demand. I explore several potential channels.

A. Military withdrawal and changes in university seats

Changes in capacity on the supply side could drive changes in demand. Available seats in public higher education institutions are allocated based on students' test scores on the university entrance exam. The number of available seats is not announced before the exam. However, supply of university seats can be measured by actual enrollment. Aggregate participation and enrollment is shown in Figure 1.8. Between 2003-2011 there is a smooth increasing trend for both participation and enrollment. In the period 2012-2015 total demand for higher education measured by participation in the university entrance exam increased substantially for both males and females. For instance, in 2012, the number of test takers increased by approximately 66% compared to 2011. In response to increased demand, Afghanistan's Ministry of Higher Education added the number of seats by establishing five provincial higher education institutions (The Daily Afghanistan-e Ma, 2012). As rise in demand continued in the following years, additional seats were created through opening evening shifts (Da Azadi Radio, 2013). However, the expansion of seats in response to rise in demand happened at the country-level and did not target specific locations. It also did not create different capacity for males and females. Therefore, the increase in participation in the aftermath of withdrawal cannot be explained by changes in the number of seats.

B. Military withdrawal and insurgent recruitment

Another reason for the differential effects of withdrawal on demand for higher education might be male recruitment by insurgent groups. As explained in Section 1.2.3, military withdrawal was accompanied by the closure of several development NGOs that recruited locals. In addition to those that were directly employed with foreign military, others lost jobs because several development programs such as the Provincial Reconstruction Teams (PRTs) stopped. This added to the unemployed labor force. Consequently, unemployed men might have joined the insurgent groups while unemployed women might have chosen to continue higher education. This is a plausible hypothesis because the level of insurgency and violence increased as a result of withdrawal (Fetzer et al., 2021).

To examine this argument, I estimate Equations 1.1 and 1.2 with per capita insurgent fatality as the dependent variable. Since insurgency increased, if recruitment of insurgent groups among employees of foreign military and NGOs who lost their jobs added to their fighters, then insurgent fatalities should also be higher in districts from which foreign military evacuated. As shown in Table 1.12, withdrawal did not change insurgent casualties. There is no evidence that withdrawal contributed to insurgents' recruitment and hence this does not explain the differential gender impact of withdrawal on university participation.

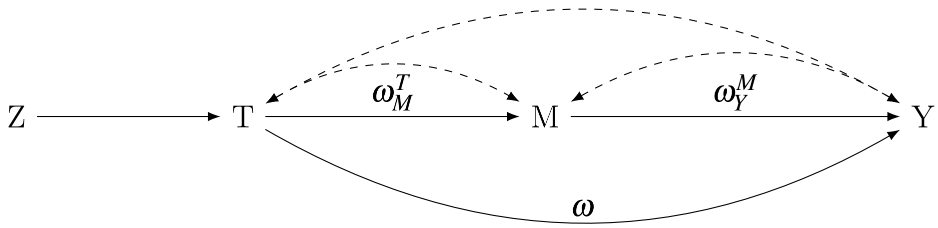
C. Military withdrawal and income

The third potential channel is shrinking local economy and migration. I use self-reported income for the period 2007-2019 from the SAP data to examine the effect of withdrawal on income. Data for years before 2007 is not available. Due to large number of missing data, I include 2016-2019 to construct a sample that provides data points for a larger number of districts for at least one year since the start of withdrawal process. Table 1.13 shows OLS and 2SLS regression results for Equation 1.2 where natural logarithm of income is the outcome of interest. OLS regression results in column (1) shows when withdrawal is directly used as independent variable, the effect of withdrawal is small in magnitude and statistically not significant. On the other hand, 2SLS regression results show that income reduced by approximately 5% as a result of withdrawal.

To examine the mediation effect of changes in income as a result of foreign military withdrawal on demand for higher education, I employ the framework introduced by Dippel et al. (2019) a long with the Stata program in Dippel et al. (2020). In this causal mediation analysis framework, the effect of a treatment T on outcome Y is estimated through an intermediary outcome M using the same instrumental variable Z. In other words, the framework allows to decompose ω in Equation 1.2 to mediation effect via the intermediary outcome M (indirect effect) and direct effect. I illustrate the decomposition process in a directed acyclical graph (DAG). Let ω denote the total effect of withdrawal T on demand for higher education Y. Similarly, let ω_M^T show the effect of withdrawal on district income M. Finally, let ω_Y^M represent the effect of changes in district income on demand for higher education. This chain reaction initiated by the exogenous instrumental variable Z (least-cost distance between districts and nearest military logistic hub) is presented in Figure 1.1.

Following the rational presented in Figure 1.1, we can apply Equation 1.1 and Equation 1.2 to estimate the mediator variable.

Figure 1.1: Mediation effect in DAG



$$Withdrawal_{i,t} = \beta_0 + \lambda Transition_{i,t} + \gamma(Distance_i \times Post_t) + \alpha_i + \eta_t + \alpha_i \times t + \theta_j X_{i,t} + \varepsilon_{i,t} \quad (1.3)$$

$$Logincome_{i,t} = \gamma_0 + \gamma_1 Transition_{i,t} + \omega_M^T(\widehat{Withdrawal}_{i,t}) + \alpha_i + \eta_t + \alpha_i \times t + \theta_j X_{i,t} + \varepsilon_T^M \quad (1.4)$$

where ω_M^T is the causal effect of foreign military withdrawal on log of district income which serves as the mediator for the final outcome. ε_T^M captures the effect of unobserved variables that affect income.

The next step is to estimate the effect of district income (the mediator) on demand for higher education. Dippel et al. (2019) show that conditional on the the treatment variable (military withdrawal) the exclusion restriction holds in a 2SLS setting in which the predicted mediator is used to estimate the outcome of interest. In this case, in the first stage, log of income is predicted as in the following equation.

$$Logincome_{i,t} = \beta_0 + \beta_1 Transition_{i,t} + \omega_M^Z (Distance_i \times Post_t) \\ + \omega_M^T Withdrawal_{i,t} + \alpha_i + \eta_t + \alpha_i \times t + \theta_j X_{i,t} + \varepsilon_T \quad (1.5)$$

where ω_M^Z is the effect of the IV on the mediator, conditional on military withdrawal.

In the second stage, the outcome of interest is estimated using the predicated mediator from Equation 1.5.

$$Participation_{i,t} = \beta_0 + \beta_1 Transition_{i,t} + \omega_Y^M (\widehat{Logincome}_{i,t}) \\ + \omega_Y^T Withdrawal_{i,t} + \alpha_i + \eta_t + \alpha_i \times t + \theta_j X_{i,t} + \varepsilon_Y \quad (1.6)$$

where ω_Y^M is the causal effect of income on demand for higher education. ω_Y^T it is the indirect effect of foreign military withdrawal on demand for higher education. Other parameters are the same as in Equation 1.1 and Equation 1.2.

Combining Equations 1.3-1.6, we can derive the following equation that disentangle the direct and indirect effect of foreign military withdrawal.

$$\underbrace{\hat{\omega}}_{\text{total effect}} = \left(\underbrace{\widehat{\omega_Y^T}}_{\text{direct effect}} + \underbrace{\widehat{\omega_M^T}\widehat{\omega_Y^M}}_{\text{indirect effect}} \right) \quad (1.7)$$

The result of the decomposition framework summarized in Equation 1.7 is presented in Table 1.14. The total effect of foreign military withdrawal on females participation in the university entrance exam is nearly 0.3 percentage points but there are no significant effects on male participation. 88% of the changes in female participation is driven by changes in income as a consequence of foreign military withdrawal. This effect is marginally significant at 90% confidence level.

D. Military withdrawal and migration

I explore the reasons for differential gender effect of changes in income in the aftermath of withdrawal on demand for higher education. One potential factor could be differences in male and female migration. I use intention to migrate as proxy for migration. Intention to migrate is a survey question that asks respondents if they would leave Afghanistan should they get a chance. Table 1.15 presents the result of OLS and 2SLS regression with intention to migrate as outcome of interest. Outcome is segregated by gender. OLS regression results show that the magnitude of effect is positive in magnitude for males and females but coefficient of withdrawal for female migration is not significant. In 2SLS regressions, female migration change both in magnitude and statistical significance. As a result of withdrawal, male migration increased by approximately 11 percentage points. The magnitude on female migration is approximately 6 percentage points but statistically insignificant. Therefore, foreign military withdrawal increased male migration more than females.

To sum up, NATO military evacuation from Afghanistan's districts increased females' demand for higher education. There were no significant effects in males' demand for higher education. In the aftermath of the NATO military withdrawal, both men and women lost jobs. While men migrated out of Afghanistan, women, because of cultural restrictions of on their mobility and hardships in the illegal migration routes for women and children, adopted the higher education path.

1.6.2 Anecdotal evidence on migration

Increase in security risks and draining of economic opportunities following the withdrawal of foreign troops in Afghanistan also intersected with the E.U. refugee crisis. In 2015 nearly 180,000 asylum seekers from Afghanistan reached Europe, which continued to increase in

2016 (see Figure 1.16). Followed by Syria, Afghanistan ranked the second largest contributor of asylum seekers in the E.U. (BBC, 2016). Due to the exodus of young men, the government of Afghanistan launched a campaign to discourage young Afghans from emigrating.

Migration from Afghanistan started to decrease when Europe imposed restrictions on the intake of asylum seekers from Afghanistan (see Figure 1.16). In an attempt to discourage new arrivals, at the Brussels Conference on Afghanistan in October 2016, the Government of Afghanistan and the E.U. signed an agreement about returning Afghan refugees whose asylum cases were rejected by the E.U. member states (European Council, 2018). Despite objections from civil society and human rights organizations (ECRE, 2020), the government of Afghanistan agreed with the deportation. In return, the E.U. committed financial assistance to address the cause of migration (The New York Times, 2016). In the aftermath of the agreement, the first group of refused Afghan asylum seekers was returned to Afghanistan in December 2016 (Gossman, 2017). The number of migrants decreased subsequent to these restrictions. It should be noted that these numbers reflect people who already entered Europe. Information on number of people that left Afghanistan is not available. Therefore, the actual state of exodus could be several times higher.

A majority of the asylum seekers from Afghanistan were male individuals below age 35 (Eurostat, 2021). Between 2008-2020, 75% of Afghan asylum applicants in 28 E.U countries were male. Approximately 90% of total applicants during this period were below age 35. Of the total registered applicants, 41% were below age 18. Consisting 39% of total unaccompanied minors, between 2008-2015, Afghanistan was ranked the largest contributor of asylum seekers below age 18 (Connor and Krogstad, 2016). Hence, a significant portion of Afghan refugees during the 2014-2015 refugee crisis was potentially high school students that would otherwise take the university entrance examination.

One reason for the large share of young migrants from Afghanistan is young demographics. The median age in Afghanistan is 18.4 years. Another reason for the migration of young people from Afghanistan could be deteriorating economic opportunities. A survey of the Afghan people shows that in 2015, approximately 40% of Afghans stated they would leave the country if they had a chance. This willingness to migrate increased by 6 percentage points over the previous year. Respondents that showed higher willingness to migrate were more likely men, young, educated, and lived in relatively peaceful and stable areas of the Northern, Central, and urban areas of the country. Afghans cited insecurity and unemployment as the major push factors for their decision to migrate (The Asia Foundation, 2015, 2016; Akseer et al., 2017).

Drawdown of foreign troops during 2011-2017 was the main contributor of large scale migration. The emigration of Afghan citizens to Europe in the magnitudes of the E.U. migration crisis is unprecedented. Afghanistan has been in conflict for four decades and it has the largest number of refugees. But before 2014 the majority of Afghan refugees were in neighboring countries such as Iran and Pakistan. Moreover, previously, refugees were mainly families and labor migrants. The 2011-2017 migration crisis is different because majority are young, male, and single. Europe as the destination is also the distinguishing feature. The anecdotal evidence about migration supports my empirical findings that explains the differential impact of withdrawal on male and female demand for higher education.

1.7 Conclusion

In this chapter, I show that foreign military withdrawal from Afghanistan's districts increased female demand for higher education but did not change male demand. Female participation in university entrance exam increased approximately by 0.3 percentage points.

I explore several mechanisms for this finding. First, I show that changes in enrollment capacity do not explain my results. Second, I examine the role of recruitment of young men by insurgent groups as another potential channel. To do so, I look at insurgent fatality. An increase in insurgent recruitment would also lead to increased fatality in an environment of intensified violence. OLS and 2SLS regressions do not support this hypothesis. I also investigate drought of economic opportunities and migration as another channel. Foreign military evacuation reduced income by approximately 5 percentage points and increased intention to migrate among men by approximately 10 percentage points. Therefore, subsequent to the departure of NATO troops local economy shrunk and as a result unemployed young men migrated. Women on the other hand, could not migrate due to hardship in illegal migration routes and cultural restriction of mobility of women. Instead, young women chose to apply for higher education.

My findings is consistent with the stylized facts. The coincidence of withdrawal timing in Afghanistan with the Syrian conflict exacerbated the outflow of young men from Afghanistan. In response to the humanitarian crisis in Syria, some of the European countries, particularly Germany, adopted an easy open-door policy towards refugees for few years. The threatening shadow of instability and deteriorating economy as a consequence of withdrawal, motivated the young Afghans to head to the E.U. borders in massive numbers. Profile of Afghan refugees in the E.U shows that majority of Afghan refugees in the E.U. were young men who might have otherwise participated in the university entrance examination and continued higher education in Afghanistan.

My findings suggest that foreign intervention creates economic dependencies. After several years of presence, expatriation of troops can have negative consequences on the host population. In the aftermath of the NATO troops downsizing in Afghanistan there was a massive youth flight. Moreover, my findings have a policy implication for the programming of military interventions. An intervention creates winners and losers. Winners of the intervention are the most affected by the withdrawal decisions. One of the consequences for the winners is the immediate reaction of the young population and their decision to immigrate. This leads to depleting young talents. A phased-out and well-planned withdrawal that considers such feedback may be intertwined with the withdrawal programming.

1.8 Bibliography

Abadie, Alberto, "Semiparametric difference-in-differences estimators," The Review of Economic Studies, 2005, 72 (1), 1–19.

Acemoglu, Daron, David H Autor, and David Lyle, "Women, war, and wages: The effect of female labor supply on the wage structure at midcentury," Journal of political Economy, 2004, 112 (3), 497–551.

Akseer, Tabasum, Mohammad Shoaib Haidary, Rebecca Miller, Sayed Masood Sadat, Christina Satkowski, Helen Seese, Mohammad Jawad Shahabi, Kris Veenstra, Zachary Warren, and Fahim Ahmad Yousufzai, "Afghanistan In 2017: A Survey of the Afghan People," 2017.

Anderson, Ben, "Withdrawal symptoms: Social and cultural aspects of the October 6 coup," Bulletin of Concerned Asian Scholars, 1977, 9 (3), 13–30.

Andersson, Linda, Johan Lundberg, and Magnus Sjöström, "Regional effects of military base closures: the case of Sweden," Defence and Peace Economics, 2007, 18 (1), 87–97.

Andrew, Beath, Fotini Christia, and Ruben Enikolopov, "Winning Hearts and Minds through Development: Evidence from a Field Experiment in Afghanistan," World Bank Policy Research Working Paper, 2013, 6129.

Assi, Abbas, Democracy in Lebanon: Political Parties and the Struggle for Power since Syrian Withdrawal, Bloomsbury Publishing, 2016.

Asteris, Michael, David Clark, and Shabbar Jaffry, "The economic effect of military facility contraction: a naval case study," Defence and peace economics, 2018, 29 (3), 268–293.

Athey, Susan and Guido W Imbens, "Design-based analysis in difference-in-differences settings with staggered adoption," Journal of Econometrics, 2021.

Azzi, Pierre, "Harsh Rule: Recognizing the Taliban," Harvard International Review, 1999, 21 (2), 13.

BBC, "Osama Bin Laden killed in top secret operation," 2011. data retrieved from https://www.bbc.com/news/av/world-south-asia-13262963.

_ , "Migrant crisis: Migration to Europe explained in seven charts," 2016. data retrieved from https://www.bbc.com/news/world-europe-34131911.

Beine, Michel, Frederic Docquier, and Hillel Rapoport, "Brain drain and human capital formation in developing countries: winners and losers," The Economic Journal, 2008, 118 (528), 631–652.

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess, "Revisiting event study designs: Robust and efficient estimation," arXiv preprint arXiv:2108.12419, 2021.

Bound, John and Sarah Turner, "Going to war and going to college: Did World War II and the GI Bill increase educational attainment for returning veterans?," Journal of labor economics, 2002, 20 (4), 784–815.

Brookings, "Afghanistan Index," 2015. data retrieved from https://www.brookings.edu/wp-content/uploads/2016/07/index20150210.pdf.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," Journal of Econometrics, 2020.

_ and _ , "Difference-in-differences with multiple time periods," Journal of Econometrics, 2021, 225 (2), 200–230.

Chaffetz, David, "Afghanistan in Turmoil," International Affairs (Royal Institute of International Affairs 1944-), 1980, 56 (1), 15–36.

Chaisemartin, Clément De and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," American Economic Review, 2020, 110 (9), 2964–96.

Congressional Research Service, "Afghanistan: Background and U.S. Policy," 2021. data retrieved from https://fas.org/sgp/crs/row/R45122.pdfn.

Connor, Philip and Jens Manuel Krogstad, "Europe sees rise in unaccompanied minors seeking asylum, with almost half from Afghanistan," 2016. data retrieved from https://www.pewresearch.org/fact-tank/2016/05/10/eu-unaccompanied-minors/.

Cortright, David, Ending Obama's War: Responsible Military Withdrawal from Afghanistan, Routledge, 2015.

Da Azadi Radio, "MOHE has increased enrollment capacity," 2013. data retrieved from http://www.dailyafghanistan.com/national_detail.php?post_id=126380.

Dippel, Christian, Andreas Ferrara, and Stephan Heblich, "Causal mediation analysis in instrumental-variables regressions," The Stata Journal, 2020, 20 (3), 613–626.

_ , Robert Gold, Stephan Heblich, and Rodrigo Pinto, "Mediation analysis in IV settings with a single instrument," Technical Report, Mimeo 2019.

ECRE, "Joint Statement: Afghanistan is Not Safe: the Joint Way Forward Means Two Steps Back," 2020. data retrieved from https://www.ecre.org/joint-statement-afghanistan-is-not-safe-the-joint-way-forward-means-two-steps-back/.

European Council, “Brussels Conference on Afghanistan, 4-5 October 2016,” 2018. data retrieved from https://www.consilium.europa.eu/en/meetings/international-summit/2016/10/04-05/.

Fetzer, Thiemo, Pedro C. L. Souza, Oliver Vanden Eynde, and Austin L. Wright, “Security Transitions,” American Economic Review, July 2021, 111 (7), 2275–2308.

Fortuna, Mário JA, João CA Teixeira, and Francisco JF Silva, “Gone with the Winds of Peace: The Regional Economic Effects of Military Base Downsizings and Closures,” Defence and Peace Economics, 2021, pp. 1–24.

Goodman-Bacon, Andrew, “Difference-in-differences with variation in treatment timing,” Journal of Econometrics, 2021.

Gossman, Patricia, “Why the European Union Shouldn’t Deport Afghans,” 2017. data retrieved from https://www.hrw.org/news/2017/01/24/why-european-union-shouldnt-deport-afghans.

Gribincea, Mihai, ““Withdrawal of the Russian Forces from Moldova–The Key to Revitalizing the CFE Treaty,” The Future of Conventional Arms Control in Europe, 2009, pp. 292–302.

Hooker, Mark A and Michael M Knetter, “Measuring the economic effects of military base closures,” Economic Inquiry, 2001, 39 (4), 583–598.

Jaworski, Taylor, ““You’re in the Army Now:” The Impact of World War II on Women’s Education, Work, and Family,” The Journal of Economic History, 2014, 74 (1), 169–195.

Kahn, Shulamit and Donna Ginther, “Women and STEM,” Technical Report, National Bureau of Economic Research 2017.

Khalilzad, Zalmay, “Afghanistan in 1994: Civil war and disintegration,” Asian Survey, 1995, 35 (2), 147–152.

Lee, Jim, “The regional economic effects of military base realignments and closures,” Defence and Peace Economics, 2018, 29 (3), 294–311.

Loven, FV, “The Logistical Challenges Confronting the Afghanistan Drawdown,” Technical Report, Technical report, Civil-Military Fusion Centre 2013.

NATO, “Logistic suppor for NATO operations,” Technical Report 2006. data retrieved from https://www.nato.int/nato_static/assets/pdf/pdf_publications/20120116_logistics-e.pdf.

_ , “Transition: Inteqal,” 2010. data retrieved from https://www.nato.int/cps/en/natohq/topics_87183.htm.

Paloyo, Alfredo R, Colin Vance, and Matthias Vorell, "The regional economic effects of military base realignments and closures in Germany," Defence and Peace Economics, 2010, 21 (5-6), 567–579.

Petřík, Jaroslav, "Provincial reconstruction teams in Afghanistan: Securitizing aid through developmentalizing the military," in "The securitization of foreign aid," Springer, 2016, pp. 163–187.

Ponzio, Richard J, "Transforming political authority: UN democratic peacebuilding in Afghanistan," Global Governance: A Review of Multilateralism and International Organizations, 2007, 13 (2), 255–275.

Rashid, Ahmed, "The Taliban: exporting extremism," Foreign Affairs, 1999, pp. 22–35.

Reuters, "Pakistan stops NATO supplies after deadly raid," 2011. data retrieved from https://www.reuters.com/article/us-pakistan-nato/pakistan-stops-nato-supplies-after-deadly-raid-idUSTRE7AP03S20111126.

Rose, Evan K, "The rise and fall of female labor force participation during World War II in the United States," The Journal of Economic History, 2018, 78 (3), 673–711.

Rubin, Barnett R, "The fragmentation of Afghanistan," Foreign Affairs, 1989, 68 (5), 150–168.

Rupesinghe, Kumar, "Building Peace After Military Withdrawal," Bulletin of Peace Proposals, 1989, 20 (3), 243–251.

Seng, Loh Kah, "10. The British Military Withdrawal from Singapore and the Anatomy of a Catalyst," in "Singapore in global history," Amsterdam University Press, 2012, pp. 195–214.

Sokov, Nikolai, "The Withdrawal of Russian Military Bases from Georgia: Not Solving Anything," PONARS Policy Memo, 2005, 363, 3.

Special Inspector General for Afghanistan Reconstruction, "Collapse of the Afghan National Defense and Security Forces: An Assessment of the Factors That Led to Its Demise," Technical Report 2022. data retrieved from https://www.sigar.mil/pdf/evaluations/SIGAR-22-22-IP.pdf.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," Journal of Econometrics, 2021, 225 (2), 175–199.

The Asia Foundation, "Afghanistan in 2015: A Survey of the Afghan People," 2015. data retrieved from https://asiafoundation.org/resources/pdfs/Afghanistanin2015.pdf.

_ , "Afghanistan in 2016: A Survey of the Afghan People," 2016. data retrieved from https://asiafoundation.org/wp-content/uploads/2016/12/2016_Survey-of-the-Afghan-People_full-survey.Apr2017.pdf.

The Daily Afghanistan-e Ma, "In 2012, 80 thousand applicants shall be enrolled," 2012. data retrieved from http://www.dailyafghanistan.com/national_detail.php?post_id=126380.

The New York Times, "Europe Makes Deal to Send Afghans Home, Where War Awaits Them," 2016. data retrieved from https://www.nytimes.com/2016/10/06/world/asia/afghanistan-eu-refugees-migrants.html.

The Wall Street Journal, "Obama Details Plan for Forces in Afghanistan," 2014. data retrieved from https://www.wsj.com/articles/obama-details-plan-for-forces-in-afghanistan-1401234489.

UNESCO, "Afghanistan in 2016: A Survey of the Afghan People," 2018. data retrieved from http://uis.unesco.org/country/AF.

UNHCR, "Afghanistan Situation," 2021. data retrieved from https://data2.unhcr.org/en/situations/afghanistan.

U.S. Department of State, "The Global War on Terrorism: The First 100 Days," 2001. data retrieved from https://2001-2009.state.gov/s/ct/rls/wh/6947.htm.

1.9 Tables

Table 1.2: The instrument and district characteristics

	Travel distance from logistic hub	
	2003	2006
	(1)	(2)
Per capita university participation	0.011	0.021
	(0.055)	(0.051)
Per capita fatality from violence	-0.081	-0.174*
	(0.276)	(0.102)
Elevation	0.315***	0.310***
	(0.047)	(0.047)
Night light luminosity	-0.219***	-0.222***
	(0.045)	(0.045)
District population	0.047*	0.048*
	(0.026)	(0.027)
Distance from Kabul City	0.548***	0.559***
	(0.062)	(0.063)
Distance from commercial airports	-0.284***	-0.272***
	(0.046)	(0.045)
Large city	-1.854***	-1.883***
	(0.257)	(0.266)
Number of Obs	391	391

Notes: This table presents results from OLS regressions. The outcome variable is least-cost travel distance from nearest military airbase "the instrument". All variables are standardized. Standard errors are heteroscedasticity robust. */**/*** denotes significance at the 10/5/1 percent.

Table 1.3: First stage regressions

	Military Withdrawal	
	(1)	(2)
Travel distance to military airport × Post 2010	0.081***	0.081***
	(0.010)	(0.010)
Security Transition	-0.003	-0.002
	(0.018)	(0.018)
Number of violence incidents		0.001
		(0.001)
F-statistic	32	25
Obs (district-year)	5083	5083
Number of Districts	391	391

Notes: This table presents the results of first stage regressions. It examines the correlation between military withdrawal ("the endogenous variable") and least-cost travel distance between districts and the nearest regional military airbase "the instrument" once security transition to local allies started ("Post2010"). The outcome of interest is military withdrawal from districts. It is a binary variable that turns on when the last military base was closed. District panel covers the period 2003-2015. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific linear trends. Column (2) includes violence incidents (ambushes, shooting, suicide attacks) divided by district population fixed at 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.4: First stage regressions with alternative instrument

	Military Withdrawal	
	(1)	(2)
Distance from province capital × Post 2010	-0.000	-0.000
	(0.000)	(0.000)
Security Transition	0.001	0.002
	(0.018)	(0.018)
Number of violence incidents		0.001
		(0.001)
F-statistic	0	0
Obs (district-year)	5083	5083
Number of Districts	391	391

Notes: This table presents the results of first stage regressions for an alternative instrument. It examines the correlation between military withdrawal ("the endogenous variable") and distance between districts and province capital "the instrument" once security transition to local allies started ("Post2010"). Province and district are level two and level three administrative divisions in Afghanistan. Several districts form a province and one district in each province serves as the provincial capital. The outcome of interest is military withdrawal from districts. It is a binary variable that turns on when the last military base was closed. District panel covers the period 2003-2015. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific linear trends. Column (2) includes violence incidents (ambushes, shooting, suicide attacks) divided by district population fixed at 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.5: First stage regressions with alternative instrument

	Military Withdrawal	
	(1)	(2)
Distance from Kabul × Post 2010	-0.000	-0.000
	(0.000)	(0.000)
Security Transition	0.002	0.002
	(0.017)	(0.017)
Number of violence incidents		0.001
		(0.001)
F-statistic	0	0
Obs (district-year)	5083	5083
Number of Districts	391	391

Notes: This table presents the results of first stage regressions for an alternative instrument. It examines the correlation between military withdrawal ("the endogenous variable") and distance between districts and Kabul City "the instrument" once security transition to local allies started ("Post2010"). Kabul is Afghanistan's Capital. The outcome of interest is military withdrawal from districts. It is a binary variable that turns on when the last military base was closed. District panel covers the period 2003-2015. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific linear trends. Column (2) includes violence incidents (ambushes, shooting, suicide attacks) divided by district population fixed at 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.6: Reduced form regressions

	Total Participation		Male Participation		Female Participation	
	(1)	(2)	(3)	(4)	(5)	(6)
Travel distance to military airport × Post 2011	0.031**	0.031**	0.005	0.005	0.026***	0.026***
	(0.012)	(0.012)	(0.007)	(0.007)	(0.007)	(0.007)
Security Transition	-0.022	-0.022*	-0.018*	-0.018*	-0.004	-0.004
	(0.013)	(0.013)	(0.009)	(0.009)	(0.007)	(0.007)
Number of violence incidents		0.000		-0.000		0.000
		(0.000)		(0.000)		(0.000)
Mean Outcome	0.272	0.272	0.219	0.219	0.053	0.053
Std Outcome	0.435	0.435	0.318	0.318	0.155	0.155
Obs (district-year)	5083.000	5083.000	5083.000	5083.000	5083.000	5083.000
Number of Districts	391	391	391	391	391	391

Notes: This table presents the results of reduced form regressions. It estimated the direct effect of the least-cost travel distance between districts and the nearest regional military airbase “the instrument” on university exam participation. The outcome of interest is participation in university entrance examination divided by the 2016 district population estimates. District panel covers the period 2003-2015. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific time trends. Columns (2), (4) and (6) include violence incidents (ambushes, shooting, suicide attacks) divided by district population fixed at 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.7: Total university entrance exam participation and military withdrawal

	OLS		2SLS	
	(1)	(2)	(3)	(4)
Military withdrawal	0.027	0.027	0.386**	0.386**
	(0.017)	(0.017)	(0.156)	(0.155)
Security Transition	-0.020	-0.020	-0.020	-0.021
	(0.013)	(0.013)	(0.015)	(0.015)
Number of violence incidents		0.000		-0.000
		(0.000)		(0.001)
Mean Outcome	0.272	0.272	0.272	0.272
Std Outcome	0.435	0.435	0.435	0.435
Weak IV statistic			64.400	65.023
Obs (district-year)	5083	5083	5083	5083
Number of Districts	391	391	391	391

Notes: This table presents results for OLS and 2SLS regressions. The outcome is total per capita participation in the university entrance examination. Results for OLS regressions are presented in columns (1) and (2) and results for 2SLS regressions are presented in columns (3) and (4). In OLS regressions withdrawal ("the endogenous variable") is the variable of interest that directly enters the OLS regression. In the first stage of two-stage regressions, withdrawal is predicted using travel cost between districts and nearest military airbase ("the instrument"); and in the second stage the predicted withdrawal from the first stage is used to estimate the effect of withdrawal on the outcome. The outcome of interest is total participation in university entrance examination normalized by 2016 district population estimates. District panel covers the period 2003-2015. All regressions include district fixed effects, year fixed effects, and district-specific time trends. Columns (2) and (4) include violence incidents (ambushes, shooting, suicide attacks) that is the annual count of armed conflict events divided by the 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.8: University entrance exam participation by gender and military withdrawal

	Panel A –– OLS			
	Male		Female	
	(1)	(2)	(3)	(4)
Military withdrawal	0.004	0.004	0.023***	0.023***
	(0.013)	(0.013)	(0.008)	(0.008)
Security Transition	-0.017*	-0.018*	-0.003	-0.002
	(0.010)	(0.010)	(0.007)	(0.007)
Number of violence incidents		-0.000		0.000
		(0.000)		(0.000)
Mean Outcome	0.219	0.219	0.053	0.053
Std Outcome	0.318	0.318	0.155	0.155
Obs (district-year)	5083.000	5083.000	5083.000	5083.000
Number of Districts	391	391	391	391

	Panel B –– 2SLS			
	Male		Female	
	(1)	(2)	(3)	(4)
Military withdrawal	0.092	0.058	0.328***	0.328***
	(0.087)	(0.087)	(0.092)	(0.091)
Security Transition	-0.007	-0.018*	-0.003	-0.003
	(0.010)	(0.010)	(0.009)	(0.009)
Number of violence incidents		-0.000		-0.000
		(0.000)		(0.000)
Mean Outcome	0.227	0.219	0.053	0.053
Std Outcome	0.319	0.318	0.155	0.155
Weak IV statistic	71.706	65.023	64.400	65.023
Obs (district-year)	5474	5083	5083	5083
Number of Districts	391	391	391	391

Notes: This table presents results for OLS and two-stage least square regressions. Results for OLS regressions are presented in Panel A and results for 2SLS regressions are presented in Panel B. In OLS regressions withdrawal ("the endogenous variable") is the variable of interest that directly enters the OLS regression. In the first stage of two-stage regressions, withdrawal is predicted using travel cost between districts and nearest military airbase ("the instrument"); and in the second stage the predicted withdrawal from the first stage is used to estimate the effect of withdrawal on the outcome. The outcome of interest is participation in university entrance examination by gender normalized by 2016 district population estimates. District panel covers the period 2003-2015. All regressions include district fixed effects, year fixed effects, and district-specific time trends. Columns (2) and (4) include violence incidents (ambushes, shooting, suicide attacks) that is the annual count of armed conflict events divided by the 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.9: University entrance exam participation and military withdrawal in each transition year

	Panel A −− OLS								
	2011			2012			2013		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Military withdrawal	0.008	-0.011	0.019	0.015	-0.004	0.019	-0.021	-0.018	-0.003
	(0.032)	(0.026)	(0.015)	(0.023)	(0.015)	(0.012)	(0.017)	(0.017)	(0.004)
Number of violence incidents	0.000	0.000	-0.000	-0.000	-0.001	0.000	0.000	0.000	0.000
	(0.002)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean Outcome	0.356	0.269	0.087	0.266	0.216	0.050	0.151	0.147	0.004
Std Outcome	0.501	0.349	0.192	0.430	0.310	0.154	0.276	0.263	0.020
Obs (district-year)	1794.000	1794.000	1794.000	2158.000	2158.000	2158.000	1131.000	1131.000	1131.000
Number of Districts	138	138	138	166	166	166	87	87	87
	Panel B −− IV								
	2011			2012			2013		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Military withdrawal	0.294*	0.073	0.221**	0.635	0.068	0.567**	-1.690	-1.882	0.192
	(0.162)	(0.088)	(0.099)	(0.425)	(0.233)	(0.261)	(2.165)	(2.354)	(0.447)
Number of violence incidents	-0.000	0.000	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001	0.000
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.000)
Mean Outcome	0.356	0.269	0.087	0.266	0.216	0.050	0.151	0.147	0.004
Std Outcome	0.501	0.349	0.192	0.430	0.310	0.154	0.276	0.263	0.020
Weak IV statistic	54.891	54.891	54.891	11.806	11.806	11.806	0.599	0.599	0.599
Obs (district-year)	1794	1794	1794	2158	2158	2158	1131	1131	1131
Number of Districts	138	138	138	166	166	166	87	87	87

Notes: This table presents results for OLS and two-stage least square regressions for each transition year. In OLS regressions withdrawal ("the endogenous variable") directly enters the regression. In 2SLS regressions, first withdrawal is predicted using travel cost between districts and nearest military airbase ("the instrument"); and in the second stage the predicted withdrawal is used to estimate the effect of withdrawal on the outcome. The outcome of interest is participation in university entrance examination normalized by 2016 district population estimates. District panel covers the period 2003-2015. The outcome is total participation in columns (1), (4) and (7), male participation in columns (2), (5) and (8), and female participation in columns (3), (6) and (9). Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.10: Arbitrary transition time. — Demand for higher education and withdrawal

	Total Participation		Male Participation		Female Participation	
	(1)	(2)	(3)	(4)	(5)	(6)
Military withdrawal	-0.024	-0.022	0.109	0.101	-0.132	-0.122
	(0.253)	(0.245)	(0.174)	(0.168)	(0.132)	(0.127)
Security Transition	-0.020	-0.020	-0.018*	-0.018*	-0.002	-0.002
	(0.013)	(0.013)	(0.010)	(0.010)	(0.007)	(0.007)
Number of violence incidents		0.000		-0.000		0.000
		(0.000)		(0.000)		(0.000)
Mean Outcome	0.272	0.272	0.219	0.219	0.053	0.053
Std Outcome	0.435	0.435	0.318	0.318	0.155	0.155
Weak IV statistic	15.463	17.261	15.463	17.261	15.463	17.261
Obs (district-year)	5083	5083	5083	5083	5083	5083
Number of Districts	391	391	391	391	391	391

Notes: This table presents the 2SLS regressions. It estimated the direct effect of the least-cost travel distance between districts and the nearest regional military airbase "the instrument" on university exam participation. Instead of $Post > 2010$ that marks the start of the transition process, I use $Post > 2013$ in the construction of the instrument to examine the arbitrariness of the interaction term in the instrument. The outcome of interest is participation in university entrance examination divided by the 2016 district population estimates. District panel covers the period 2003-2015. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific time trends. Columns (2), (4) and (6) include violence incidents (ambushes, shooting, suicide attacks) divided by district population fixed at 2016 population estimates. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.11: Demand for higher education and withdrawal with Callaway and Sant'Anna (2021)

	Simple DD			IPW DD		
	Total	Male	Female	Total	Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Military withdrawal	0.045	0.012	0.033**	0.068*	0.036	0.032**
	(0.032)	(0.022)	(0.014)	(0.035)	(0.025)	(0.014)
Number of Observations	5083 5083	5083	5083	5083	5083	5083
Number of Districts	391 391	391	391	391	391	391

Notes: This table presents results from non-parametric event study method of Callaway and Sant'Anna (2020). It shows the average treatment effects of military withdrawal on treated districts. Estimates in columns (1), (2) and (3) do not account for baseline differences. Estimates in columns (4), (5), and (6) includes time-invariant least-cost travel distance from nearest military hub that is used for inverse probability weighting based on Abadie (2005). Standard errors are clustered at province level instead of district because in this estimation method the treatment effect is estimated in each period separately and in one period districts being the unit of observation cannot be used as cluster. */**/*** denotes significance at the 10/5/1 percent.

Table 1.12: Troop withdrawal and insurgent fatality

	OLS (1)	2SLS (2)
Military withdrawal	0.005	-0.008
	(0.007)	(0.026)
Security Transition	0.019***	0.019***
	(0.006)	(0.006)
Number of violence incidents	0.005***	0.005***
	(0.001)	(0.001)
Mean Outcome	0.023	0.023
Std Outcome	0.089	0.089
Weak IV statistic		65.023
Obs (district-year)	5083	5083
Number of Districts	391	391

Notes: This table presents results of OLS and 2SLS regressions. The outcome of interest is district-level insurgent fatalities measured by number of insurgents died or injured during air attacks, face-to-face fights, and suicide attacks. In the OLS regression (column 1), military withdrawal directly enters the regression. In 2SLS regression (column 2), first military withdrawal ("the endogenous variable") is predicted using using travel cost between districts and nearest military airbase ("the instrument"); and in the second stage the predicted withdrawal is used to estimate the effect of withdrawal on insurgent fatality. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific time trends. District panel covers the period 2007-2015. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.13: Troop withdrawal and self-reported income

	OLS (1)	2SLS (2)
Military withdrawal	0.071	-5.278**
	(0.418)	(2.252)
Security Transition	-0.319	-0.228
	(0.645)	(0.659)
Number of violence incidents	-0.001	-0.001
	(0.009)	(0.009)
Mean Outcome	5.151	5.151
Std Outcome	5.360	5.360
Weak IV statistic		66.287
Obs (district-year)	4158	4158
Number of Districts	384	384

Notes: This table presents results of OLS and 2SLS regressions. The outcome of interest is district-level average self-reported income. Because of unavailability of income data, district panel covers the period 2007-2019. In the OLS regression (column 1), military withdrawal directly enters the regression. In 2SLS regression (column 2), first military withdrawal ("the endogenous variable") is predicted using using travel cost between districts and nearest military airbase ("the instrument"); and in the second stage the predicted withdrawal is used to estimate the effect of withdrawal on average income. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects, year fixed effects, and district-specific time trends. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.14: Effect of foreign military withdrawal on demand for higher education via income

	Exam Participation		
	Total	Male	Female
	(1)	(2)	(3)
Total effect	0.393**	0.105	0.289***
	(0.183)	(0.110)	(0.090)
Direct effect	0.067*	0.031	0.036*
	(0.036)	(0.021)	(0.022)
Indirect effect	0.326	0.073	0.253*
	(0.236)	(0.109)	(0.153)
% mediation effect	82.903	70.023	87.576
F-statistic (T on Z)	74.679	74.679	74.679
F-static (M on Z | T)	6.524	6.524	6.524
Obs (district-year)	4157	4157	4157
Number of Districts	389	389	389

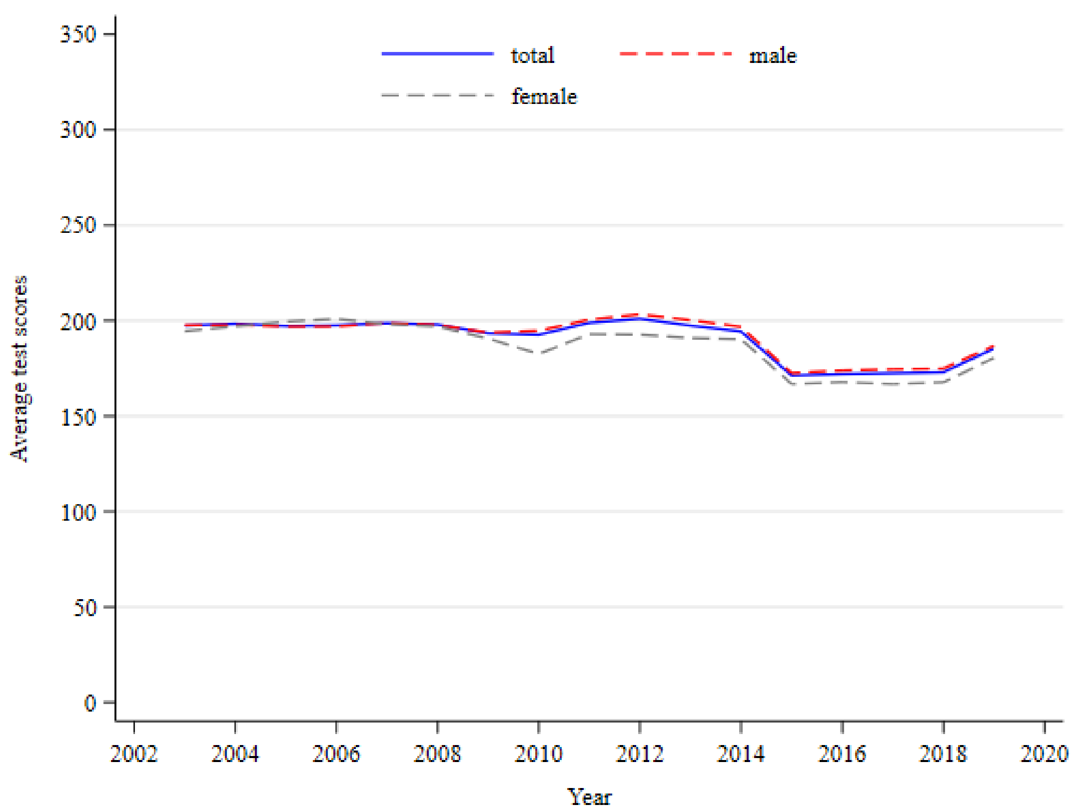
Notes: This table presents 2SLS regressions for mediation effects. The mediator variable is district-level average self-reported income. Because of unavailability of income data, district panel covers the period 2007-2019. The purpose of mediation analysis is to show the effect of for military withdrawal ("the endogenous variable") on outcome (demand for higher education) is mediated by changes in income ("the mediator"). Military withdrawal is predicted using using travel cost between districts and nearest military airbase ("the instrument"). All regressions include security handover, district fixed effects, year fixed effects, and district-specific time trends. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 1.15: Intention to migrate and military withdrawal

	OLS			2SLS		
	Total (1)	Male (2)	Female (3)	Total (4)	Male (5)	Female (6)
Military withdrawal	3.175**	3.081	3.797**	7.797**	10.900**	6.009
	(1.545)	(2.083)	(1.900)	(3.447)	(5.443)	(4.185)
Security Transition	-2.573	-1.218	-3.082	-6.328*	-7.569	-4.880
	(2.351)	(3.180)	(3.038)	(3.350)	(5.162)	(4.004)
Number of violence incidents	0.075	0.037	0.119*	0.058	0.008	0.111
	(0.057)	(0.077)	(0.066)	(0.061)	(0.084)	(0.068)
Mean Outcome	35.019	38.080	32.210	35.019	38.080	32.210
Std Outcome	17.158	22.519	21.468	17.158	22.519	21.468
Weak IV statistic				207.420	207.420	207.420
Obs (district-year)	1290	1290	1290	1290	1290	1290
Number of Districts	223	223	223	223	223	223

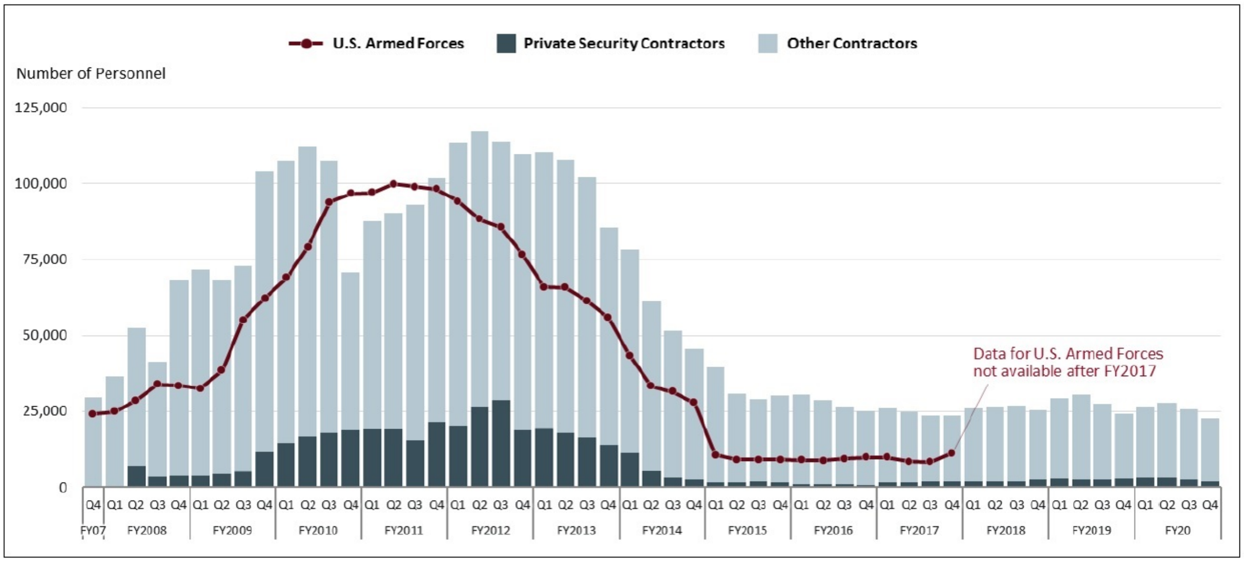
Notes: This table presents results of OLS and 2SLS regressions. The outcome of interest is intention to migrate that is the share of respondents that indicated they would leave Afghanistan if they had a chance. Because of unavailability of sufficient survey data, district panel covers the period 2007-2019. Panel A and Panel B present results from OLS and 2SLS regressions, respectively. Each panel has separate columns for total migration, male migration, and female migration. In OLS regressions, military withdrawal directly enters the regression. In 2SLS regression, first military withdrawal ("the endogenous variable") is predicted using using travel cost between districts and nearest military airbase ("the instrument"); and in the second stage the predicted withdrawal is used to estimate the effect of withdrawal on migration. Security Transition that marks the beginning of handover of security responsibilities to local allies and official start of physical withdrawal, is included in all regressions. All regressions include district fixed effects. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

1.10 Figures



Notes: This figure demonstrates average test scores in Afghanistan's university entrance exam over time. The solid blue line shows average scores of all test takers and dashed lines represent average scores by gender.

Figure 1.9: Average test score over time

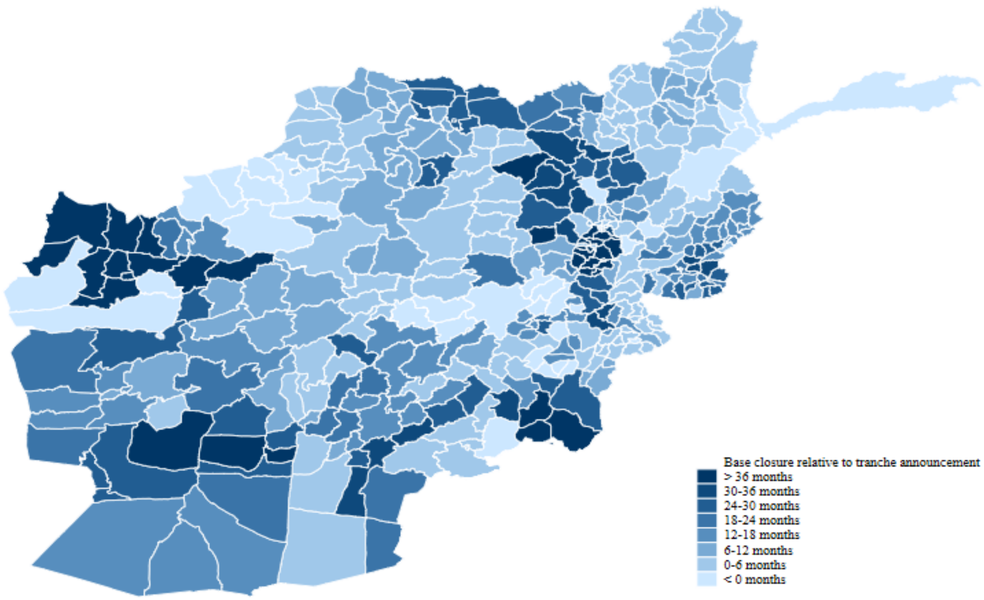


(a) Troops Level Based on Congressional Research Services

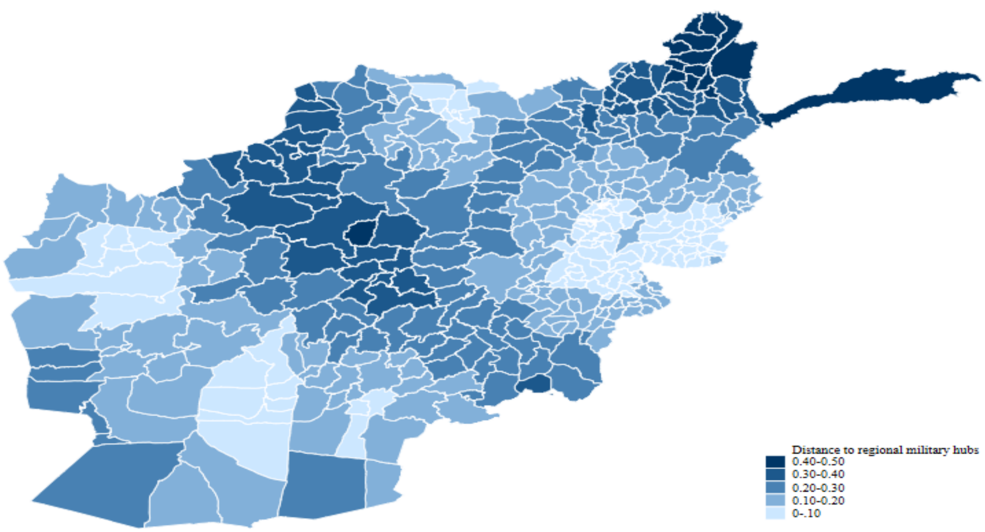
(b) Troops Level Based on BBC Report

Notes: This figure presents the total number of U.S military and contractor personnel present in Afghanistan between 2002-2020. Panel (a) has been copied from page 7 of the Congressional Research Services updated as of February 22, 2021. The figure is available at (https://fas.org/sgp/crs/natsec/R44116.pdf). For convenience, I have removed the figure notes from the original figure. The notes in the figure states that the figure was created by CRS. Panel (b) is copied from https://www.bbc.com/news/world-41014263. The figures also mentions the US Department of Defense as the data source. Figures on the vertical axis show thousand troops. The data in this panel does not include contractor personnel. It provides information on troops level for the years that is missing in the CRS presented in Panel (a).

Figure 1.2: U.S. Armed Forces and DOD-Funded Contractor Personnel in Afghanistan (Sources: CRS & BBC)



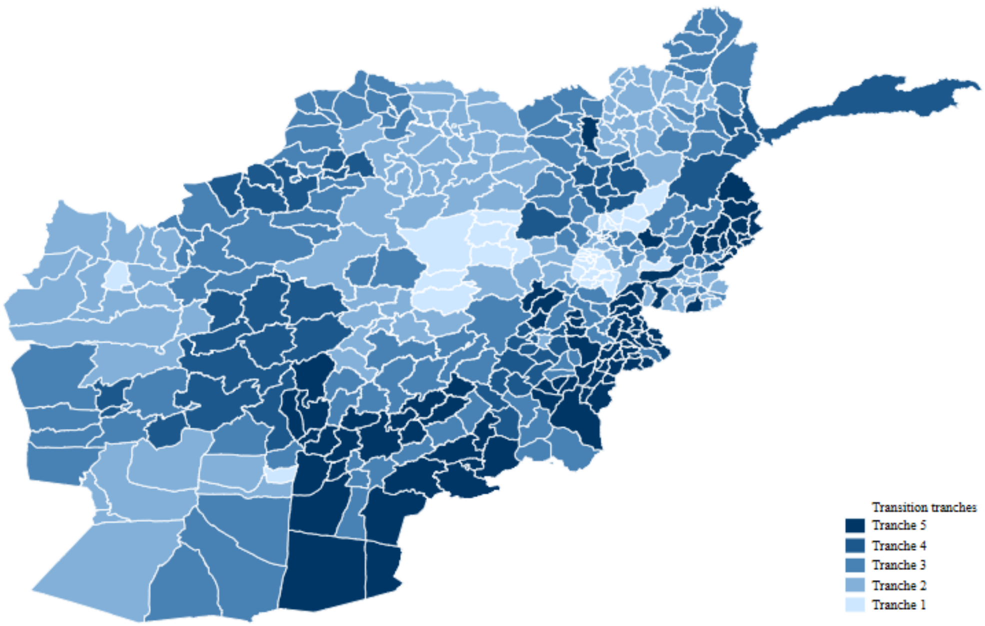
(a) Withdrawal time relative to transition announcement



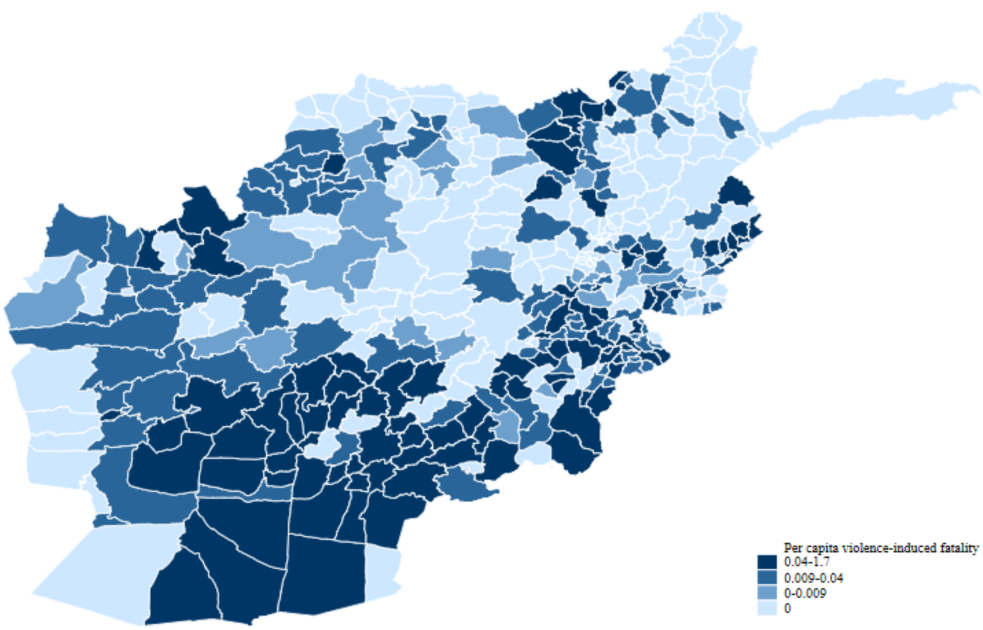
(b) Least-cost travel distance from nearest hub

Notes: Panel (a) shows withdrawal time relative to transition announcement. Transition announcement was made five phases called “transition tranche”. Withdrawal is measured by closure of last military base in the district. In other words, Panel (a) demonstrates time distance between the date that transition tranche was announced and date that the last military base was closed. Panel (b) depicts least-cost travel distance from one of the 10 regional military hubs identified and calculated in Fetzer et al. (2021). Those military airbases were maintained and operated by the coalition forces for training of ANSF and supply of logistics to remote combat locations.

Figure 1.3: Total participation in university exam and military withdrawal



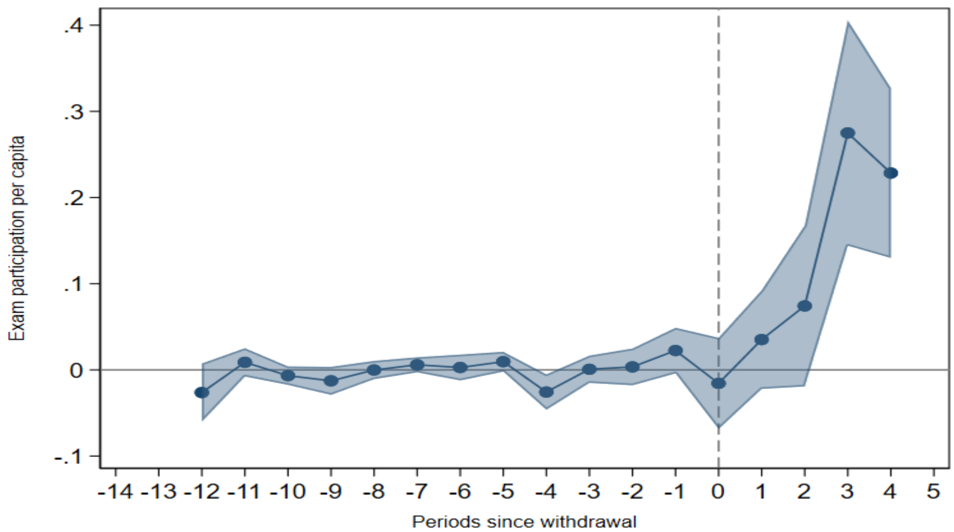
(a) Transition tranche



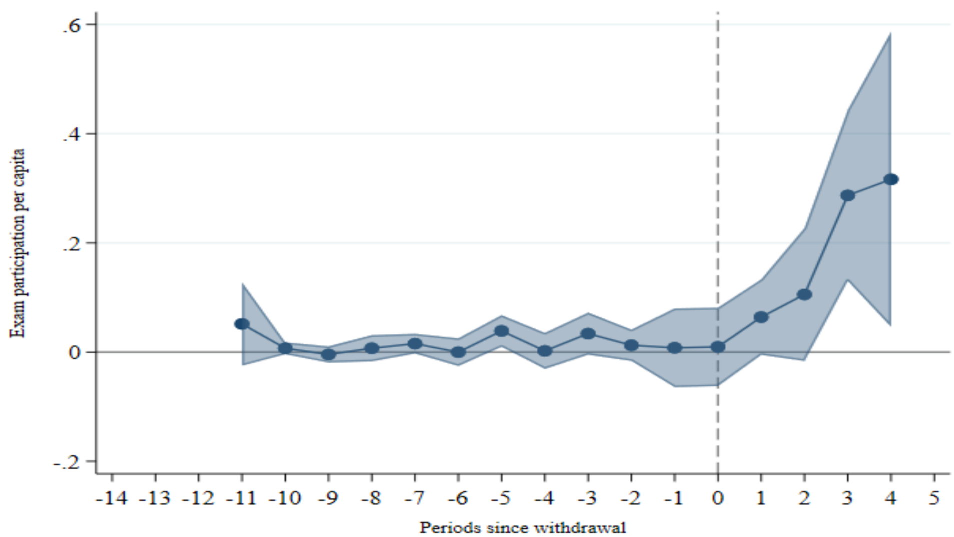
(b) Per capita casualty from violence

Notes: Panel (a) demonstrates five transition tranches that refers to a group of districts that were considered for a transition phase. Color becomes darker early to late tranches. Panel (b) shows per capita fatalities due to insurgent attacks and counter attacks by the NATO and Afghan National Security Forces.

Figure 1.4: Transition tranches and per capita fatality



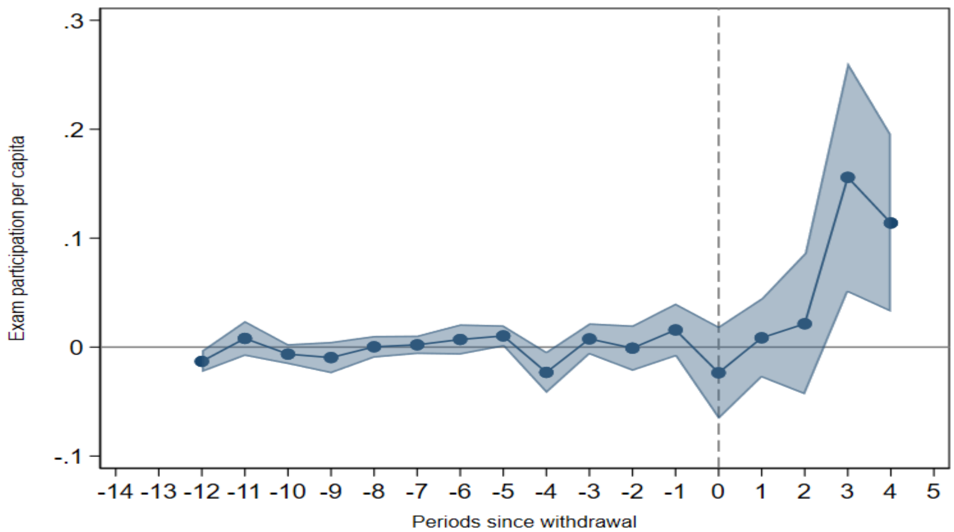
(a) Simple DD



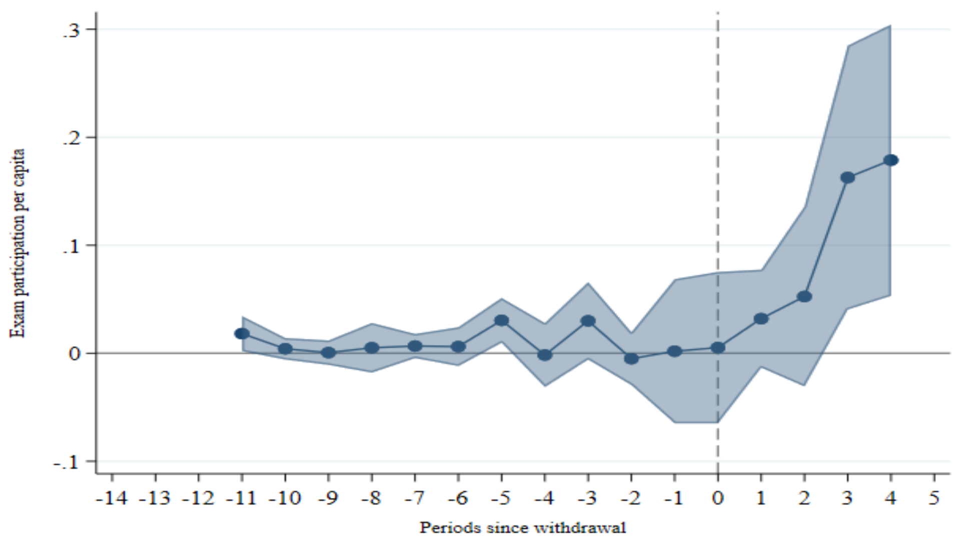
(b) DD with IPW

Notes: This Figure presents aggregate effects for each event time in Callaway and Sant'Anna (2020). The the outcome of interest is total demand for higher education measured by total participation in the university entrance exam. The outcome is normalized by 2016 district population estimates. In Panel (a) Callaway and Sant'Anna (2020) is estimated without controls and in Panel (b) time-invariant controls are included that is used for inverse probability weighting.

Figure 1.5: Total participation in university exam and military withdrawal



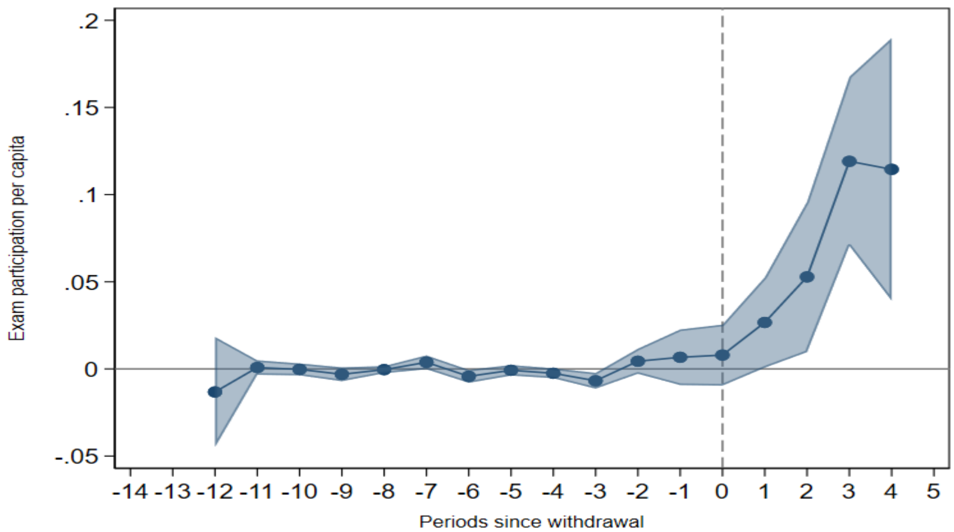
(a) Simple DD



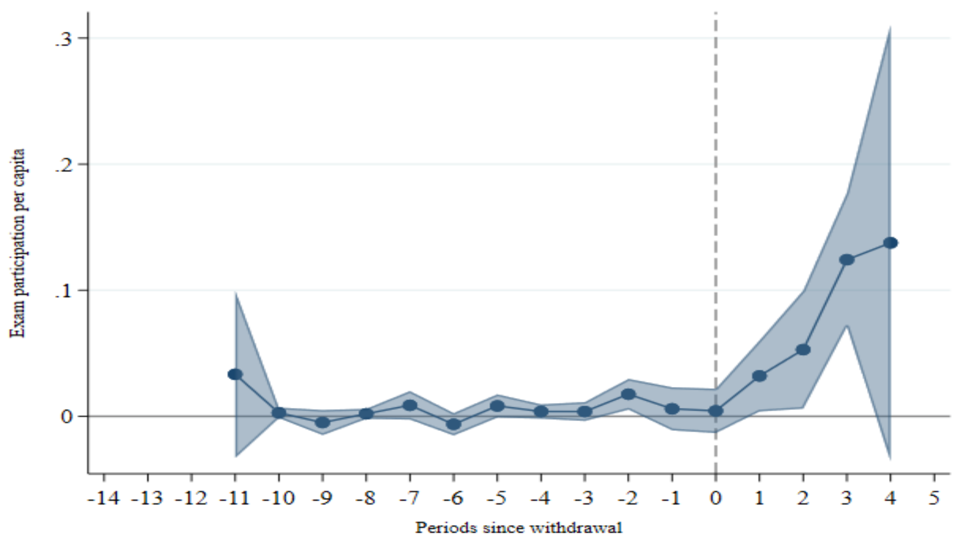
(b) DD with IPW

Notes: This figure presents aggregate effects for each event time in Callaway and Sant'Anna (2020). The the outcome of interest is male demand for higher education measured by total participation in the university entrance exam. The outcome is normalized by 2016 district population estimates. In Panel (a) Callaway and Sant'Anna (2020) is estimated without controls and in Panel (b) time-invariant controls are included that is used for inverse probability weighting.

Figure 1.6: Male participation in university entrance exam and military withdrawal



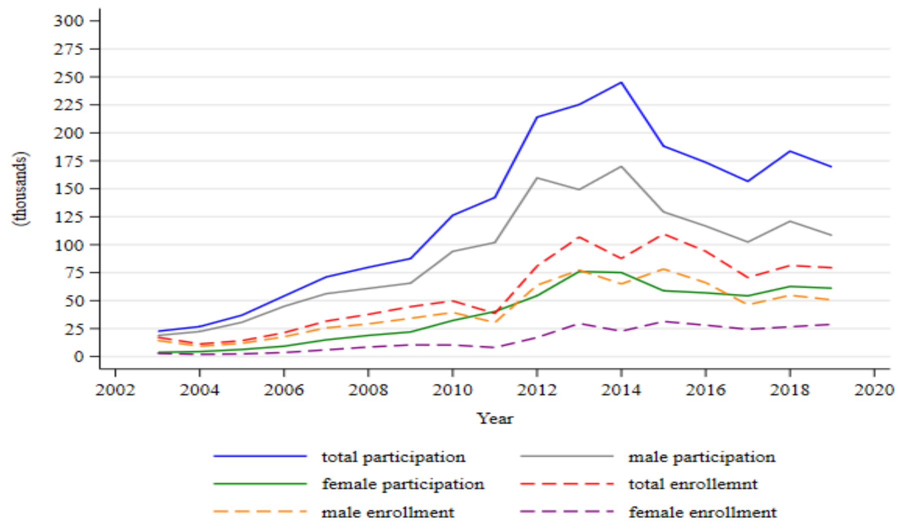
(a) Simple DD



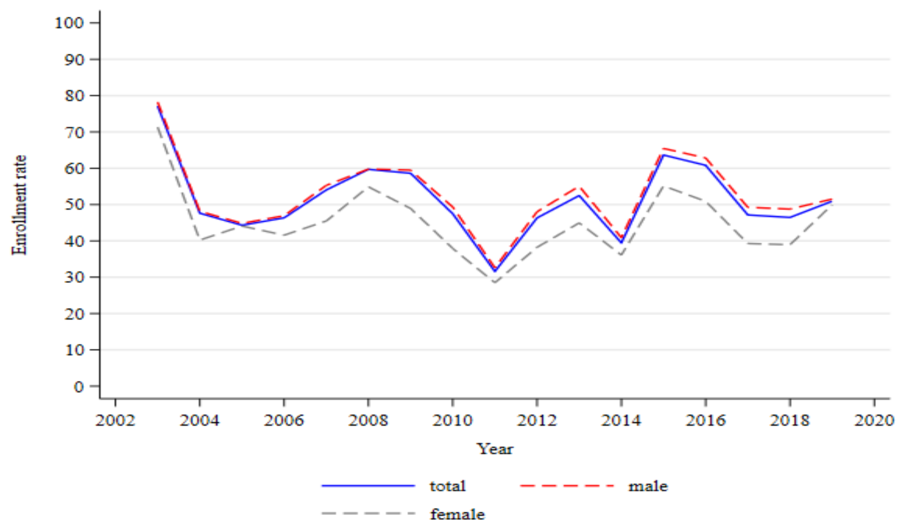
(b) DD with IPW

Notes: This figure presents aggregate effects for each event time in Callaway and Sant'Anna (2020). The the outcome of interest is female demand for higher education measured by total participation in the university entrance exam. The outcome is normalized by 2016 district population estimates. In Panel (a) Callaway and Sant'Anna (2020) is estimated without controls and in Panel (b) time-invariant controls are included that is used for inverse probability weighting.

Figure 1.7: Female participation in university entrance exam and military withdrawal



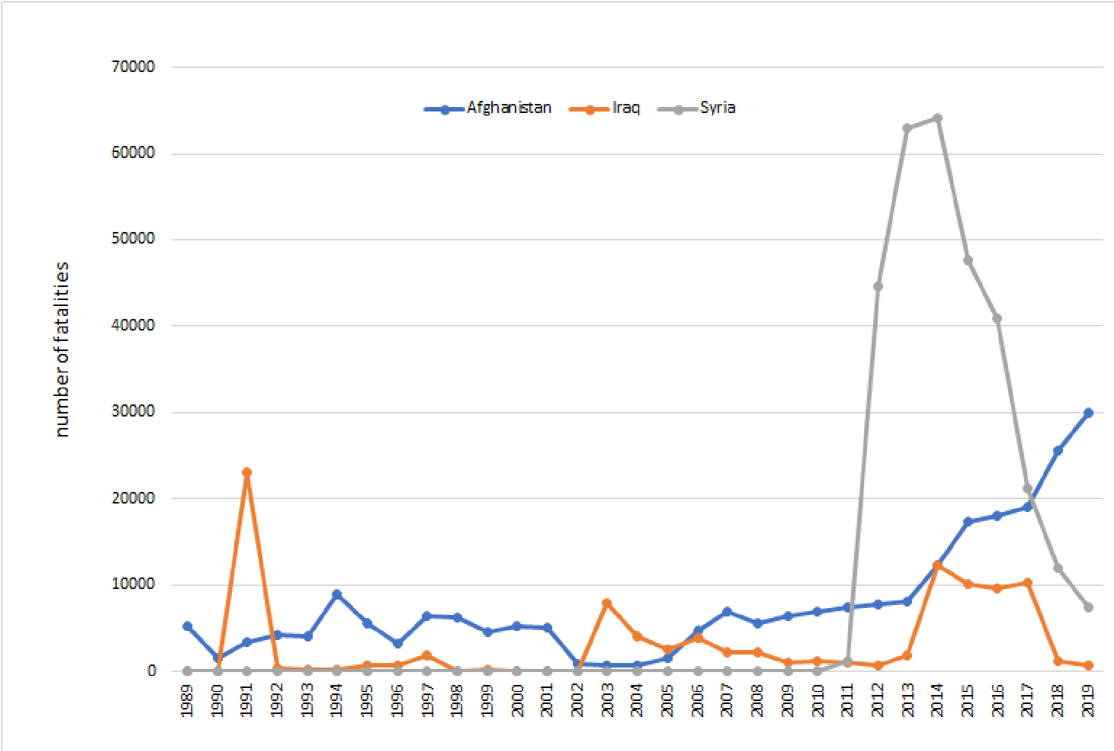
(a) Participation and enrollment



(b) Enrollment rate

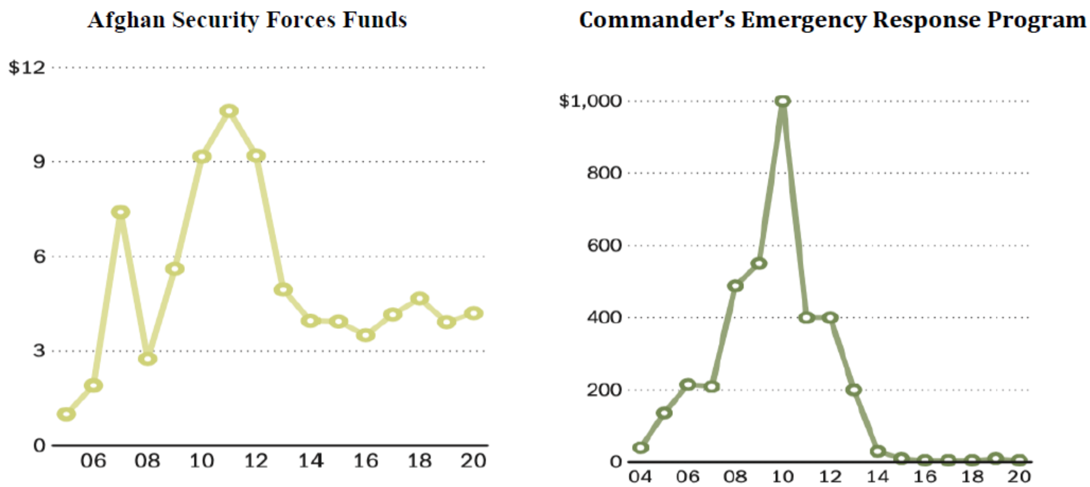
Notes: This figure demonstrates participation in university entrance examination and enrollment. Panel (a) shows total participation and enrollment by gender where solid lines represent participation in university entrance exam and dashed lines exhibit enrollment. Enrollment is the total number of applicants that got an offer from public four-year and two-year university programs. Panel (b) depicts enrollment rate by gender.

Figure 1.8: Exam participation and enrollment over time

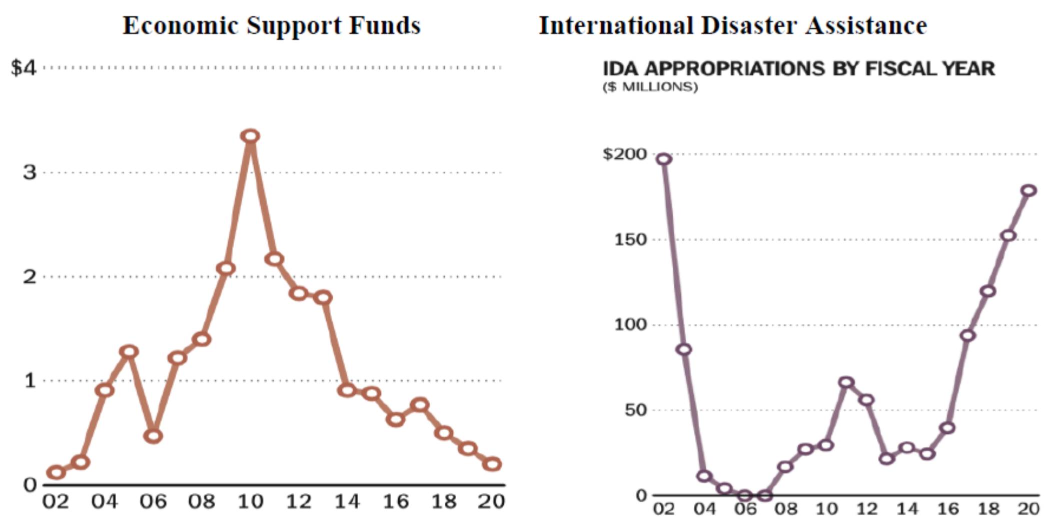


Notes: Data source: World Development Indicators, The World Bank. Syrian crisis started in 2011 as a result of ISIS rise. The country had a relatively better political stability before that. Iraq experienced an unstable period during Iran-Iraq and Iraq-Kuwait wars but was mostly stable between 1992-2002 until the U.S carried attacks on the Sadam Hussain regime. Afghanistan, on the other hand, has been in turmoil since the 1979. Civilian fatalities surged in the aftermath of a reduction in the U.S and ISAF troops since 2012.

Figure 1.10: Conflict Induced Fatalities in Afghanistan, Iraq and Syria (Source: The World Bank)



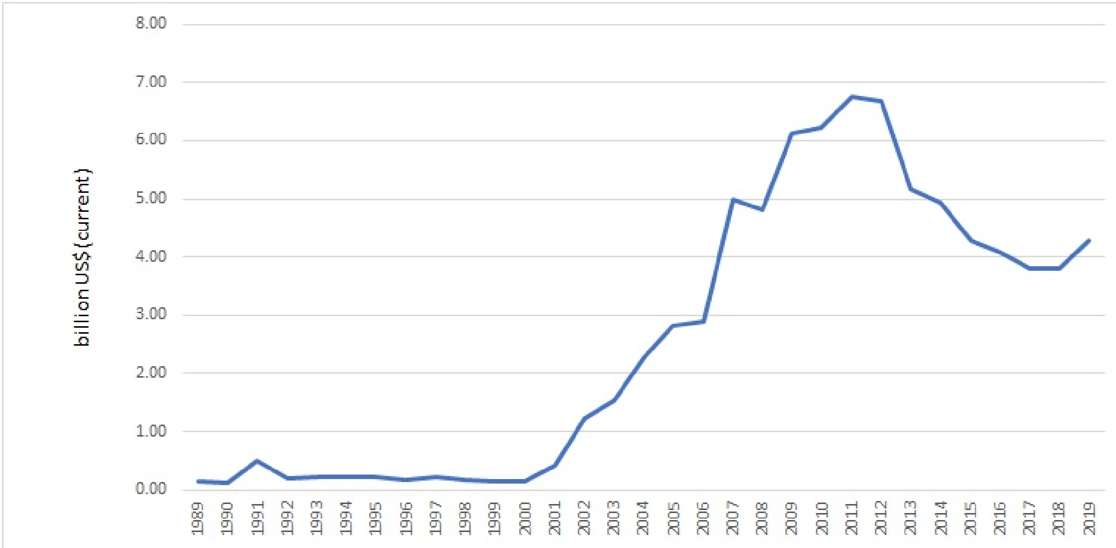
(a) Security Spending



(b) Economic and Disaster Relief Spending

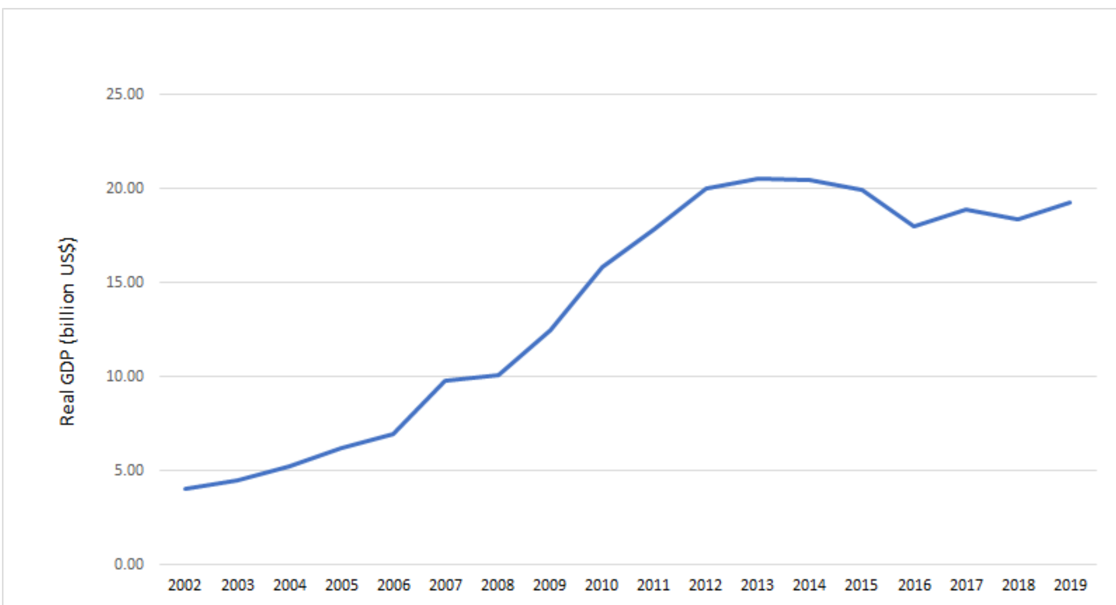
Notes: This figure has been retrieved from the Center for Strategic and International Studies (CSIS) without any editing. The figure is available at https://www.csis.org/analysis/biden-transition-and-real-impact-us-force-cuts-afghanistan. The figures mentions Special Inspector General for Afghan Reconstruction (SIGAR) reports, and Report to the United States Congress, October 30, 2020, pp. 50, 54, 56, 57 and 61, as the data source for this figure. I have removed the figure notes for convenience. The vertical axis for Commander's Emergency Response Program and International Disaster Assistance are in million US dollars. The numbers for the Afghan Security Forces Funds, and Economic Support Funds are in billion US dollars.

Figure 1.11: U.S and International Spending in Afghanistan (Source: Center for Strategic and International Studies)



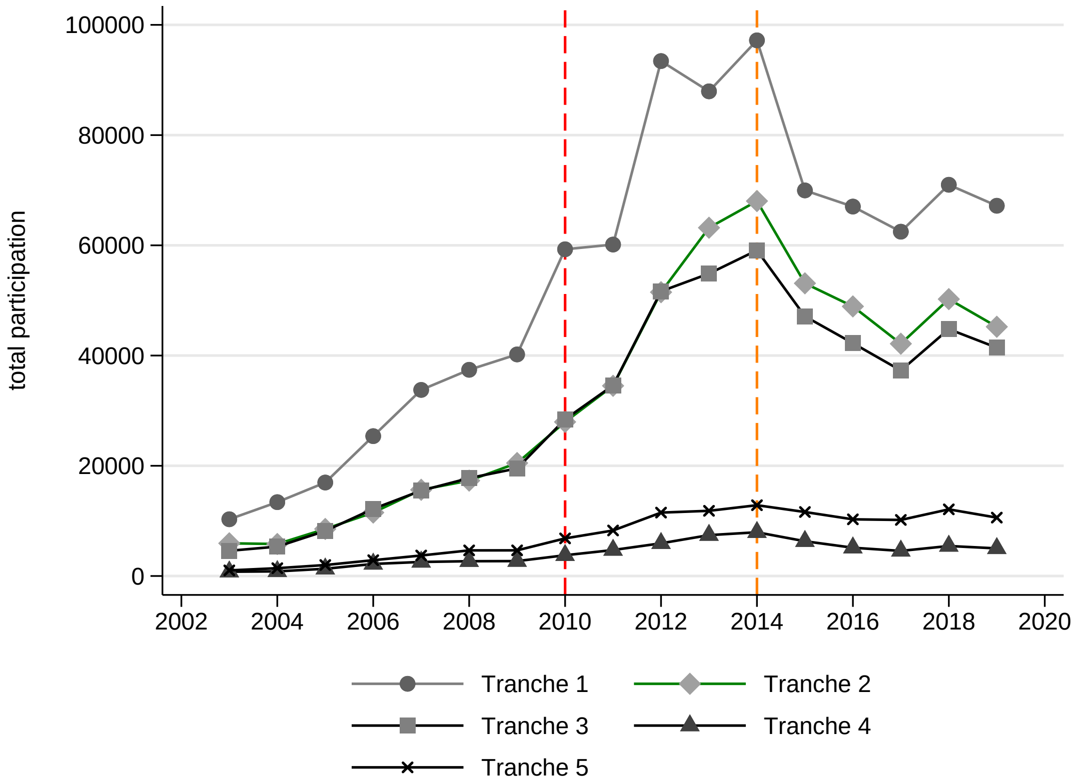
Notes: Data source: World Development Indicators, The World Bank. This figure shows the aggregate annual foreign aid received by Afghanistan. It includes military and civilian assistance.

Figure 1.12: Total Annual Aid Spending in Afghanistan



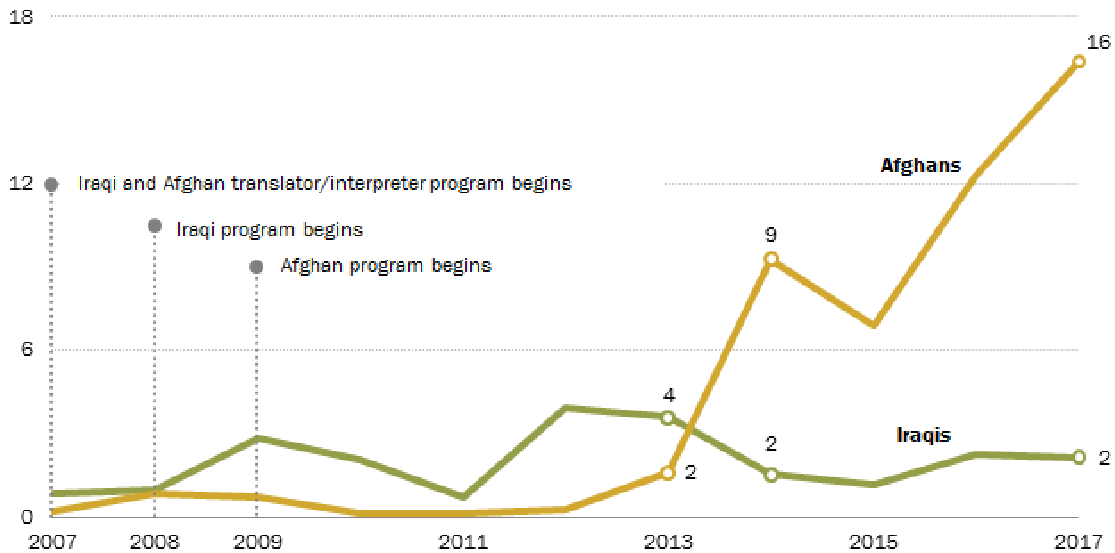
Notes: Data source: World Development Indicators, The World Bank.

Figure 1.13: Afghanistan Real GDP Trend



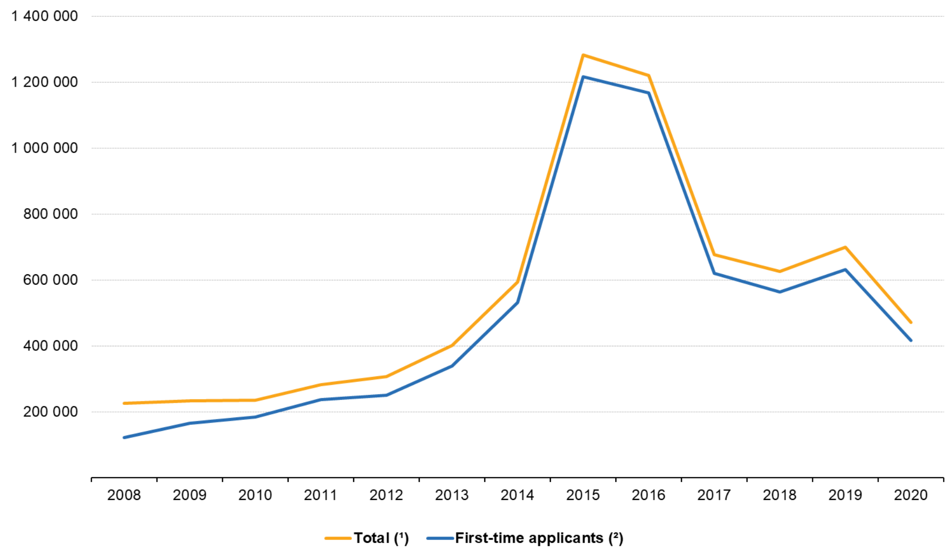
Notes: This figure compares the average test score of students admitted to STEM and non-STEM majors. STEM majors broadly include majors such as agriculture, computer science, engineering, earth science, mathematics, medicine, and natural science. All social sciences, arts, education and veterinary education are classified as non-STEM majors. As we can see, there has always been a gap between the two categories. But the gap seems parallel.

Figure 1.14: Trenches Compared Based on Average Participation

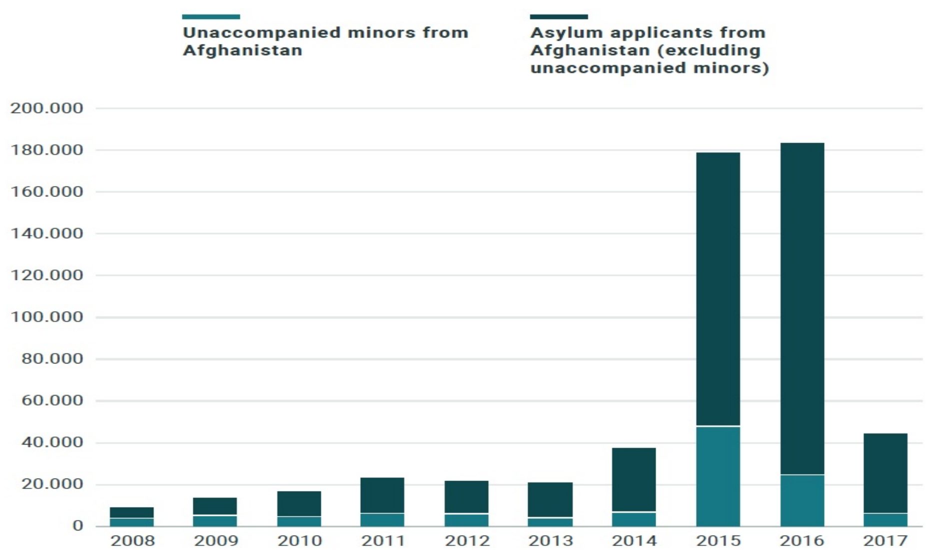


Source: https://www.pewresearch.org/fact-tank/2017/12/11

Figure 1.15: SIV Application



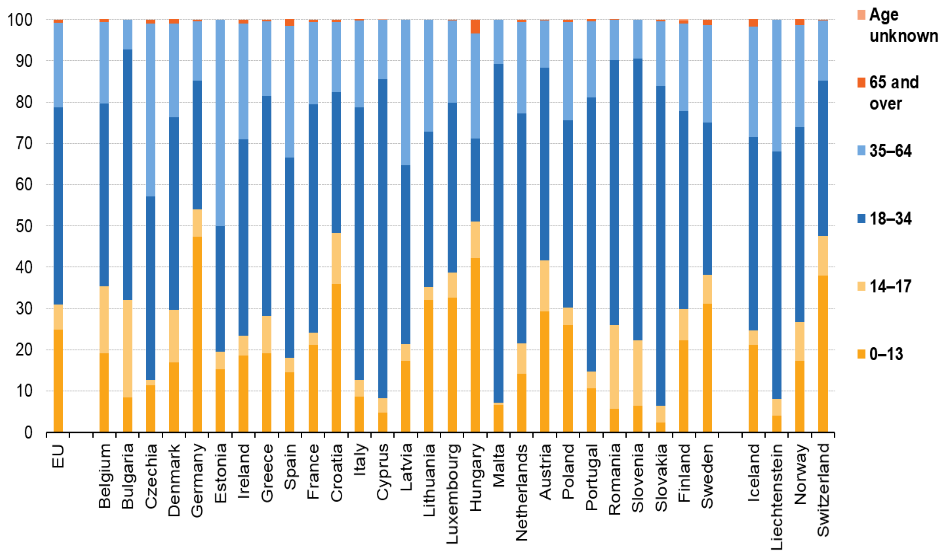
(a) Number of Asylum Applicants (non-EU citizens) in EU



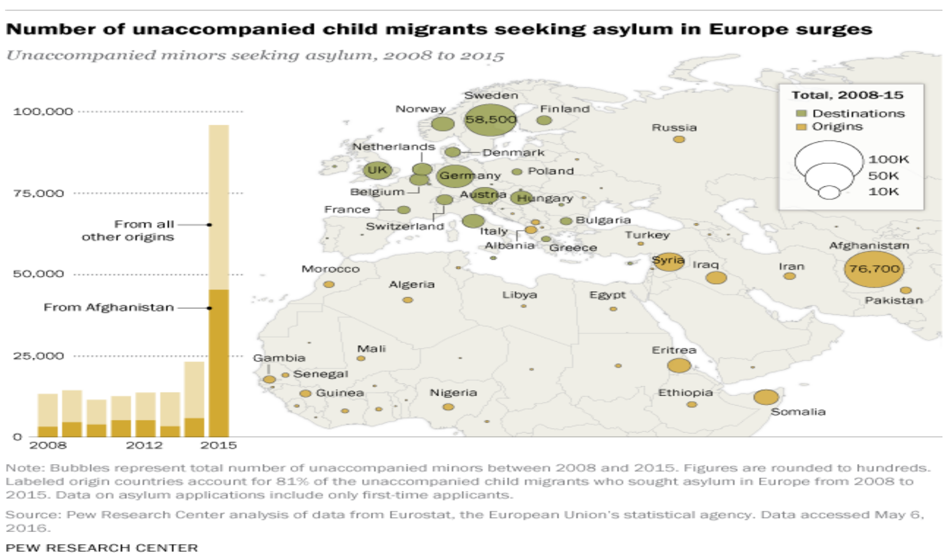
(b) Number of Asylum Applicants from Afghanistan in EU

Notes: This figure has been retrieved from the eurostat without any editing. For convenience, I have removed the figure title and notes. Both panels are available at https://ec.europa.eu/eurostat/statistics-explained/index.php. Panel (a) presents the total number asylum applicants from non-EU countries in the EU. In the period 2014-2020 citizens of Syria, Afghanistan, and Iraq formed the largest three asylum seekers, in order. Panel (b) presents the number of asylum applications from Afghanistan in the EU.

Figure 1.16: Pattern of First Time Asylum Seekers in the EU (Source: Eurostat)



(a) Total Asylum Seekers by Age



(b) Contributors of Unaccompanied Minors

Notes: This figure has been copied from the eurostat and Pew Reseach Center without any editing. For convenience, I have removed the figure title and notes. Panel (a) presents the total number asylum applicants from non-EU countries in the EU by age during 2008-2020. Panel (b) describes the largest contributors of unaccompanied minors between 2008-2015.

Figure 1.17: Age Breakdown of Refugee Applications in the EU (Source: Pew Research Center)

Chapter 2. Exam Preparation and Access Inequalities in Higher Education: Evidence from Rural Afghanistan

Ahmad Shah Mobariz Arya Gaduh

2.1 Introduction

Rural high school graduates are less academically prepared to compete in college enrollment and sustain satisfactory academic performance. This is true for developed and developing countries. For instance, in the United States, the enrollment rate of rural high school graduates in tertiary education is 4 percentage points lower than their urban counterpart. The degree completion rate among students from rural communities is 7 percentage points lower than those of suburban areas (Wells et al., 2019). In China, a considerable gap exists between rural and urban students, especially in the first two years of college (Zhao, 2020). A 22.5 percentage points lower enrollment in higher education among rural students as compared to urban enrollment in India indicates a wider rural-urban disparity in South Asia (Tilak, 2015). Rural-urban gap in education, particularly, tertiary education, is a major determinant of welfare dispersion between rural and urban areas (Teng et al., 2007; Young, 2013; Fang and Sakellariou, 2013; Agyire-Tettey et al., 2018; Azam, 2019).

The gender gap in general and among rural communities, in particular, is another issue in many parts of the world. In many developing countries, the gender gap in education is pervasive and presents a problem for development. Klasen (2002), for instance, argued that between 0.4-0.9 percentage points of the difference in per capita growth rate between East Asia and Sub-Saharan Africa, and South Asia and the Middle East could be attributed to a difference in inequality in educational attainment between the male and female population in the observed regions. In low-income countries, female enrollment in undergraduate programs is one-third of male enrollment. In countries such as Afghanistan, Benin, Central African Republic, Chad, and Niger the female to male student ratio in tertiary education is below 40 percent (UNESCO International Institute for Education Planning, 2017). Despite its importance, research on effectiveness of policies to address the problem of gender disparity in higher education in low-income countries is rare.

In this chapter, we study Programe-e Erteqai Zarfiyat wa Amadagi Kankor dar Manatiq Mahroom (PEZAK) which sent students from Afghanistan's top universities to help high school students in underdeveloped areas of Afghanistan to prepare students to compete in the national university entrance examination (Kankor). The program was female-inclusive since it includes female participation as a condition to receive the program. In areas with a female school, the program is allocated to female schools instead of male schools. The program was the private initiative of a former member of the Parliament of Afghanistan in a collaboration with local businesses and the receiving communities.

The treatment assignment in PEZAK was non-staggered. Before 2008 no school received treatment. Between 2008-2019, in total 184 schools received the program at least once. During this period, treatment turned on and off several times. On average, 62 schools received the program each year. Furthermore, on average, in each successive year, approximately 20% of schools treated in the previous year were replaced with new schools. Treatment in each period affected at least three cohorts: grade 12, grade 11 and grade 10. Students in grade 12 graduated from high school and took the university entrance exam immediately or few months after completing the PEZAK, while those in grades 11 and 10 took the test one year and two years later, respectively.

We employ the estimator developed by de Chaisemartin and D'Haultfoeuille (2022) to estimate the treatment effects on test scores, enrollment in top-rank universities, enrollment in low-rank universities and passing rate. This estimator is robust to heterogeneous treatment effects across cohorts and over time. It estimates several two-by-two difference-in-differences

(2x2 DD) by comparing several combinations of treatment and control groups over time $\boldsymbol{DID}_{\mathrm{l}}$ and aggregates them in one average total effect. In other words, the average effect from this estimator is the weighted average of DID estimates across all time periods and possible values of treatment. The DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those that remains 0 in $t-1$ and t. Similarly, DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those whose treatment is equal to 1 in both periods. The possibility of negative weight occurring is high in the later case, especially when treatment effect transmits from one period to another period. $\boldsymbol{DID}_{\mathrm{l}}$ estimates the effect of having switched treatment for the first time l periods ago. This method also allows to test for parallel trends with placebo estimates that compares switchers and non-switchers before treatment. Estimates from this estimator are not prone to bias due to heterogeneity in treatment effects and negative weights raised in recent DD literature (de Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). It also accommodates non-staggered designs such as the PEZAK.

Our results show that the program had a positive impact on the test scores, enrollment in top universities, and passing rate. In the first year of exposure, passing rate increased by 0.14 (s.e.=0.07) standard deviations. The dynamic effects on test scores and enrollment in top-rank universities in the second and third year after the first exposure, also kicked off. The average total effect of the the program was 0.17 (s.e. = 0.14), 0.26 (s.e. = 0.12), 0.07 (s.e. = 0.07) and 0.10 (s.e. = 0.11) on test scores, enrollment in top-rank universities, enrollment in low-rank universities and overall passing rate, respectively. Among all outcomes, the average effect on test scores has the highest statistical significance. The average total effects maybe interpreted as effect per unit of the program for receiving a weakly higher number of treatment.

Our heterogeneous treatment analysis shows that treated female students had lower likelihood of enrollment in low-rank universities as compared to treated male students. The difference in test scores and passing rate are negative but they are statistically insignificant. This result could be interpreted as increasing female students' ambitions for high-rank programs instead of low-quality provincial public universities. While male students have more flexibility to travel, mobility for female students is restricted to their area of residence. Male students may travel to the provincial centers and large cities to prepare for the entrance exam, while female students do not have that advantage. Therefore, a program such as the PEZAK program helps female students enroll in good-quality programs.

The PEZAK program in Afghanistan embodies several characteristics of similar interventions (Fack and Grenet, 2015; Dynarski, 2003; Seftor et al., 2009; Abbiati et al., 2018; Cunha et al., 2017; Loyalka et al., 2013; Bettinger et al., 2012). It is similar to Teach for America (Xu et al., 2011) and Summer Bridge Program (Strayhorn, 2011) in the United States, and Pratham in India (Banerjee et al., 2010). It has three of the features seen in the literature: academic preparation, information and outreach and, nudging through peer effect. The PEZAK program is entirely managed by local communities which distinguishes it from other interventions. With few exceptions, most of the interventions towards enabling disadvantaged students to enroll and complete a university degree, are set in the developed countries, particularly the United States (Herbaut and Geven, 2019). Our paper contributes to the literature by presenting the impact of an informal community education program that addresses the problem of continuation in the context of a low-income country.

This chapter is organized as follows. In section 2.2 we provide background about the higher education system in Afghanistan and describe the program. Section 2.3 describes the data. In section 2.4 we present our empirical strategy. The results of our analysis are provided in sections 2.5. In section Section 2.6 we present a benefit-cost analysis of the program. 2.7 concludes the chapter.

2.2 Background

2.2.1 Inequalities in Afghan Higher Education

One of the remarkable achievements of the post-Taliban (post-2001) Afghanistan (until its demise in 2021) was how the country rebuilt her education system (Aturupane, 2013). Three decades of war had inherited the post-Taliban government with dysfunctional educational institutions and shattered infrastructure, resulting in low literacy rates, especially among women (Roof, 2015). The US-backed Islamic Republic of Afghanistan (IRA) invested heavily in revamping the education system. In the fiscal year 2003, immediately after installation of the UN-mandated interim authority, 16% of Afghanistan's national budget equivalent to 8.3% of the country's GDP was allocated to education (International Monetary Fund, 2005). In later years, because of reduction in foreign aid and pressure from security sector, the share of education spending in the national budget reduced. For example, In 2018, the education sector formed 10.15% of the national budget, equivalent to 3.2% of Afghanistan's GDP (Ministry of Finance, 2019). Indeed, over 70 percent of the existing school infrastructure, was established in the post-2001 era and based on the World Bank estimates, gross enrollment increased from 20.8% in 2001 to 106.7% in 2019 (World Bank, 2021a). However, despite the remarkable progress, only 21% of girls and 43% of boys at school-going age attended school, and 51% of schools lack adequate building structures (Ministry of Education, 2016).

The government implemented reforms in higher education. During their six-years tenure, the Taliban bared women from education at all levels. Depriving women of education and employment, led to an exodus of female teachers and professionals. They also changed curriculum by replacing standard subjects with theology and Islamic Sharia. Afghanistan's constitution (2004) guarantees access to public education for all citizens up to an undergraduate degree without any charge (The Ministry of Justice of Afghanistan, 2005). The post-Taliban government built physical infrastructure, introduced standard school curriculum, and more importantly, opened schools to women. In 2009, 19 higher education institutions were functional while only five universities with limited number of majors and minimal physical infrastructure were active in the 2001. In the later years, a higher education institution was established in 34 provincial capitals. The higher education law passed in 2007 also legalized private tertiary education. (Abdulbaqi, 2009).

University admission was competitive and students were placed based on the national university entrance test known as Kankor.[1] Despite efforts towards improving access to education, enrollment rates in higher education in Afghanistan remained low. Gross enrollment in tertiary education increased from nearly 1% in 2001 to 10% in 2020 (World Bank, 2021b). Based on the World Bank estimates, in 2020, the average gross enrollment in tertiary education was 40% in the world, 26% in South Asia, and 9% in low-income countries.

There were two major sources of inequalities in Afghanistan's higher education system. First, there was the urban/rural divide. Most university students came from Afghanistan's capital, Kabul, and three other major cities. Based on our calculations using Afghanistan's university entrance exam data, Kabul, Balkh, Herat and Nangarhar accounted for 45% percent of university students in 2018 while in the same year, those provinces constituted 32% of Afghanistan's population (National Statistics and Information Authority, 2019). This inequality was partly driven by unequal distribution of educational inputs. For instance, rural schools have 15% lower number of classrooms per student, 34% less computer labs, and 19% less electricity supply (World Bank, 2018). Similarly, in 2018, there was a substantial learning gap between rural and urban students (World Bank, 2018). Students could take the Kankor test upon their graduation from grade 12. A combination of the number of available seats, demand and the distribution of applicants' test scores would determine the cutoff score for admission to a particular program in a particular year.[2] The high stake associated with

[1]Kankor is a standard multiple-choice test that covers high school materials similar to the Scholastic Aptitude Test (SAT). Math and science constitute a major portion of the test.

[2]A program is a combination of the field and university. Every year, the number of seats available for each

the Kankor exam scores led to the proliferation of private tutoring facilities and these were more prevalent in urban centers. Combined with better quality schools in urban areas, the relative availability of tutoring centers exacerbated the educational gap between urban and rural areas.

Then, there was the gender gap. Under the Taliban regime, women were prohibited from going to school. During the U.S. presence in Afghanistan, the government of Afghanistan and donor community placed a significant attention on woman empowerment. Between 2002-2020, the USAID alone spent $787.4 million on women empowerment programs, in addition to hundreds of other programs that included a gender component (SIGAR, 2021). Promote, the largest USAID women empowerment program, had a component to increase women enrollment in higher education (Feminist Majority Foundation, 2014; SIGAR, 2021). Female students received preferential treatment with dormitory facilities in public universities. However, the impacts of those policies on access to higher education had been limited, especially in rural areas. Consider the need for Kankor preparation. Female students in rural areas were disadvantaged because cultural norms limited their mobility to urban centers without proper dormitory facilities. Hence, female students from disadvantaged communities received lesser share of the available seats at higher education institutions.

2.2.2 Tutor For Rural Afghanistan: The PEZAK Program

Evidence from high-income countries shows that academic preparation and outreach programs can bridge the gap between relatively better-off and disadvantaged students. For instance, Upward Bound, one of the largest federal programs in the United States that was

program was announced but the cutoff score was endogenously determined by the demand for a program, and was therefore unknown at the time of the test. Students ranked their three preferred program. Students may know the cutoffs for previous years but they do not know the standards for the year in which they are taking the test. Admission was based on competition (in terms of the exam scores) among those applying to the same program. If a student's score failed to qualify for her first choice, the allocation algorithm considers her second and third choices. Usually, medical fields, engineering, law, and economics were highly competitive. Furthermore, programs in universities in large cities such as Kabul, Herat, Balkh, and Kandahar were very competitive.

designed towards the academic preparation of disadvantaged students to enter and complete college, has had a positive impact on enrollment to a post-secondary program and college completion (Seftor et al., 2009). PEZAK was a private initiative aimed at reducing access inequalities in higher education in Afghanistan.

PEZAK was a female-inclusive university-entrance exam preparation program that targeted high school students (grades 10-12) in Central Afghanistan's remote communities. Established in 2008, its objective was to help these students succeed in the Kankor exam. PEZAK recruited bright university students from Afghanistan's top universities as exam trainers. These trainers would be dispatched to these remote areas to tutor high school students in math and science — the two most important subjects in Kankor. They also provided information about university degrees, program selection, and logistics on the entrance test. On average, the trainers would stay on their assigned destination during the winter season for around 3 months.

The program — which was initiated by a former Afghan Minister of Parliament (MP) — was an informal collaboration between the various spectrum of the community. The MP's office managed the program, the recipient community provided accommodation and food to the trainers, and the local businesses provided financial support to the program. Financing was provided in part by Afghan residents in Canada and Europe who hailed from the same communities that the program served. Throughout the lifetime of the program, PEZAK paid an average of $400 per trainer for the three months of training, which covered the service fee and transportation costs.

PEZAK operated in rural districts with majority Hazara ethnic group because the MP belongs to the same ethnicity (see Figure B.1). Interested communities in these districts could submit a proposal to the MP office. Proposals would be evaluated based on the number of potential beneficiaries (i.e., students), accessibility to female students, and the community commitment to host the trainer. If successful, the community would be assigned a trainer and the training would be held either in a school or a mosque in the village. Successful communities would still need to apply in the following year to renew their participation. At the end of a cycle, program administrators would hold a meeting where trainers would report the extent of community support for the program. They would then decide whether to renew an existing program based on the community's proposal and the trainer's report. Communities with an existing program that was not renewed could reapply in the following year. The number of communities selected varied over the year: Between 2008–2019, 184 PEZAK schools in 46 districts received the program at least once (see Figure B.2). On average, 60 schools received the program each year. The replacement rate of PEZAK schools in two consecutive year was 20%. On average, two years and three years later, 30% and 40% of treated schools were new recruits, respectively. In other words, the rate of turn-off between two years is 20%. However, since grades 10-12 participate in the program, even a school was not in the program for two years after it was treated, the school could still be considered treated because the potential test takers benefited from the program.

Trainings were typically conducted during winter (between the months of December and March) when schools were off for three months. Winter in the Central regions of Afghanistan is harsh with temperatures ranging between -12.4 °C and + 2.7 °C. Snowfalls are usually present for about four months. Extreme weather conditions combined with an underdeveloped road system and inadequate modes of transportation make traveling across distant villages difficult, especially for women and children. This constraint minimized potential spillover effects, since students from a non-PEZAK school could not travel to a PEZAK school to benefit from the training. This was particularly true for female students.

2.3 The Data

We combined the data from the PEZAK administrators with two large administrative datasets obtained from the government of Afghanistan. The PEZAK database includes the list of schools that received the PEZAK program, the trainers' education qualifications, and the fee paid to the trainers. We collected information from the 2008–2019 period from the program administrator's archive; however, some information from the program's early years of implementation was missing. Although PEZAK started in 2008, they did not have records of the list of schools that participated between 2008 and 2010.

Our outcome variables were constructed from a large administrative dataset of the Kankor results between 2003 and 2019. The dataset came from the Ministry of Higher Education and contained the universe of exam takers. It contains information on student gender, year of graduation from high school, year of taking the test, school name, test scores and test result which indicates the field of study and the collage that the student was admitted to. 30% of the test takers between 2003-2019 were female.

Finally, to obtain the school characteristics variables, we linked these datasets with the school census database. The data was obtained from the Ministry of Education for 2011–2018. It provides information on school physical characteristics, human resources, and enrollment statistics for the universe of 17,702 schools in Afghanistan.

We matched the Kankor and school administrative datasets using school names. First we matched the two datasets with the help of STATA software. String matches with a similarity score of 100% was separated as perfect match. Because of the multiple patterns in the way names are written in Afghanistan's local languages, Farsi and Pasho, and lack of standard reporting of names in the two datasets, we were able to match only 30% of the data. To match the remaining data, we employed a manual data matching approach. For this purpose, in collaboration with Porsesh Research and Studies Organization (PRSO) in Kabul, we hired nine university students during June-August 2019 and 2020 to read names in two datasets and record their confidence level in the match on a scale of $1-3$, which represents perfect match in names, 50% match and no match. In the matching procedure, school names in the two datasets were matched at year-province level to make sure schools in one province were not matched with the same-name schools in a different province. In addition to school names, we also used district and village names as additional guiding identifiers. Each validator worked on one province. We rigorously trained data validators. Furthermore, we employed three layers of quality control. First, a data set validated by one validator was randomly assigned to another validator who would check the quality of match and record errors. On average, the error rate in this round was 5%. Second, all validated data was re-distributed for a second validation. Third, we also validated their work randomly notified the research assistants about the errors committed. Combined with software matching, we were able to match 97% of data.

2.4 Empirical Strategy

2.4.1 Identification

To identify the treatment effects of PEZAK program on student outcomes, we exploit variations in the timing of the program roll out. The number of treated schools between 2008-2019 is shown in Figure 2.1. On average, each year the planner maintained 62 treated schools by rotating the treatment assignment among 184 schools. Treatment switches on and off multiple times. To illustrate the nature of fluctuations in treatment, let us consider four time periods T, $T+1$, $T+2$ and $T+3$. Let the share of new treated schools in each period be the ratio of number of schools that are treated in periods $T+1$, $T+2$ and $T+3$ but were not treated in period T to the total number treated schools in each of the successive periods. This is presented in Figure 2.2. On average, the share of new treated schools in successive years is approximately 20%. This means that in period $T+1$, 80% of treated schools are those that were treated in period T as well and 20% are new schools. Two years later in $T+2$, on average 31% of treated schools were not treated in period T. Similarly, three years later in $T+3$, 44% of the treated schools are new schools relative to period T.

The replacement rate after one year, two years, and three years for each period is shown in Panel (b) of Figure 2.2. 2008-2011 was a pilot period in which the same 54 schools were treated each year. The switching rates in each year resemble the average numbers discussed earlier.

Furthermore, treatment in one period affects three cohorts. As discussed in Section 2.2, PEZAK classrooms are consist of grades 10-12. Students in grade 12 take the university entrance examination immediately or few months after the training. Students in grades 11 and 10 take the university entrance exam one year and two years after exposure to the program. Therefore, if a school received the program in year T but switched out in years $T+1$ and $T+2$ we consider them in the treatment group because the grades 11 and 10 cohorts who graduate in $T+1$ and $T+2$, respectively, were already exposed to the program in period T. By applying this rule, the program converges to a staggered design. To illustrate this, in Figure 2.3 we present the average treatment frequency between 2008-2019 in three scenarios. In the first scenario, average treatment frequency is computed considering treatment as it happens. In the second scenario, in $T+1$ treatment is applied to schools that received the program in year T. Similarly, in the third scenario, once schools receive treatment in year T treatment is applied to them in the following two years as well. Schools that were treated throughout the period should have 12 years treatment. In the third scenario the average treatment frequency is approximately 8 times. The third scenario is more plausible and we use it for our identification.

To estimate the treatment effects of PEZAK on student outcomes, we employ a fully flexible fixed effects estimator that computes instantaneous and dynamic treatment effects. Let y_{st} denote student outcomes averaged at school level for school s at time time t, and μ_s and λ_t denote school fixed effects and year fixed effects. Also, consider E_s to be the time school s was exposed to treatment for the first time and μ_l denote time distance from first treatment year. We estimate the following dynamic event study specification:

$$y_{s,t} = \sum_{l} \mu_l 1\{t - E_s = l\} + \mu_s + \lambda_t + \varepsilon_{s,t} \tag{2.1}$$

$\varepsilon_{s,t}$ captures unobserved time-varying school characteristics. The relative times $l = t - E_s$ covers all possible dynamic effects. At $l = 0$ the specification gives instantaneous effects and for all $l > 0$ coefficients provides the dynamic effects l periods after first exposure to treatment. The coefficients for the relative times before first exposure to treatment ($l < 0$) Equation 2.1 presents placebo estimates which compares treatment and control groups l periods before first treatment.

There are two challenges with the above specification. First, Sun and Abraham (2021) shows that estimate of μ_l is a linear combination of cohort-specific relative periods l and other relative periods. Due to potential treatment heterogeneity across cohorts and varying dynamic effects, treatment effects from other relative periods contaminates the estimate of μ_l. Therefore, unless strong assumptions about homogeneity of treatment effects are satisfied, estimates of μ_l is likely to be biased. Second, Equation 2.1 requires a staggered design where treatment stays until the end of the panel once it occurs. As explained earlier, PEZAK does not follow a clean sequential treatment assignment. The program design is such that treatment turns on and off multiple times.

To overcome both concerns, we employ the generalized estimator introduced by de Chaisemartin and D'Haultfoeuille (2022) that is robust to heterogeneous treatment effects. This approach accommodates all forms of treatment assignment including non-staggered binary treatment, which is the case in our design. It computes the $\boldsymbol{DID}_1$ estimates which is the weighted average of difference-in-difference (DID) estimates across all time periods and possible values of treatment. The DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those that remains 0 in $t-1$ and t. Similarly, DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those whose treatment is equal to 1 in both periods. The possibility of negative weight occurring is high in the later case, especially when treatment effects transmits from earlier periods to the next periods. $\boldsymbol{DID}_{\text{l}}$ estimates the effect of having switched treatment for the first time l periods ago. These estimators are robust to heterogeneity and dynamic effects.

Another utility of the estimator that we use in this paper is clear interpretation for all estimates. Coefficients for periods before the first treatment (relative periods with negative sign) are placebo estimates that compares switchers and non-switchers before the switch. Those can be used to test for parallel trends before the intervention. Period 0 shows instantaneous treatment effects and the rest of the periods show dynamic treatment effects. The dynamic effects are interpreted as the effects of of having weakly higher number treatment between first exposure and l periods after the first exposure. Furthermore, as in two-way fixed effects estimators (TWFE), this estimator also allows to compute the overall effects per unit of treatment by aggregating the instantaneous and dynamic effects. This could be interpreted as the average total effects per unit of treatment, which may be used for cost-benefit analysis.

In the following section we examine potential threats to our identification strategy. First, we explore outcome evolution in the raw data. Second, we inspect potential sources of endogeneity. Third, as a validity check, we explore risks of contamination because of negative weight, a concern raised by several authors including Goodman-Bacon (2021), de Chaisemartin and d'Haultfoeuille (2020), Callaway and Sant'Anna (2021), and Sun and Abraham (2021).

2.4.2 Identification Validity Checks

A. Parallel Trends

First, we compare school characteristics with the available data. School characteristics data is not available for the years 2003-2010. We use the data for 2011, the closest available data to the start of the program, to compare treated and non-treated schools. As presented in Table 2.1, treated and control schools are comparable in terms of location, climate, instruction language, and access to drinking water. In Table 2.2, we also compare continuous school characteristics in 2011. Most of these characteristics are time-invariant or may not vary for several years. Some imbalances were present in characteristics such as distance from a hospital, school age, average class size, teacher-to-student ratio, total enrollment, and toilet facilities. The average distance from a hospital was lesser for treated schools, treated school age was older and larger. Since those characteristics are fixed in nature, we expect that the school fixed effects account for time-invariant heterogeneity.

Second, we explore differences in outcome trends. Figure 2.4 and Figure 2.5 compares evolution of test scores, passing rate, enrollment in top universities, and enrollment in low-rank universities in PEZAK schools and non-PEZAK schools, all averaged at the school level. In both figures, two lines distinguish never-treated schools (non-PEZAK schools) and 184 PEZAK schools that were treated at least once during 2008-2019. It is worth noting that these comparison are based on the raw data. A formal test of difference in trend before the intervention is provided with placebo tests in the results section.

Panel (a) of Figure 2.4 shows school-level average test scores. Raw data shows that from 2003 to 2007 school-level average test scores in PEZAK and non-PEZAK schools was similar. It started to diverge in 2008. Panel (b) of Figure 2.4 shows the trend for average passing rate. Passing rate is the ratio of students that were enrolled in any public university program to total test-takers. PEZAK schools seem to have higher average overall passing rate throughout the panel. But there does not seem any change the pattern after the start of the program.

Figure 2.5 compares enrollment in top-rank and low-rank public universities between

PEZAK and non-PEZAK schools. Panel (a) of Figure 2.5 demonstrates the pattern for average enrollment in top universities, defined by the ratio of students enrolled in top universities to total test-takers. Similar to test scores, until 2007 the evolution of enrollment in top-rank universities seems similar before the intervention and PEZAK schools start to have higher enrollment rate in 2008 and beyond. Panel (b) of Figure 2.5 plots the trend in average enrollment in low-rank universities, defined by the ratio of students enrolled in low-rank universities to total test-takers. There is no clear path but it seems PEZAK schools, on average, had a higher enrollment rate in low-rank universities.

B. Self Selection

Self-selection to PEZAK schools could be a threat to our identification. As explained in the description of the program, the program is implemented in the Central region of Afghanistan during the winter season. With heavy snowfall and poor infrastructure, in that region, commute between villages far apart is challenging. However, with anticipation of the program, students could switch from an untreated school to a treated school. To address this concern, we regress student enrollment on time and interaction of treatment with time. Our test rejects the hypothesis that treatment had a positive impact on enrollment. Treatment did not change high school enrollment and grade 12 enrollment. The result of this regression is presented in Table 2.3. We confined this analysis to the years for which enrollment data is available.

C. Exogenous Time of Treatment

An important characteristic of the PEZAK program is non-staggered implementation. There is no clear information about the reasons for dropping schools once they have been in the program. Available data about the program shows that the program selected relatively larger schools. This is consistent with the goal of the program to benefit more students. If the time of the program is selected deliberately, that might bias our results. To address this concern, we run a probit regression with the treatment as the outcome on the school characteristics including two lags of school characteristics. The idea is to check if the observed past school characteristics affect treatment assignment. As shown in Table 2.4, we do not see any significant effect of past characteristics on treatment assignment. This indicates that the treatment time was random.

D. Heterogeneity of Treatment Effects and Negative Weights

The estimator that we use to estimate treatment effects of PEZAK on student outcomes is a weighted average of several combinations of two-by-two canonical difference-in-differences (2x2 DD). In the case of TWFE it has been shown that the estimate might be biased due to heterogeneity of treatment across cohorts and over time (Goodman-Bacon, 2021; de Chaisemartin and d'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021). We employ a heterogeneity-robust estimator that alleviates such concerns. Yet, we examine the extent that negative weights might be a concern.

First, we test for potential sources of bias due to negative weights with Bacon decomposition introduced in Goodman-Bacon (2021). Bacon decomposition estimates all possible 2x2 DD estimates and their associated weights. It estimates 2x2 DDs for three sets of treatment and control groups. The first set is a comparison of early treated cohorts versus late cohorts where groups treated late serve as controls. The second set of comparison is between ever treated groups and never treated groups. The last comparison is between late treated versus early treated cohorts where early treated groups serve as controls. The possibilities of having negative weight is high in the later case because of taboo controls. If the treatment effect is not constant across cohorts, earlier treated groups can introduce a negative weight such that the sign of individual treatment effects is opposite of sign of overall treatment effects. It is worth noting that Bacon decomposition is possible only with staggered treatment. For the purpose of this decomposition, we make an assumption that treatment is non-decreasing so PEZAK design resembles a staggered roll out.

Figure 2.6 and Figure 2.7 plots 2x2 DD estimates against their respective weights from Bacon decomposition. In Figure 2.6 average DD estimates for three sets of DDs mentioned above is presented for test scores. The average coefficient for early versus late treated groups is 0.05 standard deviations and its associated weight is 0.06. Similarly, the average estimate for later versus early DDs is 0.09 and its corresponding weight is 0.04. With an average estimate of 0.17 and weight of 0.90, the largest share comes from comparison of treated versus never treated groups. Figure 2.7 plots Bacon decomposition for enrollment in top universities, enrollment in low-rank universities and overall passing rate, all standardized. For all outcomes, some individual weights are negative but very small. This indicates that even with the use of TWFE with a staggered design, the level of bias is likely to be very small in our estimates.

Finally, in Figure 2.8 we present weight decomposition using the decomposition method of de Chaisemartin and d'Haultfoeuille (2020), in which weights that each treatment cohort carries over time are compared in staggered and non-staggered designs. Cohorts are tracked over time. Cohort is defined based on the year in which school received treatment for the first time. The Y-axis represents weights that each cohort carries over time. In Panel (a) weights are calculated from a staggered adoption design where treatment is non-diminishing. In other words, once units are treated, they remain treated for ever. In this design, under the common trends assumption, TWFE estimate is a weighted sum of 48 ATTs such that 39 ATTs receive a positive weight, and 9 receive a negative weight. The sum of the negative weights is equal to -0.38. Panel (b) presents weights from a binary treatment where treatment turns on and off. Dynamic effects are estimated based on the first treatment. Under the common trends assumption, TWFE estimates a weighted sum of 45 ATTs where 32 ATTs receive a positive weight, and 13 receive a negative weight. The sum of the negative weights is equal to -0.12. The minimal value of the standard deviation of the treatment effect across treated groups and time periods under which beta and the ATT could be of opposite signs is equal to 0.03 in Panel (a) and 0.05 in Panel (b). In both panels, the 2008 cohort has the largest share of negative weights. Therefore, the TWFE estimates are more biased in non-staggered design than staggered design.

The non-staggered design with binary treatment is our preferred design. Based on the weight analysis described above, negative weight might pose some threats and bias our results. But the estimator that we employ mitigates such concerns. In the following section we present results from estimator developed by de Chaisemartin and D'Haultfoeuille (2022).

2.5 Results

2.5.1 Main Results

We estimate the effects of PEZAK on student outcomes using the heterogeneity-robust estimator of de Chaisemartin and D'Haultfoeuille (2022) described in Section 2.4. Instantaneous and dynamic effects are computed and presented in event study graphs similar to Equation 2.1. All estimates are relative to -1, the period immediately before first exposure which is omitted from the graph to avoid multi-collinearity. Period 0 shows the instantaneous treatment effect. Periods to the right of 0 demonstrate dynamic effects and periods to the left plot the placebo estimates, which is the difference in outcome between treated and untreated schools one period before anyone was exposed to the program.

In Figure 2.10 estimates are presented for four outcomes: test scores, enrollment in top-rank universities, enrollment in low-rank universities, and overall passing rate. Originally, test scores is a continuous variable that ranges between 0 and 360. Enrollment in top-rank universities is the percentage of total test takers who secured seats in one of the top-five universities in Afghanistan. Similarly, enrollment in low-rank universities is the percentage of test takers who were admitted to low-rank provincial universities. Passing rate is the share of test takers who were enrolled in any public university. However, in our estimates standardized versions of outcome variables are used.

The instantaneous effects of PEZAK on test scores, enrollment in top-rank universities and enrollment in low-rank universities is positive in magnitude but statistically insignificant. In the year of exposure to the program, on average, PEZAK schools scored 0.06 (s.e.=0.078) standard deviations higher on the test. However, the 95% confidence interval for this estimate is $(-0.08 - +0.22)$ and the corresponding t value is 0.85. Similarly, in the first year of exposure to the program, the estimate for the effect of PEZAK on enrollment in top-rank universities is 0.17 (s.e.=0.13) standard deviations. The corresponding 95% confidence interval for this estimate is $(-0.09 - +0.43)$ and the computed t score is 1.3. The treatment effect on enrollment in low-rank universities is 0.05 (s.e.=0.07) standard deviations. The 95% confidence interval and t score associated with this estimate is $(-0.08 - +0.19)$ and 0.71. On the contrary, the instantaneous treatment effect of the program on passing rate is 0.14 (s.e.=0.07) standard deviations and statistically significant. The corresponding 95% confidence interval with this estimate is $(+0.004 - +0.29)$ which gives a t score of 2.

The dynamic effects of PEZAK on student outcomes is positive and strongly significant for test scores. One year after the first exposure, on average, treated schools had 0.15 (s.e.=0.05) standard deviations higher test scores. The improvement in test scores continues to 0.19 (s.e.=0.07) standard deviations and 0.20 (s.e.=0.08) standard deviations two years and three after the first exposure, respectively. The magnitude of the estimates on enrollment in top-rank universities also increase after the first year of exposure but the effect is statistically significant only in the third period (estimate = 0.24 SD; s.e. = 0.11). The magnitude of the estimate for enrollment in low-rank universities increases to 0.08 (s.e.=0.11) and 0.15 (s.e.=0.13) one period and two periods after the first treatment, respectively. It reduces to 0.04 (s.e.=0.09) in the third year. Passing rate also follows similar trajectory.

One period after the first exposure the estimate on passing rate is 0.15 (s.e.=0.10) standard deviations. In the second and third period after first treatment, the treatment effect on passing rate falls back to 0.04 (s.e.=0.11) and 0.01 (s.e.=0.09), respectively. Therefore, with some differences in the time lag, the dynamic effects of the program was positive on all outcomes. For test scores and enrollment in top universities this effect was stronger and statistically more precise.

The placebo estimates are very small and insignificant for all outcomes. Two years before the first exposure the placebo estimate is 0.003 (s.e. = 0.069) for test scores, 0.02 (s.e. = 0.13) for enrollment in top-rank universities, 0.14 (s.e. = 0.10) for enrollment in low-rank universities, and -0.05 (s.e. = 0.12) for passing rate. The 95% confidence intervals are large which make the estimated differences fall in the null hypothesis range. The insignificant and small placebo estimates supports our parallel trends assumptions discussed in Section 2.4.

The average total effects of the program are presented in Table 2.5. In this table the instantaneous treatment effect (period 0) and dynamic effects (6 periods after the first exposure) is averaged. The estimate of interest in this table is the third row, which is the results for binary and non-staggered design in a scenario that treatment affects grades 10-12. The average total effect of the the program is 0.17 (s.e. = 0.14) standard deviations for test scores, 0.26 (s.e. = 0.12) for enrollment in top five universities, 0.07 (s.e. = 0.07) for enrollment in low-rank provincial universities, and 0.10 (s.e. = 0.11) for overall passing rate. Among all outcomes, the average effect on test scores has the highest statistical significance. The average total effects maybe interpreted as effect per unit of the program for receiving a weakly higher number of treatment. The other three rows will be discussed in Section 2.5.3 in detail. But briefly, the first row presents average total estimates for a situation where treatment is flexible. In the second row, treatment assignment is such that treatment at T $\Rightarrow$ treatment at T+1. The last row is a staggered adoption assignment where treatment is non-diminishing once it occurs. The estimates are positive in all scenarios.

2.5.2 Heterogeneous Treatment Effect

In this section, we examine whether the PEZAK program had any differential treatment effects on male and female students. In Afghanistan, due to cultural norms, certain freedoms are proscribed for women. They have less or no flexibility to travel alone and live away from their family. While male students in both treated and untreated schools may have other options to prepare for the university entrance examination, such as traveling to larger cities, the PEZAK program that provides tutoring facilities close to students' homes, might benefit female students more than male students. Moreover, the program places special attention on female participation.

To evaluate the potential differential impacts of the program, we use a TWFE estimator because of its flexibility to use interaction terms. With the heterogeneity-robust estimator of de Chaisemartin and D'Haultfoeuille (2022) that we utilized to estimate the main results, one needs to aggregate the data at the level that intervention occurs, school in our case. Even with the use of student data in separate samples, testing the differences is not possible. Acknowledging some level of downward bias due to negative weights, we use a TWFE estimator to estimate the differential treatment effects of the program on female and male students. For this purpose, we make an assumption that treatment does not perish and stays once it is introduced. We assume a staggered design. Moreover, we use the student data and cluster the standard errors at the school level. Therefore, we estimate the following TWFE specification:

$$y_{i,s,t} = \beta_0 + \beta_1 D_{s,t} + \beta_2 D_{s,t} \times Female_i + \gamma X_{i,s,t} + \mu_s + \lambda_t + \varepsilon_{i,s,t} \tag{2.2}$$

where $y_{i,s,t}$ is the outcome for student i in school s at time t. $D_{s,t}$ is the treatment indicator for schools s at year t. $D_{s,t} = 1$ if school s was treated at time t. This is in fact the $Post \times Treatment$ in the canonical two-period difference-in-differences framework. $Female_i = 1$ if student is female and 0 he is male. $X_{i,s,t}$ is a vector of student characteristics for student i at time t. In our data student gender and test retaking status are the two student-level characteristics. μ_s represent school fixed effects, and λ_t denote year fixed effects. $\varepsilon_{i,s,t}$ captures the unobserved student features that we can not capture. The coefficient of interest in Equation 2.2 is β_2, which is the difference in outcome for female treated students.

Table 2.6 presents the results for Equation 2.2. In this table, $Post \times Treatment$ and $Post \times Treatment \times Female$ stand for β_1 and β_2, respectively. On average, female students in general have substantially lower outcome as compared to male students. Students who retake the test have significantly higher success on the test. The term $Post \times Treatment \times Female$ is negative for enrollment in low-rank universities and significant at 5% significance level. However, we do not see evidence for differential impact for other outcomes. It should be noted that enrollment and passing rate are binary variables in this specification and we run a linear probability regression to estimate the differential treatment effects. Enrollment in top-rank universities is equal to 1 if student i secured a seat in one of top five universities. Enrollment in low-rank universities is coded the same way. Passing is equal to 1 if student i was enrolled in any public universities.

Therefore, we do not observe a substantial differences in treatment effects between male and female students. The only instance of a difference is enrollment in low-rank universities. Female students had lower probability of enrolling in low-rank universities.

Table A.1 and Table A.2 in the Appendix present results from our main strategy for male and female separately. Results for male students are consistent across different estimators. However, the average total effects for female students are different in de Chaisemartin and D'Haultfoeuille (2022) and TWFE. In the former, the coefficient on test scores, enrollment in top rank universities and overall passing rate is positive and negative for enrollment in low-rank universities. However, all of the results are statistically insignificant. On the hand, TWFE with school average outcomes are small in magnitude and statistically insignificant.

The source of differences in results for female students could be unbalanced panel. The number of female students before 2008 was substantially low. In such circumstances, TWFE with student-level data is more suitable.

2.5.3 Robustness of Results

A. Staggered Design with CH Estimator

In this section we examine the robustness of our results with the CH Estimator, i.e., estimator of de Chaisemartin and D'Haultfoeuille (2022), by changing the treatment assignment to a staggered design. In Section 2.5.1 we discussed results for a treatment regime where treatment at period $\boldsymbol{T}$ affects three cohorts and hence treatment at $\boldsymbol{T}$ implies treatment at $\boldsymbol{T}+1$ and $\boldsymbol{T}+2$. In this regime, treatment can turn on and off after the first exposure. On the other hand, in a staggered design once units receive treatment they do not forget it, which implies subjects will remain in treatment after their first exposure to the intervention. Some degrees of downward bias is expected from years that units are not treated but otherwise considered treated.

Figure 2.11 presents results for this strategy. The instantaneous and dynamic effects presented in this figure is similar to results in Figure 2.10. In the first period of exposure, the school-level average outcomes experienced an increase. However, among the four outcomes, the instantaneous effect is statistically significant only for passing rate. For test scores, enrollment in top-rank universities and enrollment in low-rank universities, the 95% confidence interval is large which includes possibilities of zero difference. One period after the first exposure, and the following periods, the effects become stronger and statistically significant. Therefore, our main results are close to estimates from a staggered design. It indicates that effects of the program persists and strengthens after the first year of exposure.

The average total effects of PEZAK program with a staggered design is presented in Table 2.7. In the first three rows different variants of binary and non-staggered designs are presented. The third row corresponds to the main estimates discussed in Section 2.5.1. In the fourth row, estimates are from implementation of de Chaisemartin and D'Haultfoeuille (2022) with a staggered roll out. The average total effect of the program is 0.14 (s.e.=0.06) standard deviations on test scores, 0.21 (s.e.=0.10) on enrollment in top-rank universities, 0.06 (s.e.=0.07) on enrollment in low-rank universities, and 0.11 (s.e.=0.08) on passing rate. As expected all estimates in the staggered design are smaller than non-staggered regime. Overall, the instantaneous, dynamic and average total estimates in this approach confirms our main estimates.

B. Staggered Design with CS Estimator

We examine our results by employing the CS estimator, i.e., the estimator by Callaway and Sant'Anna (2021), to estimate the effects of PEZAK program on student outcomes. This approach estimates the treatment effects in three steps: First, all group-time average treatment effects, i.e., the average treatment effect for group g at time t is identified and computed. Groups refer to cohorts based on their first treatment. Similar to de Chaisemartin and D'Haultfoeuille (2022), Callaway and Sant'Anna (2021) also estimate treatment effects in all possible 2x2 DDs. Never treated or not-yet treated units can be used as control for all cohorts and years. Second, this approach proposes several aggregation schemes: aggregation by cohort, by event period, by calendar time, and overall average treatment effects. Lastly, the estimator uses a bootstrap method to compute standard errors. This estimator is also heterogeneity-robust but it is different from de Chaisemartin and D'Haultfoeuille (2022) for it requires a staggered design. In our implementation of this method, similar to the approach in Section 2.5.3, we consider a staggered roll out.

Figure 2.12 presents results from this estimator. To make results comparable with the main results we aggregate the estimates for each relative period. Relative periods are time distance from the first treatment period. In the first year of exposure, estimates for all outcomes demonstrate an increase but the estimate is statistically significant only for passing rate. In the following years, the effect becomes stronger for test scores and enrollment in top-rank universities. The average total effects of the program on student outcomes are shown in the fifth row of Table 2.7. The average effect is 0.15 (s.e.=0.06) standard deviations on test scores, 0.23 (s.e.=0.09) on enrollment in top-rank universities, 0.06 (s.e.=0.07) on enrollment in low-rank universities, and 0.10 (s.e.=0.08) on passing rate. These estimates are similar to estimates from staggered roll out with the estimator of de Chaisemartin and D'Haultfoeuille (2022) and smaller than the results discussed in Section 2.5.1 because of the potential downward bias in a staggered design.

C. TWFE Estimator

We also apply a two-way fixed effects (TWFE) estimator to estimate the treatment effects of PEZAK program on student outcomes. In this method, assuming that parallel trends before the intervention hold, in the most basic form, the treatment effect is estimated by including time fixed effects, unit fixed effects, and an indicator for treatment. Equation 2.1 discussed in Section 2.4 is one form of TWFE that estimates treatment effects in all relative times after the first exposure. We estimate Equation 2.1 to compare the dynamic treatment effects with our main results. Furthermore, using Equation 2.3 described below we estimate the average treatment effects of the program.

$$y_{s,t} = \beta_0 + \beta_1 D_{s,t} + \mu_s + \lambda_t + \varepsilon_{s,t} \tag{2.3}$$

where $D_{s,t}$ is the treatment indicator which is equal to 1 if if school s received the program at time t. μ_s and λ_t are school fixed effects and year fixed effects, respectively.

Figure 2.13 shows results from estimating Equation 2.1 in a TWFE event study study framework. With TWFE estimator, the instantaneous treatment effect seems weaker as compared to our main results. In the first year of exposure none of the estimates are statistically significant. One year later, the magnitude of the effects increase for most of the outcomes but only test scores and passing rate are marginally significant. The confidence intervals are larger than those in our main estimates.

Average total effects from applying Equation 2.3 are given in rows 6 through 9 of Table 2.7. We apply the TWFE estimator with several treatment regimes. In row 6 treatment at time T only affects cohort T. In the row 7 treatment at $T-1$ implies treatment at T which means exposure to treatment affects at least two cohorts. In row 8 treatment at $T-2$ affects treatment at T which implies exposure to treatment affects at least three cohorts. In row 9 treatment is staggered which means once schools are treated they will not forget the intervention. Among several variants of treatment regimes, the staggered design is the closest to our main estimates. With a TWFE implemented on a staggered design, the average total effect is 0.16 (s.e.=0.03) standard deviations on test scores, 0.20 (s.e.=0.06) on enrollment in top-rank universities, -0.03 (s.e.=0.07) in enrollment in low-rank universities, and 0.09 (s.e.=0.06) on passing rate.

Recent literature on difference-in-differences has shown that the estimate from TWFE estimators are prone to bias because they average several combinations of 2x2 DDs, some of which might compare treatment groups with bad controls that produce negative weights. Hence the individual DDs might be of an opposite sign as compared to the overall estimate (de Chaisemartin and d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun and Abraham, 2021). In Section 2.4.2 using Bacon decomposition and weight decomposition proposed by de Chaisemartin and d'Haultfoeuille (2020) we show that out of 48 DDs 9 comparisons carry negative weights. This could be exacerbated with non-staggered design where treatment turns on and off. Therefore, the closest natural estimate from a TWFE estimator to our main estimates is when TWFE is applied to a staggered design.

Overall, the estimates from de Chaisemartin and D'Haultfoeuille (2022), Callaway and

Sant'Anna (2021) and TWFE estimators implemented on a staggered design are similar to our main results from de Chaisemartin and D'Haultfoeuille (2022) applied on non-staggered treatment. This confirms that our results are robust to several variants of new difference-in-difference estimators.

2.6 Cost Effectiveness

In 2018, the program cost USD 500 per trainer or USD 8 per student. The cost of the program is split between the MP's office and the community. This includes USD 400 compensation for tutoring services paid by the MP's office and approx. USD 100 for heating and food, born by the community. This expenditure is estimated based on the price of bread (the main food staple in Afghanistan) and other essential commodities. The program increased test scores, on average, by 0.17 standard deviations. This implies that the program costs USD 4.7 per student per 0.1 standard deviation increase in test scores.

The program pays slightly above the market rate. The average entry-level monthly salary in Afghanistan's public sector, for individuals with an undergraduate degree, is USD 111.32. The community contribution towards food and accommodation may be considered equivalent to the non-monetary benefits to the public services employees. The base payment without community contribution to trainers is approx. 20% higher than public service salary scale for equivalent skills. Based on the 2016 Survey of the Afghan People, the average monthly household income in Afghanistan was estimated at USD 165 (Asia Foundation, 2016). This means the PEZAK tutors are paid 24% less than the average household income. Therefore, the remuneration is competitive enough to recruit talented university students for the program.

The financial opportunity cost to the trainers is likely low. We do not have statistics about temporary employment and employment sectors for college students. But information on the broad state of the economy may substantiate our argument about the opportunity cost. In

2020, the unemployment rate in Afghanistan was 11.73% (The World Bank, 2020). In 2016 51.5% of Afghan cited unemployment as the main reason to leave the country if they could. This means that finding a temporary job for college students could be challenging. Hence, a well-organized program can mobilize resources that would otherwise remain unemployed, to increase enrollment of disadvantaged communities in the universities.

In 2013, the government of Afghanistan started an alternative program to bridge the academic gap between rural and urban students. She organized a tutoring program by hosting students from the provinces at training centers in the capital. The PEZAK program reached 5,166 students in 2013. In a hypothetical situation, if the government would bring the same number of students to the capital, in a very conservative estimation at the lower bound, multiplying the monthly contribution made by the community to the PEZAK program (USD 33.34) with the number of students would cost USD 172,200. This is 85.19% higher than the total cost of the PEZAK program. If we consider prices in urban areas, the estimation increases considerably. Moreover, if we add trainer's salaries and other expensive logistics in urban areas such as Kabul, the actual cost of the government program would be much higher.

Hence, recruiting college students in the months that schools are off is the best consideration for similar contexts. Its cost-effectiveness is reasonably justifiable.

2.7 Conclusion

In this chapter, we present an impact evaluation of PEZAK, an education intervention that recruited college students to provide math and science tutoring to high school students in Afghanistan. The goal of the program was to prepare college aspirants for the national university entrance examination.

We applied heterogeneity-robust estimator to obtain unbiased estimates of treatment effects. The treatment assignment was such that exposure to the program affected students in grade 10-12. Treatment regime is binary and non-staggered where treatment turns on an off. Three years after the first exposure schools could stay in the program while some leave the program and come back in the later periods. Our identification strategy estimates treatment effects in all individual difference-in-differences and aggregates them.

We find that the program had a positive impact on test scores, enrollment in top-rank universities, and overall passing rate. The instantaneous effect of the program is positive for all outcomes but significant only for passing rate. The dynamic effects are positive and significant for test scores, enrollment in top-rank universities and passing rate. The average total effect of the program is positive for these outcomes as well.

There are some gender differences in the impact of the program. Female students in the PEZAK program has lower probability of enrollment in low-rank universities as compared to male students. However, there are no differences in other outcomes. It could be that the program had a positive effect on student ambitions such that they did not select low-rank schools.

The unique context and features of the PEZAK program make it an interesting intervention. It focuses on students in underdeveloped communities and women. The goal of the intervention is to increase the representation of disadvantaged students in higher education. It embodies the features of similar interventions that have been conducted mainly in developed countries. The focus of the program on women makes our study a novel contribution to the available literature.

The result of our study provides valuable policy insights. The program is conducted for three months. Structured programs of this kind for a longer period could potentially have a remarkable impact on beneficiaries. The evidence from this program could be used to scale up the program in other regions.

2.8 Bibliography

Abbiati, Giovanni, Gianluca Argentin, Carlo Barone, and Antonio Schizzerotto, "Information barriers and social stratification in higher education: evidence from a field experiment," British Journal of Sociology, 2018, 69 (4), 1248–1270.

Abdulbaqi, Misbah, "Higher education in Afghanistan," Policy perspectives, 2009, pp. 99–117.

Agyire-Tettey, Frank, Charles Godfred Ackah, and Derek Asuman, "An unconditional quantile regression based decomposition of spatial welfare inequalities in Ghana," The Journal of Development Studies, 2018, 54 (3), 537–556.

Asia Foundation, "A Survey of the Afghan People: Afghanistan in 2016," 2016.

Aturupane, Harsha, "Higher education in Afghanistan: An emerging mountainscape," Technical Report, The World Bank 2013.

Azam, Mehtabul, "Accounting for growing urban-rural welfare gaps in India," World Development, 2019, 122, 410–432.

Banerjee, Abhijit V, Rukmini Banerji, Esther Duflo, Rachel Glennerster, and Stuti Khemani, "Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India," American Economic Journal: Economic Policy, feb 2010, 2 (1), 1–30.

Bettinger, Eric P, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu, "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment," The Quarterly Journal of Economics, 2012, 127 (3), 1205–1242.

Callaway, Brantly and Pedro HC Sant'Anna, "Difference-in-differences with multiple time periods," Journal of Econometrics, 2021, 225 (2), 200–230.

Cunha, Jesse M, Trey Miller, and Emily Weisburst, "Information and College Decisions: Evidence From the Texas Go Center Project," Educational Evaluation and Policy Analysis, nov 2017, 40 (1), 151–170.

de Chaisemartin, Clément and Xavier d'Haultfoeuille, "Two-way fixed effects estimators with heterogeneous treatment effects," American Economic Review, 2020, 110 (9), 2964–96.

_ and Xavier D'Haultfoeuille, "Difference-in-differences estimators of intertemporal treatment effects," Technical Report, National Bureau of Economic Research 2022.

de Chaisemartin, Clément, Xavier D'Haultfoeuille, and Yannick Guyonvarch, "***DID_MULTIPLEGT***: Stata module to estimate sharp Difference-in-Difference designs with multiple groups and periods," Statistical Software Components, Boston College Department of Economics May 2019.

Dynarski, Susan M, "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion," The American Economic Review, 2003, 93 (1), 279–288.

Fack, Gabrielle and Julien Grenet, "Improving College Access and Success for Low-Income Students: Evidence from a Large Need-Based Grant Program," American Economic Journal: Applied Economics, 2015, 7 (2), 1–34.

Fang, Zheng and Chris Sakellariou, "Evolution of Urban–rural Living Standards Inequality in T hailand: 1990–2006," Asian Economic Journal, 2013, 27 (3), 285–306.

Feminist Majority Foundation, "Largest USAID Women's Empowerment Program in the World Launched in Afghanistan," 2014. data retrieved from Feminist Newswire, https://feminist.org/news/largest-usaid-womens-empowerment-program-in-the-world-launched-in-afghanistan/.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," Journal of Econometrics, 2021, 225 (2), 254–277.

Herbaut, Estelle and Koen Martijn Geven, "What Works to Reduce Inequalities in Higher Education? A Systematic Review of the (Quasi-) Experimental Literature on Outreach and Financial Aid," 2019.

International Monetary Fund, Reconstructing Afghanistan, USA: International Monetary Fund, 2005.

Klasen, Stephan, "Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the Effect of Gender Inequality in Education on Economic Development," The World Bank Economic Review, 2002, 16 (3), 345–373.

Loyalka, Prashant, Yingquan Song, Jianguo Wei, Weiping Zhong, and Scott Rozelle, "Information, college decisions and financial aid: Evidence from a cluster-randomized controlled trial in China," Economics of Education Review, 2013, 36, 26–40.

Ministry of Education, "National Education Strategic Plan," Technical Report 2016.

Ministry of Finance, "National Budget for Fiscal Year 2018," 2019.

National Statistics and Information Authority, "Afghanistan Statistical Yearbook 2018-19," Technical Report 2019. data retrieved from Regulatory Indicators for Sustainable Energy, https://rise.esmap.org/data/files/library/afghanistan/Electricity%20Access/Afghanistan_Statistical-Yearbook-2018-19_compressed.pdf.

Roof, David J, "Day-by-Day: Higher Education in Afghanistan.," in "FIRE: Forum for International Research in Education," Vol. 1 ERIC 2015, pp. 64–80.

Seftor, Neil S, Arif Mamun, and Allen Schirm, "The Impacts of Regular Upward Bound on Postsecondary Outcomes Seven to Nine Years after Scheduled High School Graduation. Final Report.," US Department of Education, 2009.

SIGAR, "Support for Gender Equality: Lessons from the U.S. Experience in Afghanistan," Technical Report 2021. data retrieved from Special Inspector General for Afghanistan Reconstruction, https://www.sigar.mil/pdf/lessonslearned/SIGAR-21-18-LL.pdf.

Strayhorn, Terrell L, "Bridging the pipeline: Increasing underrepresented students' preparation for college through a summer bridge program," American Behavioral Scientist, 2011, 55 (2), 142–159.

Sun, Liyang and Sarah Abraham, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," Journal of Econometrics, 2021, 225 (2), 175–199.

Teng, S, X Yue, G Bjorn et al., "The Urban-rural Income Gap and Inequality in China," Review of Income and Wealth, 2007, 53 (5).

The Ministry of Justice of Afghanistan, "The constitution of the Islamic Republic of Afghanistan," 2005.

The World Bank, "Development Indicators," 2020. data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=AF.

Tilak, Jandhyala BG, "How inclusive is higher education in India?," Social Change, 2015, 45 (2), 185–223.

UNESCO International Institute for Education Planning, "Six ways to ensure higher education leaves no one behind," 2017.

Wells, Ryan S, Catherine A Manly, Suzan Kommers, and Ezekiel Kimball, "Narrowed gaps and persistent challenges: Examining rural-nonrural disparities in postsecondary outcomes over time," American Journal of Education, 2019, 126 (1), 1–31.

World Bank, "The Learning Crisis in Afghanistan," Technical Report 2018. data retrieved from, https://documents1.worldbank.org/curated/en/588881536147087211/AUS0000428-REVISED-SABER-SD-Afghanistan-digital-9-27.pdf.

_ , "Development Indicators," 2021. data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS?locations=AF.

_ , "Development Indicators," 2021. data retrieved from World Development Indicators, https://data.worldbank.org/indicator/SE.TER.ENRR?locations=AF.

Xu, Zeyu, Jane Hannaway, and Colin Taylor, "Making a difference? The effects of Teach For America in high school," Journal of Policy Analysis and Management, jun 2011, 30 (3), 447–469.

Young, Alwyn, "Inequality, the urban-rural gap, and migration," The Quarterly Journal of Economics, 2013, 128 (4), 1727–1785.

Zhao, Kai, “Rural-urban gap in academic performance at a highly selective Chinese university: variations and determinants,” Higher Education Research & Development, 2020, pp. 1–16.

2.9 Tables

Table 2.1: School Characteristics in 2011

	Non-MP School		MP School		Total	
	No.	%	No.	%	No.	%
School Location						
Rural	1,864	82.84	176	95.65	2,040	83.81
Urban	386	17.16	8	4.35	394	16.19
Total	2,250	100.00	184	100.00	2,434	100.00
Climate						
Cold	1,792	79.64	147	79.89	1,939	79.66
Very Cold	300	13.33	33	17.93	333	13.68
Warm	158	7.02	4	2.17	162	6.66
Total	2,250	100.00	184	100.00	2,434	100.00
Instruction Language						
Dari	1,902	84.53	178	96.74	2,080	85.46
Dari and Pashto	61	2.71	3	1.63	64	2.63
Pashto	287	12.76	3	1.63	290	11.91
Total	2,250	100.00	184	100.00	2,434	100.00
School Gender						
Girls	381	16.93	21	11.41	402	16.52
Girls with few boys classes	174	7.73	13	7.07	187	7.68
Boys	612	27.20	50	27.17	662	27.20
Boys with few girls class	1,083	48.13	100	54.35	1,183	48.60
Total	2,250	100.00	184	100.00	2,434	100.00
School has a library						
Yes	38	2.01	1	0.56	39	1.88
No	1,854	97.99	179	99.44	2,033	98.12
Total	1,892	100.00	180	100.00	2,072	100.00
Study Place						
Yes	281	14.85	29	16.11	310	14.96
No	1,611	85.15	151	83.89	1,762	85.04
Total	1,892	100.00	180	100.00	2,072	100.00
Tap Water						
Tap Water	1	0.05	0	0.00	1	0.04
Other Sources	2,116	98.92	179	98.35	2,295	98.88
Water Pump	22	1.03	3	1.65	25	1.08
Total	2,139	100.00	182	100.00	2,321	100.00

Notes: This table compares the categorical variables in Non-PEZAK and PEZAK schools. We present the number of schools and share of schools in each category for non-PEZAK, PEZAK and all schools in PEZAK districts. The numbers on the left shows the frequency of each category and numbers on the right of each column shows percentage of each sub category. It is worth mentioning that due to missing values for some of the variables, the total across all variables may not be the same. For example, variables library, study place, and tap water have lesser total count values due to missing values.

Table 2.2: School Characteristics in 2011 before Matching
Continuous Variables

	(1) Mean Non-MP	(2) Mean MP	(3) Diff	(4) Total
Distance from the Nearest Road (Km)	7.49	5.02	2.47	7.27
Distance from the Nearest River (Km)	51.30	46.41	4.89	50.86
Distance from the Nearest Hospital (Km)	60.66	51.29	9.38**	59.84
School Age	9.85	21.50	-11.65***	10.73
Average Class Size	34.96	37.50	-2.55***	35.17
Teacher with an Undergraduate Degree or Higher	0.41	0.53	-0.12	0.42
Teacher to Student Ratio	0.03	0.02	0.00***	0.03
Total Enrollment	343.85	580.99	-237.14***	361.77
Number of Male Toilets	1.05	2.17	-1.12***	1.13
Number of Female Toilets	0.25	0.18	0.07	0.25
Number of Mixed Toilets	0.01	0.00	0.01**	0.01
Observations	2250	184	2434	2434

Notes: This table compares continuous variables in PEZAK and Non-PEZAK schools. As we can see, there are some imbalances in school characteristics. However, as shown in the robustness checks, the imbalance does not affect our results.

Table 2.3: Exogeneity of Enrollment

	(1) Total Enrollment	(2) High School Enrollment	(3) Grade 12 Enrollment
Treatment × 2012	-18.2458	1.9421	0.6373
	(58.51)	(13.34)	(3.95)
Treatment × 2013	-43.6319	2.8618	4.1807
	(58.51)	(13.34)	(3.95)
Treatment × 2014	-47.0171	4.9815	3.9167
	(58.51)	(13.34)	(3.95)
Treatment × 2015	-65.1042	4.1892	2.1825
	(58.51)	(13.34)	(3.95)
Treatment × 2016	-67.0972	4.5120	2.7438
	(58.51)	(13.34)	(3.95)
Treatment × 2017	-68.2907	4.4530	3.4953
	(58.51)	(13.34)	(3.95)
Treatment × 2018	-82.1928	1.2755	2.8585
	(58.67)	(13.38)	(3.97)
Avg Enrollment in 2011	449.2373***	43.0188***	8.8060***
	(13.89)	(3.17)	(0.94)
Total Obs	12,909	12,909	12,909

Notes: In this table we regress school enrollment on the treatment. We include year fixed effects and year fixed effects interacted with treatment. The coefficients of interest are the interaction terms. We have excluded other terms from the table. The goal is to show that enrollment, specially class 12 enrollment, was not affected by treatment. In other words, we examine if school choice was endogenous. We do not find any evidence for self-selection. Values in the brackets represent standard errors. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 2.4: Randomness of Treatment Time

	Probability of Treatment
Total Enrollment	-0.0001
	(0.00)
Total High School Enrollment	0.0001
	(0.00)
Total Grade 12 Enrollment	0.0000
	(0.00)
Teacher to Student Ratio	2.2572
	(1.54)
Average Class Size	0.0008
	(0.00)
Boys' School	-0.0866***
	(0.03)
Girls' School	-0.1593***
	(0.04)
Total Enrollment (-1)	-0.0000
	(0.00)
Total High School Enrollment (-1)	-0.0004
	(0.00)
Total Grade 12 Enrollment (-1)	0.0001
	(0.00)
Teacher to Student Ratio (-1)	2.6007
	(1.66)
Average Class Size (-1)	-0.0001
	(0.00)
Total Enrollment (-2)	0.0000
	(0.00)
Total High School Enrollment (-2)	0.0005
	(0.00)
Total Grade 12 Enrollment (-2)	-0.0010
	(0.00)
Teacher to Student Ratio (-2)	-0.4206
	(1.58)
Average Class Size (-2)	0.0001
	(0.00)
Total Obs	1263

Notes: We regress the treatment indicator on school characteristics, and one year and two years lag for time varying school characteristics. The goal is to examine if past characteristics affected future treatment. The result shows that none of the lag terms are significant. Therefore, treatment time was fairly exogenous. Values in the brackets represent standard errors. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 2.5: Average Total Effects of PEZAK Program

	Test Scores	Top Five	Prov Uni	Passing Rate
	(1)	(2)	(3)	(4)
dC-DH (2021a), flexible treatment	0.245	0.370	0.109	0.190
	(0.117)	(0.184)	(0.131)	(0.143)
dC-DH (2021a), treatment at T-1 implies treatment at T	0.197	0.298	0.087	0.153
	(0.094)	(0.148)	(0.106)	(0.115)
dC-DH (2021a), treatment at T-2 implies treatment at T	0.172	0.260	0.076	0.134
	(0.082)	(0.129)	(0.078)	(0.101)
dC-DH (2021a), staggered adoption	0.143	0.216	0.063	0.111
	(0.068)	(0.107)	(0.076)	(0.084)

Notes: This table shows estimated average effects of the PEZAK program on test scores, enrollment in top five universities, enrollment in low-rank provincial universities, and overall passing rate 0-6 years after adoption. All outcomes are standardized. Estimates are calculated using ***did_multipleGT*** STATA package of de Chaisemartin et al. (2019) based on de Chaisemartin and d'Haultfoeuille (2020) and de Chaisemartin and D'Haultfoeuille (2022). We apply the ***did_multipleGT*** in four scenarios. First, treatment is flexible where some schools remain treated between 2008 (first treatment) and 2019 (latest year) and for others treatment turns on and off several times. Second, treatment status is redefined by considering year T's treatment based on year T-1 treatment status. Treatment at T-1 implies treatment at T. Third, treatment at T-2 implies treatment at T. Fourth, a staggered design is considered where schools remain treated after the first treatment. The treatment effect is the sum of instantaneous (period 0) and dynamic (1-6) effects. The parameter can be interpreted as the average total effect per unit of treatment or average effect of receiving one unit higher number of treatments. Standard errors presented in the brackets below the estimates are calculated with a bootstrap approach with 100 replications.

Table 2.6: Heterogeneous Treatment Effects

	Test Score (1)	Top-rank University (2)	Low-rank University (3)	Passing Rate (4)
Post × Treatment × Female	-0.01	0.01	-0.03**	-0.01
	(0.06)	(0.02)	(0.01)	(0.02)
Post × Treatment	0.14***	0.04***	-0.00	0.00
	(0.04)	(0.01)	(0.01)	(0.01)
Female	-0.47***	-0.12***	-0.06***	-0.11***
	(0.02)	(0.01)	(0.00)	(0.01)
Retaking the Test	0.17***	0.04***	0.03***	0.04***
	(0.02)	(0.00)	(0.00)	(0.01)
Constant	0.27***	0.21***	0.17***	0.64***
	(0.01)	(0.00)	(0.00)	(0.00)
N	232,092	232,092	232,092	232,092
School FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Clustering	Yes	Yes	Yes	Yes

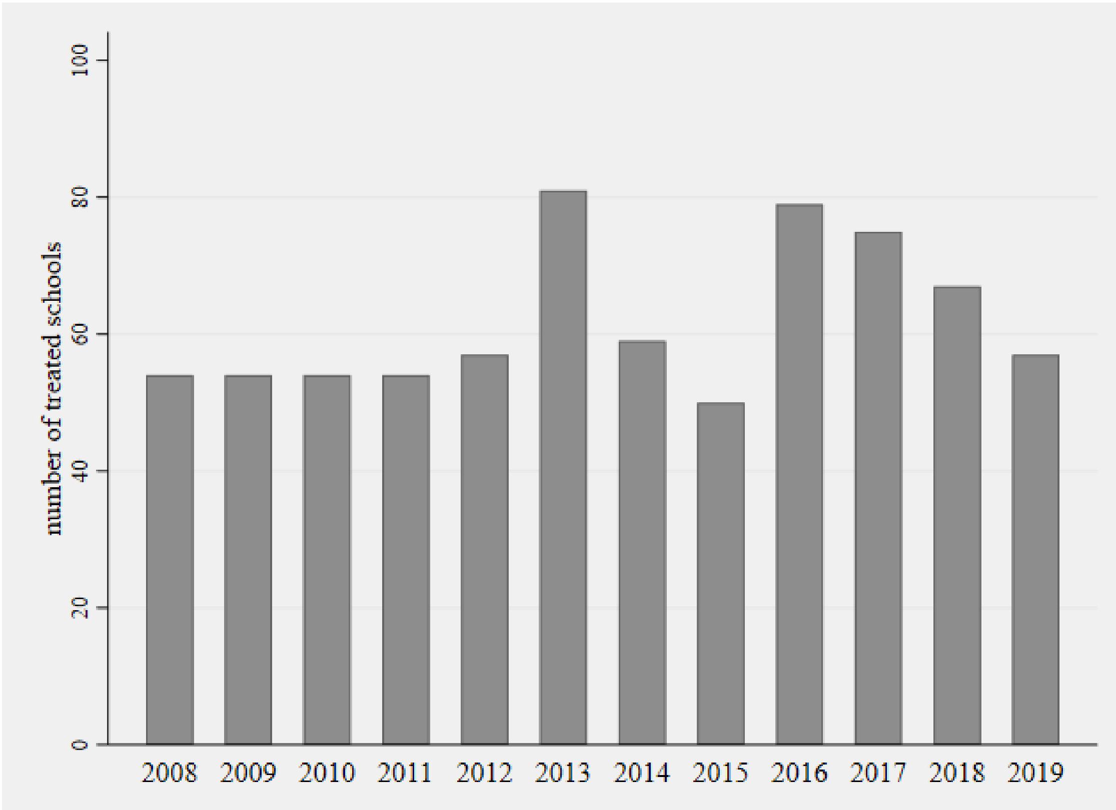
Notes: This table presents estimates from TWFE estimators with student-level data. The triple interaction term is the coefficient of interest. The interaction term between student gender ($female = 1$) and treatment show the differential treatment effect on female students. Column (1) is the result of an OLS regression on Equation 2.2. Columns (2)-(4) are linear probability regressions of Equation 2.2. In columns with controls, we include student gender and retaking the test as additional regressors. Values in the brackets represent standard errors. Standard errors are clustered at the school level. */**/*** denotes significance at the 10/5/1 percent.

Table 2.7: Average Total Effects of PEZAK Program

	Test Scores	Top Five	Prov Uni	Passing Rate
	(1)	(2)	(3)	(4)
dC-DH (2021a), flexible treatment	0.245	0.370	0.109	0.190
	(0.117)	(0.184)	(0.131)	(0.143)
dC-DH (2021a), treatment at T-1 implies treatment at T	0.197	0.298	0.087	0.153
	(0.094)	(0.148)	(0.106)	(0.115)
dC-DH (2021a), treatment at T-2 implies treatment at T	0.172	0.260	0.076	0.134
	(0.082)	(0.129)	(0.078)	(0.101)
dC-DH (2021a), staggered adoption	0.143	0.216	0.063	0.111
	(0.068)	(0.107)	(0.076)	(0.084)
CS (2021), staggered adoption	0.151	0.239	0.062	0.105
	(0.062)	(0.092)	(0.077)	(0.083)
TWFE, flexible treatment	0.035	0.098	0.073	0.094
	(0.033)	(0.063)	(0.057)	(0.049)
TWFE, treatment at T-1 implies treatment at T	0.025	0.061	0.050	0.055
	(0.031)	(0.056)	(0.049)	(0.047)
TWFE, treatment at T-2 implies treatment at T	0.042	0.071	0.015	0.068
	(0.032)	(0.056)	(0.051)	(0.046)
TWFE, staggered adoption	0.160	0.203	-0.033	0.098
	(0.038)	(0.065)	(0.073)	(0.060)

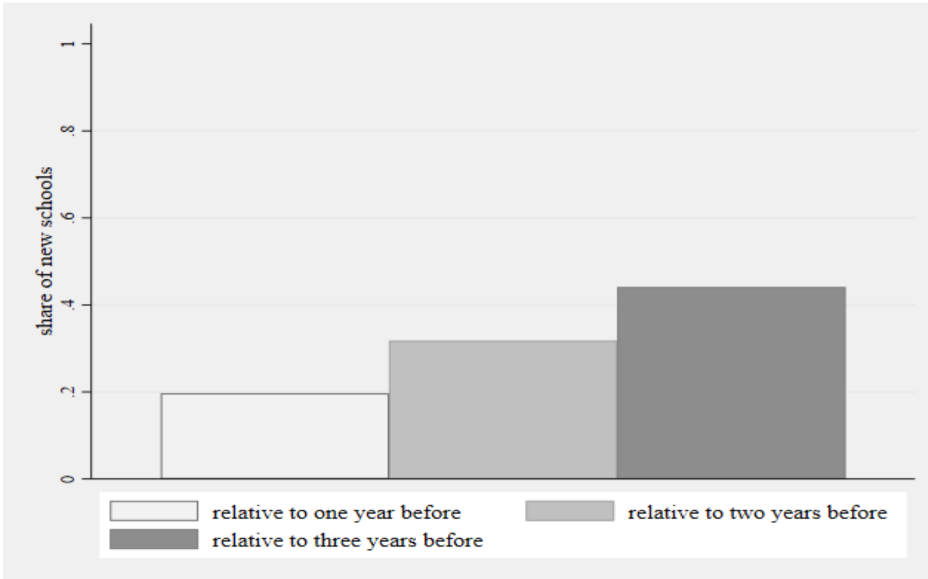
Notes: This table shows estimated average effects of the PEZAK program on test scores, enrollment in top five universities, enrollment in low-rank provincial universities, and overall passing rate 0-6 years after adoption. All outcomes are standardized. Estimates are calculated using ***did_multipleGT*** STATA package of de Chaisemartin et al. (2019) based on de Chaisemartin and d'Haultfoeuille (2020) and de Chaisemartin and D'Haultfoeuille (2022). We apply the ***did_multipleGT*** in four scenarios. First, treatment is flexible where some schools remain treated between 2008 (first treatment) and 2019 (latest year) and for others treatment turns on and off several times. Second, treatment status is redefined by considering year T's treatment based on year T-1 treatment status. Treatment at T-1 implies treatment at T. Third, treatment at T-2 implies treatment at T. Fourth, a staggered is considered where schools remain treated after the first treatment. Fifth, treatment effect in a staggered design scenario is estimated with Callaway and Sant'Anna (2021) estimator. Sixth, scenario 1-4 are estimated using a TWFE estimator. The treatment effect is the sum of instantaneous (period 0) and dynamic (1-6) effects. The parameter can be interpreted as the average total effect per unit of treatment or average effect of receiving one unit higher number of treatments. Standard errors presented in the brackets below the estimates are calculated with a bootstrap approach with 100 replications.

2.10 Figures

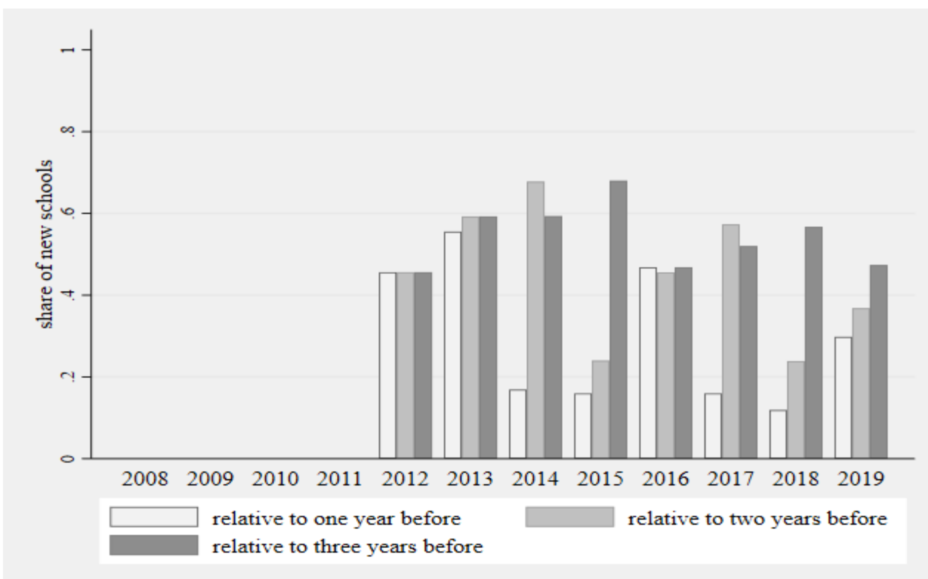


Notes: This Figure plots the number of schools that received treatment. From 2008 to 2011 there were 54 PEZAK schools each year.

Figure 2.1: Number of Treated Schools Over Time



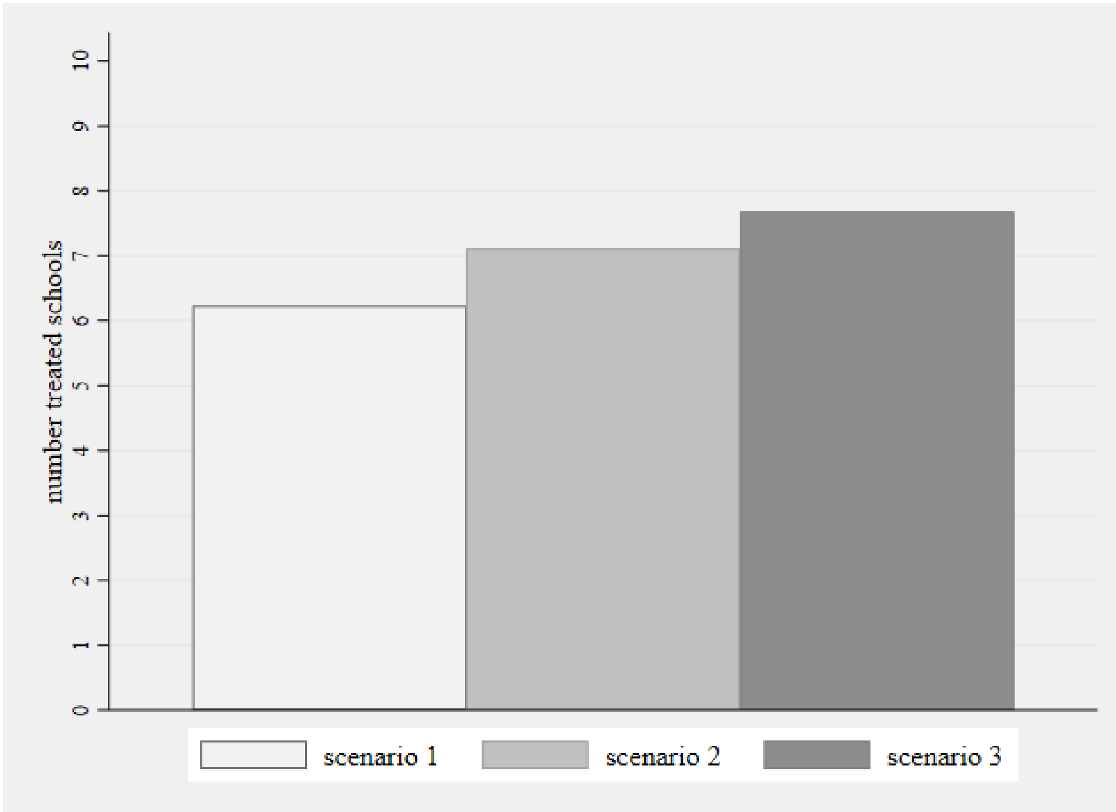
(a) Average Share of New Schools



(b) Share of New Schools in Each Year

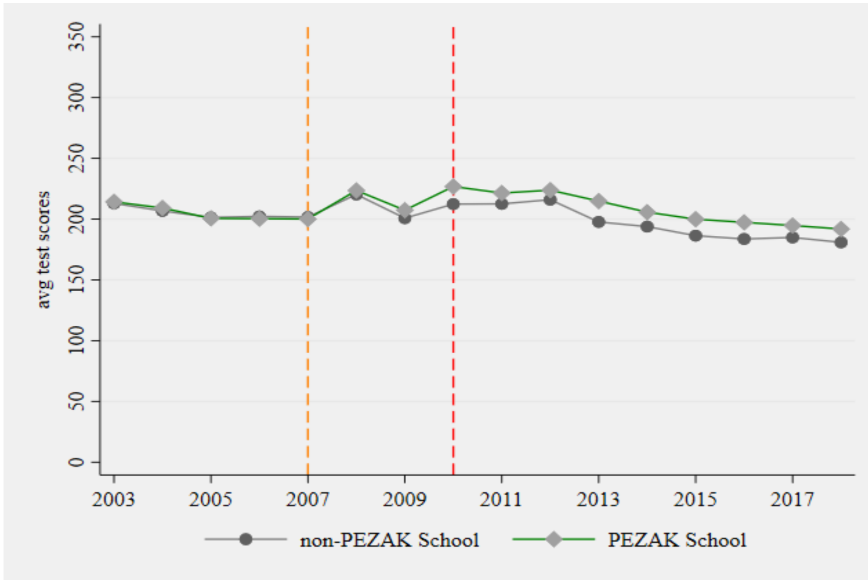
Notes: This Figure presents the switching rate of treated schools relative to previous years. Panel (a) shows the average switching rate between 2008-2019 relative to one year, two years, and three years before. Panel (b) presents the switching rate over time. There is a positive correlation between time distance and average switching rate.

Figure 2.2: Share of New PEZAK Schools Relative to Reference Period T

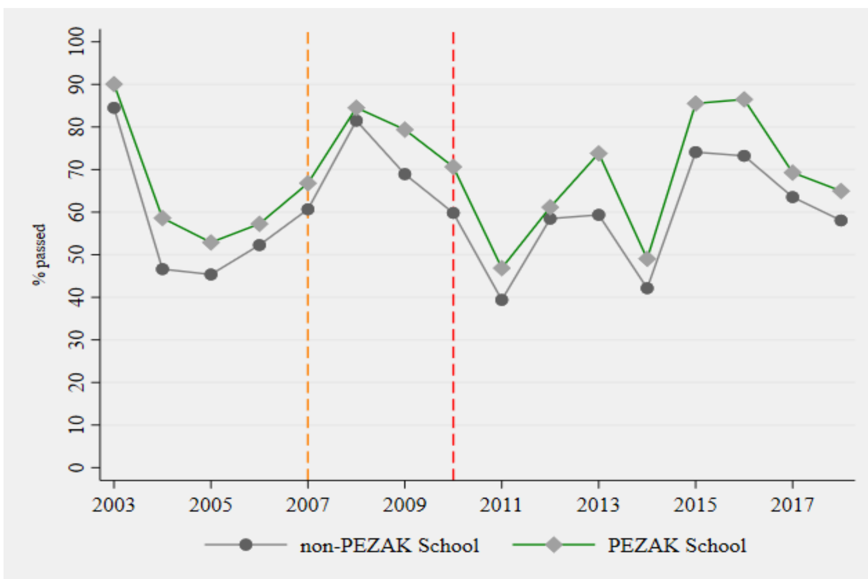


Notes: This Figure plots the number of schools in three scenarios. In scenario 1 treatment in period T affects only cohort T. In scenario 2 once schools receive treatment in period T, treatment in period T+1 is implied. In scenario 3 treatment in period T implies treatment in period T+1 and T+2.

Figure 2.3: Average Treatment Frequency in Three Scenarios



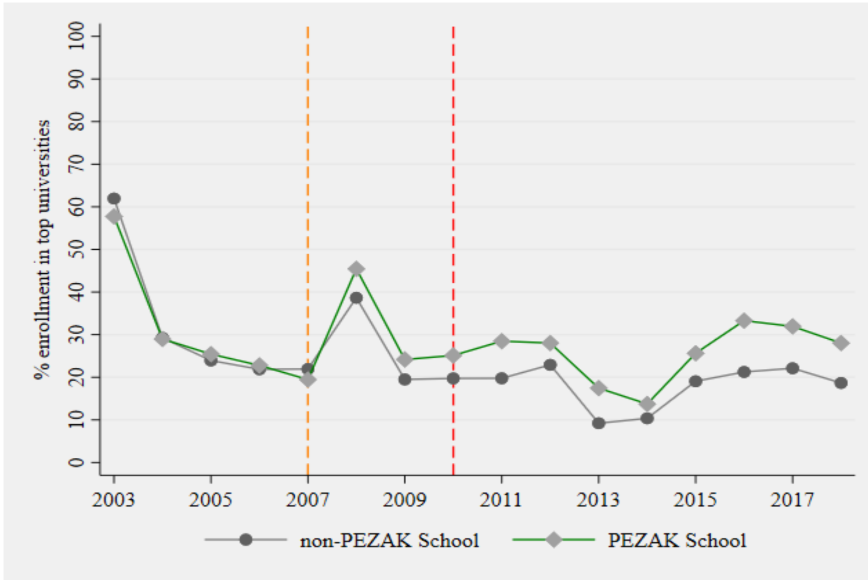
(a) Test Scores



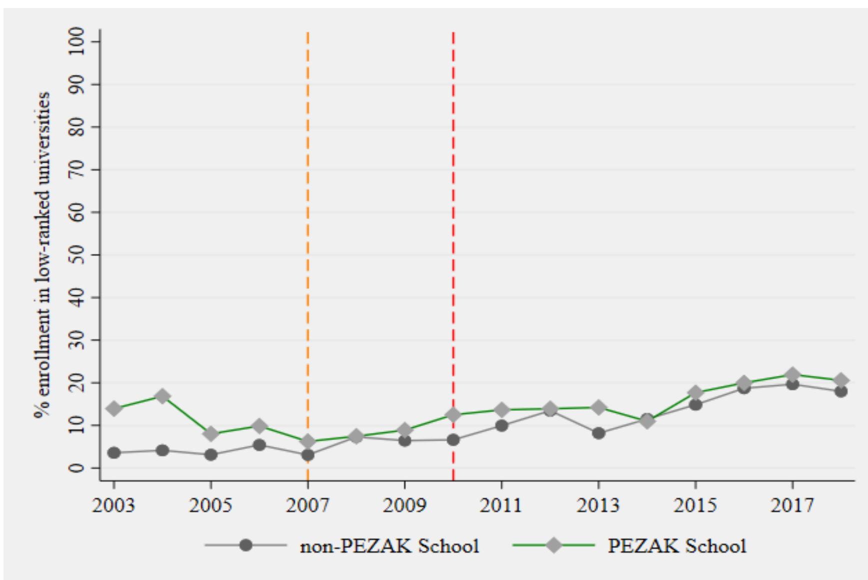
(b) Passing Rate

Notes: This Figure compares evolution of test scores and passing rate between 184 PEZAK and non-PEZAK schools. Until 2007 no school received the program. In 2008-2011, 54 schools received the program. The rest of PEZAK schools received the program in the following years. Panel (a) presents evolution of school-level average test scores. Panel (b) compares school-level average passing rate. Passing rate is the school-level ratio of students that were enrolled to any public university program to total test takers.

Figure 2.4: Trend Analysis: Test Score and Average Passing Rate



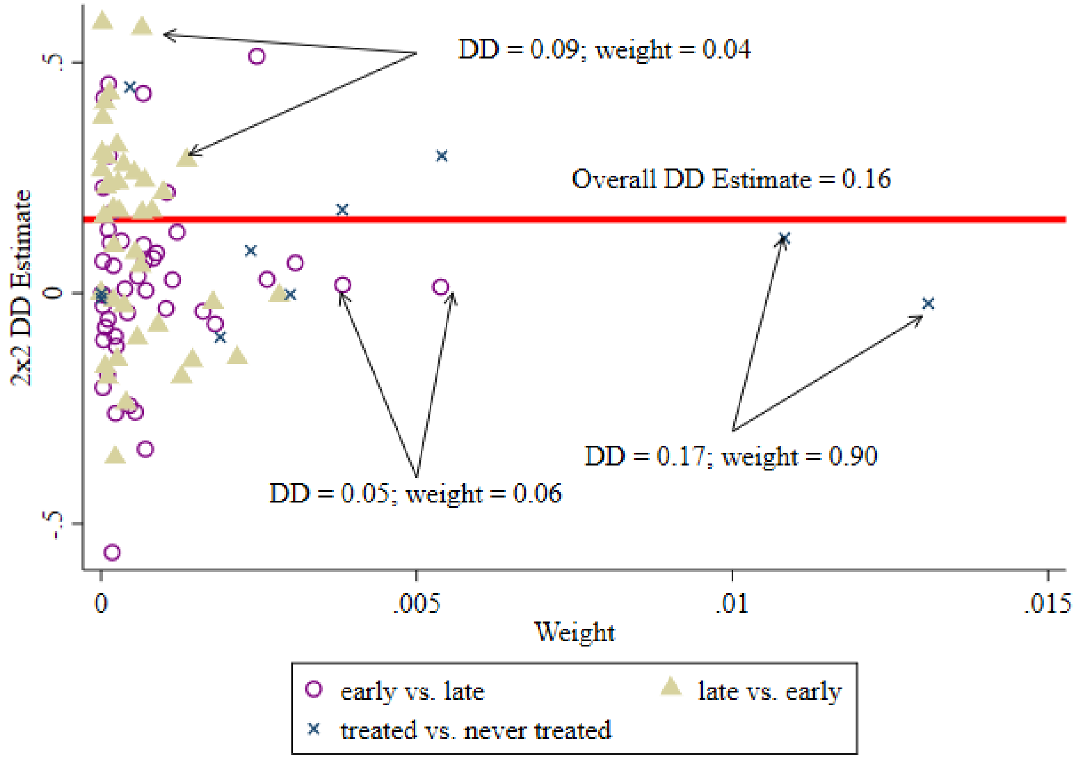
(a) Enrollment in Top Universities



(b) Enrollment in Provincial Universities

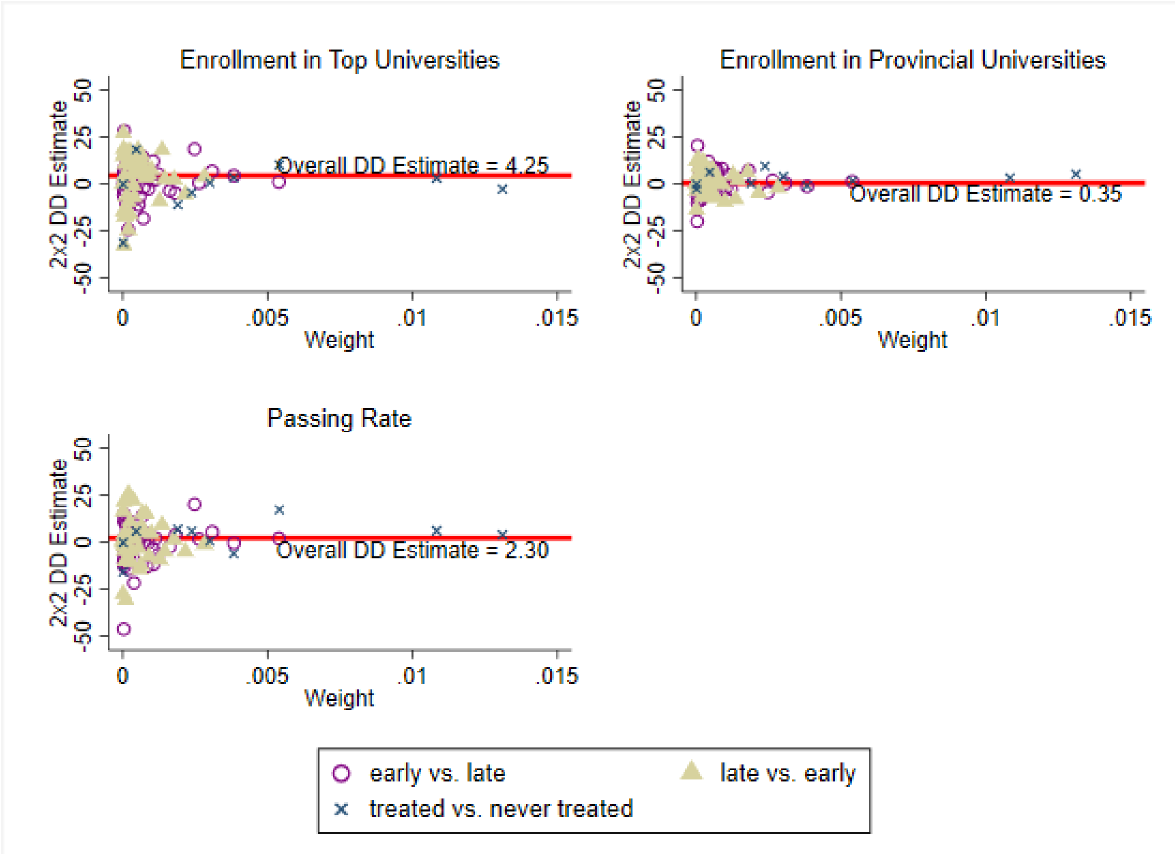
Notes: This Figure compares evolution of enrollment in top-rank universities and enrolment in low-rank provincial universities between 184 PEZAK and non-PEZAK schools. Until 2007 no school received the program. In 2008-2011, 54 schools received the program. The rest of PEZAK schools received the program in the following years. Enrollment is the share of total test takers who secured seats in top-rank(low-rank) universities. Panel (a) presents evolution of school-level average enrollment in top-rank universities. Panel (b) compares school-level average enrollment in low-rank universities.

Figure 2.5: Trend Analysis: Enrollment in Top Universities and Enrollment in Low-ranked Universities



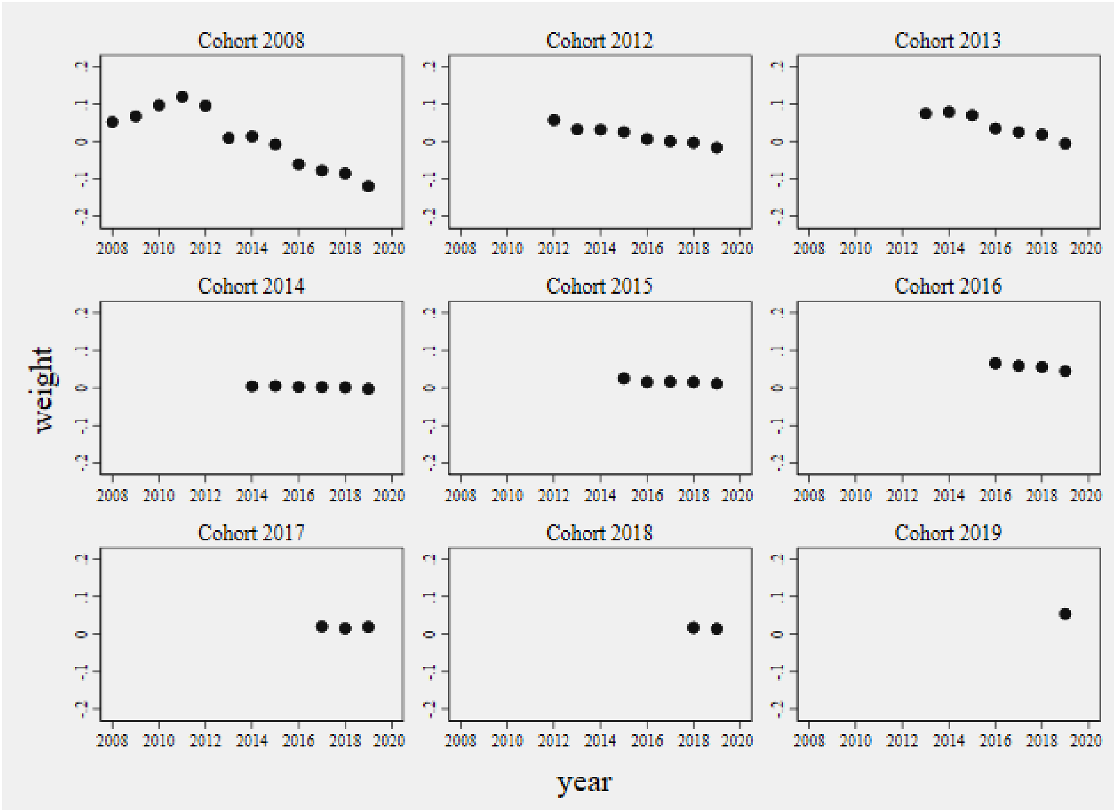
Notes: This Figure presents Bacon Decomposition using bacondecomp STATA package based on Goodman-Bacon (2021). The outcome is test scores. All 2×2 DID estimates are plotted against their respective weights. There are mainly three groups of 2×2 canonical DIDs: early treated unites as treatment and late treatment units as control group, late treated units as treatment and early treated units as control group, and unites treated any time as treatment group and never treated units as control groups. It is assumed that treatment is staggered in nature. The weighted average treatment effect and total weight for each group is shown by arrows represented by two DIDs in each category. The average treatment effect for all possible DIDs is shown on the red horizontal line.

Figure 2.6: Bacon Decomposition on Test Scores



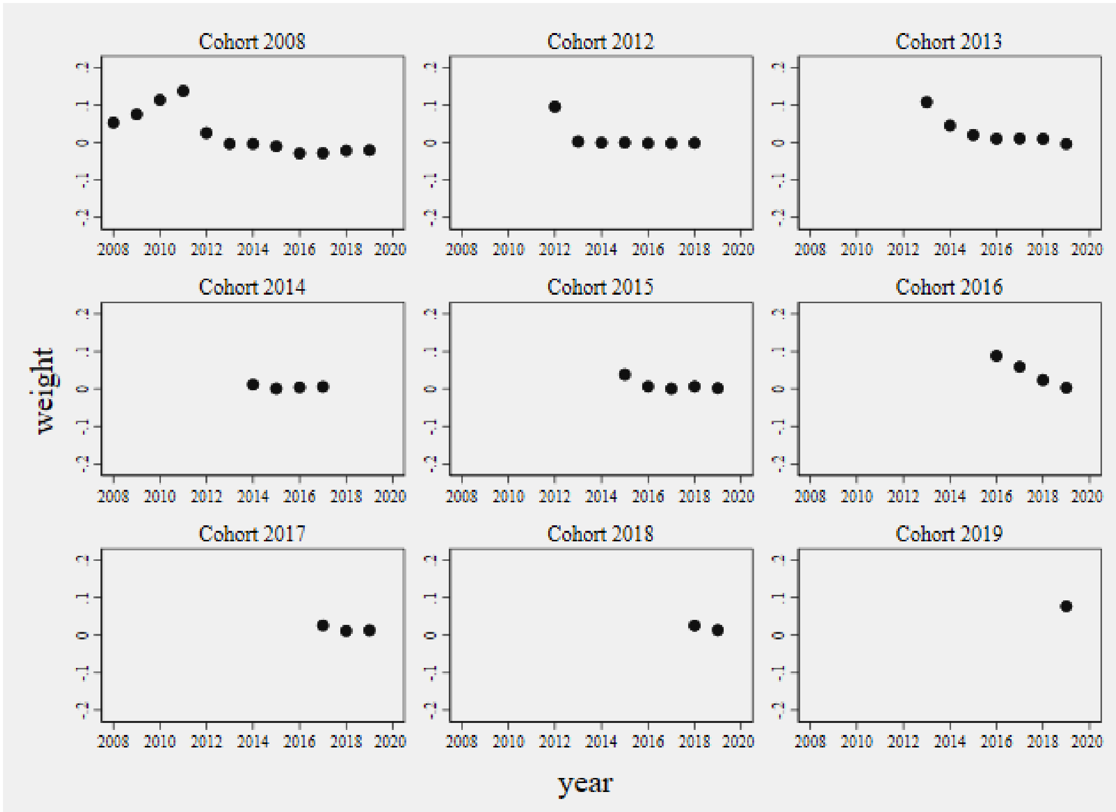
Notes: This Figure presents Bacon Decomposition using bacondecomp STATA package based on Goodman-Bacon (2021). The outcome is enrollment in top universities, low-rank provincial universities, and overall enrollment. All 2×2 DID estimates are plotted against their respective weights. There are mainly three groups of 2×2 canonical DIDs: early treated unites as treatment and late treatment units as control group, late treated units as treatment and early treated units as control group, and unites treated any time as treatment group and never treated units as control groups. It is assumed that treatment is staggered in nature. The weighted average treatment effect and total weight for each group is shown by arrows represented by two DIDs in each category. The average treatment effect for all possible DIDs is shown on or below the red horizontal line.

Figure 2.7: Bacon Decomposition on University Enrollment



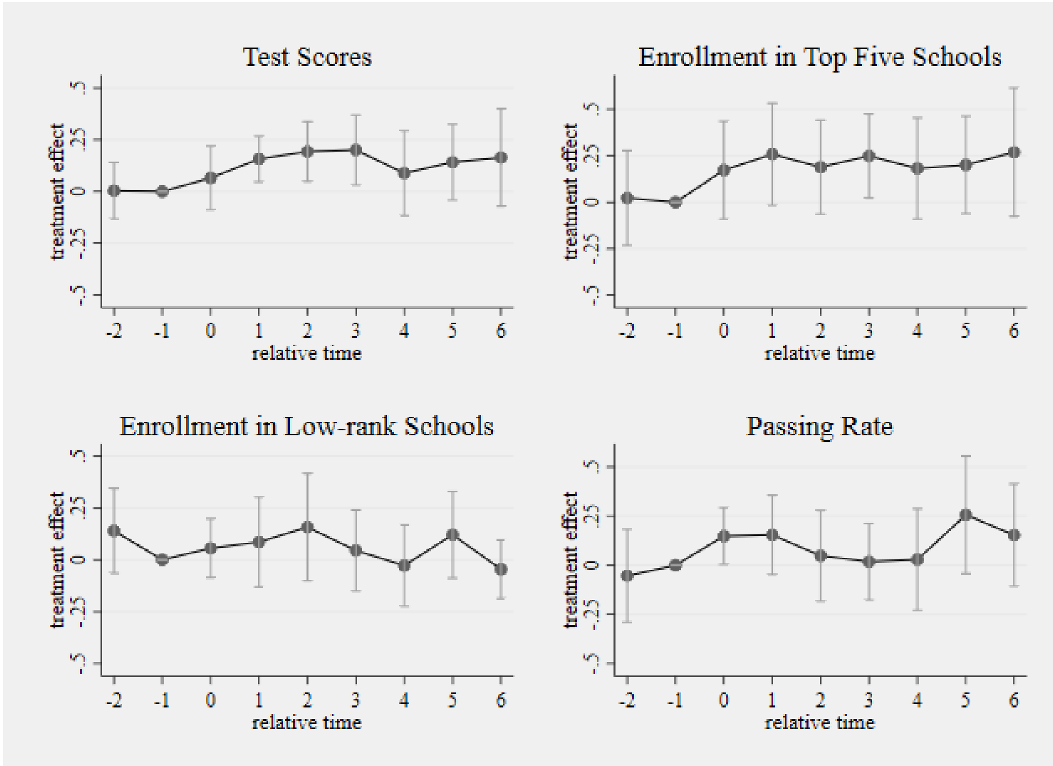
Notes: This Figure plots the weights attached to each group-time cell using twowayfeweights STATA command based on de Chaisemartin & D'Haultfoeuille (2021a). Cohorts are tracked over time. Cohort is defined based on the year in which school received treatment for the first time. The Y-axis represents weights that each cohort carries over time. Weights are calculated from a staggered adoption design where treatment is non-diminishing. In other words, once units are treated, they remain treated for ever. In this design, under the common trends assumption, TWFE estimates a weighted sum of 48 ATTs such that 39 ATTs receive a positive weight, and 9 receive a negative weight. The sum of the negative weights is equal to -0.38. The minimal value of the standard deviation of the treatment effect across treated groups and time periods under which beta and the ATT could be of opposite signs is equal to 0.03. The 2008 cohort has the largest share of negative weights.

Figure 2.8: de Chaisemartin and D'Haultfoeuille (2021a) Weight Decomposition, Staggered Treatment



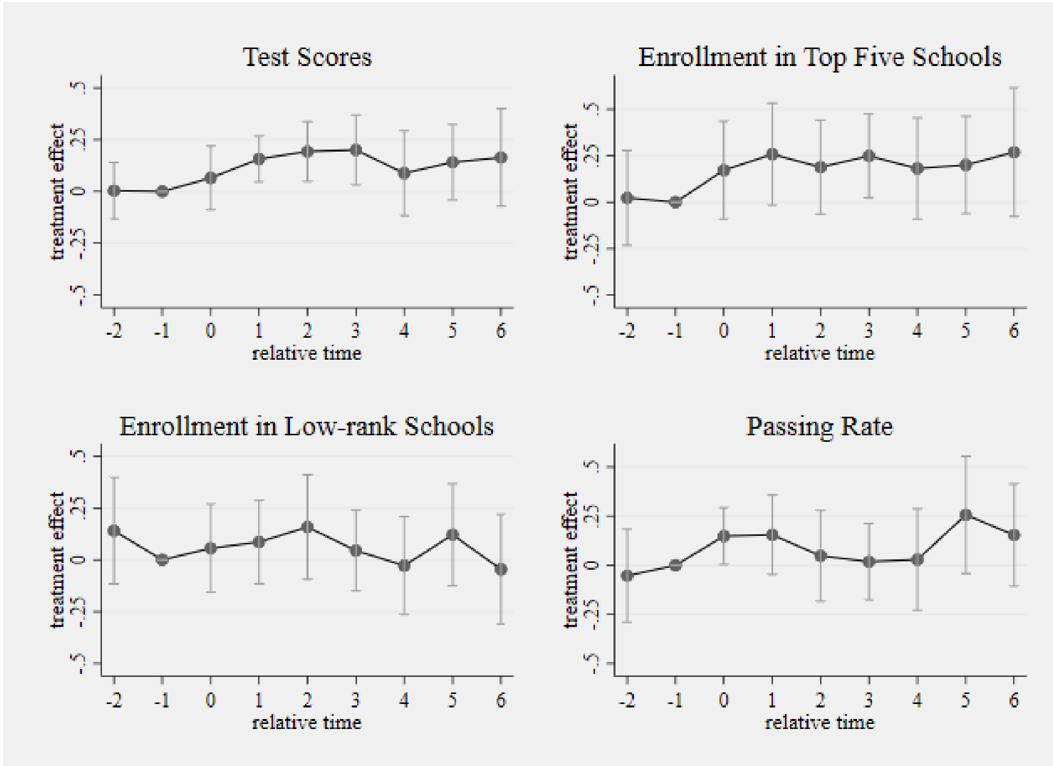
Notes: This Figure plots the weights attached to each group-time cell using twowayfeweights STATA command based on de Chaisemartin & D'Haultfoeuille (2021a). Cohorts are tracked over time. Cohort is defined based on the year in which school received treatment for the first time. The Y-axis represents weights that each cohort carries over time. Weights are calculated from a binary treatment where treatment turns on and off. Dynamic effects are estimated based on the first treatment. Under the common trends assumption, TWFE estimates a weighted sum of 45 ATTs where 32 ATTs receive a positive weight, and 13 receive a negative weight. The sum of the negative weights is equal to -0.12. The minimal value of the standard deviation of the treatment effect across treated groups and time periods under which beta and the ATT could be of opposite signs is equal to 0.03 0.05. The 2008 cohort has the largest share of negative weights.

Figure 2.9: de Chaisemartin and D'Haultfoeuille (2021a) Weight Decomposition, non-Staggered Treatment



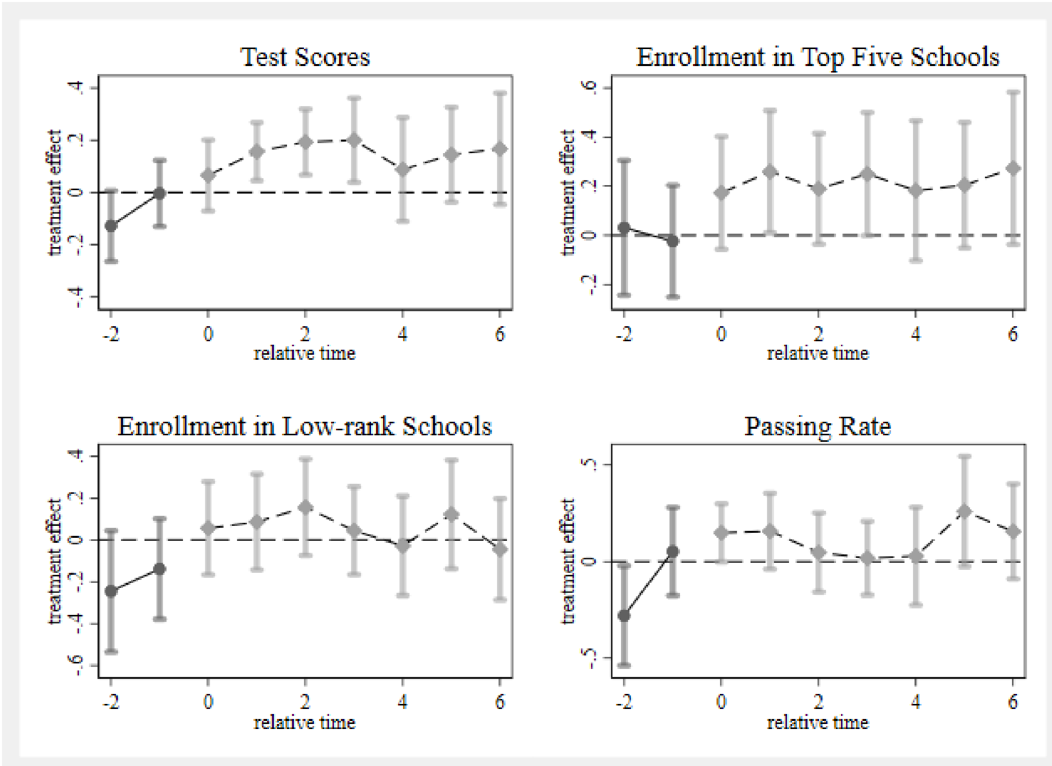
Notes: This Figure presents dynamic treatment effects in a treatment regime where treatment in period T-2 implies treatment at period T, meaning that treatment stays at least for three periods. Treatment can turn on and off after each three periods. The treatment effects shows intention to treat effect or the effects of having received a weakly higher number of treatments. The treatment effect is estimated using *did_multiplegt* STATA package which implements the estimator in (de Chaisemartin and D'Haultfoeuille, 2022). It computes the DID_l estimates which is the weighted average of DID estimates across all time periods and possible values of treatment. The DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those that remains 0 in $t-1$ and t. Similarly, DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those whose treatment is equal to 1 in both periods. The possibility of negative weight occurring is high in the later case, especially when treatment effects transmit from earlier periods to the next periods. DID_l estimates the effect of having switched treatment for the first time l periods ago. These estimators are robust to heterogeneity and dynamic effects. Coefficients for periods before the first treatment (relative periods with negative sign) are placebo estimates that compare switchers and non-switchers before the switch. Period 0 shows instantaneous treatment effects and the rest of the periods show dynamic treatment effects. Standard errors are estimated using a bootstrap approach with 100 replications. Vertical lines show 95% confidence intervals.

Figure 2.10: Treatment Effects with de Chaisemartin & D'Haultfoeuille (2021a) Estimator: Treatment at T-2 ⇒ Treatment at T



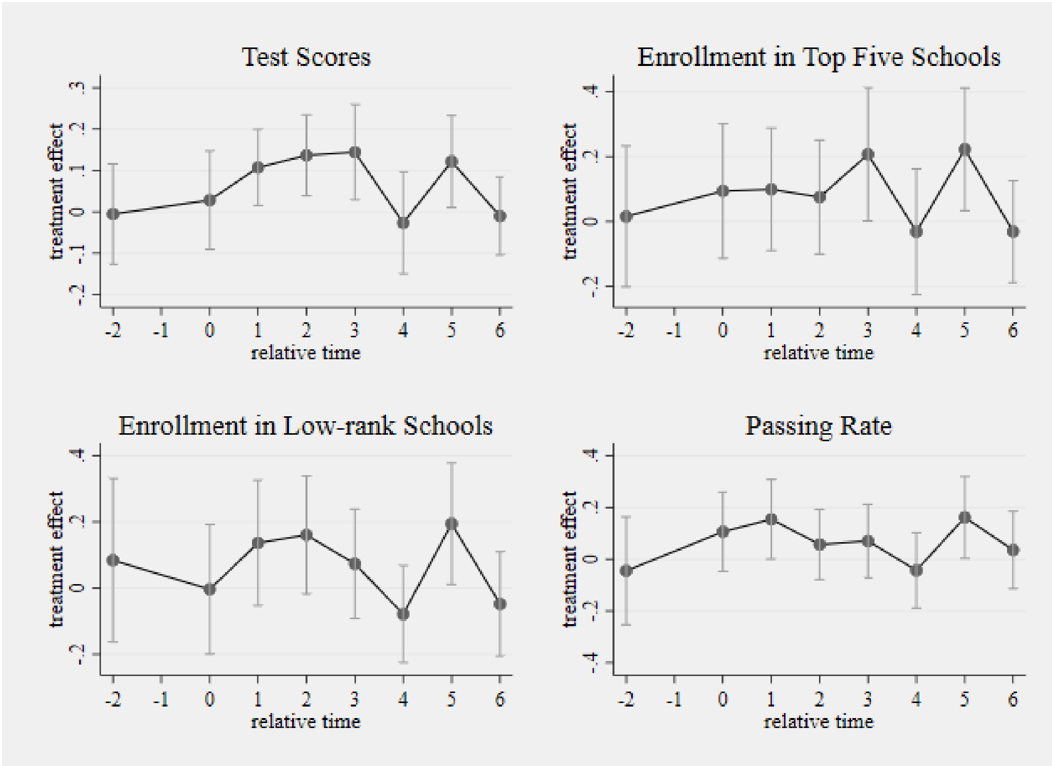
Notes: This Figure presents dynamic treatment effects in a staggered treatment regime where treatment resembles an immunization program such that once a unit is treated it remains treated for ever. The design is staggered because units received their first treatment at different times. The treatment effects shows intention to treat effect or the effects of having received a weakly higher number of treatments. The treatment effect is estimated using ***did_multiplegt*** STATA package which implements the estimator in (de Chaisemartin and D'Haultfoeuille, 2022). It computes the $\boldsymbol{DID_l}$ estimates which is the weighted average of DID estimates across all time periods and possible values of treatment. The DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those that remains 0 in $t-1$ and t. Similarly, DIDs compare outcome evolution from $t-1$ to t in groups whose treatment switches to 1 in period t and those whose treatment is equal to 1 in both periods. The possibility of negative weight occurring is high in the later case, especially when treatment effects transmit from earlier periods to the next periods. $\boldsymbol{DID_l}$ estimates the effect of having switched treatment for the first time l periods ago. These estimators are robust to heterogeneity and dynamic effects. Coefficients for periods before the first treatment (relative periods with negative sign) are placebo estimates that compare switchers and non-switchers before the switch. Period 0 shows instantaneous treatment effects and the rest of the periods show dynamic treatment effects. Standard errors are estimated using a bootstrap approach with 100 replications. Vertical lines show 95% confidence intervals.

Figure 2.11: Treatment Effects with de Chaisemartin & D'Haultfoeuille (2021a) Estimator: Staggered Treatment



Notes: This Figure presents dynamic treatment effects in a staggered treatment regime where treatment resembles an immunization program such that once a unit is treated it remains treated for ever. The design is staggered because units received their first treatment at different times. The treatment effects show intention to treat effect or the effects of having received a weakly higher number of treatments. The treatment effect is estimated using *csdid* STATA package which implements the estimator in Callaway and Sant'Anna (2021). Coefficients for periods before the first treatment (relative periods with negative sign) are placebo estimates that compare switchers and non-switchers before the switch. Standard errors are estimated using a bootstrap approach with 100 replications. Vertical lines show 95% confidence intervals.

Figure 2.12: Treatment Effects with Callaway and Sant'Anna (2021) Estimator: Staggered Treatment



Notes: This Figure presents dynamic treatment effects in a staggered treatment regime where treatment resembles an immunization program such that once a unit is treated it remains treated for ever. The design is staggered because units received their first treatment at different times. The treatment effects show intention to treat effect. The treatment effects are estimated using $TWFE$. Coefficients for periods before the first treatment (relative periods with negative sign) are placebo estimates that compare switchers and non-switchers before the switch. Standard errors are clustered at school level over time. Vertical lines show 95% confidence intervals.

Figure 2.13: Treatment Effects with TWFE Estimator: Staggered Treatment

2.11 Appendix

A Tables

Table A.1: Average Total Effects of PEZAK Program on Male Students

	Test Scores (1)	Top Five (2)	Prov Uni (3)	Passing Rate (4)
dC-DH (2021a), flexible treatment	0.283	0.436	0.084	0.203
	(0.119)	(0.182)	(0.134)	(0.156)
dC-DH (2021a), treatment at T-1 implies treatment at T	0.229	0.352	0.068	0.164
	(0.096)	(0.147)	(0.108)	(0.126)
dC-DH (2021a), treatment at T-2 implies treatment at T	0.200	0.308	0.059	0.143
	(0.084)	(0.129)	(0.095)	(0.110)
dC-DH (2021a), staggered adoption	0.168	0.260	0.050	0.121
	(0.071)	(0.109)	(0.080)	(0.093)
CS (2021), staggered adoption	0.179	0.282	0.055	0.127
	(0.066)	(0.095)	(0.082)	(0.092)
TWFE, flexible treatment	0.011	0.111	0.068	0.064
	(0.036)	(0.064)	(0.060)	(0.051)
TWFE, treatment at T-1 implies treatment at T	0.002	0.063	0.050	0.044
	(0.033)	(0.056)	(0.052)	(0.048)
TWFE, treatment at T-2 implies treatment at T	0.020	0.071	0.012	0.032
	(0.034)	(0.058)	(0.052)	(0.047)
TWFE, staggered adoption	0.155	0.239	-0.028	0.087
	(0.042)	(0.068)	(0.074)	(0.064)

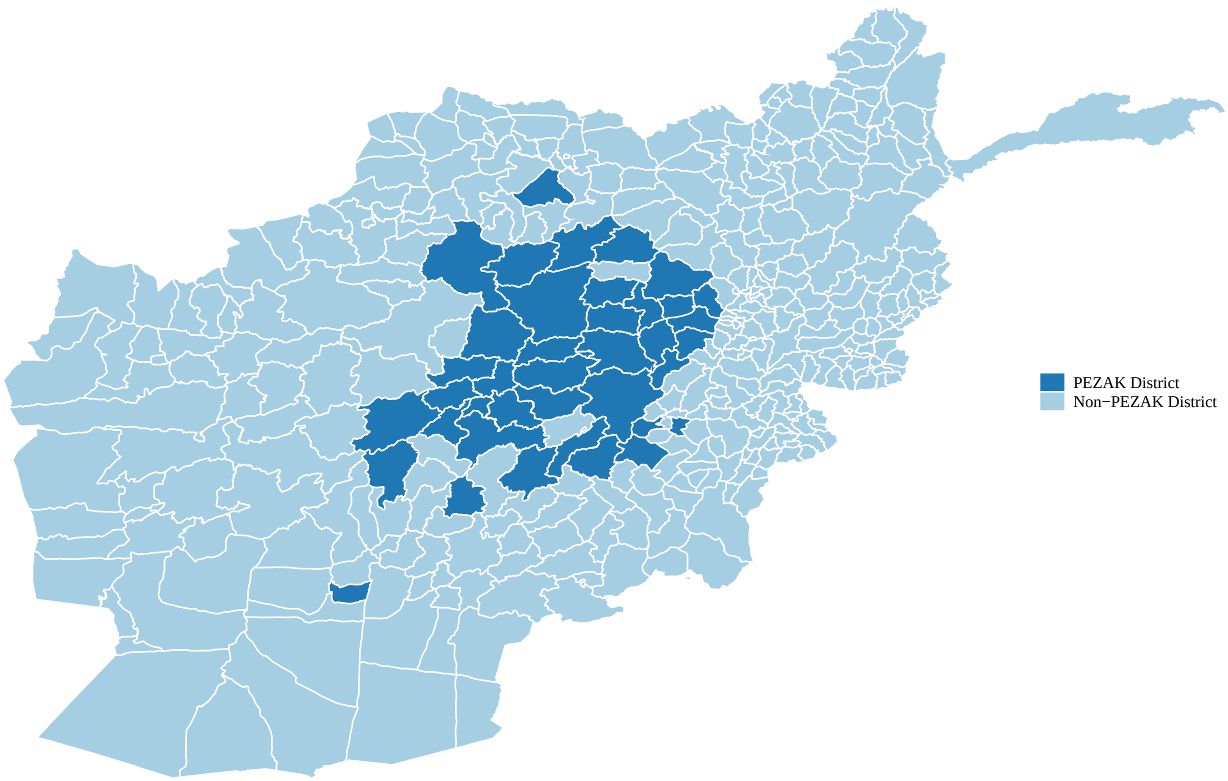
Notes: This table shows estimated average effects of the PEZAK program on male test scores, enrollment in top five universities, enrollment in low-rank provincial universities, and overall passing rate 0-6 years after adoption. All outcomes are standardized. Estimates are calculated using ***did_multipleGT*** STATA package of de Chaisemartin et al. (2019) based on de Chaisemartin and d'Haultfoeuille (2020) and de Chaisemartin and D'Haultfoeuille (2022). We apply the ***did_multipleGT*** in four scenarios. First, treatment is flexible where some schools remain treated between 2008 (first treatment) and 2019 (latest year) and for others treatment turns on and off several times. Second, treatment status is redefined by considering year T's treatment based on year T-1 treatment status. Treatment at T-1 implies treatment at T. Third, treatment at T-2 implies treatment at T. Fourth, a staggered is considered where schools remain treated after the first treatment. Fifth, treatment effect in a staggered design scenario is estimated with Callaway and Sant'Anna (2021) estimator. Sixth, scenario 1-4 are estimated using a TWFE estimator. The treatment effect is the sum of instantaneous (period 0) and dynamic (1-6) effects. The parameter can be interpreted as the average total effect per unit of treatment or average effect of receiving one unit higher number of treatments. Standard errors presented in the brackets below the estimates are calculated with a bootstrap approach with 100 replications.

Table A.2: Average Total Effects of PEZAK Program on Female Students

	Test Scores (1)	Top Five (2)	Prov Uni (3)	Passing Rate (4)
dC-DH (2021a), flexible treatment	0.073	-0.270	0.197	0.128
	(0.168)	(0.265)	(0.187)	(0.236)
dC-DH (2021a), treatment at T-1 implies treatment at T	0.058	-0.216	0.158	0.103
	(0.134)	(0.212)	(0.150)	(0.189)
dC-DH (2021a), treatment at T-2 implies treatment at T	0.051	-0.189	0.138	0.090
	(0.118)	(0.186)	(0.131)	(0.165)
dC-DH (2021a), staggered adoption	0.044	-0.162	0.118	0.077
	(0.101)	(0.159)	(0.112)	(0.142)
CS (2021), staggered adoption	0.032	-0.160	0.097	0.059
	(0.088)	(0.135)	(0.137)	(0.137)
TWFE, flexible treatment	0.080	0.109	0.007	0.182
	(0.054)	(0.078)	(0.098)	(0.077)
TWFE, treatment at T-1 implies treatment at T	0.072	0.110	0.059	0.120
	(0.048)	(0.071)	(0.081)	(0.072)
TWFE, treatment at T-2 implies treatment at T	0.089	0.088	0.076	0.183
	(0.047)	(0.070)	(0.082)	(0.072)
TWFE, staggered adoption	-0.006	-0.029	0.003	-0.003
	(0.059)	(0.086)	(0.117)	(0.091)

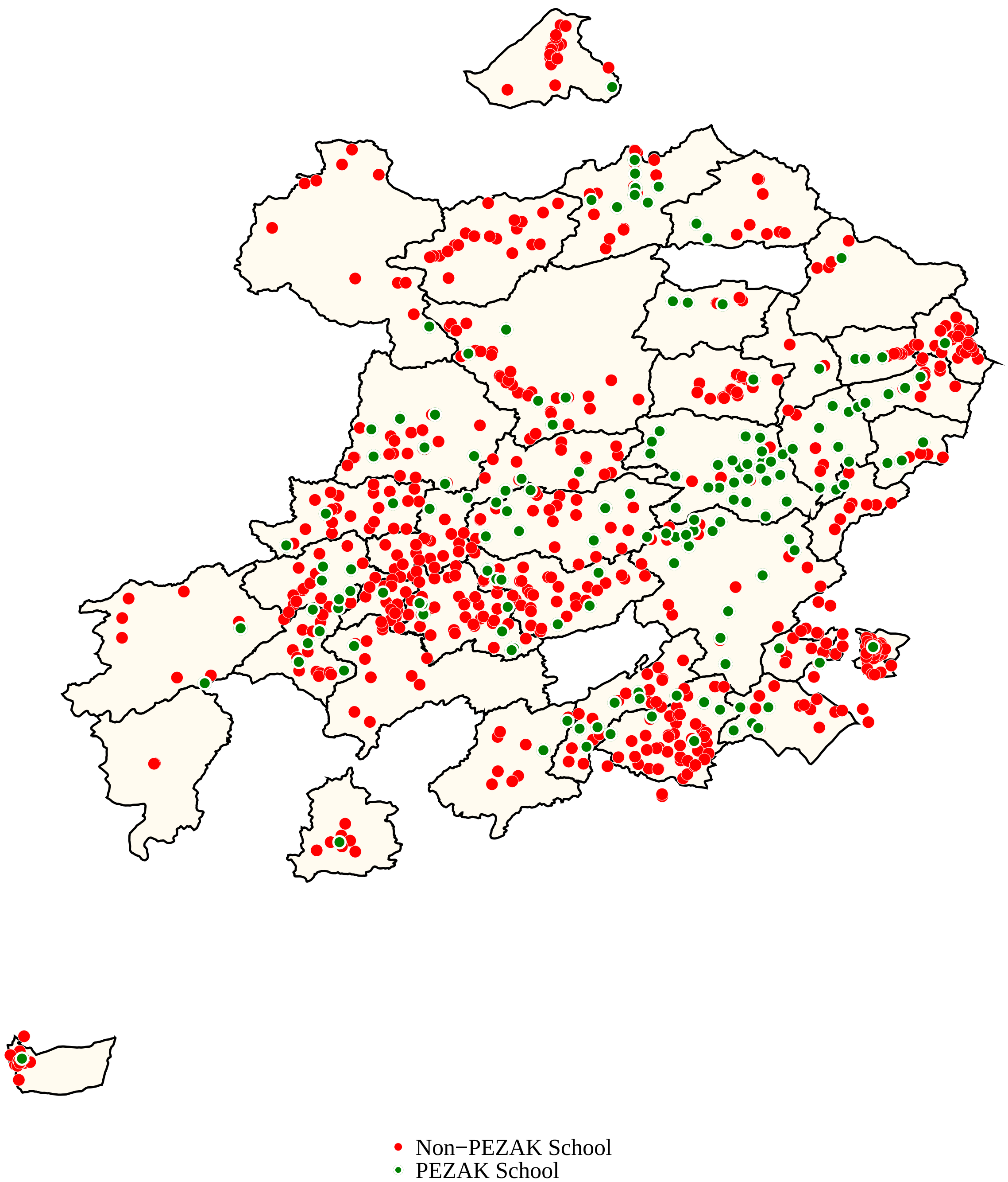
Notes: This table shows estimated average effects of the PEZAK program on female test scores, enrollment in top five universities, enrollment in low-rank provincial universities, and overall passing rate 0-6 years after adoption. All outcomes are standardized. Estimates are calculated using ***did_multipleGT*** STATA package of de Chaisemartin et al. (2019) based on de Chaisemartin and d'Haultfoeuille (2020) and de Chaisemartin and D'Haultfoeuille (2022). We apply the ***did_multipleGT*** in four scenarios. First, treatment is flexible where some schools remain treated between 2008 (first treatment) and 2019 (latest year) and for others treatment turns on and off several times. Second, treatment status is redefined by considering year T's treatment based on year T-1 treatment status. Treatment at T-1 implies treatment at T. Third, treatment at T-2 implies treatment at T. Fourth, a staggered is considered where schools remain treated after the first treatment. Fifth, treatment effect in a staggered design scenario is estimated with Callaway and Sant'Anna (2021) estimator. Sixth, scenario 1-4 are estimated using a TWFE estimator. The treatment effect is the sum of instantaneous (period 0) and dynamic (1-6) effects. The parameter can be interpreted as the average total effect per unit of treatment or average effect of receiving one unit higher number of treatments. Standard errors presented in the brackets below the estimates are calculated with a bootstrap approach with 100 replications.

B Figures



Notes: This figure shows the location of PEZAK districts on the map of Afghanistan. PEZAK districts are districts in which at least one school have received the PEZAK program. Those are mainly the districts in the Central regions of the country.

Figure B.1: Location of PEZAK Districts



Notes: This figure plots PEZAK and Non-PEZAK schools using GPS coordinate points. It is worth noting that GPS information is not available for all of the schools. Some districts have larger number missing districts as compared to others. As we can see some districts also have larger number of PEZAK school assignment. The selection of districts is entirely decided by the program organizers.

Figure B.2: Plot of PEZAK and Non-PEZAK schools

Chapter 3. Long-term Impacts of Gender-Balanced Local Development Councils on Female Education

Ahmad Shah Mobariz Andrew Beath

3.1 Introduction

Women in decision-making roles improve policy outcomes for women. For instance, village councils led by women in India choose public goods that are most favorable to women (Chattapodhyay and Duflo, 2004). Similarly, women in legislative positions has a positive role in determining inheritance rights, land reform, health, and education in favor of women (Clots-Figueras, 2011). In India, one percentage point increase in the number of women in the legislative positions increased the fraction of female school teachers by 5% (Clots-Figueras, 2012). Moreover, women in leadership positions positions affect aspirations of female students. For example, women in village councils in rural India increased female students' aspiration about themselves and also parents' aspiration about their daughters (Beaman et al., 2012).

However, the effects of women in leadership capacities on people's attitudes towards women in the labor force and general role of women in the family and society is not clear. In India, despite increasing young women's aspiration, women in influential posts did not change labor market opportunities for women (Beaman et al., 2012). On the other hand, Beath et al. (2013a) find that, in Afghanistan, equal membership of women in village development councils increased women's income in female-specific economic activities and improved mobility of women, but it did not have any impact on the role of women in decision-making within household or society. This indicates that depending on the context, the role of women in local leadership can produce different outcomes for women.

Most papers that have investigated the effects of women in leadership roles on policy outcomes, have focused on the immediate impacts of such interventions. Economic experiments are usually conducted for a specific period of time and results are measured during the experiment or at the conclusion of the experiment. But changes in behavior and economic interactions happen over a longer period of time (Normann et al., 2014). Studies that traces experimental subjects after the intervention are rare.

In this chapter, we examine the long-term impacts of gender-balanced community development councils (CDCs) on female enrollment. For this purpose, we combine current administrative education data with Afghanistan's National Solidarity Program Impact Evaluation (NSP IE). The NSP IE was a randomized experiment that was conducted between 2008-2012 to measure the effects of Afghanistan's flagship development program (Beath et al., 2015b). The experiment assigned 250 villages to treatment and 250 villages to the control group in 10 districts. Treatment villages elected village councils, half of them women, to prioritize development projects that would be funded by the National Solidarity Program (NSP), the largest community development program in Afghanistan during 2002-2017. CDCs received upto $\$60,000$ block grant to spend on development projects such as irrigation, roads, power supply, school construction, and literacy programs.

The NSP IE experiment employed a matched-pair cluster randomization procedure. Village clusters were selected such that treatment and control villages located within 1 kilometer fall in one cluster. Village pairs were selected using an optimal greedy matching algorithm based on village characteristics. One village in each pair was randomly assigned to treatment and the other pair served as control.

We assess the impact of the NSP on female enrollment at the school level. The intervention happened at the village level but due to unavailability of village level enrollment data, we use school level information. For this purpose, we construct a school sharing dataset that connects villages with the corresponding schools. We obtained a school-village pairs dataset from Afghanistan's Ministry of Education (MOE) that provides schools and names of their associated villages. We construct a school-sharing dataset using village names in the school-village pairs and the NSP IE villages. 70% of schools and villages were matched using village names. We also obtained another administrative school survey database from the MOE that provides a range of variables such as location, enrollment measures and school characteristics for the universe of schools in Afghanistan. Using the school location identifiers and geographic coordinate points in the school survey and similar identifiers for villages in the NSP IE data, we were able to complete the remaining 30% of school-sharing dataset.

We define school treatment status from the corresponding villages' treatment status. In rural Afghanistan, schools are shared between several villages due to unavailability of resources. Treated and untreated villages could share the same school. We confine our analysis to schools that were either entirely shared by treated villages or control villages. We call such schools "homogeneous" schools. School shared by treated villages serve as treatment schools and those shared by untreated villages serve as control schools. It should be noted that both treated and control schools also have a number villages that were not experimental villages (neither treatment nor control). We call these villages "non-IE villages". On average, treated and control schools has 4 and 3.4 non-IE villages, respectively. Non-IE villages did not receive NSP during NSP-IE experiment. Since the distribution of those schools is similar for treated and control villages, those villages do not affect our analysis.

We estimate the treatment effects using OLS. Since the underlying treatment assignment in the NSP IE was random, OLS estimates provide unbiased treatment effects. The outcome of interest is share of female in total school enrollment. In our basic model we consider the average enrollment during 2013-2018. To control for immediate effects of the NSP on female enrollment we include the average enrollment in the last two years of the NSP IE as baseline control. Our estimates measure the sustainability of an intervention that established gender-balanced village councils.

We find no evidence to support any long-term effects of the NSP on female enrollment. In the short run, the coefficient on treatment is insignificant at conventional levels for all outcomes. In the long run, after accounting for potential baseline differences, we fail to reject the hypothesis that there was no difference between treated and untreated schools. The NSP established CDCs with an objective to maintain them as institutions of local governance. Based on the literature about the role of women in decision-making positions on policy outcomes in favor of women, and evidence of positive effects in other settings, one would expect that gender-balanced development councils might translate into a positive effect on female education. But we do not find evidence to support this premise.

There are some heterogeneity in the treatment effects. We breakdown the total enrollment to primary, secondary and high school enrollment and estimate the treatment effect on share of female enrollment in each levels. We observe that the share of women in secondary education in the treated schools reduced by 7 percentage points. We also conduct geographic heterogeneity analysis for primary enrollment and secondary enrollment. It seems that the negative effect is driven by the province of Herat with nearly 12.5 percentage points reduction in female enrollment in secondary education. On the other hand, in the province of Daikundi the treated schools had approximately 6 percentage points higher share of female enrollment in secondary school.

Provinces in the NSP IE were selected to represent Afghanistan's ethno-linguistic and religious diversity (Beath et al., 2013a). Hesarak and Shirzad districts in Nangarhar province are predominantly Pashton and the NSP IE districts in Herat province are ethnic Tajik. The province of Daikundi is ethnic Hazara. A perception survey of the Afghan people show that in Daikundi province the attitude towards women's role in the society are relatively positive. The same survey shows negative attitudes towards women in Herat province. Therefore, we have suggestive evidence that in the long run, the NSP produced positive outcomes for female education in areas that were culturally favorable to women and negative effects in culturally conservative communities. The underlying cultural difference may have caused differences in the effects of gender-balanced village councils.

We conduct a several robustness checks. First, instead of averaging outcome data during 2013-2018, we use the 2018. This address the concerns that years close to the final stages of the NSP IE might underestimate the long-term effects. Second, we relax the definition of treated schools by assigning treatment status to schools if more than 50% of the schools that share them were treatment villages. Third, instead of removing schools whose associated villages received the NSP after the NSP IE, we control for post-IE treatment status. Finally, in another effort to increase the sample size, we combine robustness checks 2 and 3 by assigning treatment status based on more than 50% share of treated villages and by controlling for NSP status during 2013-2018 instead of dropping observations, which increase the sample size by 67%. Results from the mentioned robustness checks are similar with our main results.

In line with findings of Beath et al. (2013a), our finding suggests that gender-balanced development councils mandated by a development program in conservative contexts might be transitory. Such interventions may not change the attitude of people towards the greater role of women in society and their participation in education. Other behavioral interventions may be applied for such purposes.

Our paper contributes to the literature on women empowerment and female in leadership. We examine the sustainability of women empowerment interventions. Duflo (2012) argues that the relationship between women empowerment initiatives and development may not be strong enough to be "self-sustaining", and continued impact may require sustainable policy commitments. In Afghanistan, it is shown by Beath et al. (2013a) that the establishment of gender-balanced village councils increased female mobility and socialization in the short run. But our findings show that those changes did not translate into improvement in female education. The effects were rather counterproductive in relatively conservative communities.

We also contribute to the emerging tradition to combine past experiments with current administrative data. Field experiments in Economics have attracted an immense amount of attention. Banerjee (2020) persuades scholars to use past RCTs to explain long-term impacts. Bouguen et al. (2018) uses past RCTs to study the long-term impact of cash transfer programs on economic productivity and living standards in poor countries. However, this practice is not common yet, at least in Economics. We combine the NSP-IE experiment with current administrative school data to answer a different relevant policy question.

The rest of the chapter is organized as follows. In section 3.2 we present an overview of the research context and background. Section 3.3 introduces the data used in this paper. In section 3.4 we discuss our empirical strategy. Section 3.5 presents the results. Finally, section 3.6 concludes the chapter.

3.2 Background

3.2.1 National Solidarity Program (NSP)

The NSP was the largest development program in Afghanistan during 2002-2017. The program started as a flagship development program after the collapse of the Taliban regime and the formation of Transitional Islamic State of Afghanistan (TISA) (Center for Public Impact, 2016). It reached nearly all of Afghanistan's 35,000 villages in 398 districts. The NSP was mainly funded from Afghanistan Reconstruction Trust Fund (ARTF), a multi-donor commitment for Afghanistan's reconstruction which was administered by the World Bank. The total spending on the NSP was nearly $1.6 billion (UN-Habitat, 2017).

The NSP had two main goals. The main goal of the program was to establish gender-balanced community development councils (CDCs) as institutions of local governance (Beath et al., 2018). For this purpose, the program created elected CDCs through a direct voting system. Facilitating partners (FPs), mainly local and international NGOs, hired by the central government, helped with the election process and formation of CDCs. The second objective of the program was to consolidate communities with the central government by winning hearts and minds (Beath et al., 2013b). It was contemplated that economic development would help achieve this goal. The elected bodies selected a set of development priorities that would be funded by the program. On average, the program spent $200 on each household. The FPs also implemented the development priorities.

Between 2002-2017, three NSP phases were implemented. First, the Emergency National Solidarity Program (ENSP) implemented development projects worth $289.2 million in 16,502 villages between 2002-2006 (World Bank, 2008). From 2007 through 2011, the second NSP (NSP II) spent $394 million in new villages and completed projects that remained incomplete under NSP I (World Bank, 2014). The third phase of NSP (NSP III) was concluded in 2017 by spending $1.256 billion (World Bank, 2017). In 2017, the NSP evolved to the Citizens' Charter which was embedded in Afghanistan National Development and Peace Framework (ANPDF) (Islamic Republic of Afghanistan, 2017).

3.2.2 NSP Impact Evaluation (NSP IE)

The NSP-IE was a randomized experiment which was designed as part of the NSP II to rigorously examine the effects of NSP on a comprehensive range of outcomes related to the goals of the program (Beath et al., 2015a). 500 villages were selected in 10 districts that did not receive NSP in the first phase. Half of the evaluation villages were randomly assigned to the program and the other 250 villages were selected as controls. All of the 10 districts had equal number of treatment and control villages. The treatment villages established village councils and received NSP grants during 2008-2012. The control villages did not receive NSP in NSP II but would receive the program in NSP III in 2013.

The NSP IE districts were selected based on three conditions. First, IE districts should not have had any NSP experience in the first phase. Second, the security situation of the districts would allow for execution of NSP projects and data collection. Third, the districts should have a minimum of 65 villages. 50 villages in each district were selected based on certain eligibility criteria such as the minimum number of households in the village. 25 pairs of villages were constructed based on a distance threshold between them. One of the villages in each pair was randomly assigned to receive NSP and the other pair would serve as control. The treatment-control split was also constrained by a clustering condition such that all treatment and control villages would fall in distinct clusters, respectively (see Figures A.1-A.10).

3.3 Data

To conduct this study, we combine four data sources. The NSP IE village survey data and project information before and after the intervention, village location and treatment treatment status for 500 villages was collected from NSP IE website. Administrative school enrollment and school characteristics data was accessed from the Ministry of Education of Afghanistan (MOE). This is a comprehensive database of all of Afghanistan's 18000 schools. We also collected school-village pairs data from the MOE which links schools with corresponding villages. The fourth database is a comprehensive database of NSP projects which includes projects implemented under three NSP phases and Citizens' Charter. This database was accessed from Afghanistan's Ministry of Rural Rehabilitation and Development (MRRD).

We matched school-sharing data with the NSP IE data and NSP project-level information using the village names. The corresponding school for 70% of NSP IE villages were identified using village names. Schools for the remaining villages were located using the least Euclidean distance between villages and schools. School-village pairs data thus constructed allows us to understand which treated and control villages share a particular school.

We identified NSP status for NSP IE villages in NSP III. Table 3.1 provides a summary of matched schools and villages. In total, there are 1835 villages in the NSP IE districts. 500 of these villages were part of the NSP IE of which 250 were treatment and 250 were control. Out of 250 treatment villages 53 villages continued to have an NSP project in NSP III, and 197 treatment villages never received NSP after the completion of NSP IE. Out of 250 control villages, however, 181 villages received NSP after IE and 69 villages did not.

There are 426 schools in the NSP IE districts of which 265 schools are shared at least by one NSP IE treatment or control village along with non-IE villages. Out of the 265 villages shared by the NSP IE villages, 164 villages were homogeneous. We define homogeneity of schools based on treatment-control composition of associated villages. Schools that are just shared by treatment or just control villages are considered homogeneous, and heterogeneous otherwise. Out of 164 homogeneous schools 90 were shared just by treated villages and 74 by control villages.

Moreover, we separate the homogeneous schools by their NSP status after NSP IE. Out of the 90 homogeneous treatment schools, 28 schools were associated with villages that received NSP in NSP III and 59 schools were shared by villages that did not receive NSP in NSP III. Similarly, among homogeneous control villages, 30 schools were associated with villages that received NSP after 2012 and 43 schools belonged to villages that did not received NSP in NSP III. In our main analysis, we work with homogeneous schools that did not receive NSP after the conclusion of NSP IE. Homogeneous treatment schools are considered as treated schools and vice verse.

Treatment assignment in the NSP IE experiment was random. At the baseline in 2007, village characteristics were fairly balanced (Beath et al., 2013a). Using the survey data from the experiment, in Table 3.2 we examine balance in additional village characteristics including indicators for education. In terms of magnitudes, the average population in control villages was larger than treatment villages. Similarly, average household monthly income in control villages was 281.4 Afghani (approx. $6) higher than treatment villages. Variables such as share of spending on children education and willingness to send additional daughter to school were approximately the same. However, none of the differences in village characteristics were not statistically significant.

At the end of NSP IE, treated villages had higher per capita income, and better access to services (Beath et al., 2015a). Those differences were due to the NSP (Beath et al., 2013a). Treated and control villages did not demonstrate any differences in male education. On average, self-reported hours spent in school by women was more in treatment villages than control villages. We examine if this difference translated in higher enrollment of women in school in the years after completion of the program.

Table 3.3 compares mean school characteristics for 102 schools based on Afghanistan's school survey data in 2011. On average, schools in both groups were established around the same time and they were not different in quality of teacher education. Similarly, school facilities such as toilets for women were not different in both groups. Therefore, the characteristics of 59 homogeneous treated "treatment schools" and 43 homogeneous control "control schools" that we use in this study were similar.

3.4 Empirical Strategy

We estimate the effects of NSP in two stages. First, we examine short-term impacts of the NSP on female enrollment. We define short-term to be four years in the program. For this purpose, we use enrollment data for 2011, i.e., the fourth year of NSP IE. Since the treatment assignment was random, and until 2012 NSP IE treatment villages were the only recipients of NSP in NSP IE districts, there is no concern about selection bias. Schools that were homogeneously shared by NSP treatment villages serve as treatment group. Similarly, schools which were homogeneously shared by control villages are considered control schools. We estimate the treatment effect with an OLS estimator as presented in the following equation.

$$Y_s = \beta_0 + \beta_1 T_s + \varepsilon_s \tag{3.1}$$

where, Y_s is the outcome of interest for school s. Outcome include share of female enrollment in three stages: primary, secondary, and high school. $T_s = 1$ if the school s belongs to a treatment village (or group of treated villages in case of shared school) and $T_s = 0$ if the school s belongs to a control village (or group of control villages). ε_s captures the unobserved school characteristics.

To estimate long-term effects of the NSP on female enrollment, we use average enrollment for years after NSP IE (2013-2018). We exclude schools whose associated villages received the NSP during 2013-2018. To account for differences that could be caused by the NSP in the short-run, in Equation 3.1, we include the average of outcomes in 2011 and 2012 which correspond to the final years of the NSP IE. We estimate Equation 3.2 below.

$$Y_s = \beta_0 + \beta_1 T_s + \beta_2 Y_0 + \varepsilon_s \tag{3.2}$$

where, Y_0 denotes short-run outcome (average of 2011 and 2012). Standard errors are clustered at district level to account for school characteristics that might be correlated with district features.

3.5 Results

3.5.1 Short-term Effects

Table 3.5 presents OLS results based on Equation 3.1. It shows short-term effects of the NSP. The outcome of interest is school-level share of girls in total enrollment, i.e., enrollment across grades averaged over 2013-2018 period. Columns (1) and (2) differ in the definition of treatment status. In column (1) schools are homogeneous in their treatment status and in column (2) the homogeneity assumption is relaxed to include schools shared by more than 50% treated villages. Boys' school is a binary variable that indicates if school is designated for boys alone. In rural Afghanistan, due to lack of school infrastructure, schools designated for boys or girls have classes for the opposite gender as well. 66% of schools in NSP IE districts have boys and girls classes.

Our regression results supports the null hypothesis. The coefficient for the treatment variable is positive in magnitude but statistically insignificant at conventional levels. Considering the mean and stand error of the control group, with a power of 0.8 and 5% significance level, one would expect to get a minimum effect size of 13 percentage points. But the effect size from our estimates is 1.5 percentage points. Beath et al. (2013a) show that the NSP increased the number of hours women spent at school. We do not have precise evidence for any impact of the NSP on female enrollment.

3.5.2 Long-term Effects

Table 3.6 presents regression results from Equation 3.2 which estimates long-term effects of the NSP on female enrollment. It is long-run because nearly 10 years has passed since the start of the program. To avoid loss of observations due to missing outcome data in some years, we average the outcome for 2013-2018. Column (1) presents results without controlling for baseline differences. In column (2), we control for baseline outcome. Enrollment data for 2008-2010 is not available. Baseline refers to the last year of the NSP IE. For baseline outcomes, we consider the average of 2011 and 2012 which are the final two years of NSP IE. It is also worth noting that in addition to homogeneity restriction, we also contain the sample size to schools whose associated villages did not receive NSP after the conclusion of the NSP IE.

In the basic model without baseline outcome, the treatment coefficient is negative in magnitude but statistically insignificant. Controlling for short-term differences in column

(2) column does not change the results. We fail to reject the null hypothesis of no difference in school enrollment between NSP treated and control schools.

3.5.3 Robustness Checks

We take several measures to evaluate the robustness of our findings. These results are presented in Table 3.7. Panel A presents results without baseline outcome and Panel B controls for differences in the final years of the NSP IE.

First, instead of averaging outcomes, we consider the data for 2018 alone. This is the latest year for which enrollment data is available. This exercise addresses the concerns that results could have been driven by years close to the final years of the NSP IE conclusion which may not reflect long-term effects. Results for this exploration is provided in column (1). It shows that the coefficient on treatment does not change.

Second, we explore if increasing the sample size affects our results. For this purpose, we consider a broader definition of treatment status such that when more than 50% of villages that share a school were treated, are considered as treated. OLS results for this exercise is given in column (2). The treatment coefficient is strategically insignificant in both panels but the sign of the coefficient are in opposite directions.

Third, instead of constraining the sample size to schools whose village did not receive NSP between 2013-2018, we control for treatment status in post NSP IE period. Results for this robustness check are presented in column (3). Similar to columns (1) and (2) results are not precisely estimated. We fail to reject the null hypothesis.

Finally, we relax restrictions on homogeneity and post NSP IE treatment status which increase the sample size by approx. 64%. Schools shared by majority (above 50%) treated villages are defined treated. Moreover, instead of dropping schools whose villages received the NSP during 2013-2018, we control for post-IE NSP status. The result is given in column (4). We do not find evidence that supports long-term effects of NSP on female enrollment.

3.5.4 Heterogeneity Analysis

We conduct three types of heterogeneity analysis. First we examine potential treatment effects by disaggregating enrollment to primary, secondary and high school and estimate Equation 3.2 with each enrollment levels. Results are presented in Table 3.8. The treatment coefficient is negative and insignificant for primary enrollment and positive but insignificant for high school enrollment. However, in the case of secondary enrollment it is negative and significant at **5%** significance level. The average share of female enrollment in secondary school (grades 7-9) in the NSP schools during 2013-2018 was 7 percentage points lower as compared to control schools.

Next, we examine spatial heterogeneity at the province level. The NSP IE was conducted in 10 districts in six provinces. In the estimation of geographic differences, to maintain degrees of freedom, instead of district we interact province with the treatment indicator. We did not get sufficient observations in case of high school enrollment to conduct a province-by-province analysis. We run separate OLS regressions for total enrollment, primary enrollment, and secondary secondary enrollment. This result is presented in Figure 3.1.

There is some geographic heterogeneity in treatment effects. In Table 3.6, we observe a statistically significant negative effect of the NSP on female enrollment in secondary school. Figure 3.1 compares differences in treatment effects in four province: Daikundi, Ghor, Herat and Nangarhar. Primary enrollment in treatment schools were approximately 2.5 percentage points lower than control schools in the province of Daikundi and Herat and nearly 6 percentage points lower in Nangarhar. More importantly, we observe a stark difference in secondary enrollment in the provinces of Daikundi and Herat. In the province of Daikundi the treatment effect is approximately 6 percentage points higher but in Herat province it is approximately 12.5 percentage points lower than control schools.

Provinces represent Afghanistan's ethnic diversity. Experimental districts in the province of Ghor and Nangarhar are ethnic Pashtoon, Herat is ethnic Tajik, and Daikundi is ethnic

Hazara. To explore potential sources of geographic heterogeneity in treatment effects, we compare people's perception about women's education, employment and marriage age in provinces of Baghlan, Balkh, Daikundi, Ghor, Herat and Nangarhar. Since the NSP IE was conducted in rural areas of these provinces, we do not include provincial capitals. Table 3.3 provides a summary of key variables that reflect public opinion towards women based on the Asia Foundation's Survey of the Afghan People between 2006-2019 (Akseer et al., 2017). Numbers in this table denote the share of respondents with certain opinion. The Central province of Daikundi has the most favorable attitude, and Ghor and Herat has the most negative attitude towards women. For instance the share of people agreeing with women traveling to a different location for higher education is 73.64% in Daikundi and 37.7% in Herat. Similarly, 78.7% of people in Daikundi had a positive opinion with women's work outside home while it was 68% in Herat. 13% of people in Daikundi agreed with complete face-covering for women while the share of people with this opinion in the province of Herat was approximately 23%. The gender-balanced village councils in the conservative setting such as Herat province may have produced counter-productive results for women in the long run while it improved gender outcomes in areas with relatively positive attitude towards women.

Finally, we examine temporal heterogeneity. As shown in Figure 3.2, the treatment effect does not change over time. In case of high school enrollment, the difference in share of female enrollment between treated and control schools is approx. 20% but with very wide confidence interval. In the case of primary enrollment, the difference is near zero in all of the years post NSP IE. The difference in secondary and high school enrollment is zero in 2012, 2013, and 2018 but negative in other years. Therefore, there seems to be some temporal heterogeneity in case of secondary and high school enrollment. The higher the distance from last year of the NSP IE, the wider differences in share of female enrollment.

3.6 Conclusion

We presented long-term effects of gender-balanced local development councils on female enrollment. We use data from a randomized impact evaluation of Afghanistan's National Solidarity Program which was conduced during 2008-2012. The NSP was the largest and flagship development program in Afghanistan between 2002-2017 that mandated gender-balanced development councils. Drawing from the literature about women in leadership and decision-making positions, we hypothesize that such program might have a positive impact on female education.

First, we presented the impact of the NSP in the short-run. For this purpose, we define the final two years of the NSP IE as short-run. We do not find any evidence to conclude any differences in share of female enrollment in primary, secondary and high school. Second, we investigate long-term effects of the program. To do so, we average outcomes for years 2013-2018 and estimate the effects using OLS. We also account for short-term differences in outcome by including baseline outcomes in the regressions. Except for enrollment in secondary school, we do not find sufficient evidence for difference between treatment control schools 10 years after the start of the program and five years after the conclusion of the program.

We observe some heterogeneity. First, we disaggregate enrollment data to primary, secondary and high school. The treatment effect on secondary enrollment was -7.06 percentage points and significant at 5% significance level. OLS results for other outcomes do not provide conclusive evidence for treatment effects. Our heterogeneity analysis in four provinces of Daikundi, Ghor, Herat and Nangarhar show some differences. The treatment effect on share of female enrollment in secondary school is approximately 6 percentage points in Daikundi province. The treatment effect on the same outcome in Herat provinces is approximately -12.5 percentage points. It shows that the long term effects of the NSP on share of female enrollment at higher school grades was positive in some places and negative in some other places. From perception survey data we have suggestive evidence that the effect was positive in areas that have culturally favorable attitude towards women and negative in relatively conservative areas. Furthermore, except for some marginally significant differences in share of female enrollment in primary school in the later years, the treatment effect is similar across years.

Our findings suggest that in conservative contexts, mandatory female in the local leadership may not be effective in changing the overall role of women in the society and attitude towards female education. For this to happen, other behavioral changing interventions may be required.

3.7 Bibliography

Akseer, Tabasum, Mohammad Shoaib Haidary, Rebecca Miller, Sayed Masood Sadat, Christina Satkowski, Helen Seese, Mohammad Jawad Shahabi, Kris Veenstra, Zachary Warren, and Fahim Ahmad Yousufzai, "Afghanistan in 2017: A survey of the Afghan people," 2017.

Banerjee, Abhijit Vinayak, "Field experiments and the practice of economics," American Economic Review, 2020, 110 (7), 1937–51.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova, "Female leadership raises aspirations and educational attainment for girls: A policy experiment in India," science, 2012, 335 (6068), 582–586.

Beath, Andrew, Fotini Christia, and Ruben Enikolopov, "Empowering women through development aid: Evidence from a field experiment in Afghanistan," American Political Science Review, 2013, pp. 540–557.

_ , _ , and _ , "Winning hearts and minds through development: Evidence from a field experiment in Afghanistan," in "APSA 2013 Annual Meeting Paper, American Political Science Association 2013 Annual Meeting" 2013.

_ , _ , and _ , "The National Solidarity Program: Assessing the Effects of Community-Driven Development in Afghanistan," 2015.

_ , _ , and _ , "The National Solidarity Programme: Assessing the effects of community-driven development in Afghanistan," International Peacekeeping, 2015, 22 (4), 302–320.

_ , _ , and _ , "Do elected councils improve governance? Experimental evidence on local institutions in Afghanistan," 2018.

Bouguen, Adrien, Yue Huang, Michael Kremer, and Edward Miguel, "Using RCTs to Estimate Long-Run Impacts in Development Economics," Technical Report, National Bureau of Economic Research 2018.

Center for Public Impact, "Building trust in government: Afghanistan's National Solidarity Programme (NSP)," 2016.

Chattapodhyay, Raghabendra and Esther Duflo, "Women as policy makers: Evidence from a randomized policy experiment in India," Econometrica, 2004, 72 (5), 1409–1443.

Clots-Figueras, Irma, "Women in politics: Evidence from the Indian States," Journal of public Economics, 2011, 95 (7-8), 664–690.

_ , "Are female leaders good for education? Evidence from India," American Economic Journal: Applied Economics, 2012, 4 (1), 212–44.

Duflo, Esther, “Women empowerment and economic development,” Journal of Economic literature, 2012, 50 (4), 1051–79.

Islamic Republic of Afghanistan, “Afghanistan National Development Strategy (ANPDF),” Technical Report 2017. retrieved from https://www.refworld.org/pdfid/5b28f4294.pdf.

Normann, Hans-Theo, Till Requate, and Israel Waichman, “Do short-term laboratory experiments provide valid descriptions of long-term economic interactions? A study of Cournot markets,” Experimental Economics, 2014, 17 (3), 371–390.

UN-Habitat, “Analytical Closure Report: National Solodiarity Programme,” Technical Report 2017. retrieved from https://unhabitat.org/sites/default/files/download-manager-files/ANALYTIC%20CLOSURE%20REPORT%20-%20National%20Solidarity%20Programme%20%28NSP%29.pdf.

World Bank, “Implementation Completion and Results Report,” Technical Report 2008. retrieved from the World Bank, https://documents1.worldbank.org/curated/en/463571475112276713/pdf/000020051-20140618090947.pdf.

_ , “Implementation Completion and Results Report,” Technical Report 2014. retrieved from the World Bank, https://documents1.worldbank.org/curated/en/423251475110479015/pdf/000020051-20140625192550.pdf2.

_ , “Implementation Completion and Results Report,” Technical Report 2017. retrieved from the World Bank, https://documents1.worldbank.org/curated/en/984941514909801500/pdf/ICR00003688-12282017.pdf.

3.8 Tables

Table 3.1: General Description of NSP Villages and Schools

Description	Value
PANEL A. VILLAGES	
NSP IE districts	10
Number of villages in NSP IE districts	1835
NSP IE villages	500
Treatment	250
Received NSP after IE	53
Did not recive NSP after IE	197
Control	250
Received NSP after IE	181
Did not recive NSP after IE	69
Non-IE villages	1335
PANEL B. SCHOOLS	
Number of Schools in IE districts	426
Shared by IE and non-IE villages	265
Homogeneously shared by treated or control IE villages	164
IE Treatment	90
treated after IE	28
not treated after IE	59
IE Control	74
treated after IE	30
not treated after IE	43
Shared by treated and control IE villages	101
Shared by non-IE villages only	161

Notes: NSP = National Solidarity Program. IE = Impact Evaluation (a randomized NSP intervention intended to evaluate the NSP programs). non-IE villages = villages that were not part of the IE program but are in the districts that IE was implemented. IE Treatment = villages under the NSP IE that were assigned to the NSP program. IE Control = villages under NSP IE that were part of the control group. During IE refers to the period 2008-12 during which NSP IE was conducted. After IE refers to the period after IE was concluded (2013 and after). NPS Schools refer to schools that were shared by NSP IE villages.

Table 3.2: Baseline comparison of NSP villages

	(1) Mean Treatment	(2) Mean Control	(3) Mean Difference	(4) Mean Total
Village population	903.01	886.86	-16.15	894.94
Number of houses in the village	93.98	82.25	-11.73	88.11
Number of households in the village	148.98	143.78	-5.20	146.38
Averge household size	9.79	9.92	0.13	9.86
Number of childern below 15 in the village	409.30	390.83	-18.46	400.06
Number of adult between 15-16 in the village	476.77	459.67	-17.10	468.22
Number of HH men between 15-60	4.56	4.59	0.03	4.58
Number of HH childern below 15	4.56	4.59	0.03	4.58
Number of HH women between 15-60	2.35	2.42	0.07	2.39
Average number childern per adult female	5.45	5.61	0.17	5.53
Average monthly houshold income	6774.23	7055.61	281.37	6914.92
Shere of spending on children's education	0.00	0.00	-0.00	0.00
Share of spending on food	0.58	0.57	-0.01	0.58
Price of wheat (Afs/Kg)	13.10	13.21	0.11	13.16
Price of flour (Afs/Kg)	16.80	17.14	0.34	16.97
Price of rice (Afs/Kg)	34.27	34.27	-0.00	34.27
Price of liquid gas (Afs/Ltr)	64.16	63.75	-0.41	63.96
Price of disel (Afs/Ltr)	45.12	45.77	0.66	45.45
Wage rate (Afs/day)	159.12	163.07	3.95	161.09
Averge adult male years of education	1.47	1.56	0.09	1.52
Number of HHs that would send additonal son to school	8.66	8.54	-0.12	8.60
Number of HHs that would send additonal daughter to school	7.87	7.82	-0.05	7.85
Observations	250	250	500	500

Notes: We use experimental data in Beath et al. (2015b). This table compares treatment and control villages in the NSP experiment. Column 1 and column 2 show average village characteristics for the treatment and control villages, respectively. Column 3 show the difference in average characteristics between treatment and control villages. Statistically significant difference in average characteristics is denoted by $\star$. Column 4 shows the mean characteristics of all villages.

Table 3.3: Comparison of NSP IE Schools

	(1) Mean Treatment	(2) Mean Control	(3) Mean Difference	(4) Mean Total
Number of non-IE villages	3.75	3.44	-0.30	3.62
School age	10.83	10.21	-0.62	10.57
Teacher with an Undergraduate degree or higher	0.00	0.00	0.00	0.00
Number of female toilets	0.22	0.21	-0.01	0.22
Number of mixed toilets	0.00	0.00	0.00	0.00
Observations	59	43	102	102

Notes: Data is from Afghanistan's Ministry of Education (MOE). This table compares treatment and control schools. Treated schools refer to schools that were entirely shared by NSP IE treated villages. Similarly, control schools are those that were entirely shared by control villages. Moreover, villages associated with treated and control schools did not receive NSP after NSP IE. Both treated and control schools were also shared by neutral, non-IE villages. Column 1 and column 2 show average school characteristics for the treatment and control villages, respectively. Column 3 show the difference in average characteristics between treated and control schools. Statistically significant difference in average characteristics is denoted by $\star$. Column 4 shows the mean characteristics of all schools.

Table 3.4: Perception Towards Women in NSP IE Provinces

	(1) Baghlan	(2) Balkh	(3) Daikundi	(4) Ghor	(5) Herat	(6) Nangarhar
Madrasa education for women	94.67	94.56	89.68	86.95	87.16	97.30
primary education for women	92.96	85.50	82.40	69.35	84.19	94.20
secondary education for women	87.85	88.30	83.89	61.22	78.02	85.75
university education for women in the same province	70.53	81.32	86.13	46.89	62.29	64.50
university education for women in different province	38.40	50.52	73.64	39.05	37.70	42.22
university education for women abroad	27.87	41.99	65.39	41.36	32.31	31.02
women working outside home	67.53	78.45	78.71	55.53	67.99	58.66
women working in government offices	62.63	84.69	84.86	55.03	66.91	72.50
women as politician	49.30	49.30	57.24	46.42	44.31	35.65
women working in NGOs	38.28	60.96	73.43	29.51	46.63	49.22
women working in schools	76.38	90.01	73.33	41.69	79.11	77.75
women working in health sector	83.24	87.72	81.71	63.40	75.64	85.63
women working in the army	24.30	41.75	54.79	33.89	35.53	40.80
women should wear Burqa	39.01	23.79	12.88	29.02	22.58	63.36
marriage age for women	19.88	19.64	18.52	18.96	18.47	19.18

Notes: This table compares people's perception towards women's education, employment and marriage in NSP IE provinces. Because of insufficient observations at the district level, we use province instead of district. Baghlan, Balkh, Daikundi, Ghor, Herat and Nangarhar are the six provinces in which the NSP IE was conducted. Since the NSP IE was conducted in rural areas, we remove provincial capitals in this comparison so that the province average is sufficiently close to the NSP IE district average. Values are the percentage of respondents in the Survey of the Afghan People (SAP) with specific opinion about women. For instance, Madrasa education for women indicates the share of respondents who agreed with religious education for women. The last variable in the table "marriage age for women" is the age at which people think women should get married.

Table 3.5: Short-term Effects of NSP on Female Enrollment

	Homogeneous Schools	Treated Villages > 50%
	(1)	(2)
Treated	1.45	0.40
	(4.42)	(2.46)
Boys School	-40.28***	-39.58***
	(9.60)	(7.21)
Constant	82.54***	82.58***
	(8.34)	(6.66)
Obs	117	195

Notes: This table presents OLS estimates. Values in the brackets represent standard errors. Standard errors are clustered at district level. */**/*** denotes significance at the 10/5/1 percent. Outcome is school-level share of total female enrollment in grades 1-12. In column (1) treated schools refer to schools that were entirely shared by NSP IE treated villages (homogeneous treatment). Similarly, control schools are those that were entirely shared by control villages (homogeneous control). In column (2), treatment status is defined based on the share of treated villages in the total number of villages that share a school. If more than 50% of the villages that share a school are treated villages, the school is considered a treated school, otherwise a control school. Non-IE villages are present in both treatment and control groups. Column (1) is a restricted sample. Boys School is an indicator variable that captures if a school is entirely boys school or boys school with some girls school. The variable is equal to 1 if a schools is designated as boys school or boys school with some girls classes, and 0 otherwise.

Table 3.6: Long-term Effects of NSP on Female Enrollment

	W/o Baseline Outcome	With Baseline Outcome
	(1)	(2)
Treated	-1.28	-0.76
	(4.61)	(0.68)
Boys School	-34.26***	-4.18*
	(6.57)	(2.05)
Baseline		0.96***
		(0.04)
Obs	81	67

Notes: This table present OLS estimates. Values in the brackets represent the standard errors. */**/*** denotes significance at the 10/5/1 percent. The outcome variable is school-level share of female in school (grades 1-12) averaged over 2013-2018. Treated schools refer to schools that were entirely shared by NSP IE treated villages (homogeneous treatment). Similarly, control schools are those that were entirely shared by control villages (homogeneous control). Villages associated with schools in this sample size did not receive NSP between 2013-2018. Non-IE villages are present in both treatment and control groups. Column (1) presents estimates without baseline outcome. In column (2) baseline outcomes are included. Baseline outcome is the average of outcome during 2011 and 2012, which is the last two years of NSP IE. Including baseline outcome reduces the sample size due to missing value. The variable Boys School controls for gender designation of schools. The variable is equal to 1 if a schools is designated as boys school or boys school with some girls classes, and 0 otherwise.

Table 3.7: Robustness Checks. Long-term Effects of NSP on Female Enrollment

	Panel A - Basic Model			
	W/o Averaging (1)	> 50% Treated Villages (2)	Control for Post-IE Status (3)	> 50% Treated Villages & Control for Post-IE (4)
Treated				
Boys School	-32.47***	-36.40***	-38.04***	-35.50***
	(5.89)	(6.93)	(7.06)	(6.53)
Received NSP after IE			-0.80	-2.27
			(2.65)	(2.95)
Constant	76.07***	76.54***	75.80***	74.54***
	(3.34)	(4.89)	(4.21)	(4.58)
Obs	76	130	128	213

	Panel B - Control for Baseline Outcomes			
	W/o Averaging (1)	> 50% Treated Villages (2)	Control for Post-IE Status (3)	> 50% Treated Villages & Control for Post-IE (4)
Treated				
Boys School	-5.16	-4.80**	-4.33	-3.18
	(3.34)	(1.98)	(3.59)	(2.29)
Received NSP after IE			0.87	-0.84
			(1.38)	(1.13)
Baseline	0.97***	0.94***	0.94***	0.92***
	(0.06)	(0.04)	(0.06)	(0.04)
Obs	64	113	111	187

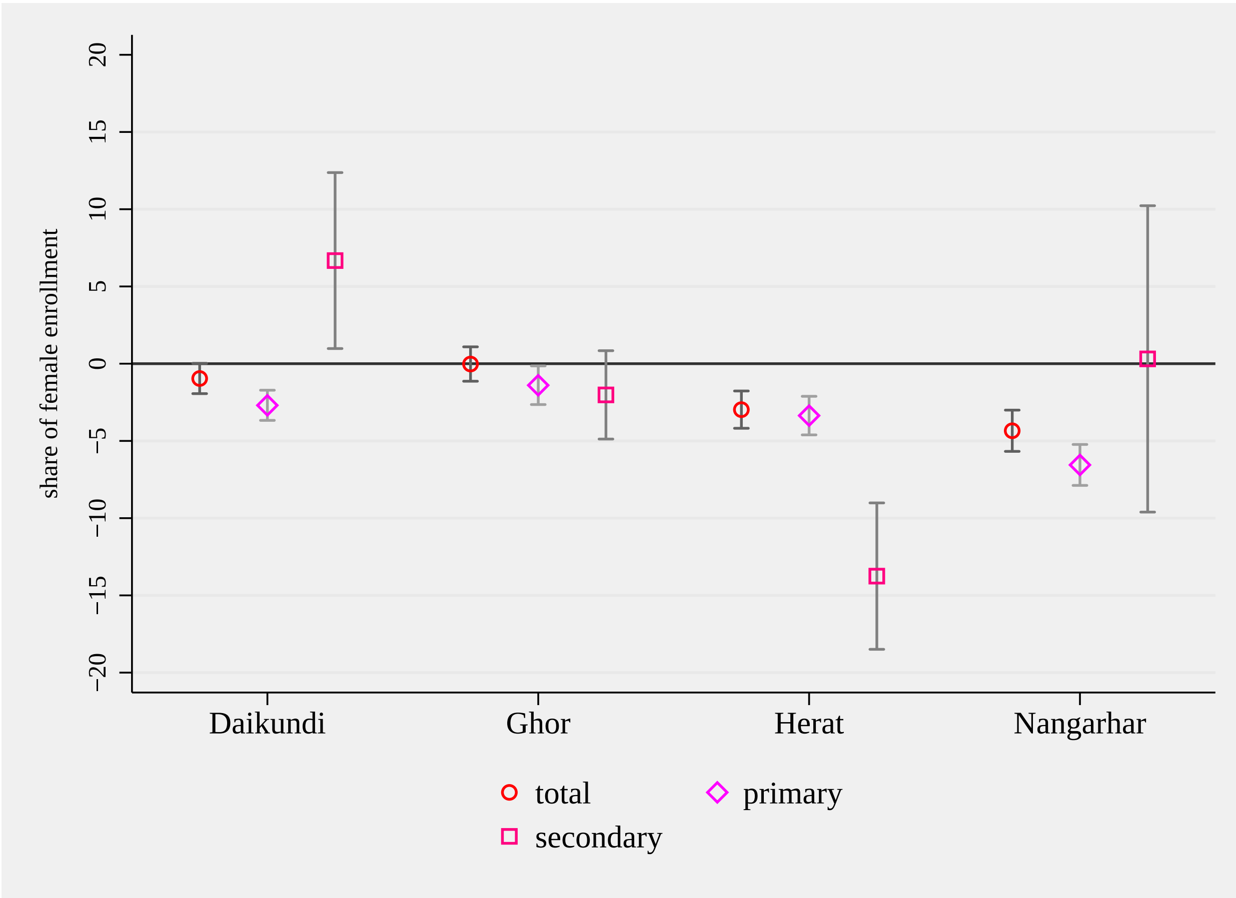
Notes: This table presents OLS estimates. Values in the brackets represent the standard errors. */**/*** denotes significance at the 10/5/1 percent. The outcome variable is school-level share of female enrollment (grades 1-12) averaged over 2013-2018. In Panel (a) baseline outcome is not included but it is included in Panel (b). In column (1), instead of averaging outcome over 2013-2018, outcome data for 2018 is considered. In column (2) the treatment status assignment to schools is relaxed. Instead of "homogeneous treatment", schools are considered treated if more than 50% of villages corresponding to school was treated. In column (3) schools are homogeneous but instead of restricting the sample to school that did not receive NSP after NSP IE, we control for NSP status during 2013-2018. In column (4) we relax both heterogeneity of treatment and post-IE NSP status by combining assumptions in columns (1) and (2). The variable Boys School controls for gender designation of schools. The variable is equal to 1 if a schools is designated as boys school or boys school with some girls classes, and 0 otherwise.

Table 3.8: Heterogeneity Analysis. Long-term Effects of NSP on Female Enrollment

	% Female Enrollment Primary	% Female Enrollment Secondary	% Female Enrollment High School
	(1)	(2)	(3)
Treated	-0.82	-7.06**	1.03
	(1.12)	(2.90)	(5.64)
Boys School	-2.67	-11.54	-51.90***
	(2.09)	(8.54)	(9.15)
Baseline			
Obs	66	39	15

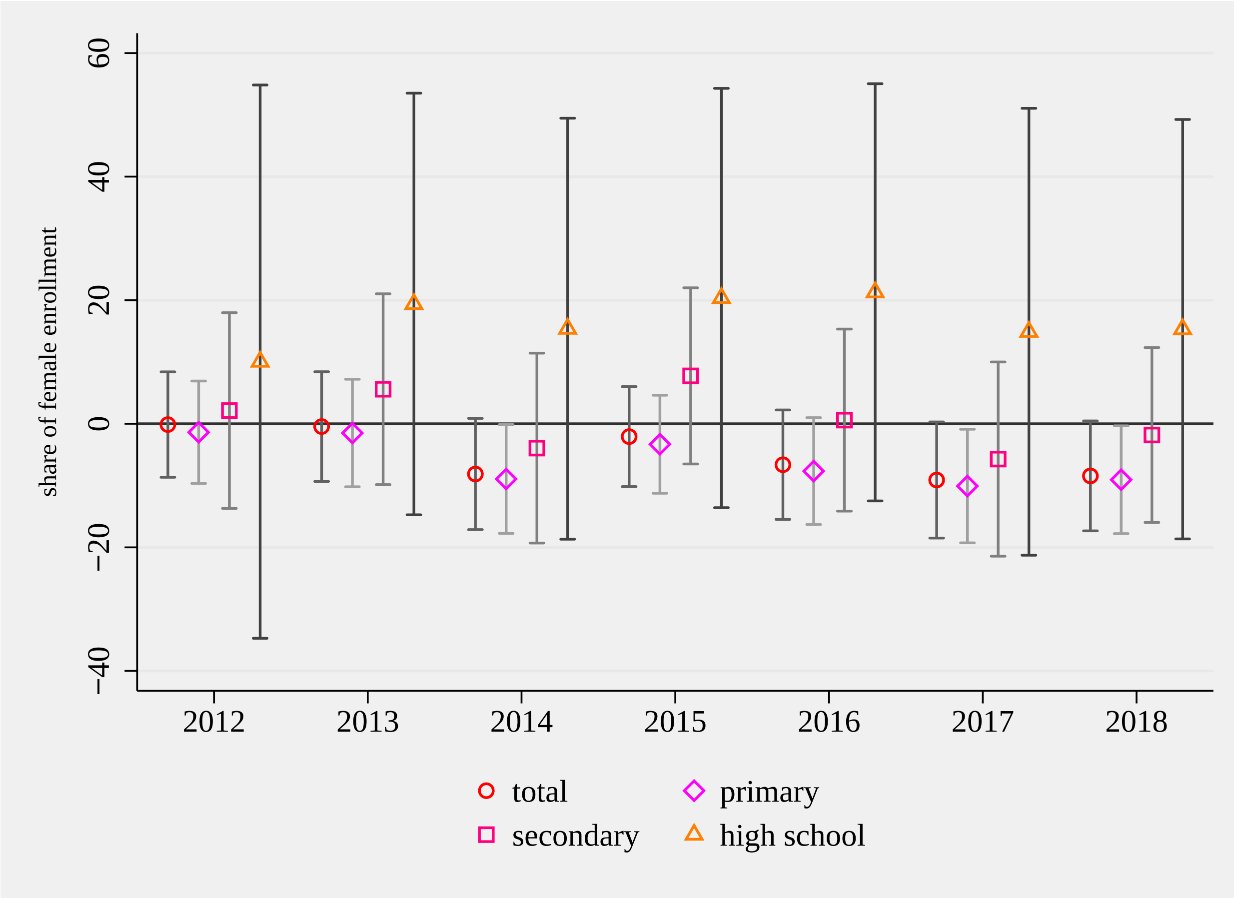
Notes: This table presents OLS estimates. Values in the brackets represent standard errors. */**/*** denotes significance at the 10/5/1 percent. The outcome variables are school-level share of female enrollment in primary, secondary and high school averaged over the 2013-2018 period. Treated schools refer to schools that were entirely shared by NSP IE treated villages (homogeneous treatment). Similarly, control schools are those that were entirely shared by control villages (homogeneous control). Villages associated with schools in this sample size did not receive NSP between 2013-2018. Non-IE villages are present in both treatment and control groups. Baseline outcomes are included in all columns. The variable Boys School controls for gender designation of schools. The variable is equal to 1 if a schools is designated as boys school or boys school with some girls classes, and 0 otherwise.

3.9 Figures



Notes: This figure presents coefficients for treatment interacted with province in OLS regressions. The outcome variables are average of 2013-2018. Treatment refers to schools that were entirely shared by NSP IE treated villages. Similarly, control schools are those that were entirely shared by treated villages. Schools whose village received NSP after NSP IE (2013-2018) are dropped. Non-IE villages are present in both treatment and control groups. The average of 2011 and 2012 is included as a baseline control. Vertical bars denote confidence interval.

Figure 3.1: Heterogeneity by Location



Notes: This figure presents coefficients for treatment interacted with year in OLS regressions. 2011 is omitted and coefficients are relative to the omitted year. Treatment refers to schools that were entirely shared by NSP IE treated villages. Similarly, control schools are those that were entirely shared by treated villages. Schools whose village received NSP after NSP IE (2013-2018) are dropped. Non-IE villages are present in both treatment and control groups. Vertical bars denote confidence interval.

Figure 3.2: Temporal Heterogeneity

3.10 Appendix

A Figures

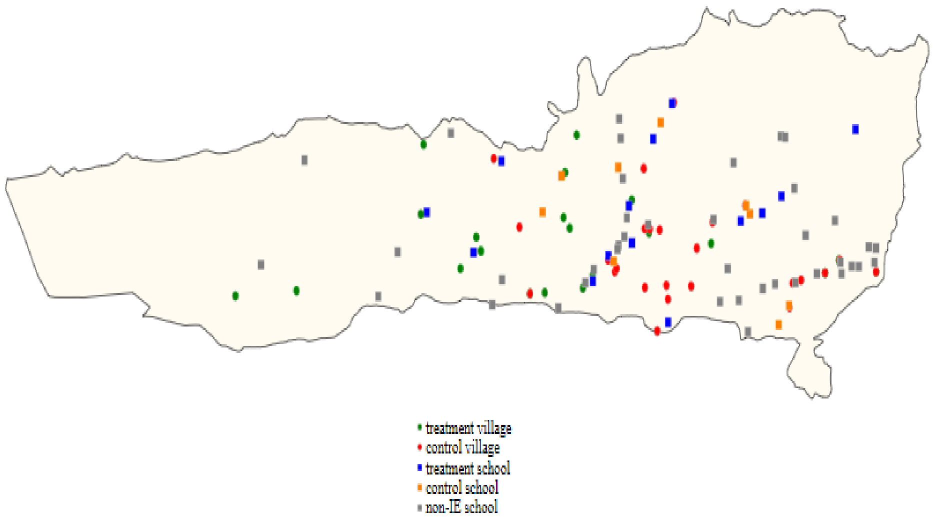


Figure A.1: Adraskan District

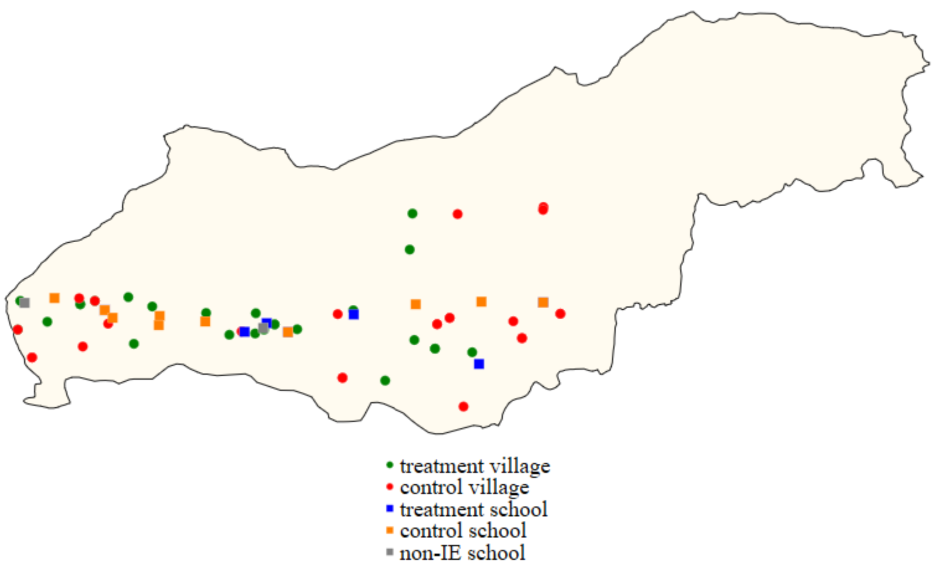


Figure A.2: Chisht-e Sharif District

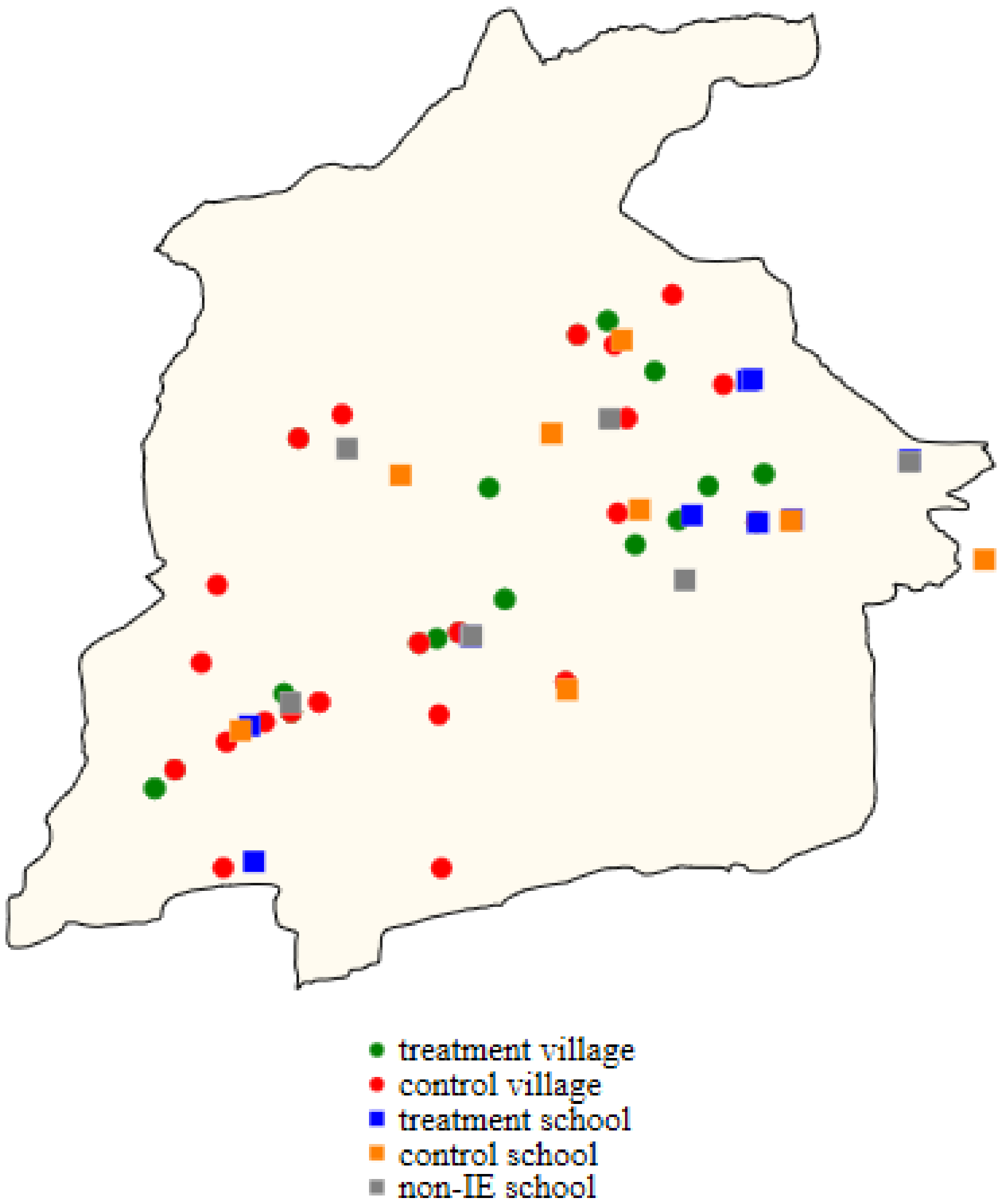


Figure A.3: Farsi District

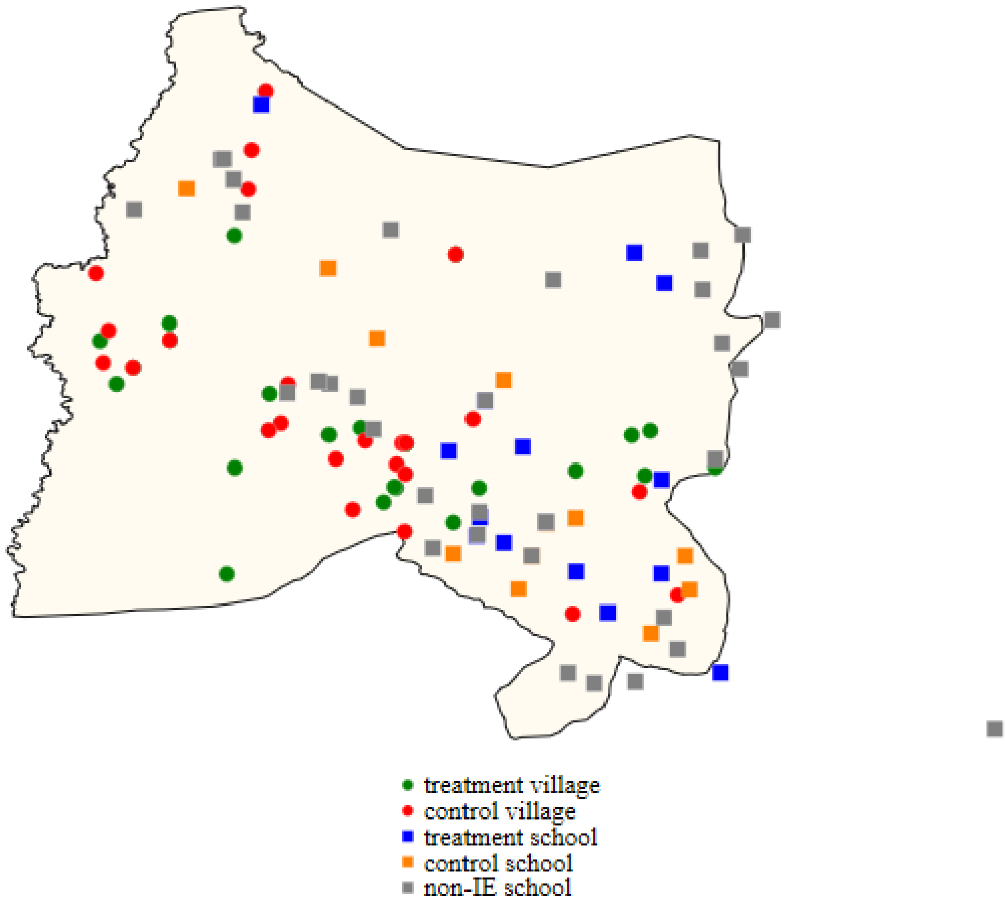


Figure A.4: Gulran District

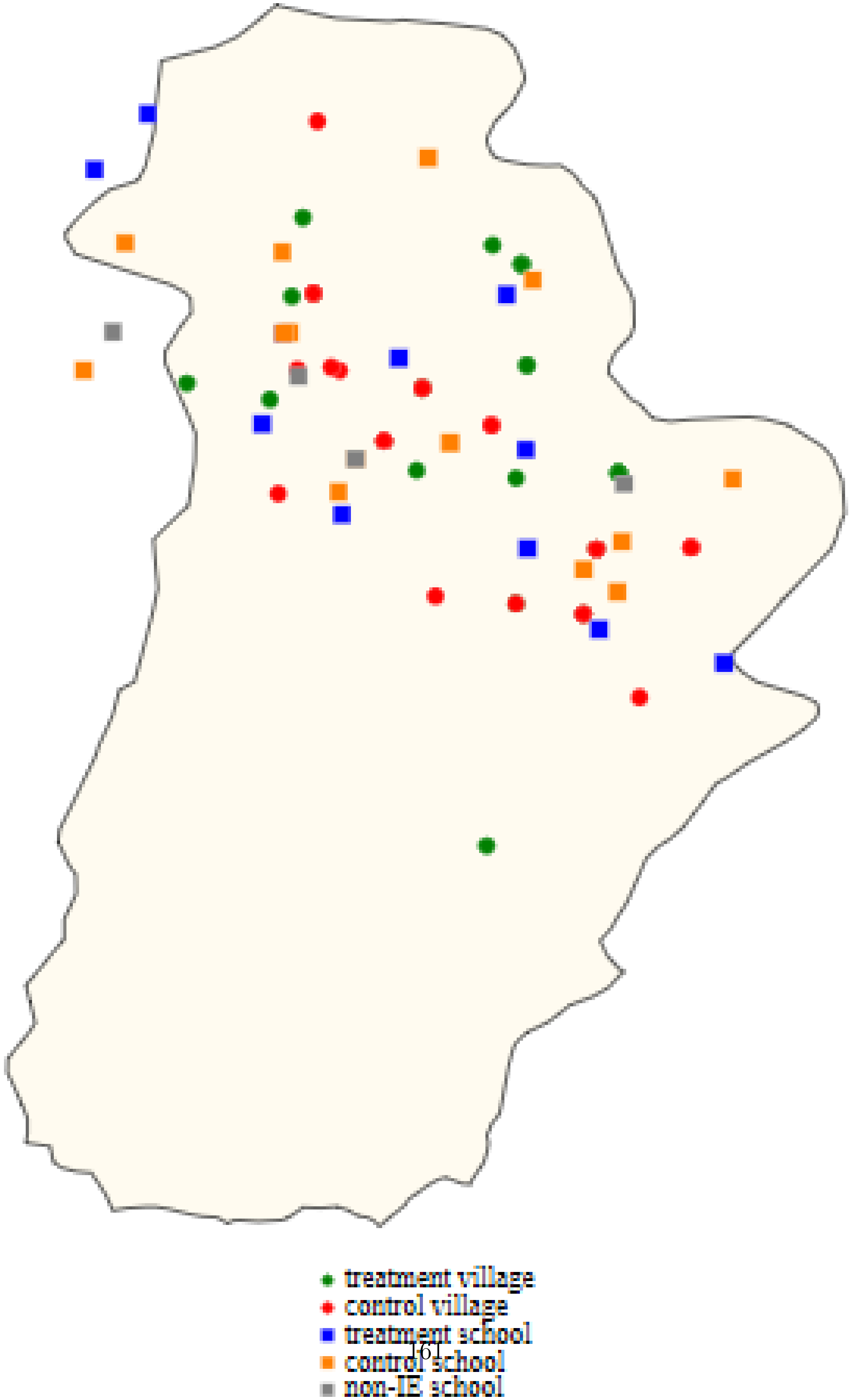


Figure A.5: Hesarak District

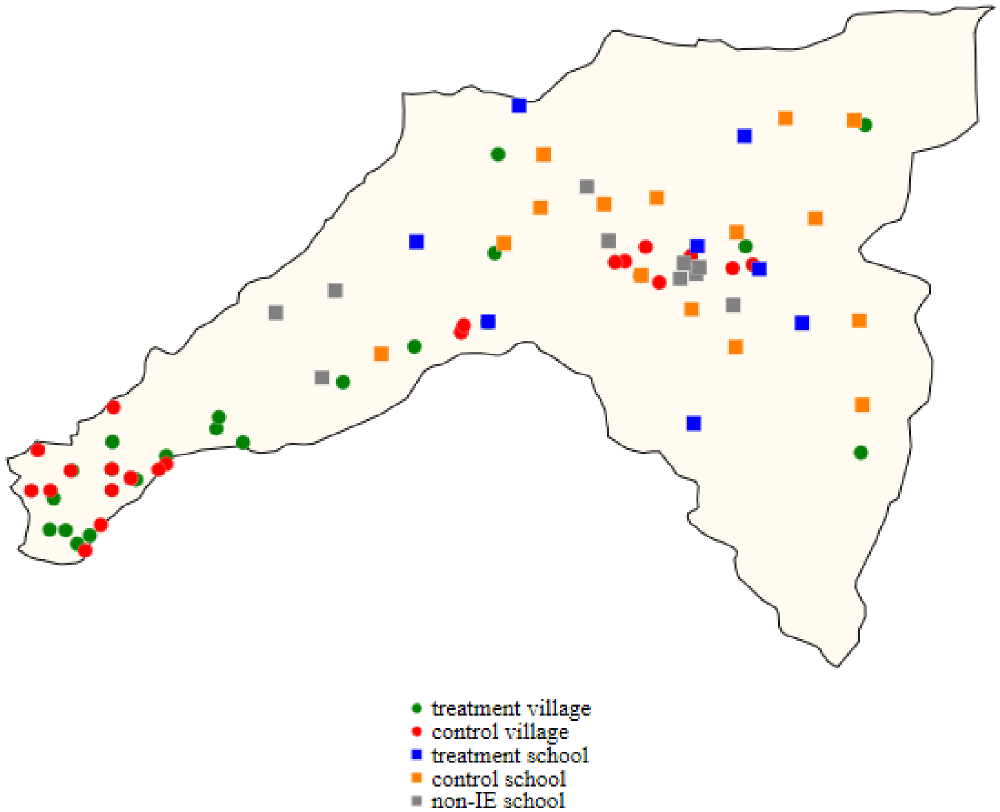


Figure A.6: Shirzad District

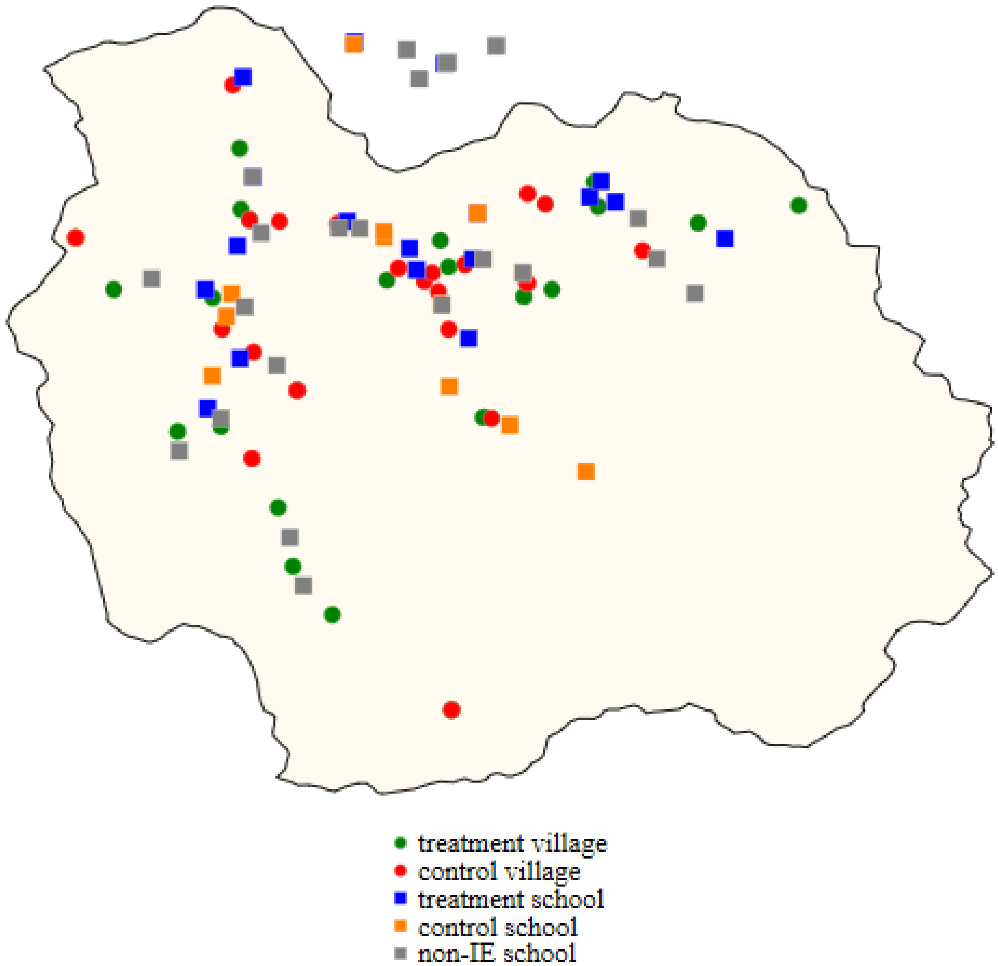


Figure A.7: Khost wa Fereng District

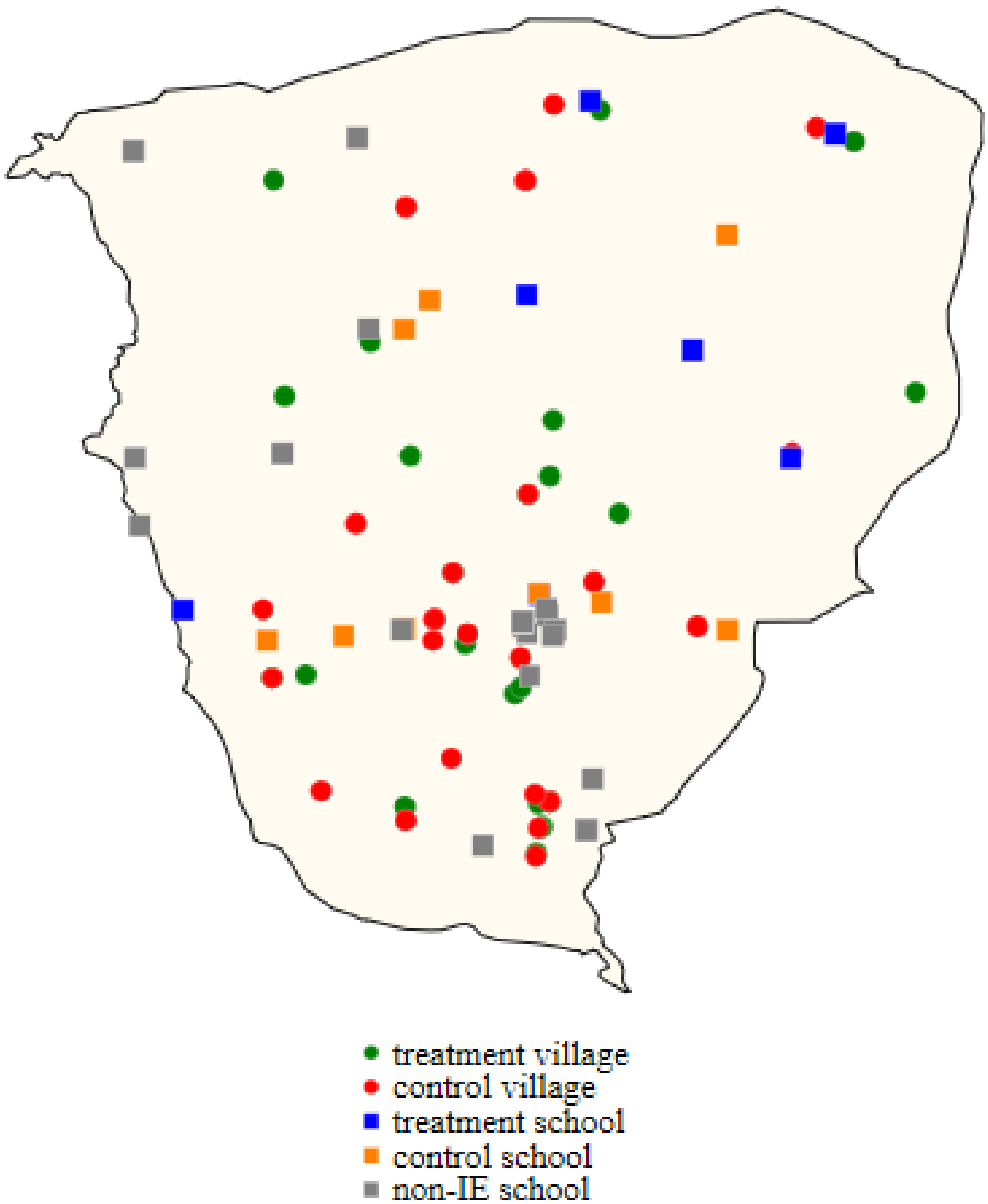


Figure A.8: Balkh District

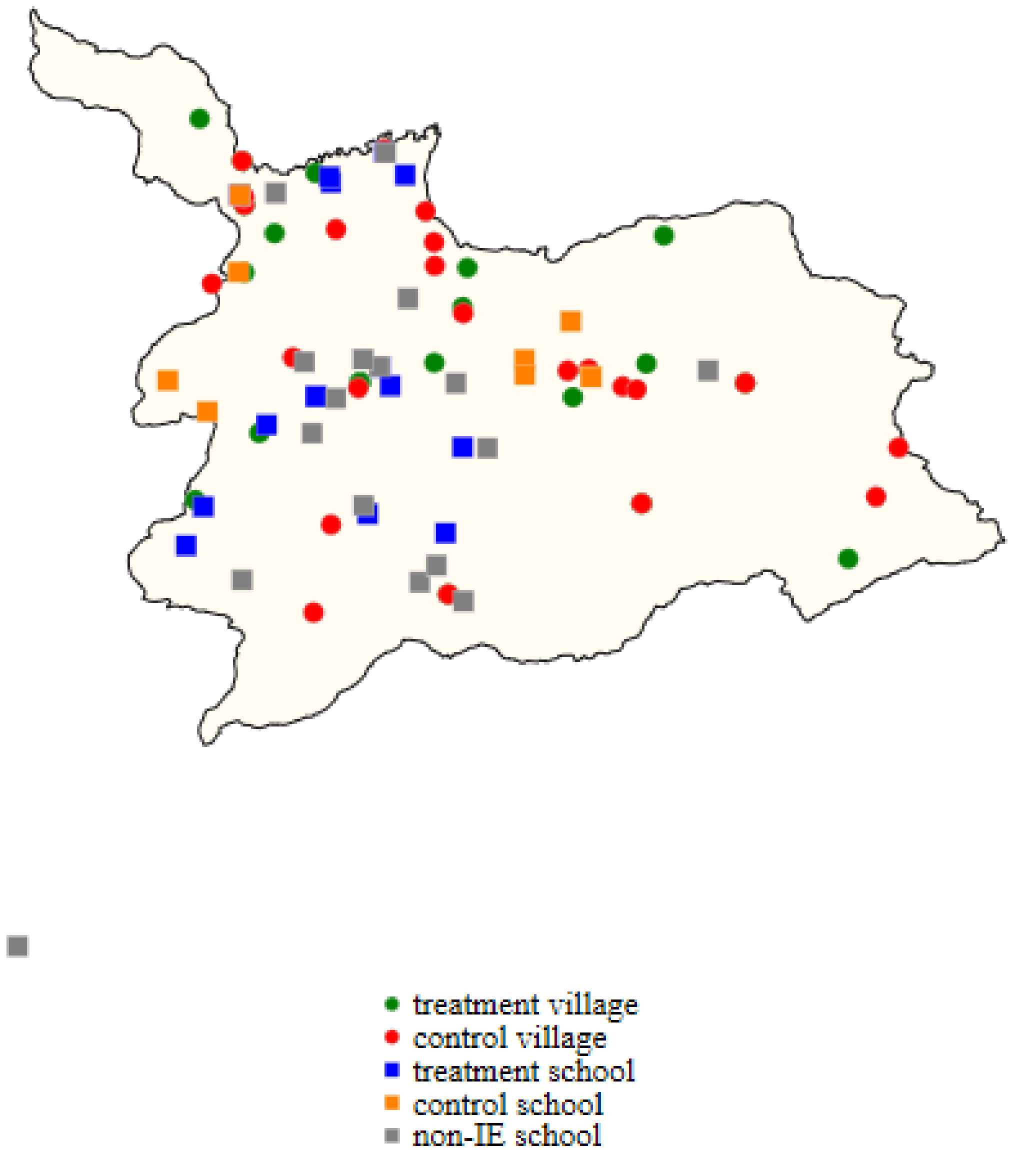


Figure A.9: Dulina District

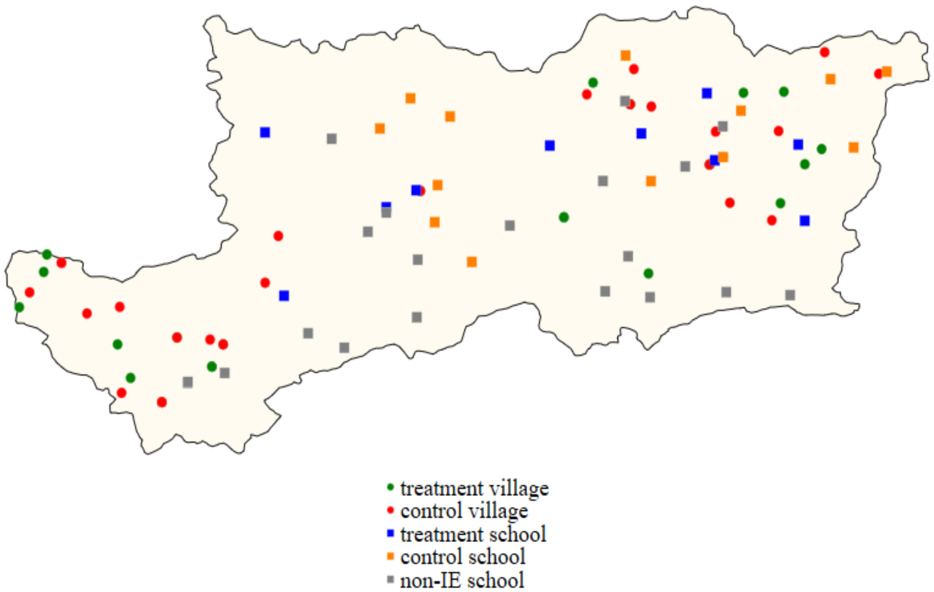


Figure A.10: Sang-e Takht District

Conclusion

This dissertation evaluated the effects of three development programs on female education in Afghanistan. The programs studied here includes an international security transition, a formal government program, and a local community-driven initiative. We focused on student enrollment and performance, particularly enrollment in higher education. Our objective was to understand the effectiveness of gender-driven programs and interventions in fragile state contexts.

The first chapter explored effects of NATO military withdrawal from Afghanistan's districts during 2011-2015, on women's demand for higher education. I employed a novel instrumental variable in this study. I used the least-cost travel distance between districts and the nearest logistic hub as an Instrumental Variable (IV) to predict the departure of foreign forces. Withdrawal of foreign forces resulted in a 0.3 percentage point increase in female university participation. I showed that declining economic opportunities and massive male emigration explain women's choice of education over migration. The findings in this study shed light on unintended consequences of a major foreign military evacuation.

The second chapter studied a community-driven education program that sends college students to schools in rural Afghanistan to prepare high school students for university entrance exam. We employed a heterogeneity-robust difference-in-differences estimator to evaluate this program. The program increased the university exam scores by 0.17 standard deviations and had a positive effect on enrollment in top universities. Moreover, the receivers of the program had higher passing rate. Importantly, female students in this program had a lower likelihood of enrollment in low-rank universities. The findings in this study highlights the success of local initiatives.

The third chapter examined long-term impacts of gender-balanced local development councils on female education. This study utilized random assignment of treatment from

Afghanistan's National Solidarity Program Intervention Evaluation (NSP IE) and current administrative data. In areas where people already have positive attitudes toward women's education and employment, the NSP increased female enrollment in higher school grades. However, in places where people hold conservative opinion about women's social interactions, the program produced negative impacts on female enrollment. Our findings informs policy about the sustainability of development interventions.

Understanding unbiased effects of development programs in conflict zones is important for future development programming. This dissertation applied rigorous empirical methods to evaluate three policy-relevant programs in Afghanistan that experienced decades of conflict. The development programs were implemented under the presence of foreign military forces in that country. Findings presented in this research point to important cultural considerations that must guide interventions in similar environments.