Massive Open Online Courses and Labor Market Outcomes: Experimental Evidence from Colombia*

Stephanie Majerowicz Román Andrés Zárate

November 28, 2023

Abstract

This paper studies the effect of free MOOC certificates on labor market outcomes. We leverage an RCT of a program offered by a large MOOC provider to public organizations during the pandemic, where around 13,000 beneficiaries among 21,000 applicants were randomly selected to receive free certificates for completing MOOC courses. Despite the free certificates, the take-up rate is low: 50% of treated beneficiaries enroll in at least one course, but only 6.2% out of the eligible participants complete them. We track participants in formal labor markets one year after the program. The treatment effects on formal employment are positive but insignificant, which results in imprecise 2SLS estimates of MOOCs' free certificates. To improve precision, we estimate an event study of free certificates, finding a significant average increase of 5.1% on formal employment. These effects are higher for low-income beneficiaries and do not vary by gender. Our results show that while MOOCs can potentially improve labor market outcomes, additional interventions are necessary to increase course completion.

*Majerowicz, Escuela de Gobierno, Universidad de los Andes, email: s.majerowicz@uniandes.edu.co. Zárate, Department of Economics, University of Toronto, email: ra.zarate@utoronto.ca. The registation record of the experiment in the AEA RCT Registry is AEARCTR-0010016 (Majerowicz and Zárate, 2022). This project was conducted in partnership with Icetex, the National Department of Planning (DNP), and the Ministry of Technology and Information (MINTIC) in Colombia. We are grateful to Alejandro Corrales from the DNP and to Lina Nuñez, Yineth Rentería, and Erika Ordoñez from Icetex . This project has benefitted from comments and suggestions from participants at the Toronto Junior Applied Micro Seminar (JAMS). Remaining errors are our own.

1 Introduction

As Massive Open Online Courses (MOOCs) entered the global educational landscape, they sparked a wave of enthusiasm for their promise to revolutionize higher education by democratizing access to courses from elite universities. For instance, the *New York Times* called 2012 the year of the MOOC (Pappano, 2012). Despite some concerns about their financial sustainability (Hoxby, 2014; McPherson and Bacow, 2015), between 2012 and 2015, MOOC enrollments exceeded a staggering 25 million (Kizilcec et al., 2017). Yet, more than one decade later, studies examining the impact of MOOCs on labor market outcomes remain scarce, with scant evidence of the long-term implications of MOOC participation (Escueta et al., 2017).

While recent studies examine the effects of virtual vs. in-person instruction (Bettinger et al., 2017; Bruhn et al., 2023), most of this evidence compares learning gains from in-person vs. online iterations of the same courses. In contrast, MOOCs offer individuals worldwide a unique opportunity to access online courses from prestigious institutions. The pertinent counterfactual, rather than an in-person version of the course, is the absence of access to such educational content.

Similar to other education interventions, there are two channels through which MOOCs can affect labor market outcomes: human capital gains and certifications that validate this knowledge. While potential human capital gains from MOOCs usually only require a time investment since auditing courses and access to most of their content are available for free, validating that knowledge typically requires participants to pay a non-trivial monetary fee. Most platforms, like EdX and Coursera, offer a "verified track", which, for a cost, provides full course access and issues a certificate upon successful course completion.[1]

This paper explores the effects of free MOOC certificates on formal labor market outcomes in Colombia by studying a program that granted participants eligibility to receive free certificates upon course completion during the pandemic. To the best of our knowledge, this is the first study linking MOOC participation and completion to the labor market. We leverage a program implemented by one of the most prominent and worldwide recognized MOOC providers that offered free certificates to public organizations in Latin America during the pandemic. In collaboration with one of these public entities, we undertook a Randomized Controlled Trial (RCT) of this initiative. Out of the 21,000 beneficiaries enrolled in the program, approximately 13,000 were randomly selected to become eligible for free certificates

[1] The fees for the verified track of EdX typically range between \$50 and \$300 USD. Meanwhile, Coursera's Professional Certificate programs are priced between \$39 and \$99 USD per month, with MasterTrack Certificate programs typically costing in the range of \$2,000 to \$5,000 USD.

upon completing MOOCs within three months.

Participants assigned to the free certificates eligibility treatment could earn free certificates for completing individual courses and specializations, a more structured degree composed of a series of related courses designed to master a specific topic. By contrast, participants assigned to the control group were not eligible for the free certificates. Due to data privacy regulations, we cannot track the activity of non-eligible participants on the platform. Given this data limitation, we focus on estimating the effects of receiving certificates for free on labor market outcomes. Yet, in a world where no participant in the control group received a certificate, a likely scenario due to the non-trivial fees of the "verified" track and the overall low MOOC completion rates, this treatment would be equivalent to the effect of receiving a MOOC certificate. Our estimates also combine the acquisition of new skills and the certificates that validate them, as we can only observe whether individuals entirely completed a course without any information about partial progress. Still, given the scarce evidence linking MOOCs to labor market outcomes, we believe estimates of treatment effects combining both channels are of general interest.

One of the primary challenges in examining the long-term impact of MOOCs on employment outcomes is the ability to track participants in the labor market over time. We overcome this challenge by combining data from the program's registration records with administrative data from Colombia's formal labor market, encompassing four years before and one year following the end of the program. This tracking allows us to study the impact of MOOCs' free certificates on formal labor employment and wages as we track monthly participation in the formal labor market for all registered participants over five years.

We first report the impacts of free certificate eligibility on course enrollment and completion. Consistent with existing evidence documenting the challenges in completing MOOCs (Banerjee and Duflo, 2014), our findings show that the take-up rate of the program is low. Despite being eligible to receive free certificates of multiple courses, including specialization certificates, only 50% of eligible beneficiaries enroll in at least one course of the program, and only around 6.2% of all treated participants completed at least one course. While there is a substantial variation in the number of completed courses, it appears that most beneficiaries did not participate actively on the platform.

The low first stage in the take-up rate of the program translates into positive but small and non-significant effects of free certificates eligibility on formal labor employment, with a clear pattern of positive impacts from six months after the end of the program onward. Being eligible to receive the free certificates increases formal employment by 0.6 percentage points (p.p.) (standard error (s.e.) 0.5 p.p.) seven months after and by 0.3 p.p. (s.e. 0.6 p.p.) one year after the program ended.

We estimate local average treatment effects (LATE) of free certificates for course completion on formal labor employment, using the eligibility random assignment as an instrument for completing the courses and receiving the free certificates in a two-stage least squares (2SLS) framework. The LATE estimates reveal the impact of free certificates for course completion on compliers (Imbens and Angrist, 1994): those who obtain the certificates for free due to the randomly assigned eligibility. In a model with one-sided compliance, as is this case, since participants in the control group could not receive the certificates for free upon course completion, this LATE will also be equivalent to the average treatment effect on the treated (ATT). As mentioned above, this will also be the ATT of earning any certificate if no one in the control group completed at least one course and paid the respective fee.

The results reveal that while the estimates of the 2SLS framework are large in magnitude, between 2.2 to 9.5 p.p., they are imprecise and not statistically different from zero. Estimates of the labor market returns of free certificates also do not show a clear pattern or statistically significant effects on daily wages.

Motivated by the large but imprecise estimates of free certificates using the 2SLS framework, we leverage the time variation before and after the program to estimate the ATT on formal employment using an event study to boost precision. The results show encouraging evidence of the impact of MOOC completion on labor market outcomes. While the estimates show that those who completed the courses were less likely to be formally employed during the program, consistent with individuals having more time to invest in finalizing the courses, from six months after the end of the program, there is a clear positive impact of receiving free certificates on formal employment. The results show statistically significant increases in the formal employment rate, with an average effect of 3.3 p.p. (p-value < 0.01) for all the post period and higher impacts (close to 5 p.p.) between 8 to 12 months after the end of the program. Notably, the estimates of the certificate effects from the event study closely align with the 2SLS framework, offering estimates that are more precise but similar in magnitude.

The event study also helps to compare the ATT of receiving free MOOC certificates vs. any certificate on labor market outcomes. In particular, as we can only track the activity on the platform for participants eligible for the free certificates, we can estimate the effect of any certificate on labor market outcomes by restricting the sample to these individuals. The cost is a lower statistical power due to the reduction in sample size. The estimates are remarkably similar to the ones including the non-eligible participants in the control group. The DiD estimate shows an average effect of certificates in the post period on formal employment of 3.5 p.p. (s.e. 0.013), which mirrors the general estimate of 3.3 p.p. (s.e. 0.013) when including non-eligible participants as part of the control group. The similarity between the two estimates supports that our results can be interpreted as the effects of any certificate on labor market outcomes, as it suggests that any participant who received a certificate during this period was in the free-certificates eligibility treatment arm.

We also explore heterogeneity in the certificates' effects on employment by income level and gender. Lower-income participants, measured by a proxy means test used to target social programs in Colombia, benefit the most from the free certificates. While the average effect in the post-period for high-income participants is only -0.7 p.p. (s.e. 1.9 p.p.), low-income participants experience a gain of 5.5 p.p. (p-value < 0.05) when they receive the free certificates. This difference between the two groups is statistically significant. By contrast, the estimates for men and women are similar in magnitude, with no statistically significant differences by the participants' gender.

Collectively, our findings emphasize the potential of MOOCs to improve labor market outcomes, especially for low-income individuals. Yet, the persistently low course completion rates, even when certificates are provided free of charge, represent a significant hurdle. Consequently, it is crucial to evaluate the effects of complementary interventions to encourage successful completion of these courses.

Our study contributes to three branches of the literature. First, it adds to the existing literature focused on assessing the effects of online education on student outcomes. While prior research has provided extensive insights into the impact of online education on learning, most of these studies have compared virtual and in-person versions of the same courses. For instance, Bettinger et al. (2017) finds a reduction in academic performance for online learners in both their current and future courses compared to in-person instruction. Recent studies of remote learning (Bruhn et al., 2023) have echoed these earlier findings. Additionally, evidence from virtual learning during the pandemic on school districts (Jack et al., 2023) and individual students (Kofoed et al., 2021) has corroborated these conclusions, with evidence also showing the effects of remote learning in widening achievement gaps (Goldhaber et al., 2023). Our study diverges from this literature, as the relevant counterfactual for MOOCs is not an in-person version of the same course. While there is recent evidence showing that online programs can increase enrollment (Goodman et al., 2019), our results suggests that online education can improve labor market outcomes by enabling individuals to access courses that would otherwise be beyond their reach.

Second, the project contributes to the literature on interventions designed to increase MOOCs completion. Most of this literature has focused on behavioral interventions with mixed evidence. For example, Patterson (2018) finds that nudges can increase students' effort and performance in MOOCs, while Oreopoulos et al. (2022) find no impacts on academic outcomes. In our case, and despite the positive effects of course completion on employment, the low completion rates suggest that even providing participants with free certificates is not incentive enough to guarantee enrollment and course finalization.

Finally, we contribute to the literature on certifications of abilities in the labor market. In particular, the primary treatment allowed participants to obtain free certificates for the completion of MOOC courses. Our data doesn't allow us to separately identify the effect of human capital from the signals of ability in the certificates. However, given that most MOOCs content is free and the main cost for participants is certificates, our positive effects on formal employment are consistent with evidence showing positive impacts of certifications (Clark and Martorell, 2014) on labor market outcomes. For the Colombian case, previous evidence (MacLeod et al., 2017) finds that firms use college reputation to signal ability in labor markets. Consistent with this evidence and in contrast to the US, where employers have a negative perception of online degrees from for-profit institutions in the US (Deming et al., 2015), our results suggest that MOOC certificates from prestigious institutions can be highly valued in the Colombian labor market.

The rest of the paper is organized as follows. Section 2 presents the setting and describes the program. Section 3 introduces the experimental design and the main data sources. Section 4 outlines our empirical methods. Section 5 presents the results, and Section 6 concludes.

2 Setting and Intervention Description

One of the leading worldwide platforms in MOOCs launched a job recovery initiative to mitigate the impact of COVID-19 on employment in Latin America, one of the most affected regions by the pandemic on both health and employment. For example, in Colombia, the setting of this study, the unemployment rate nearly doubled between July 2019 and July 2020, increasing from 10.7% to 20.2% (Dane, 2020), with higher impacts on the unemployment of vulnerable populations, including women and youth. In this context, the large MOOC provider designed a program to promote new skill acquisition and avoid the deterioration of existing skills in the labor force.

The initiative allowed public agencies across the region to apply for a specific number of slots, allowing users to receive free certificates for completed courses on the MOOC provider's platform. Each public agency had the autonomy to select its target audience and set the criteria for participation. Typically, while most MOOC content, such as video lectures, is freely accessible, obtaining full course access and certificates—which serve as proof of completion—can be expensive. Prices range from \$40 to \$300 USD per course. In Colombia, for instance, $100 USD approximates half the monthly minimum wage, posing a substantial financial obstacle for many. Participation in this initiative allowed government agencies not only to inform people about the MOOC platform and its courses but also to offer them a chance to earn certificates, validating their skills for future employment opportunities. The free certificates provided an additional incentive for enrolling in and completing the courses.

Numerous public agencies from different Latin American nations engaged in the initiative, each with its own set of eligibility and selection standards. In Colombia, for example, the Ministry of Information Technologies and Communications (Mintic) secured 50,000 slots. These were made available to any unemployed individual who registered for the program. Given the high number of slots, the program experienced an undersubscription, resulting in the acceptance of all applicants. In Costa Rica, the initiative was a collaborative venture involving the MOOC provider, the Ministries of Labor and Foreign Trade, and CINDE (the Costa Rican Investment Promotion Agency). Mirroring Colombia's Mintic initiative, Costa Rica's implementation aimed to reach 50,000 unemployed individuals, offering them access to the platform and free certificates over a period of six months.

We partnered with one public agency, the Colombian Institute for Educational Credit and Technical Studies (Icetex), which offers higher-education loans and scholarships. The Icetex only applied for 10,000 slots and targeted the program towards any of their current or former beneficiaries who had received a loan or a scholarship since 2010. The Icetex promoted the program on its website and through social media. As Icetex received over 23,000 applications for its 10,000 slots, they decided to allocate them through random assignment among the applicants, following advice from the research team, who provided technical assistance in the program's evaluation.

Applicants who were part of the program could enroll in as many courses as they wanted from a catalog of over 3,800 courses determined by the MOOC provider. As long as they completed all the course requirements by the program deadline, participants could obtain the certificate of completion for free. Eligible participants could also enroll in "specializations," which are higher-level certificates, usually composed of three to four related courses designed to master a specific skill. The program started in October 2020, and the deadline to obtain a certificate was December 31st, 2020, giving eligible participants approximately three months to complete the courses.

The MOOC provider encouraged all partner institutions to develop campaigns to promote course enrollment and completion. Icetex sent numerous emails to participants encouraging them first to enroll, and then to complete the courses they had enrolled in. As part of this program, Icetex also promoted the most-demanded courses in the region to all eligible participants through emails highlighting specific courses. The MOOC provider tracked and sent reports with enrolment statistics to all the participating institutions.

3 Experimental Design

As the program was oversubscribed, Icetex allocated the slots by random assignment. The experimental design was a simple randomization at the individual level, without any stratification. The research team suggested to the Icetex to re-randomize to avoid chance imbalances, following Banerjee et al. (2017). The randomization was run 100 times, and balance checks were performed on variables including demographic characteristics, eligibility for safety net programs, debt in student loans, and education variables for a total of 36 variables. The code used a max-min p-value criteria, keeping the randomization with the largest minimum p-value among the 36 variables used for balance checks.

The registration form and the randomization took place in September 2020, with the program starting during the first week of October. Among 21,000 participants in the registry, 10,000 participants received an offer to join the program. Two weeks after the first randomization, given that the program's take-up rate was low among the treated participants, Icetex decided to perform a second randomization: 3,000 out of the 11,000 participants first allocated to the control group received a second-round offer to join the program. In our main results, we combine both rounds of offers in a single treatment variable, but our results are similar when considering both offers separately.

3.1 Data

To characterize the participants, perform the balance checks, measure MOOCs enrollment and completion, and track participants' into the labor market, we combine different administrative data sets. The project comprises a total of five main data sets.

First, we use the registry form for the program. When participants registered for the program, they had to fill out a form where they accepted the terms and conditions. In the form, they also answered some questions, including their employment status, main course interests, and objectives with the program. They also reported some demographic characteristics, such as age, gender, and education level.

The second data set, the administrative information of Icetex loans and scholarships, is mainly used for balance checks of the randomization and treatment effect heterogeneity. As one of the eligibility conditions was to be a current or former Icetex beneficiary, we have information about the type of loan and scholarship the participants had, including the loan size, interest rates, payment schemes, and whether payments were overdue.

This data also contains information about the participants' sociodemographic characteristics, such as age, gender, and socioeconomic level, which we used for treatment effect heterogeneity. To characterize individuals' socioeconomic status, we use the SISBEN level when they applied for the loan or scholarship. The SISBEN is a unified vulnerability assessment and identification system for social assistance used by the Colombian Government (Camacho and Conover, 2011). The SISBEN is a proxy-means census that classifies the population into different brackets to determine eligibility for social programs, with lower levels representing a higher vulnerability and need for social assistance. We use the latest version of the SISBEN level to classify participants' income level, with those located in the first bracket, who are eligible for most social programs in Colombia, defined as low-income.

The third data set, the *Saber 11* scores, is also mainly used for balance checks. The data contains information on nationwide comparable performance in math and reading and additional rich socioeconomic characteristics. This data set is only available for participants who graduated from high school between 2010 and 2020 (around 60% of the sample). We cannot observe scores for participants who graduated high school before 2010. However, the proportion of students we find in the *Saber 11* data is balanced across the treatment and control groups.

As a fourth data source, the large MOOC provider shared course enrollment and completion records of participants in the treatment group with Icetex. We have access to the list of courses treated participants enrolled in and completed for which they received a free certificate. Unfortunately, due to the terms and conditions of the program, this data is unavailable for participants in the control group. For this reason, we focus our estimates on the impact of free certificates on labor market outcomes. However, given the low take-up rates among the treated participants and the non-trivial cost of the certificates, it is unlikely that participants in the control group have received course completion certificates during this period. If no one in the control group received a certificate, a likely scenario for the above reasons, our estimates would be equivalent to the effect of course completion and their respective certificates on labor market outcomes.

The main challenge in studying the effects of MOOCs is having reliable labor market data that allows one to link MOOC enrollment and completion with employment and wages. The final data set enables us to perform this tracking of participants into formal labor markets in Colombia. The primary outcomes come from the *Planilla Integrada de Liquidación de Aportes* (PILA), an administrative database administered by the Ministry of Health that records all workers' social security contributions, reporting the universe of all formally employed Colombians. The main advantage of this data set is that it allows us to track all formally employed workers every month between January 2017 and December 2021. This tracking enables us to observe any changes in the participant's employment status, sector, and wages and observe their formal employment history before the intervention. As this data set only reports formal employment, we cannot discriminate whether participants who are not formally employed are unemployed or in the informal sector.

3.2 Balance

The final experimental sample comprises 21,675 participants, 8,687 non-eligible for the free certificates, and 12,988 randomly assigned to the free certificates eligibility treatment throughout the two randomizations. All sample participants had an active or previous loan or scholarship with the ICETEX. Table 1 reports sample averages for the control and treatment groups of the free certificates eligibility and tests for balance across different demographic characteristics and employment status at baseline.

Columns 1 and 3 of Table 1 present sample averages for the control and the treatment groups, respectively. Participants are, on average, 29 years old, and 62% are female. Regarding education level, around 15% report only completing high school, and 72% have received a bachelor's degree, with the remaining 13% having a technical degree. As for employment, 57% of the participants reported being unemployed in the registration form in September 2020. The administrative data shows that contrary to the self-reported information, around 48% of them were formally employed during this month. When looking at the subset of participants (roughly 60%) with available high school exit exams (balanced across treatment and control groups), we see that around 58% of them graduated from a public school, and roughly two-thirds are first-generation post-secondary education students (with their parents having completed at most secondary education).

We check the experimental validity by showing that the variables at baseline are balanced between the eligible and non-eligible groups. Column 5 of Table 1 reports the difference between the two groups, and column 6 the standard error. While the randomization was re-run 100 times to reduce chance imbalances (following Banerjee et al. (2017)), we provide additional checks by adding baseline characteristics from the PILA data set, which was not included in the original randomization balance checks. Overall, we find that both groups are balanced at baseline on a large number of characteristics. There is only a small imbalance in one out of 24 characteristics, and the conventional p-value on the joint F-test on all of these variables is 0.83.

4 Empirical Strategy

In this section, we present the empirical strategy we follow to estimate the impact of free certificates on labor market outcomes. First, we estimate a straightforward reduced form specification of the effect of being eligible for free certificates on participants' outcomes.

$$y_{it} = \alpha + \beta z_i + \delta' X_{i0} + \varepsilon_{it}, \tag{1}$$

where y_{it} is the outcome of interest for individual i in period t (either formal employment or daily wages), and z_i is a dummy variable indicating whether individual i was assigned to the eligibility treatment or the control group. To increase the precision of the estimates, we control for a set of baseline characteristics X_{i0} selected using the double-post-lasso covariate selection method proposed by (Belloni et al., 2013). This set of covariates includes the value of the dependent variable from January 2017 to September 2020, sociodemographic characteristics such as age and gender, whether students have available standardized tests, and their math and reading scores if these are available. Lastly, ε_{it} is an error term. The parameter of interest in equation 1 is β, the treatment effect of free-certificate eligibility on participants' outcomes.

Equation 1 is of interest by itself but is also the first stage of a model where we estimate the effect of receiving free certificates on participants' outcomes, using the eligibility treatment assignment as an instrument for course completion. The following system of equations describes such a model:

$$y_{it} = \alpha + \gamma c_i + \delta_2' X_{i0} + \nu_{it} \tag{2a}$$

$$c_i = \alpha + \beta z_i + \delta_1' X_{i0} + \epsilon_{it}. \tag{2b}$$

In equations 2a and 2b, c_i is a dummy variable indicating whether participant i has completed at least one course and received a free certificate as part of the program. Equation 2a is the second stage of the model, with parameter of interest γ, the ATT of free certificates on labor market outcomes. The 2SLS model usually estimates the LATE on the compliers, but in this case, it is also the ATT, as participants in the control group cannot receive free certificates, and therefore, there is one-sided non-compliance. Equation 2b is the first stage equation, with parameter β capturing the impact of the eligibility treatment on free certificates. The terms ϵ_{it} and ν_{it} are the error components of the first and second-stage equations, respectively. The other variables are as in equation 1.

As we cannot track the activity on the platform for participants in the control group of the eligibility, we cannot interpret the parameter γ as the ATT of MOOC certificates on labor market outcomes. However, in case no one in the control group completed a course during this period and paid for the verified track to receive a certificate, a likely scenario due to the low participation rates and the non-trivial fees, the parameter γ in equation 2 would be equivalent to the ATT of MOOC certificates on labor market outcomes. In this case, all the participants with a MOOC certificate would have received it for free.

While we can estimate equations 1 to 2b for each period, we can also leverage the time variation nature of our data to improve the precision of our estimates of the ATT of free certificates. In particular, by observing the formal labor market outcomes of the participants monthly, the high frequency of our data allows us to estimate a difference-in-difference (DiD) model and an event study to increase the precision of the 2SLS while addressing selection concerns on course completion. Conditional on satisfying the parallel trend assumptions, this identification strategy allows us to overcome the limitations of the low enrollment and completion rates of the RCT by comparing the formal labor market trajectories of those who received free certificates against those who did not.

The following model specifies the DiD estimating equation:

$$y_{it} = \alpha + \theta post_t + \varphi(c_i \times post_t) + \psi_i + \varepsilon_{it}. \tag{3}$$

where $post_t$ is a dummy variable equal to one for the post period (from January 2021 onwards), and ψ_i are individual fixed effects. The parameter of interest is φ, the average difference between those who completed at least one course versus those who did not or were not eligible for the free certificates after the end of the program, after accounting for the difference between these two groups in the pre-period.

We also extend equation 3, and estimate an event study specification, where we interact the relevant treatment variable with time-period dummies. For this purpose, we estimate the following specification:

$$y_{it} = \alpha + \sum_{t \neq 0}^{T} \varphi_t(c_i \times post_t) + \tau_t + \psi_i + \varepsilon_{it}. \tag{4}$$

Here, τ_t represents dummy variables for each month, and the other terms are as in equation 3. The parameters of interest are the vector φ_t, which captures the difference between those who completed courses and those who did not or were not eligible for the free certificates in period t compared to a reference period $t = 0$, which is excluded from the estimation. We also estimate equation 4 only among participants eligible for the free certificates, as discussed above.

As there could be a differential trend between those who completed courses and those who did not in the months before the program due to the pandemic, we take March 2020 as our baseline period. As usual with event studies, and leveraging that we observe monthly labor market outcomes before and after the program, we would expect $\varphi_t = 0$, for $t < 0$, if the parallel trend assumption holds during that period. On the other hand, the estimates of φ_t for periods after September and December 2020 show the additional differences between the two groups during and after the program, respectively. When estimating equation 4, we also control for the interaction between the period dummies with income level and gender. The results are similar without these interactions.

The DiD and event study strategies also allow us to consider two different control groups to compare the ATT of receiving free MOOC certificates vs. any certificates on labor market outcomes. We can consider a general control group that includes the non-eligible participants for the free certificates and the eligible ones who did not complete courses. The estimates from this comparison group would mirror the ATT of the 2SLS strategy: the effect of receiving free certificates on labor market outcomes. We can also consider a second control group to estimate the ATT of receiving any certificate on labor market outcomes. In particular, we can estimate this ATT by restricting the sample to eligible participants, as we can track all their activity on the platform. While this strategy has less statistical power due to the smaller sample size, by comparing those who completed courses vs. those who did not among the eligible participants, we can leverage the time variation of the data to estimate the ATT of any certificate. Comparing the estimates of these two control groups can shed light on whether our estimates can be interpreted as the effect of receiving any MOOC certificate.

Finally, a concern when estimating equations 1 to 4 on daily wages is that we only observe wages for those that are formally employed, generating non-random sample selection in the estimation (Heckman, 1974). Such an issue is particularly problematic when the relevant treatment affects the likelihood of observing the daily wage of the participant. This will occur when the treatment impacts the likelihood of being formally employed. For these reasons, we refrain from estimating effects on wages when we find significant effects on formal employment. While we could estimate a Heckman selection model, such strategy would require an additional instrument for formal employment.

5 Results

5.1 2SLS Model

Our results start by reporting the treatment effects of being eligible to receive free certificates on course enrollment and completion. Table 2 reports the estimates of parameter β in equation 1 on the likelihood of enrolling in at least one course, the number of enrollments, the likelihood of completing at least one course and the number of courses completed. Notice that these estimates are equivalent to the first stage (equation 2b) in the 2SLS model described by the system of equations 2.

A first finding consistent with the literature on MOOCs is that take-up and especially completion rates are generally low. While we can only observe enrollment and completion for the eligible participants, the estimates show that only 54.4% of those offered free certificates enrolled in at least one course, and very few, only 6.2%, completed at least one course during the span of the program. Participants eligible for free certificates enrolled on an average of 3 courses and completed only 0.13 courses. While such estimates are low, this is an upper bound of the free certificates eligibility effect on MOOC participation. If control students enrolled or completed a course due to the information about the program, then the first-stage estimates reported in Table 2 would be lower.

Panel B of Table 2 reports the estimates of equation 1, but splits the eligibility treatment variable between those applicants who receive a first-round offer to join the program and those who receive a second-round offer. We find that those assigned to the first round have slightly higher enrollment (column 1) and completion rates (column 3), which translates into a higher number of enrolled (column 2) and completed courses (column 4). In fact, the p-value at the bottom of the table shows that the small differences in take-up are statistically significant. We pool both offers into a single variable for our main results, but our results are similar when we separate the two rounds of the eligibility treatment assignment.

Next, we explore the estimates of equation 1 on labor market outcomes. We estimate this equation for each period after the intervention. Figure 1 reports the reduced form eligibility effects of equation 1 and the 2SLS estimates of the ATT of free certificates, the parameter γ in equation 2a, on formal labor employment. Each marker represents a different regression as we separately estimate the impact on formal employment at each period after the program ended. Panel A presents the results of the reduced form treatment effects of eligibility. Panel B reports the 2SLS estimates on the impacts of course completion and the subsequent earning of free certificates. Appendix Table A.1 reports the same estimates in a table format.

While the results in Panel A show slightly negative or positive non-significant eligibility effects for the first five months after the program ended (up to May 2021), the eligibility effects on formal employment are clear positives from July 2021 onward. Furthermore, the 2SLS estimates reported in Panel B show large but non-significant effects of free certificates from June until December 2021. According to these results, receiving a free certificate has an average impact during these six months between 4.3 to 11.5 percentage points. Despite the non-significance, these numbers reveal a large potential effect of certificates on formal employment. These impacts represent an increase between 8.9% to 23.8% of the baseline formal employment rate for the non-eligible control group before the intervention, which is 48.3%. Appendix Figure A.1 reports the same estimates using both rounds of offers as instruments for completion. Despite some estimates being lower in magnitude, the main conclusion remains, with large positive but non-significant effects on formal employment six months after the program ended.

As there are no significant effects on formal employment, reducing any concern about differential attrition rates in the likelihood of observing wages, Figure 2 reports the eligibility treatment effects and the 2SLS estimates of free certificates on the log of daily wages. Similar to Figure 1, each marker represents a different regression with the dependent variable varying across the twelve months after the end of the program. The results in Panel A show small and precise null effects on wages. On average, being eligible to receive free certificates affects wages between -1.2% to 0.7%. None of the effects are statistically significant, but the low standard errors show that the estimates are precise. Panel B reports the 2SLS estimates of receiving free certificates on wages. Appendix Table A.2 reports the same estimates in a table format. Some of the estimates are negative and large in magnitude, such as the estimate in August 2021, but none are statistically significant. We refrain from deriving general conclusions from these estimates as there is not a consistent pattern over time.

5.2 Event Study

Motivated by the positive but imprecise estimates of the 2SLS model, we exploit the time variation in the data by estimating an event study to identify the ATT of free certificates on formal employment. Panel A of Figure 3 reports the estimates of equation 4 of free certificates on formal employment for all participants. Column 1 of Appendix Table A.3 reports the analogous estimates of the DiD model (equation 3); Panel IA pools together all the pre-periods, while Panel IIA separates the pre-period into three: (i) January 2017 to February 2020 (the excluded period) (ii) March to September 2020, the six months of the pandemic before the program, and (iii) October to December 2020, the span of the program when participants could complete MOOCs.

The results in Figure 3 provide promising evidence of the potential impact of MOOCs free certificates on formal employment. First, the pre-trends assumption holds, as there are no statistically significant differences between the two groups in the pre-period spanning from January 2017 to February 2020 (the joint significance test p-value is 0.155). Although both groups have a similar employment trajectory during the initial months of the pandemic, individuals who successfully completed these courses are less likely to be formally employed the month before and during the three months of the program. The results in column 1 of Panel IIA in Appendix Table A.3 indicate a negative pooled estimate spanning October to December 2020 of -2.5 percentage points (p-value < 0.10). This evidence suggests that individuals out of the formal sector during the months of the program may have had more time to complete courses.

The estimates in Figure 3 (Panel A) indicate that after the program's conclusion, starting in January 2021 and onward, individuals who successfully completed the courses have an upward trajectory in their formal employment. It is noteworthy that the pattern illustrated in the event study figure closely mirrors the 2SLS estimates of completion detailed in Figure 1. In fact, the event study estimates for each period are in the ballpark of the 2SLS estimates for each period. The results reveal an average post-program effect of 3.3 percentage points (p-value < 0.05). This average estimate masks the notable rise in employment evident in Panel A of Figure 3 from August 2021 to December 2021, with a statistically significant impact of approximately 5 p.p. The formal employment rate for participants in the eligibility control group in December 2021 is 63.7%, with these effects representing an increase ranging from 5.1% to 7.8% in formal employment.

The event study also allows comparing the ATT of free certificates vs. any certificate on formal employment. In particular, we can estimate the ATT of any certificate by restricting the sample to eligible participants, as we can track their course completion on the platform. Appendix Figure A.2 reports the estimates restricting the sample to eligible participants, comparing those who completed vs. those who did not complete courses. While these estimates are less powered, the results are similar to the general estimates. Panel IB of Appendix table A.3 confirms this, as the average post effect is 3.5 p.p. (s.e. 0.013), which is reasonably similar to the general estimate of 3.3 p.p. Likewise, comparing the estimates in Panel IIA vs. IIB shows remarkably similar estimates. Overall, this comparison suggests that our estimates of the free certificates effects can be interpreted as any certificates effects.

Next, we explore heterogeneous treatment effects by income level and gender. Panels B and C within Figure 3 present the event study estimates for low- and high-income participants, respectively. Our findings indicate that low-income individuals benefit more from free certificates than their higher-income counterparts. Specifically, while the estimates for high-income participants hover around zero with high precision, low-income participants experience effects roughly twice the average effect outlined in Panel A. One year after the program's conclusion, low-income participants who earned the free certificates saw a statistically significant gain of approximately 10 p.p. in formal employment. Notably, the negative trajectory in formal employment observed during the program period, spanning from October to December 2020, primarily stems from high-income participants, even though they do not ultimately reap the benefits of the certificates. Columns 2 and 3 in Panel I of Appendix Table A.3 confirm this result, with an average post-treatment free certificates effect of 5.5 p.p. (p-value < 0.05) for low-income participants and a -0.7 p.p. effect for high-income participants; this difference between the two groups is statistically significant (p-value $= 0.034$).

We also explore heterogeneity by gender, with no statistically significant differences in the effect of free certificates on formal employment between men and women. In Panels D and E of Figure 3, we present event study estimates for men and women, respectively. Notably, the effects appear pretty consistent for both genders, with a similar impact in the final months of the post-period, hovering around 5 p.p. for both groups. Estimates in columns 4 and 5 of Appendix Table A.3 reaffirm this parity. The effects are consistently positive and similar in magnitude for men and women, with no statistically significant difference between the two groups. The reduced sample size in these gender-specific analyses results in less statistical power than the overall effects in column 1. However, the overall picture remains consistent, with results of similar magnitude regardless of the participant's gender.

6 Conclusion

This paper explores the impact of free MOOC certificates on labor market outcomes. Our study took advantage of a randomized control trial (RCT) conducted during the pandemic, where a worldwide recognized MOOC provider offered free certificates to public organizations. We collaborated with one such organization in Colombia and implemented a random experiment to assess the impact of free certificate eligibility on all registered program participants.

To the best of our knowledge, this study is one of the first to evaluate the impact of MOOC completion on labor market outcomes. We achieved this by merging registration data from the program with administrative records on formal labor market outcomes. A 2SLS model leveraging the random variation in the eligibility reveals positive, albeit somewhat imprecise, estimates of free certificates on employment. The limited precision can be attributed to the relatively modest effect of free certificate eligibility on course completion, as only 6% of eligible participants completed at least one course within the three-month program period.

To improve the precision of the estimates, we exploited temporal variations in the data by employing an event study approach. The results from this event study unveiled positive and statistically significant estimates, which were consistent with the ballpark of the 2SLS findings. This compelling evidence suggests that MOOCs can have a favorable impact on labor market outcomes, with low-income participants benefitting the most from completing MOOCs. The event study also allows estimating the ATT of any certificate by restricting the sample to eligible participants for whom we can track all their activity on the platform.

The results show similar estimates of the ATT of free vs. any certificate, suggesting that our results likely capture the effect of any certificate on labor market outcomes.

While our study provides encouraging evidence on the benefits of MOOC completion for labor market outcomes, several questions linger, which we aspire to explore in future research. First, a deeper understanding of the specific courses participants complete is necessary to discern how the fields of these courses mediate the observed positive impacts. While we currently don't have data on specific courses completed by the participants, we expect to access such information in the future. Additionally, our estimates aggregate the effects of course completion with the certificates without distinguishing between the acquisition of human capital from MOOCs and the signaling value of the certificates. Differentiating the impact of the actual learning gains and the perceived value of the certificates is essential to designing effective public policies to leverage MOOCs to enhance the skill set of the labor force.

References

Banerjee, A., Chassang, S., and Snowberg, E. (2017). Decision Theoretic Approaches to Experiment Design and External Validity. In Banerjee, A. V. and Duflo, E., editors, *Handbook of Field Experiments*, volume 1 of *Handbook of Economic Field Experiments*, pages 141–174. North-Holland.

Banerjee, A. V. and Duflo, E. (2014). (dis)organization and success in an economics mooc. *American Economic Review*, 104(5):514–18.

Belloni, A., Chernozhukov, V., and Hansen, C. (2013). Inference on Treatment Effects after Selection among High-Dimensional Controls†. *The Review of Economic Studies*, 81(2):608–650.

Bettinger, E. P., Fox, L., Loeb, S., and Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review*, 107(9):2855–75.

Bruhn, J., Campos, C., and Chyn, E. (2023). Who Benefits from Remote Schooling? Self-Selection and Match Effects. Technical report, Bravo Working Paper.

Camacho, A. and Conover, E. (2011). Manipulation of social program eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65.

Clark, D. and Martorell, P. (2014). The signaling value of a high school diploma. *Journal of Political Economy*, 122(2):282–318.

Deming, D. J., Goldin, C., Katz, L. F., and Yuchtman, N. (2015). Can Online Learning Bend the Higher Education Cost Curve? *American Economic Review*, 105(5):496–501.

Escueta, M., Quan, V., Nickow, A. J., and Oreopoulos, P. (2017). Education technology: An evidence-based review. Working Paper 23744, National Bureau of Economic Research.

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E., Patterson, T., and Staiger, D. O. (2023). The educational consequences of remote and hybrid instruction during the pandemic. *American Economic Review: Insights*, 5(3):377–92.

Goodman, J., Melkers, J., and Pallais, A. (2019). Can online delivery increase access to education? *Journal of Labor Economics*, 37(1):1–34.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, 42(4):679–694.

Hoxby, C. M. (2014). The Economics of Online Postsecondary Education: MOOCs, Nonselective Education, and Highly Selective Education. *American Economic Review*, 104(5):528–33.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Jack, R., Halloran, C., Okun, J., and Oster, E. (2023). Pandemic schooling mode and student test scores: Evidence from us school districts. *American Economic Review: Insights*, 5(2):173–90.

Kizilcec, R. F., Saltarelli, A. J., Reich, J., and Cohen, G. L. (2017). Closing global achievement gaps in moocs. *Science*, 355(6322):251–252.

Kofoed, M. S., Gebhart, L., Gilmore, D., and Moschitto, R. A. (2021). Zooming to class?: Experimental evidence on college students' online learning during covid-19. *SSRN Electronic Journal*.

MacLeod, W. B., Riehl, E., Saavedra, J. E., and Urquiola, M. (2017). The big sort: College reputation and labor market outcomes. *American Economic Journal: Applied Economics*, 9(3):223–61.

Majerowicz, S. and Zárate, R. A. (2022). Online learning and labor market outcomes: Experimental evidence from colombia. (0010016).

McPherson, M. S. and Bacow, L. S. (2015). Online Higher Education: Beyond the Hype Cycle. *Journal of Economic Perspectives*, 29(4):135–54.

Oreopoulos, P., Patterson, R. W., Petronijevic, U., and Pope, N. G. (2022). Low-Touch Attempts to Improve Time Management among Traditional and Online College Students. *Journal of Human Resources*, 57(1):1–43.

Pappano, L. (2012). The year of the MOOC. 2013.

Patterson, R. W. (2018). Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course. *Journal of Economic Behavior & Organization*, 153:293–321.

Table 1: Summary Statistics and Balance

	Eligibility for free certificates					
	Non-eligible		Eligible			
	Mean	SD	Mean	SD	Diff	SE
	(1)	(2)	(3)	(4)	(5)	(6)
Age	29.48	7.35	29.52	7.59	0.041	0.103
Male	0.36	0.48	0.37	0.48	0.009	0.007
Completed High School	0.15	0.36	0.15	0.36	-0.003	0.005
Completed Bachelors	0.72	0.45	0.72	0.45	0.008	0.006
Unemployed (baseline)	0.57	0.50	0.57	0.50	-0.001	0.007
Has taken online course before	0.66	0.47	0.64	0.48	-0.017***	0.007
Program goal: acquire knowledge	0.26	0.44	0.25	0.44	-0.002	0.006
Program goal: improve job opportunities	0.65	0.48	0.65	0.48	0.003	0.007
Program goal: improve business or start-up	0.10	0.30	0.10	0.30	-0.001	0.004
Interest in Arts and Humanities	0.27	0.44	0.27	0.44	0.001	0.006
Interest in Data Science	0.29	0.45	0.29	0.45	-0.001	0.006
Interest in Computer Science	0.31	0.46	0.30	0.46	-0.003	0.006
Interest in Social Science	0.22	0.42	0.22	0.41	-0.007	0.006
Interest in personal development	0.38	0.49	0.38	0.49	-0.002	0.007
Interest in Math	0.15	0.35	0.15	0.35	0.001	0.005
Interest in Business	0.46	0.50	0.46	0.50	-0.004	0.007
Interest in Health	0.25	0.43	0.24	0.43	-0.005	0.006
Interest in IT	0.40	0.49	0.40	0.49	-0.005	0.007
In SABER11 Sample	0.61	0.49	0.61	0.49	-0.007	0.007
Female (Saber11 Sample)	0.64	0.48	0.63	0.48	-0.005	0.010
Public school	0.58	0.49	0.58	0.49	0.004	0.010
HS Exit Exam Math Score	57.99	10.48	57.85	10.43	-0.140	0.186
HS Exit Exam Reading Score	57.96	9.25	57.87	9.21	-0.092	0.164
Mother's education	4.24	3.01	4.20	3.00	-0.036	0.057
Father's education	4.64	2.53	4.59	2.52	-0.052	0.047
Formal Work 2020 m1	0.48	0.50	0.48	0.50	-0.002	0.007
Formal Work 2020 m2	0.49	0.50	0.49	0.50	0.000	0.007
Formal Work 2020 m3	0.50	0.50	0.50	0.50	-0.001	0.007
Formal Work 2020 m4	0.47	0.50	0.47	0.50	-0.003	0.007
Formal Work 2020 m5	0.47	0.50	0.47	0.50	-0.003	0.007
Formal Work 2020 m6	0.47	0.50	0.47	0.50	-0.004	0.007
Formal work 2020 m7	0.47	0.50	0.47	0.50	0.001	0.007
Formal work 2020 m8	0.47	0.50	0.47	0.50	0.003	0.007
Formal work 2020 m9	0.48	0.50	0.49	0.50	0.003	0.007
Observations	8,687		12,988			
F-stat of joint orthogonality	0.84					
Conventional p-value	0.83					

Note: This table reports summary statistics of the treatment and control groups and balance tests for sociodemographic characteristics and formal employment at baseline. All the analysis is conducted at the participant level. The F-stat of joint orthogonality is carried out on the full set of variables, including indicators of formal work each month from 2017 to mid-2020. The table only reports balance tests nine months before the program's implementation. Robust standard errors are reported in columm 7; * $p < 0.10$, ** $p<0.05$, *** $p < 0.01$.

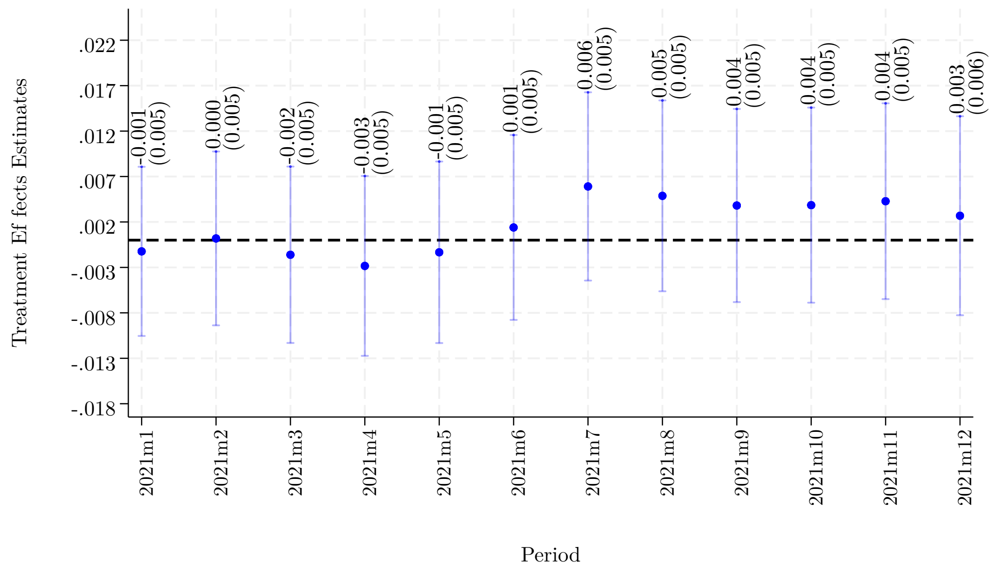
Table 2: First Stage on Course Enrollment and Completion

	Enrollment		Completion	
	Indicator	Courses	Indicator	Courses
	(1)	(2)	(3)	(4)
A. Eligibility for free certificates				
Treated	0.544***	3.101***	0.062***	0.130***
	(0.004)	(0.105)	(0.002)	(0.008)
Formal job at baseline	0.002	-0.030	-0.005*	-0.008
	(0.005)	(0.127)	(0.003)	(0.009)
Control mean	0.00	0.00	0.00	0.00
F-stat	15,494.34	867.93	859.32	281.96
N	21,675	21,675	21,675	21,675
B. Eligibility for free certificates by treatment round				
Treated 1st round	0.550***	3.033***	0.064***	0.136***
	(0.005)	(0.112)	(0.002)	(0.009)
Treated 2nd round	0.524***	3.325***	0.055***	0.109***
	(0.009)	(0.261)	(0.004)	(0.012)
Formal job at baseline	0.002	-0.031	-0.005*	-0.008
	(0.005)	(0.126)	(0.003)	(0.009)
Control mean	0.00	0.00	0.00	0.00
F-stat	7,754.57	446.71	429.77	147.84
p-value treat 1st= treat 2nd	0.014	0.304	0.049	0.074
N	21,675	21,675	21,675	21,675

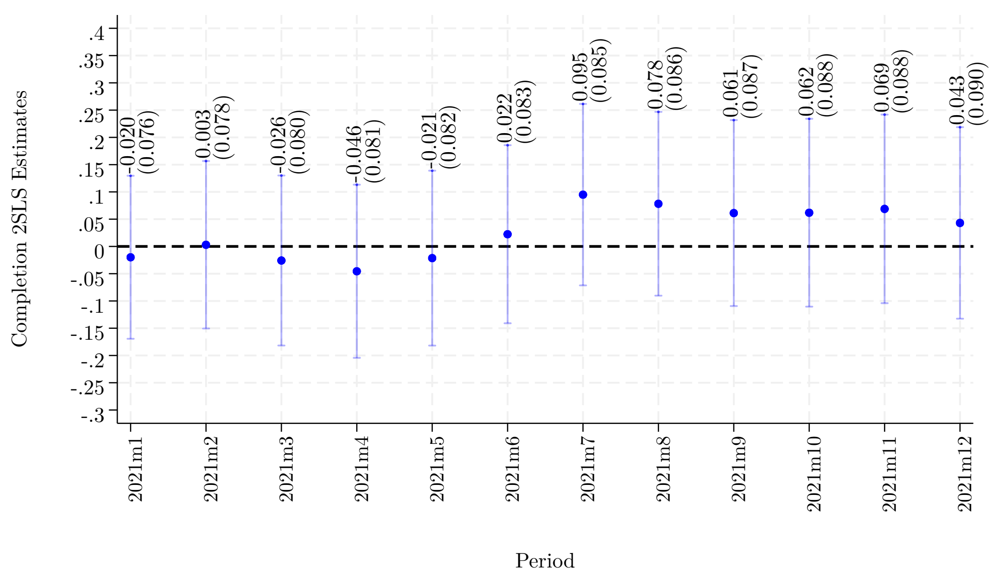
Note: This table reports treatment effects of free certificates eligibility on MOOCs' enrollment and completion. Formal job at baseline corresponds to September 2020. Panel A reports the estimates pooling together in one group the applicants who receive a 1st- and 2nd-round offer, while Panel B splits the two groups. Robust standard errors are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Reduced Form and 2SLS Effects on Formal Labor Employment

A. Reduced Form Eligibility Effects



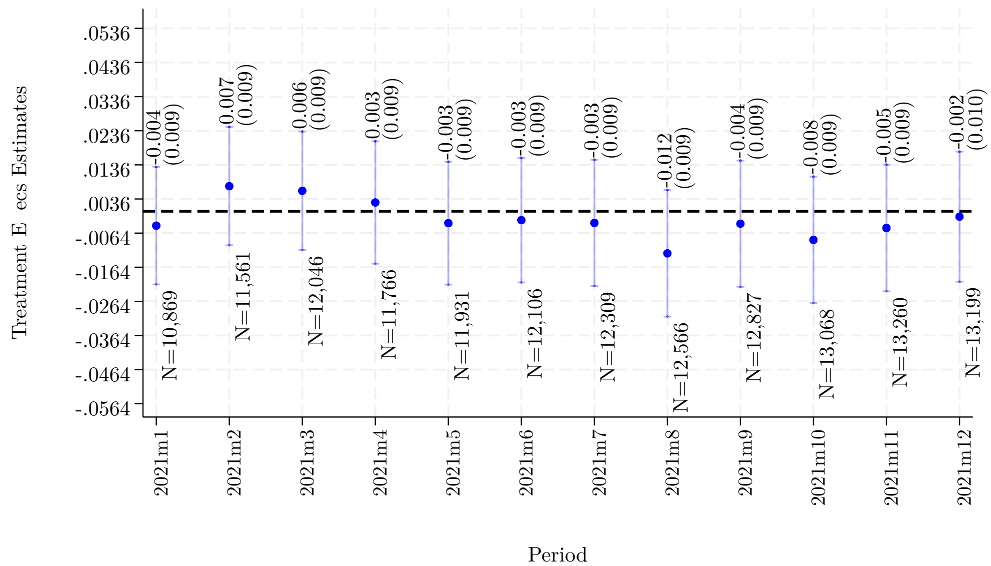
B. 2SLS Estimates of Free Certificates



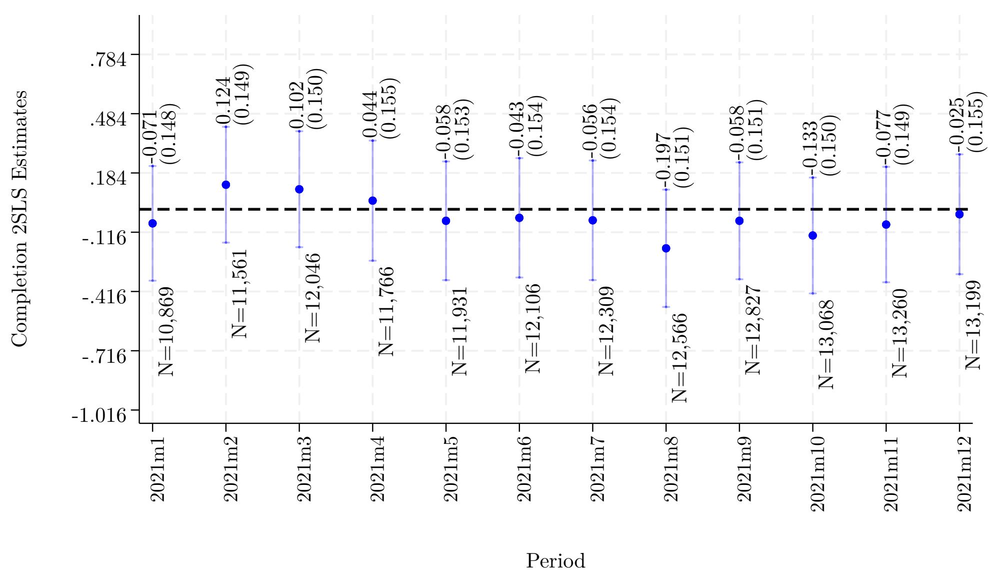
Note: This figure reports the treatment effects of free certificate eligibility and free certificates on formal labor employment for each period after the intervention. Panel A reports the estimates of equation 1, the effects of being eligible for free certificates. Panel B reports the estimates of equation 2a, the 2SLS estimates of free certificates ATT. Each marker represents a different regression. The figure shows the point estimate for each period, and robust standard errors are presented in parentheses. All regressions include a set of covariates at baseline selected using the double-post-lasso covariate selection method proposed by (Belloni et al., 2013). This set of covariates includes the value of the dependent variable for all periods before the randomization, sociodemographic characteristics such as age and gender, whether participants have available standardized test scores, and their math and reading scores. The number of observations for all estimations is 21,675.

Figure 2: Reduced Form and 2SLS Effects on Daily Wages

A. Reduced Form Eligibility Effects

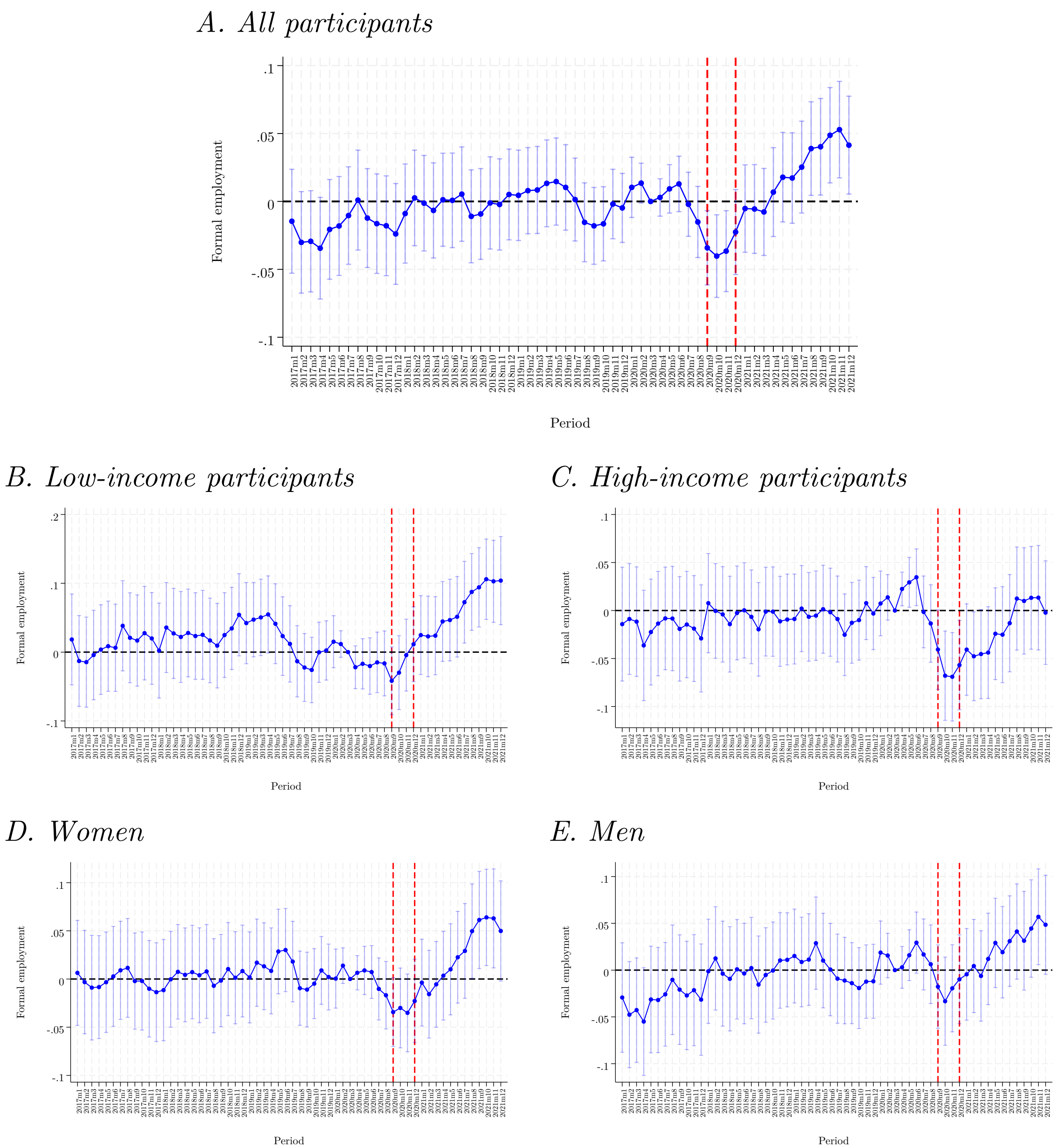


B. 2SLS Estimates of Free Certificates



Note: This figure reports the treatment effects of free certificate eligibility and free certificates on daily wages for each period after the intervention. Panel A reports the estimates of equation 1, the effects of being eligible for free certificates. Panel B reports the estimates of equation 2a, the 2SLS estimates of free certificates ATT. Each marker represents a different regression. The figure shows the point estimate for each period, and robust standard errors are presented in parentheses. All regressions include a set of covariates at baseline selected using the double-post-lasso covariate selection method proposed by (Belloni et al., 2013). This set of covariates includes the value of the dependent variable for all periods before the randomization, sociodemographic characteristics such as age and gender, whether participants have available standardized test scores, and their math and reading scores.

Figure 3: Event Study Free Certificates Effects on Formal Employment



Note: This figure reports the event study estimates of equation 4 of course completion effects on formal employment. Panel A reports the estimates for all participants, panels B and C by income level, and panels D and E by gender. The number of observations is analogous to the ones reported in Table A.3. All regressions control for the interactions between income level and gender with time dummies. The lines around each estimate represent 95% confidence intervals with standard errors clustered at the participant level.

Appendix

A Additional Tables and Figures

Table A.1: Reduced Form and 2SLS Effects on Formal Labor Employment

Period	Control	ITT (eligibility)		2SLS free certificates		
	mean	estimate	s.e.	estimate	s.e.	F-stat FS
	(1)	(2)	(3)	(4)	(5)	(6)
2020m1	0.479	-0.001	(0.004)	-0.012	(0.061)	858.59
2020m2	0.493	0.001	(0.004)	0.013	(0.069)	857.76
2020m3	0.497	-0.001	(0.005)	-0.018	(0.073)	857.75
2020m4	0.473	-0.003	(0.005)	-0.043	(0.076)	856.98
2020m5	0.468	-0.003	(0.005)	-0.046	(0.078)	857.03
2020m6	0.473	-0.004	(0.005)	-0.061	(0.079)	857.62
2020m7	0.471	0.001	(0.005)	0.019	(0.081)	856.89
2020m8	0.472	0.003	(0.005)	0.041	(0.082)	857.36
2020m9	0.483	0.002	(0.005)	0.040	(0.083)	857.31
2020m10	0.505	-0.003	(0.003)	-0.056	(0.050)	858.69
2020m11	0.520	-0.002	(0.004)	-0.025	(0.061)	858.40
2020m12	0.528	-0.004	(0.004)	-0.071	(0.069)	859.36
2021m1	0.502	-0.001	(0.005)	-0.020	(0.076)	859.62
2021m2	0.533	0.000	(0.005)	0.006	(0.078)	859.67
2021m3	0.556	-0.002	(0.005)	-0.026	(0.080)	859.26
2021m4	0.569	-0.003	(0.005)	-0.046	(0.081)	859.78
2021m5	0.576	-0.001	(0.005)	-0.022	(0.082)	859.30
2021m6	0.584	0.001	(0.005)	0.022	(0.083)	859.33
2021m7	0.592	0.006	(0.005)	0.095	(0.085)	859.50
2021m8	0.605	0.005	(0.005)	0.078	(0.086)	859.26
2021m9	0.618	0.004	(0.005)	0.061	(0.087)	859.32
2021m10	0.630	0.004	(0.005)	0.062	(0.088)	859.48
2021m11	0.639	0.004	(0.005)	0.069	(0.088)	859.39
2021m12	0.637	0.003	(0.006)	0.043	(0.090)	859.71

Note: This table reports the treatment effects of free certificate eligibility and the ATT of free certificates on formal labor employment for each month between January 2020 and December 2021. The program took place between October and December 2020. Columns 2-3 report the estimates of equation 1, the effects of being eligible for free certificates. Columns 4-6 report the estimates of equation 2a, the 2SLS estimates of free certificates ATT. Each row represents a different regression, with robust standard errors presented in columns 3 and 5. All regressions include a set of covariates at baseline selected using the double-post-lasso covariate selection method proposed by (Belloni et al., 2013). This set of covariates includes the value of the dependent variable for all periods up to 2019 for the estimations in 2020 and for all periods up to September 2020 for all estimations in 2021. The covariates also include sociodemographic characteristics such as age and gender, whether participants have available standardized test scores, and their math and reading scores. The number of observations for all estimations is 21,675.

Table A.2: Reduced Form and 2SLS Effects on Daily Wages

Period	N	Control mean	ITT (eligibility) estimate	s.e.	2SLS free certificates estimate	s.e.	F-stat FS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2020m1	10,356	14.239	-0.004	(0.008)	-0.068	(0.136)	402.53
2020m2	10,697	14.186	-0.001	(0.007)	-0.022	(0.117)	416.46
2020m3	10,748	14.535	0.004	(0.011)	0.070	(0.181)	406.01
2020m4	10,221	14.589	0.004	(0.008)	0.060	(0.128)	388.25
2020m5	10,117	14.215	0.001	(0.011)	0.013	(0.174)	389.50
2020m6	10,201	14.443	0.009	(0.010)	0.145	(0.158)	396.17
2020m7	10,225	14.519	0.003	(0.011)	0.046	(0.182)	383.56
2020m8	10,266	14.261	-0.006	(0.008)	-0.113	(0.140)	374.46
2020m9	10,500	14.269	-0.009	(0.011)	-0.157	(0.200)	367.15
2020m10	10,931	14.445	0.001	(0.009)	0.010	(0.157)	378.82
2020m11	11,265	14.555	0.000	(0.008)	0.003	(0.140)	395.98
2020m12	11,408	15.151	0.000	(0.007)	0.003	(0.129)	414.81
2021m1	10,869	15.410	-0.004	(0.009)	-0.071	(0.148)	408.84
2021m2	11,561	15.732	0.007	(0.009)	0.124	(0.149)	436.09
2021m3	12,046	15.472	0.006	(0.009)	0.102	(0.150)	452.58
2021m4	11,766	15.846	0.003	(0.009)	0.044	(0.155)	440.78
2021m5	11,931	16.185	-0.003	(0.009)	-0.058	(0.153)	454.92
2021m6	12,106	16.569	-0.003	(0.009)	-0.043	(0.154)	465.42
2021m7	12,309	16.578	-0.003	(0.009)	-0.056	(0.154)	480.04
2021m8	12,566	17.063	-0.012	(0.009)	-0.197	(0.151)	502.47
2021m9	12,827	17.086	-0.004	(0.009)	-0.058	(0.151)	514.60
2021m10	13,068	17.216	-0.008	(0.009)	-0.133	(0.150)	529.97
2021m11	13,260	17.564	-0.005	(0.009)	-0.077	(0.149)	541.97
2021m12	13,199	18.507	-0.002	(0.010)	-0.025	(0.155)	530.27

Note: This table reports the treatment effects of free certificate eligibility and the ATT of free certificates on the log of daily wages for each month between January 2020 and December 2021. The program took place between October and December 2020. Columns 3-4 report the estimates of equation 1, the effects of being eligible for free certificates. Columns 5-7 report the estimates of equation 2a, the 2SLS estimates of free certificates ATT. Each row represents a different regression, with robust standard errors presented in columns 4 and 6. All regressions include a set of covariates at baseline selected using the double-post-lasso covariate selection method proposed by (Belloni et al., 2013). This set of covariates includes the value of the dependent variable for all periods up to 2019 for the estimations in 2020 and for all periods up to September 2020 for all estimations in 2021. The covariates also include sociodemographic characteristics such as age and gender, whether participants have available standardized test scores, and their math and reading scores.

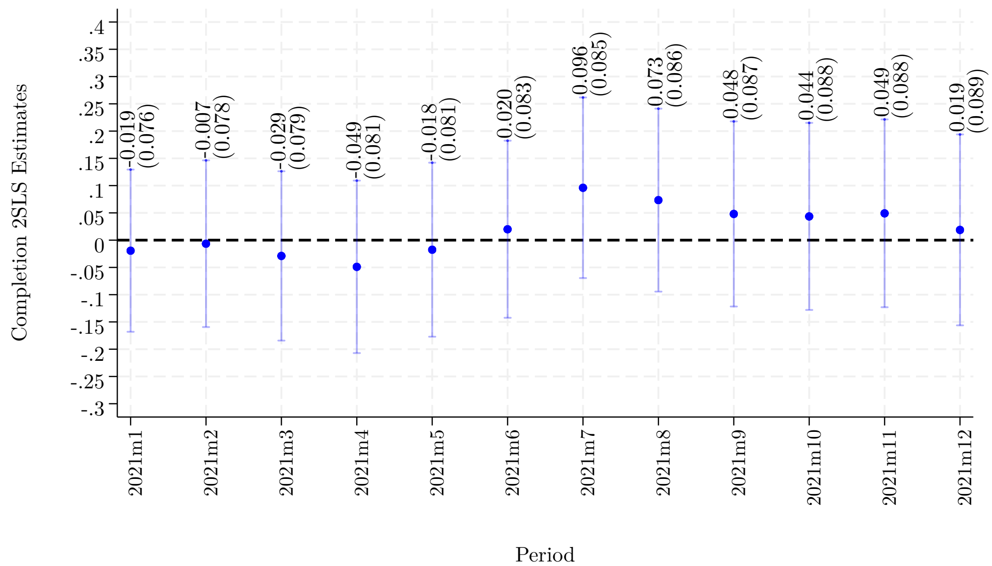
Table A.3: DiD Free Certificates and Certificates Effects on Formal Employment

Sample:	All participants	By Income Level		By Gender	
		Low-income	High-income	Women	Men
	(1)	(2)	(3)	(4)	(5)
I. Single post-period indicator					
A. All participants					
Completed x post	0.033***	0.055**	-0.007	0.029	0.036**
	(0.013)	(0.022)	(0.019)	(0.019)	(0.018)
Control mean	0.44	0.47	0.50	0.48	0.52
Equal effects test (p-value)		0.034		0.795	
N	1,300,500	519,660	516,600	770,400	470,100
B. Participants eligible for free certificates					
Completed x post	0.035***	0.045*	0.008	0.029	0.039**
	(0.013)	(0.024)	(0.019)	(0.019)	(0.018)
Control mean	0.44	0.47	0.49	0.48	0.51
Equal effects (p-value)		0.226		0.705	
N	779,280	273,360	344,940	458,880	284,520
II. Different periods					
A. All participants					
Completed x pandemic	0.002	-0.040*	0.014	-0.010	0.021
	(0.012)	(0.023)	(0.018)	(0.018)	(0.019)
Completed x during	-0.025*	-0.025	-0.056***	-0.031	-0.009
	(0.014)	(0.026)	(0.022)	(0.021)	(0.022)
Completed x post	0.032**	0.048**	-0.009	0.026	0.038*
	(0.014)	(0.024)	(0.021)	(0.021)	(0.020)
Control mean	0.44	0.47	0.50	0.48	0.52
Equal effects test (p-value)		0.075		0.676	
N	1,300,500	519,660	516,600	770,400	470,100
B. Participants eligible for free certificates					
Completed x pandemic	0.004	-0.048*	0.018	-0.007	0.020
	(0.013)	(0.026)	(0.017)	(0.019)	(0.019)
Completed x during	-0.024	-0.038	-0.042**	-0.029	-0.011
	(0.015)	(0.028)	(0.021)	(0.021)	(0.022)
Completed x post	0.034**	0.036	0.008	0.026	0.041**
	(0.014)	(0.026)	(0.020)	(0.021)	(0.020)
Control mean	0.44	0.47	0.49	0.48	0.51
Equal effects (p-value)		0.391		0.618	
N	779,280	273,360	344,940	458,880	284,520

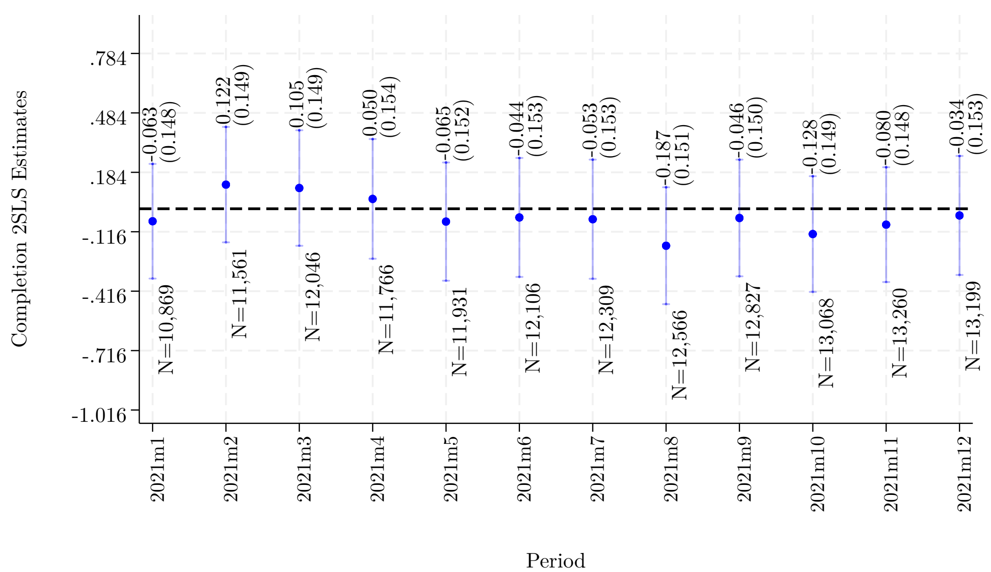
Note: This table reports the DiD estimates of equation 3. As we can only track platform activity for treated participants, Panels I.A. and II.A. report the ATT of free certificates as the control group includes eligible participants for free certificates who did not complete MOOCs and participants who were not eligible to receive the free certificates. Panels I.B. and II.B., on the other hand, report the ATT of obtaining a certificate by restricting the sample to eligible participants for free certificates. Column 1 reports the estimates for all participants, columns 2 and 3 classify participants by income level, and columns 4 and 5 by gender. Some observations are missing the income (20.3%) and the gender (4.61%) information. Standard errors clustered at the participant level are reported in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure A.1: 2SLS Free Certificates Effects with Multiple Instruments

A. Formal employment

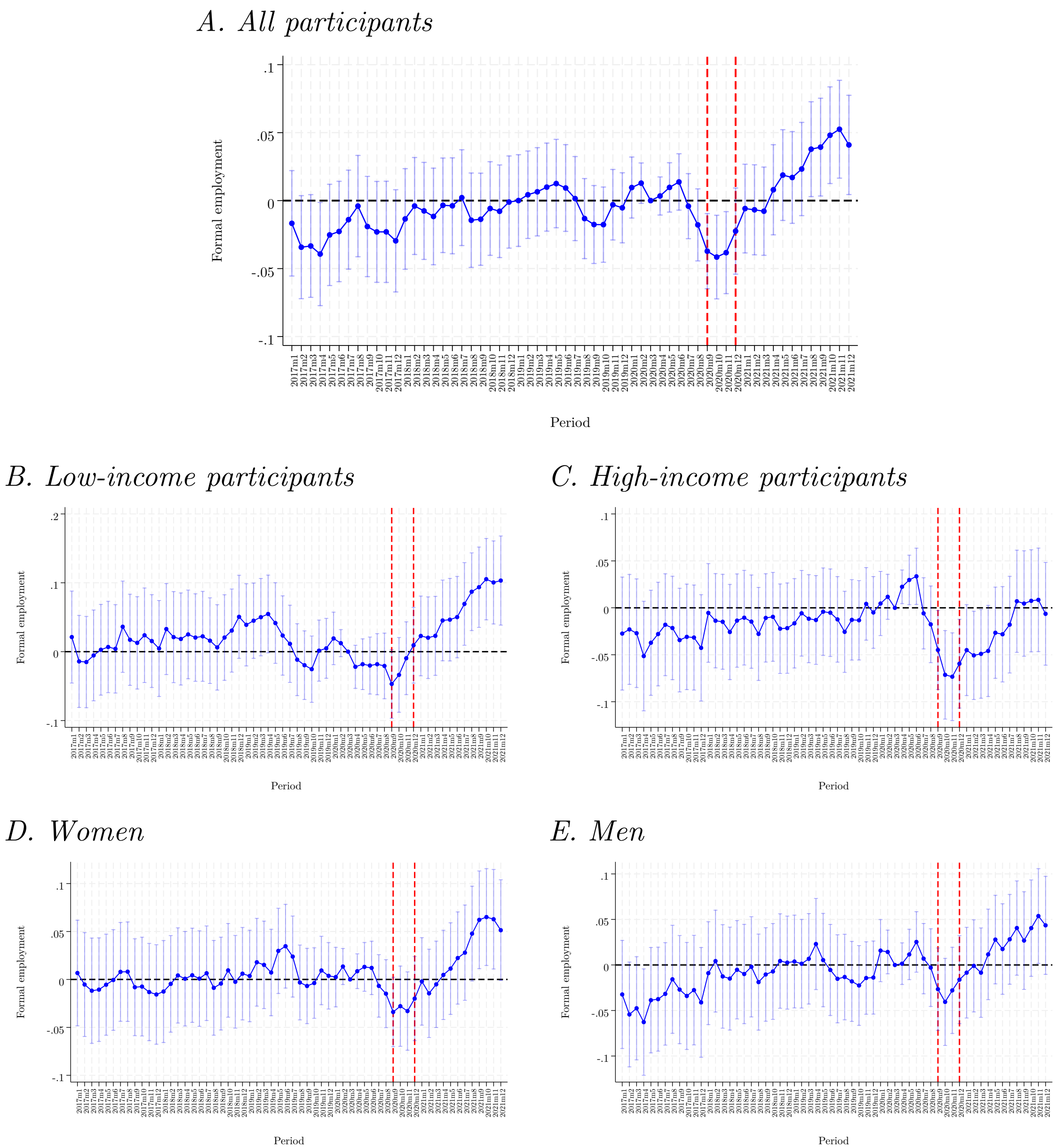


B. Daily wages



Note: This figure reports 2SLS effects of course completion on formal labor employment and wages using a multiple-instruments model with both rounds of the randomization instruments being the instruments for course completion. Panel A reports the estimates of equation 2a, the 2SLS estimates of completing at least one course on formal labor employment for each period after the intervention, and Panel B on wages. Each marker represents a different regression. The figure shows the point estimate for each period, and standard errors are presented in parentheses. All regressions include a set of covariates at baseline selected using the double-post-lasso covariate selection method proposed by (Belloni et al., 2013). This set of covariates includes the value of the dependent variable between January 2017 and September 2020, sociodemographic characteristics such as age and gender, whether participants have available standardized test scores, and their math and reading scores. The number of observations for all estimations in Panel A is 21,675.

Figure A.2: Event Study Certificates Effects on Formal Employment



Note: This figure reports the event study estimates of equation 4 of any certificate on formal employment by restricting the sample to eligible participants. Panel A reports the estimates for all participants, panels B and C by income level, and panels D and E by gender. The number of observations is analogous to the ones reported in Table A.3. All regressions control for the interactions between income level and gender with time dummies. The lines around each estimate represent 95% confidence intervals with standard errors clustered at the participant level.