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

Stephanie Majerowicz Román Andrés Zárate

November 7, 2023

Abstract

This paper studies the effect of 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% 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' completion. To improve precision, we estimate an event study of course completion, finding a significant average increase of 7.5% 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, complementary interventions to increase their completion are needed.

*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. 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) made their debut in the global education landscape, they generated significant initial enthusiasm, fueled by the promise of revolutionizing higher education through widespread access to university-level courses at minimal marginal costs. Although this prediction of an online learning takeover appears not to have immediately materialized, the Covid-19 pandemic gave online learning an unexpected boost, with most traditional universities forced to go online temporarily due to the lockdowns. While most traditional universities returned to in-person learning relatively quickly, the pandemic appears to have accelerated the demand for online courses and degrees.

Even prior to the onset of the pandemic, both public and private institutions worldwide were actively advocating the adoption of online learning, with a particular emphasis on MOOCs as a means to enhance human capital and improve the employability of individuals. For instance, 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). While MOOCs allow students to audit courses at no cost, thus potentially enabling them to acquire new skills without a financial burden, they must pay a non-trivial fee for the certificates that validate their competencies.[1]

A substantial body of evidence exists examining the effects of virtual vs. in-person instruction (Bettinger et al., 2017; Bruhn et al., 2023). However, most of these studies compare in-person and 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. Nevertheless, the efficacy of these concise online courses in skill development and their potential to signal abilities to the labor market remain uncertain. As highlighted by Escueta et al. (2017), most research on MOOCs has primarily focused on assessing whether and how various behavioral interventions can enhance MOOC completion rates, with scant evidence regarding the long-term implications of MOOC participation.

This paper aims to bridge this gap by studying the impact of MOOC certificates on formal labor market outcomes in Colombia. To the best of our knowledge, this is the first study to explore how MOOC completion affects labor market outcomes. We leverage a program implemented by a prominent MOOC provider that offered free certificates to public organizations 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 who

[1] For instance, 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 demanding costs in the range of $2,000 to $5,000 USD.

enrolled in the program, approximately 13,000 were randomly selected to become eligible for free certificates upon successfully completing MOOCs within three months. Participants could earn certificates for individual courses and for specializations, which are more structured degrees composed of a series of courses.

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 completion on formal labor employment and wages.

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% of them completed at least one course. While there is a substantial variation in the number of completed courses, it appears that most beneficiaries didn't 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 the treatment on formal labor employment, with a clear pattern of positive effects 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. 0.006) one year after the program ended. We estimate local average treatment effects (LATE) of course completion on formal labor employment, using the treatment assignment as an instrument for finishing the courses in a two-stage least squares (2SLS) framework. The LATE estimates reveal the impact of course completion on compliers (Imbens and Angrist, 1994): those who obtain the certificates due to the treatment of being eligible for the free certificates. While the estimates 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 course completion also do not show a clear pattern or statistically significant effects on daily wages.

Motivated by the large but imprecise estimates of course completion using the 2SLS framework, we leverage the time variation before and after the program to estimate the impact of course completion 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 course completion 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 completion effects estimates from the event study closely align with the 2SLS effects of course completion, offering estimates that are more precise but similar in magnitude.

We also explore heterogeneity in the completion effects on employment by income level and gender. We find that low-income participants, measured by proxy means test used to target social programs in Colombia, benefit the most from course completion. 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 complete the courses. 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

The COVID-19 Pandemic was an unprecedented shock to the world's economy. Latin America was one of the regions most affected by the pandemic, with more than 28 million cases and almost 1 million deaths as of May of 2021. In Colombia, the unemployment rate nearly doubled between July 2019 and July 2020, increasing from 10.7% to 20.2% (Dane, 2020). The unemployment of women and youth has always been higher than the population average, but it was also disproportionately affected by the pandemic. The youth (younger than 28 years old) unemployment rate increased from 17.5% to 29.7% during the same period, while women's unemployment rose to an unprecedented rate of 25.5% for all women and 37.7% for young women in the country.

In this context, the Government searched for potential employment policies that could soften the pandemic's impact on labor markets. Recent evidence documents that the length of unemployment diminishes a worker's future job market opportunities (Kroft et al., 2013).

This could be driven, in part, by a deterioration in skills over the course of an unemployment spell. Hence, it is critical to assess which policies can improve current and future labor market opportunities, as well as help individuals develop skills that are valued in the labor market.

Acquiring new skills is a potential mechanism by which workers affected by the pandemic or other negative external shocks can mitigate their impact and improve their future job opportunities. There is broad evidence documenting the positive impact of years of schooling on wages and the positive and high returns to both cognitive and non-cognitive skills in the labor market.

One of the leading platforms in MOOCs launched an initiative to mitigate the impact of COVID on employment in Latin America. The initiative allowed government agencies across the region to apply for a number of slots that would allow users to enroll in courses and receive course completion certificates from the MOOC provider. While users can normally access most of the content at the MOOC provider for free, certificates of completion are costly, with costs ranging between 40 and 300 USD for a single course. In the Colombian context, 100 USD is equivalent to about half of a monthly minimum wage, so the cost could represent a significant barrier to acquiring the certificate. Formal certificates allow workers to signal their abilities in the labor market (Spence, 1973; MacLeod et al., 2017). With this initiative, government agencies could provide participants not only with information about available courses, but also with certificates that validate their acquired skills to potential employers, which would provide an additional incentive to sign up for and complete the courses.

Several institutions across Latin American countries participated in the initiative. In Colombia, the Colombian Institute for Educational Credit and Technical Studies (Icetex), which provides higher-education loans and scholarships, received 10,000 slots to allocate to students . The Icetex offered their assigned slots to current or past beneficiaries of student loans.

The Icetex received over 23,000 applications for its 10,000 slots. They decided to allocate the slots through random assignment among the applicants, following advice from the research team who was providing technical assistance on how to evaluate the program. Students who were offered a slot 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 requirements of the course 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 four different courses. 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.

Before the randomization and as specified in the pre-registered experiment (Majerowicz and Zárate, 2022), we performed power calculations for employment in the formal sector and the average daily earnings as the primary outcome variables. Given the large sample size of the experiment, we have enough power to detect reasonable small effects. At the conventional power level of 0.8, we can detect an impact of 1.82 p.p. on formal employment and of 0.027 log points ($\approx$ 0.03 s.d.) for average daily earnings.

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 they enrolled in and the completed courses for which they received a certificate. Unfortunately, due to the terms and conditions of the program, this data is unavailable for participants in the control group. Given the low take-up rates among the treated participants and the cost of the certificates, we believe it is unlikely for participants in the control group to have completed any online course during this period.

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 students, 8,687 of whom are in the control group and 12,988 of whom were randomized into the treatment group 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, and test for balance across different demographic characteristics and employment status at baseline.

Columns 1 and 3 of Table 1 present sample averages for the treatment and the control group. 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 during the same period, around 47% of them were formally employed. 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 treatment and control 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 the treatment and control 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 0.84.

4 Empirical Strategy

In this section, we present the empirical strategy that we follow to estimate the impact of free certificates eligibility on labor market outcomes. First, we estimate a straightforward reduced form specification of the effect being assigned to the treatment 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 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 comprehension scores when available. Lastly, ε_{it} is an error term. The parameter of interest in equation 1 is β, the treatment effect of having 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 course completion on students' outcomes, using the treatment assignment as an instrument for 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 as part of the program. Equation 2a is the second stage of the model, with parameter of interest γ, the effect of course completion on participants' labor market outcomes. Equation 2b is the first stage equation, with parameter β capturing the impact of the treatment on course completion. 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.

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. 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 treatment effects estimates and address selection concerns of 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 completed a course against those who did not.

First, we can estimate the following DiD specification:

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

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 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:

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

Here, τ_t represent 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 the treatment and the control groups in period t compared to a reference period $t = 0$, which is excluded from the estimation. 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.

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.

5 Results

5.1 First Stage

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 the effect of being assigned to the treatment on the probability of enrolling in at least one course, the number of enrollments, the probability 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 treated participants, the estimates show that only 54% of those offered free certificates enroll in at least one course, and very few, only around 6%, complete at least one course. Students in the treatment group 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. 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. However, given the low completion rates among treated students and the non-trivial cost of the certificates, it is unlikely that participants in the control group completed courses during this period.

Panel B of Table 2 reports the estimates of equation 1, but splits the 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. For our main results, we pool both offers into a single variable, but our results are similar when we separate the two.

5.2 Treatment and Completion Effects

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 treatment effects of equation 1 and the 2SLS estimates of course completion effects 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 reports the reduced form treatment effects, and Panels B the 2SLS estimates of course completion effects.

While the results in Panel A show slightly negative or positive non-significant effects for the first five months after the program ended (up to May 2021), the treatment effects on formal employment are clear positives from June 2021 onward. Furthermore, the 2SLS estimates of course completion reported in Panel B show large but non-significant effects from June until December 2021. According to these results, course completion 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 course completion on formal employment. These impacts represent an increase between 8.9% to 23.8% of the baseline formal employment rate for the 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 insignificant 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 treatment effects and the 2SLS estimates of completion on the log of 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 course completion on wages. 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.

5.3 Event Study Estimates

Motivated by the positive but imprecise estimates of completion effects on formal employment, we exploit the time variation in the data with an event study analysis. Panel A of Figure 3 reports the estimates of equation 4 of completion effects on formal employment for all participants. Column 1 of Appendix Table A.1 reports the analogous estimates of the DiD model (equation 3); the estimates in Panel A pool together all the pre and the post-period, and the estimates in Panel B separate the pre-period between January 2017 to February 2020 (the excluded period), the six months of the pandemic before the program, March to September 2020, and the period of the program when participants were eligible to receive free certificates, October to December 2020.

The estimates in Figure 3 provide promising insights into the potential impact of completing MOOCs and, therefore, earning the 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 B in Appendix Table A.1 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 2020 to December 2020, with a statistically significant impact of approximately 5 p.p. From a baseline formal employment rate of 44%, these estimates translate to an increase ranging from 7.5% to 11.4% in formal employment.

As we are only fully certain about course completion for the treated participants, these estimates assume that completion rates among the control groups are zero. As a robustness check Appendix Figure A.2 report the estimates restricting the sample to treated participants, comparing those who completed vs. those who did not complete courses. While these estimates are less powered, the results are fairly similar. Overall, these results suggest that MOOC completion has positive impacts on formal employment.

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 derive greater benefits from course completion compared to their higher-income counterparts. Specifically, while the estimates for the impact of course completion on 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 completed the courses 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 course completion. Columns 2 and 3 of Appendix Table A.1 confirm this result, with an average post-treatment completion effect of 5.5 p.p. (p-value < 0.05) for low-income participants, and a -0.7 p.p. completion effect for high-income participants; the difference in the completion effect between the two groups is statistically significant (p-value $= 0.034$).

We also explore heterogeneity by gender with no statistically significant differences in the completion effect 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.1 reaffirm this parity. The effects are consistently positive and similar in magnitude for men and women, with no statistically significant difference in the completion effects between the two groups. The reduced sample size in these gender-specific analyses derives into less statistical power compared to the overall effects presented 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 certificate eligibility for MOOC completion on labor market outcomes. Our study took advantage of a randomized control trial (RCT) conducted during the pandemic, where a prominent 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 this 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 for free certificates reveals positive, albeit somewhat imprecise, estimates. 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.

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*.

Kroft, K., Lange, F., and Notowidigdo, M. J. (2013). Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment*. *The Quarterly Journal of Economics*, 128(3):1123–1167.

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.

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.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Table 1: Summary Statistics and Balance

	Control		Treatment			
	Mean	SD	Mean	SD	Diff	SE
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
Observations	8,687		12,988			
F-stat of joint orthogonality	0.83					
Conventional p-value	0.84					

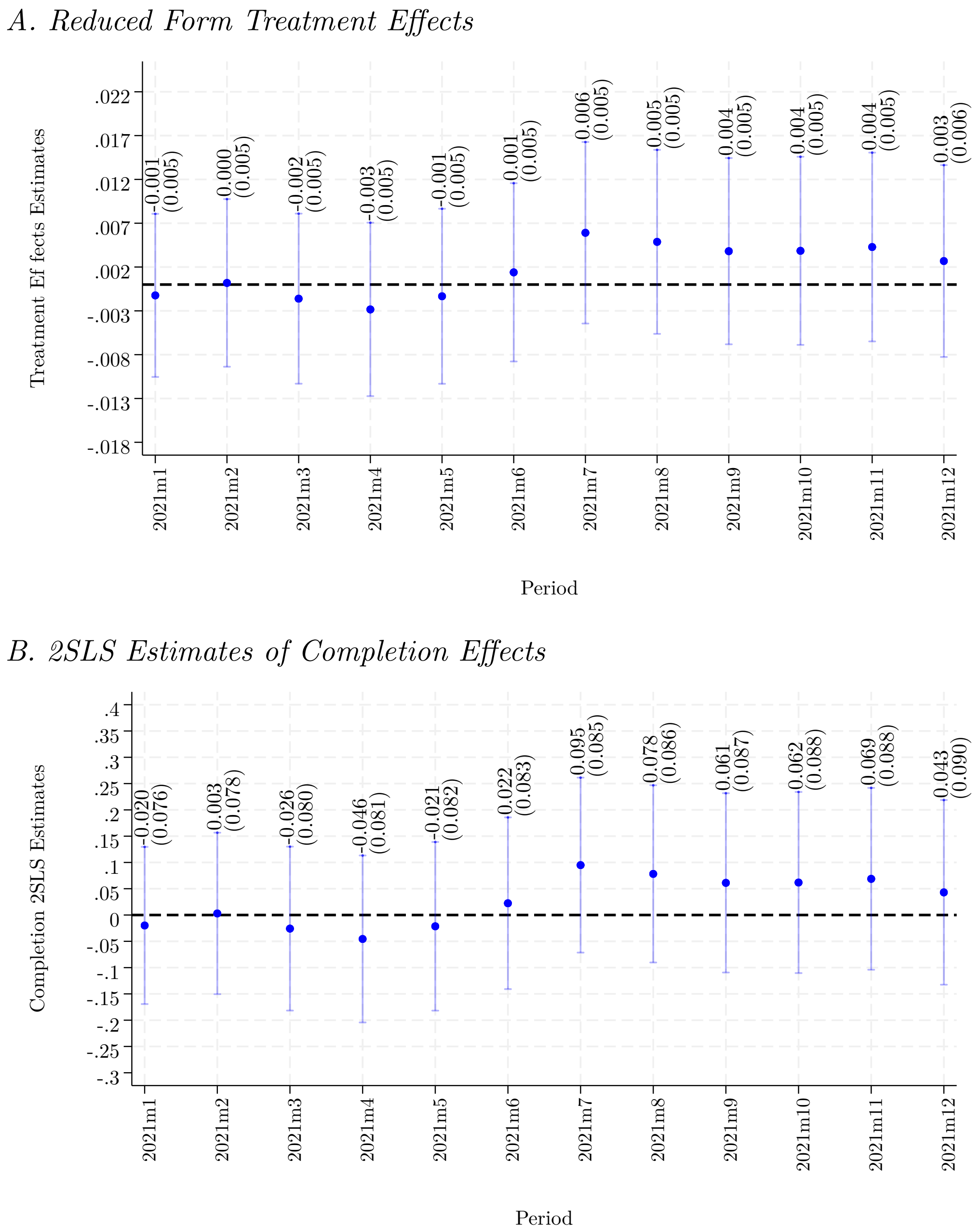
Note: The F-stat of joint orthogonality is carried out on the full set of covariates which include dummys for formal work each month from 2018 to mid-2020. In this table we show the 6 months prior to treatment, the rest are displayed in Figure X. Robust standard errors are reported in columm 7; * $p < 0.10$, ** $p<0.05$, *** $p < 0.01$.

Table 2: First Stage on Course Enrollment and Completion

	Enrollment		Completion	
	Indicator	Courses	Indicator	Courses
	(1)	(2)	(3)	(4)
A. Any treatment				
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. 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; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

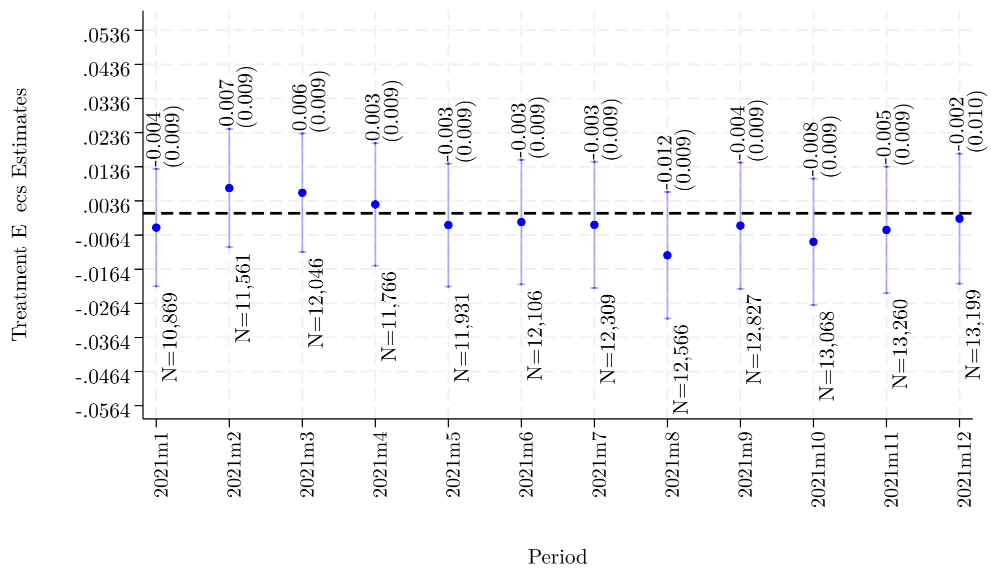
Figure 1: Reduced Form and 2SLS Effects on Formal Labor Employment



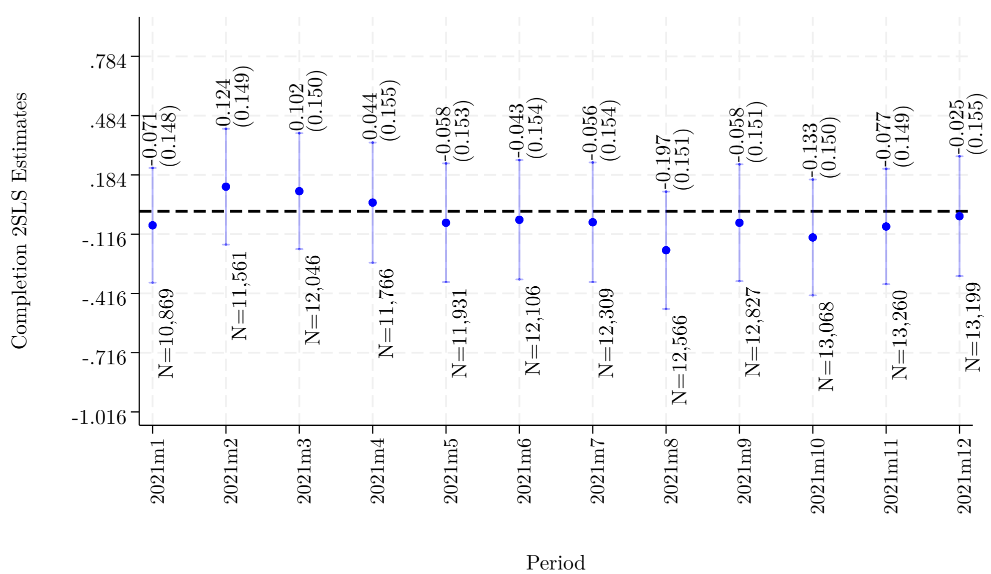
Note: This figure reports treatment and 2SLS effects of course completion on formal labor employment. Panel A reports the estimates of equation 1, the effects of receiving the treatment (eligibility for free certificates) on formal labor employment for each period after the intervention. Panel B reports the estimates of equation 2a, the 2SLS estimates of completing at least one course on formal labor employment for each period after the intervention. 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 for six months 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 Treatment Effects



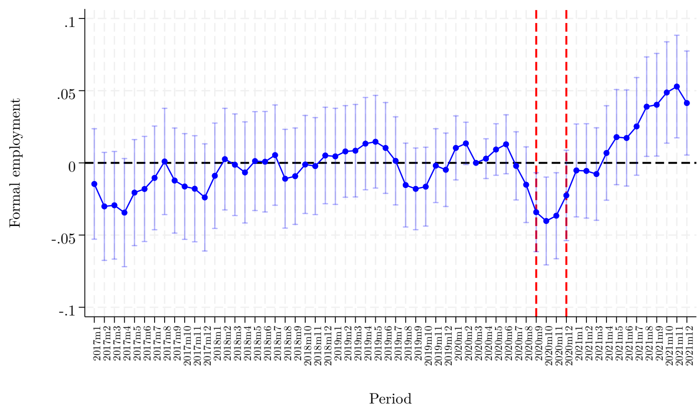
B. 2SLS Estimates of Completion Effects



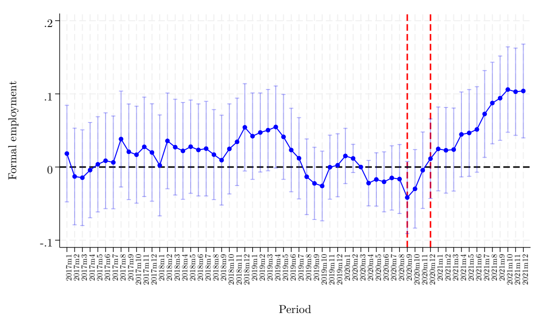
Note: This figure reports treatment and 2SLS effects of course completion on wages. Panel A reports the estimates of equation 1, the effects of receiving the treatment (eligibility for free certificates) on the log of the wage for each period after the intervention. Panel B reports the estimates of equation 2a, the 2SLS estimates of completing at least one course on the log of the wage for each period after the intervention. Each marker represents a different regression. The figure shows the point estimate for each period, and standard errors are presented in parentheses. The number of observations is reported at the bottom of the figure for each period. 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 six months 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: Estimates of Completion Effects on Formal Employment using the Event Study

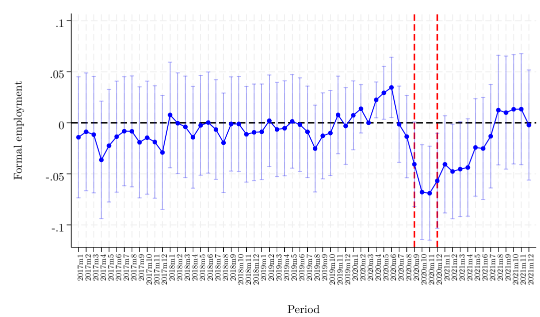
A. All participants



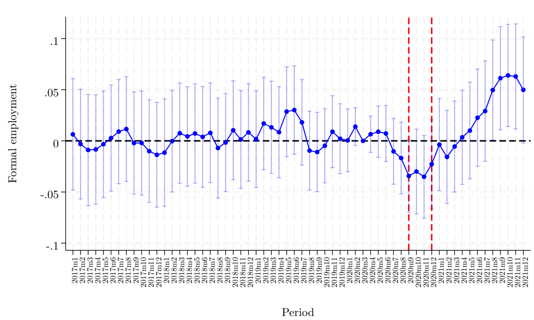
B. Low-income participants



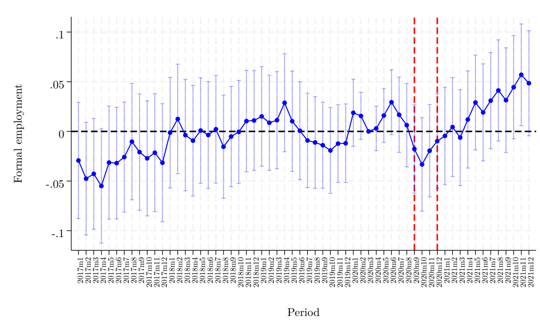
C. High-income participants



D. Women



E. Men



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.1. 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

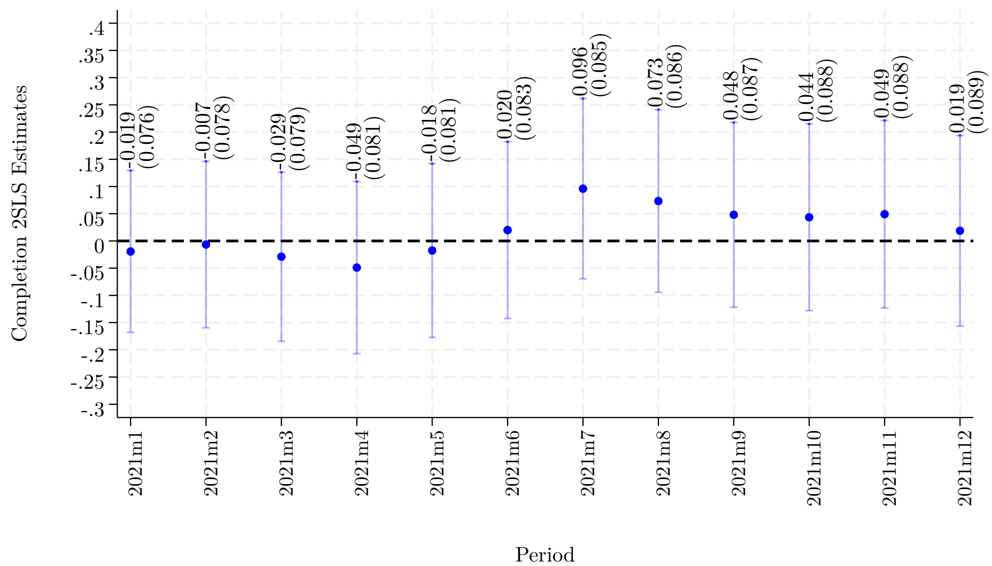
Table A.1: DiD Estimates of Completion on Formal Employment

Sample:	All participants	By Income Level		By Gender	
		Low-income	High-income	Women	Men
	(1)	(2)	(3)	(4)	(5)
A. Single post-period indicator					
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. Different periods					
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

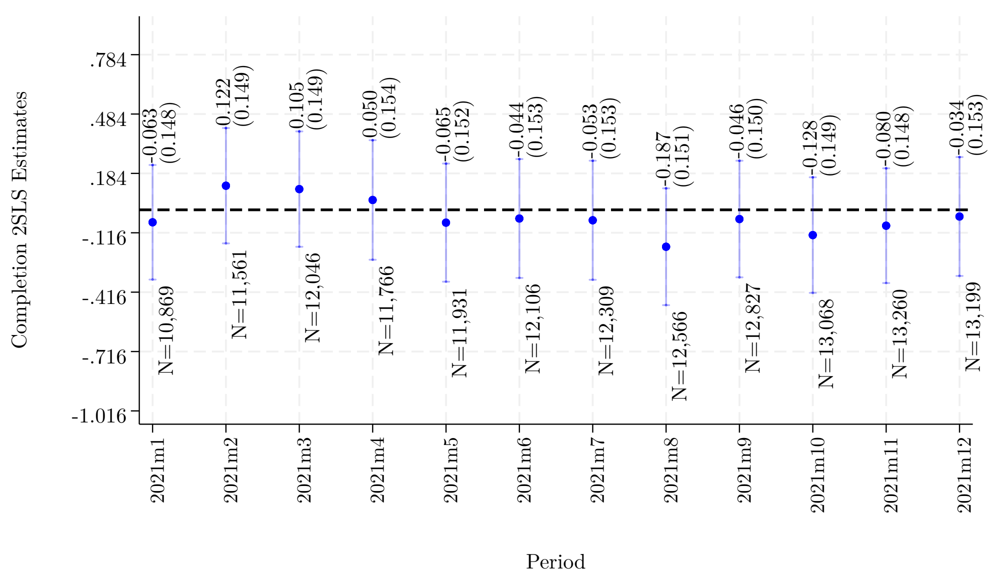
Note: This table reports the DiD estimates of equation 3 of course completion effects on formal employment. Column 1 reports the estimates for all participants, columns 2 and 3 classify participants by income level, and columns 4 and 5 by gender. Some observations are missing the income (20.3%) and the 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: Multiple-Instruments 2SLS Estimates of Completion on Labor Market Outcomes

A. Dependent Variable is Formal Employment



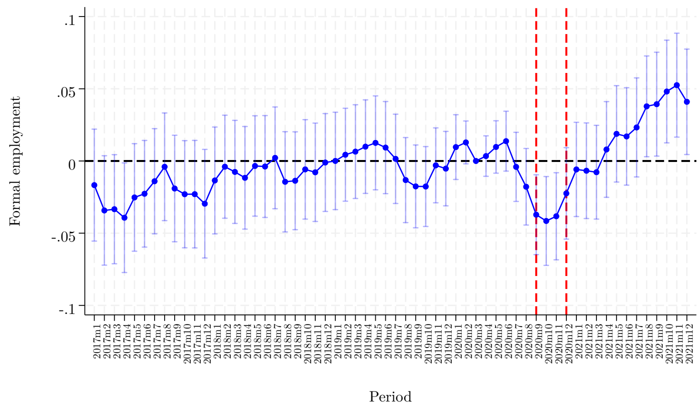
B. Dependent Variable is Daily Wages



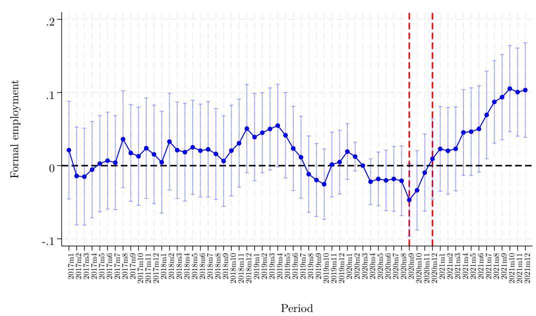
Note: This figure reports 2SLS effects of course completion on formal labor employment and wages using a multiple-instruments model with both rounds of the randomization instruments being the instruments for course completion. Panel A reports the estimates of equation 2a, the 2SLS estimates of completing 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: Estimates of Completion Effects on Formal Employment with Sample Restriction

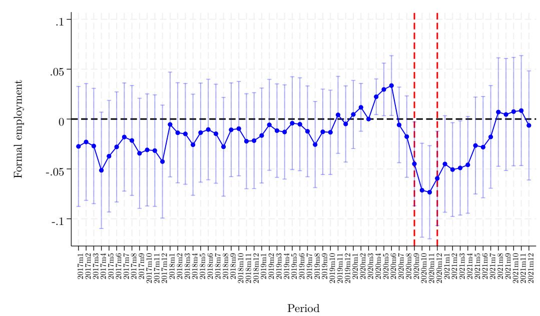
A. All participants



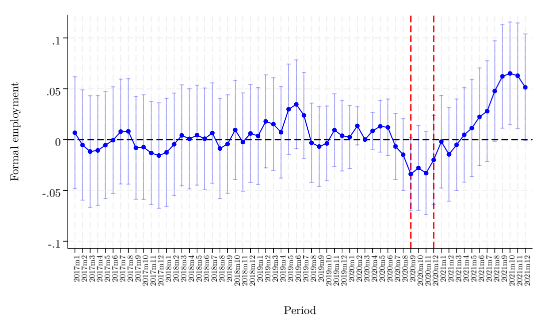
B. Low-income participants



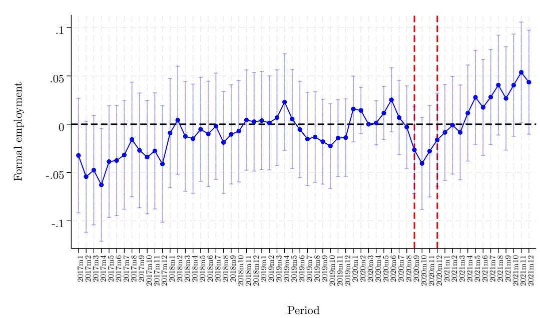
C. High-income participants



D. Women



E. Men



Note: This figure reports the event study estimates of equation 4 of course completion effects on formal employment restricting the sample to participants in the treatment group. 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.1. 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.