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

Stephanie Majerowicz Román Andrés Zárate

August 20, 2024

Abstract

We leverage an RCT to study the impact of offering free MOOC certificates during the pandemic on labor market outcomes. Despite the free certificates, the take-up rate is low, with only 6.2% of treated participants finishing at least one course. While positive, treatment effects on formal employment are not statistically significant, resulting in imprecise 2SLS estimates of course completion. An event study with more precise estimates reveals a significant average effect of 5.1% of completion on formal employment, with higher impacts for low-income participants. While MOOCs can improve employment outcomes, additional interventions are necessary to increase their 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 with their promise to revolutionize higher education by democratizing access to courses from elite universities. The *New York Times* called 2012 the year of the MOOC (Pappano, 2012), and despite some concerns about their financial sustainability (Hoxby, 2014; McPherson and Bacow, 2015), between 2012 and 2015, MOOC enrollments exceeded a staggering 25 million (René F. Kizilcec and Andrew J. Saltarelli and Justin Reich and Geoffrey L. Cohen , 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 the 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 certificates that validate this knowledge. While potential human capital gains from MOOCs usually only require a time investment since course auditing is free, validating that knowledge typically requires participants to pay a 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 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 labor market outcomes. We leverage a program implemented by one of the most prominent MOOC providers, Coursera, which offered free certificates to public organizations in Latin America during the pandemic. In collaboration with one of these entities, we undertook a Randomized Controlled Trial (RCT) of this initiative. Around 13,000 out of 21,000 participants were randomly selected to become eligible for free certificates upon completing MOOCs within three months.

One of the primary challenges in examining long-term impacts of MOOCs on employment

[1] The fees for the verified track of EdX 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 prices ranging from \$2,000 to \$5,000 USD.

outcomes is the ability to track participants in the labor market. 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 after the program. This tracking allows us to study the impact of MOOC certificates on labor outcomes as we track monthly participation in formal labor markets for all participants over five years.

Participants in the treatment group could earn free certificates by completing individual courses or specializations —a structured sequence of courses designed to provide comprehensive mastery over a specific subject—. In contrast, participants in the control group were not eligible to receive free certificates but could obtain them by paying a fee and fulfilling course requirements. Due to data privacy rules, we cannot monitor the platform activities of the control group. While it is unlikely that control participants earned a certificate due to the high fees and the low completion rates, if any did, our estimates would be a lower bound of MOOC certificate effects.

We first report treatment effects on take-up and completion rates. Consistent with existing evidence documenting the challenges of completing MOOCs (Banerjee and Duflo, 2014), our findings show that the course completion rates are low. Despite being eligible to receive free certificates of multiple courses, including specialization, only 50% of eligible beneficiaries enrolled in at least one MOOC, and only around 6.2% of all treated participants completed at least one course.

The low first stage in the take-up rate of the program translates into positive but small and non-significant treatment effects on formal labor employment, with a clear pattern of positive impacts starting six months after the end of the program. Being eligible for 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 MOOC certificates on formal labor employment using the treatment assignment as an instrument for course completion in a two-stage least squares (2SLS) framework. The LATE estimates reveal the impact on compliers (Imbens and Angrist, 1994): those who obtain the certificates due to the randomly assigned treatment. In a model with one-sided compliance, this LATE will also be equivalent to the average treatment effect on the treated (ATT). The results reveal that while the effects on formal employment are large in magnitude, between 2.2 to 9.5 p.p., they are imprecise and not statistically different from zero. Estimates on wages do not show a clear pattern and are also statistically indistinguishable from zero.

Motivated by the large but imprecise estimates 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 those who completed the courses were less likely to be formally employed during the program, consistent with individuals having more time to invest in completing the courses, from six months after the end of the program, there is a clear positive impact of 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 from the event study closely align with the 2SLS in magnitude but are more precise.

The DiD strategy also allows us to consider two different control groups to assess whether not observing control participants receiving certificates outside of the program could bias our estimates. In particular, we compare the DiD and event study estimates, including and excluding the control group from the estimating sample. If no one in the control group received a certificate outside of the program, we expect the estimates from these two samples to be similar.

The results show remarkably similar estimates in the two samples. The DiD excluding the control group 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). The similarity between the two estimates suggests that the data limitation of not observing course completion for participants in the control group has little impact on our estimates.

We also explore heterogeneity by income level and gender. Lower-income participants—characterized by a proxy means test used to target social welfare programs—benefit the most from MOOCs 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. (s.e. 2.2 p.p.) on formal employment. 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 gender.

Our findings highlight MOOCs' significant potential to enhance labor market outcomes, particularly for low-income individuals. Although our estimates reflect both skill acquisition and the signaling value of certificates, they provide meaningful insights as the evidence linking MOOCs to employment outcomes is scarce. The consistently low completion rates, even when certificates are free, also underscore the need for additional interventions that support and motivate learners to complete the courses.

Our study contributes to three branches of the literature. First, we contribute to the literature 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 (Bettinger et al., 2017; Kofoed et al., 2021; Bruhn et al., 2023; Jack et al., 2023; Goldhaber et al., 2023). Our study diverges from this literature as the relevant counterfactual for MOOCs rather than an in-person version of the same course is not accessing such educational content. While some recent evidence shows that online programs can increase enrollment (Goodman et al., 2019), our results suggest that online education can improve labor market outcomes by enabling individuals to access courses that otherwise would be unavailable.

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 course enrollment and completion.

Finally, we contribute to the literature on certifications of abilities in the labor market. While we cannot separately identify the effect of human capital from signaling, our positive effects on formal employment are consistent with the evidence of 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, our results suggest that MOOC certificates from prestigious institutions are valued in the Colombian labor market.

2 Setting and Intervention Description

One of the leading worldwide MOOCs platforms, Coursera, launched a job recovery initiative to mitigate the impact of COVID-19 on employment in Latin America. The initiative allowed public agencies across the region to apply for specific slots, allowing users to receive free certificates for completed courses on the platform. Each public agency had the autonomy to set the criteria for participation and eligibility. Prices for certificates range from $40 to $300 USD per course, and the initiative allowed obtaining them for no cost. In Colombia, for instance, $100 USD is approximately half the monthly minimum wage. Participation in this initiative allowed government agencies to inform people about the MOOC platform and its courses and offer them a chance to earn certificates, with these providing an additional incentive for course enrollment and completion.

Numerous public agencies from different Latin American nations participated in the initiative, each with its own eligibility and selection standards. We partnered with the Colombian Institute for Educational Credit and Technical Studies (Icetex), which offers higher-education loans and scholarships. The Icetex applied for 10,000 slots and targeted the program towards any current or former beneficiaries who had received a loan or a scholarship since 2010. Icetex promoted the program on its website and social media. Icetex received over 21,000 applications for its 10,000 slots and decided to allocate them through random assignment after advice from the research team.

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

Coursera encouraged all partner institutions to develop campaigns to promote course enrollment and completion. Icetex sent numerous emails encouraging participants to enroll and complete the courses. Icetex also promoted the most-demanded courses in the region to all eligible participants through emails highlighting specific courses. Coursera shared enrollment and completion reports with 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 re-randomizing to avoid chance imbalances, following Banerjee et al. (2017). The randomization was run 100 times, and balance checks were performed on 36 variables, using a max-min p-value criteria.

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

We combine different administrative data sets to characterize the participants, perform balance checks, measure MOOC enrollment and completion, and track participants' entry into the labor market. The project comprises five main data sets.

First, we use the program registry form. When participants registered for the program, they had to fill out a form accepting 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. The data contains information about participants' sociodemographic characteristics, such as age, gender, and socioeconomic level. To characterize individuals' socioeconomic status, we use the SISBEN level when they apply for a loan or scholarship. 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 (Camacho and Conover, 2011). We classify participants in the first bracket, which determines eligibility for most social programs in Colombia, as low-income.

The third data set, the *Saber 11* scores, is 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). The proportion of students 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 program's terms and conditions, this data is unavailable for participants in the control group.

The final data set allows us to track participants in formal labor markets, linking MOOC enrollment and completion with employment and wages. 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 data tracks changes in the participant's employment status, sector, and wages and observes their formal employment history before the intervention. As this data set only reports formal employment, we cannot distinguish whether not-formally employed participants 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 product 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, 29.5 years old, and 36% are male. Regarding their 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 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.

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 part of the original balance checks. Overall, we find that both groups are balanced at baseline on many characteristics. There is only a small imbalance in one out of 24 variables, and the conventional p-value on the joint F-test is 0.83.

4 Empirical Strategy

In this section, we present the empirical strategy we follow to estimate the impact of 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 also the reduced-form 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:

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

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

where 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 the parameter of interest being γ: the ATT of free certificates on labor market outcomes. The 2SLS model usually estimates the LATE on the compliers. In this case, however, it also coincides with the ATT due to one-sided compliance, as no participants in the control group could receive a free certificate. 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 eligibility control group, we cannot interpret the parameter γ as the ATT of MOOC certificates on labor market outcomes. However, if no one in the control group received a certificate during this period, a likely scenario as discussed above, the parameter γ in equation 2 would be equivalent to the ATT of MOOC certificates on labor market outcomes.

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, we observe monthly formal labor market outcomes of the participants, and 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 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 after accounting for the difference between these two groups in the pre-period.

We also extend equation 3 to 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 parameter of interest is 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.

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 use two different samples to test whether control group participants earning the certificates outside the program could be biasing our estimates. In particular, we can compare the DiD estimates with and without the eligibility control participants for whom we cannot track activity on Coursera. As mentioned earlier, if control group participants paid for the certificates, we would underestimate the effect of MOOC certificates on labor market outcomes, and the estimates from the two samples would differ. On the other hand, similar estimates from the two samples would suggest that it is unlikely that control participants earned certificates outside the program and, hence, that the data limitation due to the program's condition has little influence in biasing our estimates.

A final concern when estimating equations 1 to 4 on daily wages is that we only observe wages for those who 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 participant's daily wage. We refrain from estimating effects on wages when we find significant effects on formal employment. While we could estimate a Heckman selection model, such a 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 3.1 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. 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). 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 MOOC certificates effects on formal labor employment, the parameter γ in equation 2a. Each marker represents a different regression as we separately estimate the impact on formal employment at each period after the program. 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. Appendix Table A.1 reports the same estimates in a table format.

While the results in Panel A are inconclusive and non-significant 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 July to 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.2 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.

As there are no significant effects on formal employment, reducing any concern about differential attrition rates in the likelihood of observing wages, Appendix Figure A.1 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 for 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 exact estimates in a table format. Some 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 no 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 the event study. Panel A of Figure 2 reports the estimates of equation 4 on formal employment for all participants. Column 1 of Appendix Table A.3 reports the analogous estimates of the DiD model (equation 3); Panel I.A pools together all the pre-periods, while Panel II.A 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 2 provide promising evidence of MOOC certificates' impact on 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 2 (Panel A) indicate that after the program's end, starting in January 2021, individuals who completed the courses increase the probability of being formally employed. Notably, the pattern illustrated in the event study figure closely mirrors the 2SLS estimates of completion detailed in Figure 1, with the event study estimates 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 from August 2021 to December 2021, evident in Panel A of Figure 2, with a statistically significant impact of approximately 5 p.p. The formal employment rate for participants in the control group in December 2021 is 63.7%, so this coefficient represents an increase ranging from 5.1% to 7.8% in formal employment.

The event study also allows us to assess the bias from not observing the control group course completion by restricting the sample to treated participants for whom we can track their activity on the platform. Appendix Figure A.3 reports the estimates restricting the sample to participants in the treatment group, 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 I.B of Appendix table A.3 confirms this, showing an average post effect of 3.5 p.p. (s.e. 0.013), which is reasonably comparable to the general estimate of 3.3 p.p. Likewise, comparing the estimates in Panel II.A vs. II.B shows remarkably similar estimates. This comparison suggests that not observing control group participation has little influence in biasing our estimates.

Finally, we explore heterogeneous treatment effects by income level and gender. Panels B and C in Figure 2 report the event study estimates for low- and high-income participants, respectively. Our findings indicate that low-income individuals benefit more from MOOC certificates than their higher-income counterparts. While the estimates for high-income participants hover around zero with high precision, low-income participants experience effects roughly twice the average effect in Panel A. One year after the program's conclusion, low-income participants have a statistically significant gain of approximately 10 p.p. in formal employment from earning the certificates. Notably, the negative trajectory in formal employment observed during the program period, from October to December 2020, primarily comes from high-income participants, even though they do not ultimately obtain the benefits. Columns 2 and 3 in Panel I of Appendix Table A.3 confirm this result, with an average post-treatment 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 with this difference statistically different from zero (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. Panels D and E of Figure 2 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, with no statistically significant difference between the two groups.

6 Conclusion

This paper explores the impact of MOOC certificates on labor market outcomes by leveraging an RCT of a program implemented by Coursera during the pandemic. 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 linking the program registration records with administrative data on formal labor market outcomes in Colombia. 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 during the three-month program period.

We use the time variation in the data with an event study to improve the precision of the estimates. The results show positive and statistically significant effects in the ballpark of the 2SLS estimates. This compelling evidence suggests that MOOCs can improve labor market outcomes, with low-income participants benefiting the most from course completion. The event study also allows us to assess the bias from data limitations by restricting the sample to treated participants for whom we can track all activity on the platform. The results show similar estimates regardless of the estimating sample, which suggests that our inability to track the platform activity of the control group participants is not severely biasing our estimates.

While our study provides encouraging evidence on the benefits of MOOC completion for labor market outcomes, several questions linger, which we expect to explore in future research. First, we need a deeper understanding of specific course effects to study how the course field mediates the observed positive impacts. Our estimates also 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.

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.

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.

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.

René F. Kizilcec and Andrew J. Saltarelli and Justin Reich and Geoffrey L. Cohen (2017). Closing Global Achievement Gaps in MOOCs. *Science*, 355(6322):251–252.

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; * $p < 0.10$, ** $p<0.05$, *** $p < 0.01$.

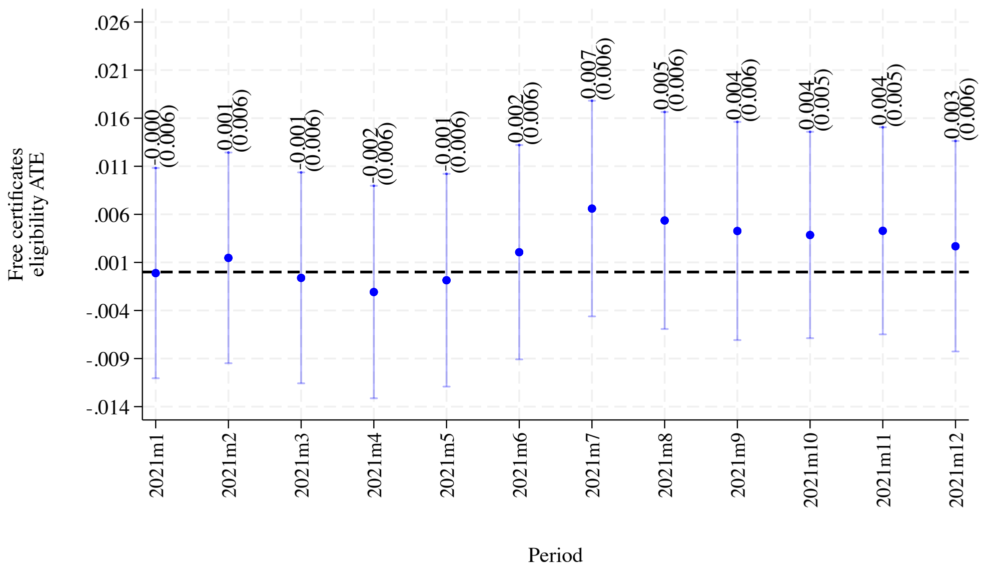
Table 2: First Stage on Course Enrollment and Completion

	Enrollment		Completion	
	Indicator	Courses	Indicator	Courses
	(1)	(2)	(3)	(4)
A. Eligibility for free certificates				
Treated	0.544***	3.100***	0.062***	0.130***
	(0.004)	(0.105)	(0.002)	(0.008)
Control mean	0.00	0.00	0.00	0.00
F-stat	15,495.57	867.70	859.25	282.38
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)
Control mean	0.00	0.00	0.00	0.00
F-stat	7,755.17	446.71	429.73	148.11
p-value treat 1st= treat 2nd	0.014	0.306	0.047	0.074
N	21,675	21,675	21,675	21,675

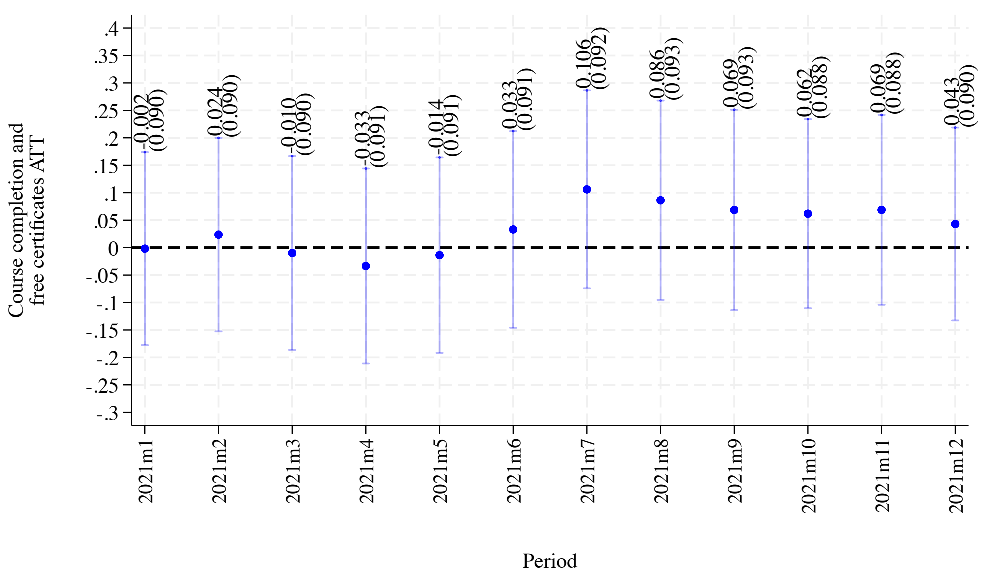
Note: This table reports treatment effects of free certificates eligibility on MOOCs' enrollment and completion. Panel A reports the estimates pooling together in one group the applicants who receive a 1st- and 2nd-round offer, while Panel B splits the two groups. Robust standard errors are reported 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 Effects



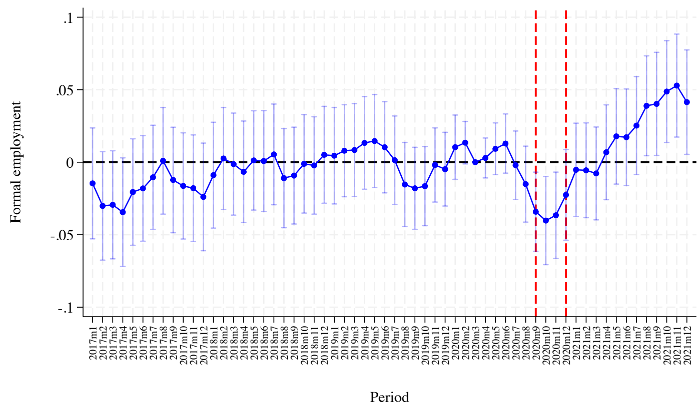
B. 2SLS Estimates



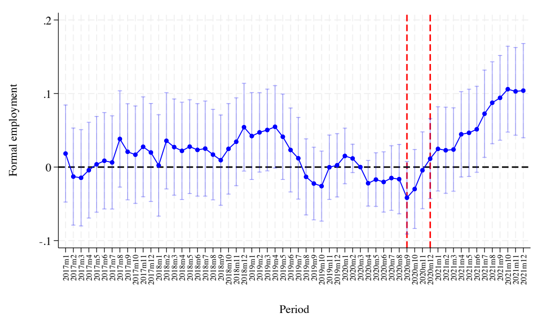
Note: This figure reports the treatment effects of free certificate eligibility and MOOCs certificates on formal labor employment for each period after the intervention. Panel A reports the estimates of equation 1, the effects of being eligible for free certificates. Panel B reports the estimates of 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: Event Study Certificates Effects on Formal Employment

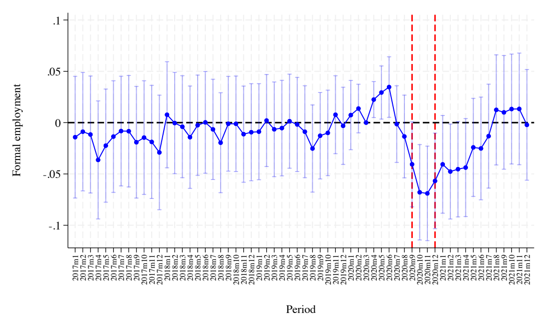
A. All participants



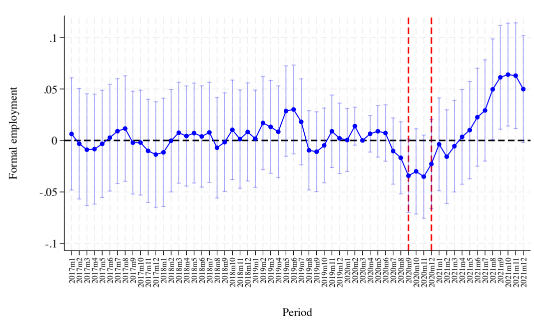
B. Low-income participants



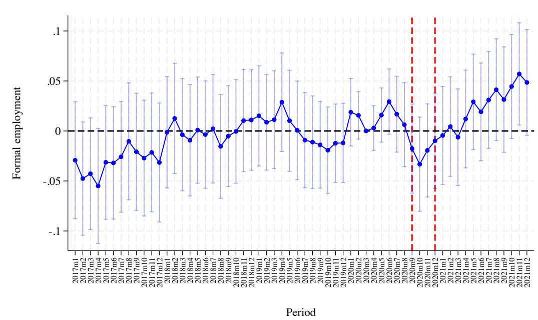
C. High-income participants



D. Women



E. Men



Note: This figure reports the event study estimates of equation 4 of MOOCs certificates 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 mean	ITT (eligibility)		2SLS free certificates		
		estimate	s.e.	estimate	s.e.	F-stat FS
	(1)	(2)	(3)	(4)	(5)	(6)
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 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	ITT (eligibility) s.e.	2SLS free certificates estimate	2SLS free certificates s.e.	2SLS free certificates F-stat FS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
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 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.

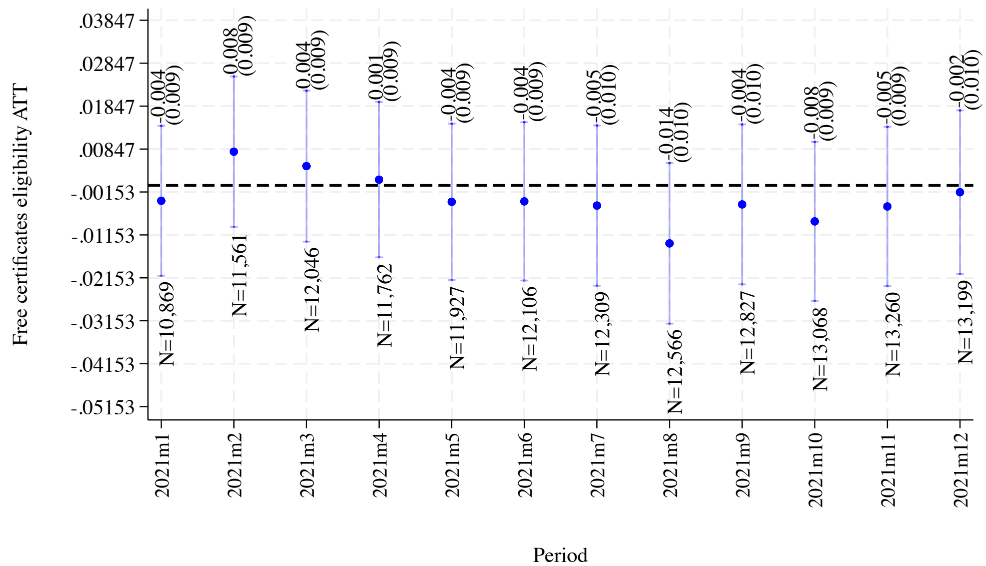
Table A.3: DiD Certificates Effects on Formal Employment

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 (p-value)		0.034		0.795	
N	1,300,500	519,660	516,600	770,400	470,100
B. Treated participants					
Completed x post	0.035***	0.054**	-0.003	0.029	0.039**
	(0.013)	(0.022)	(0.020)	(0.019)	(0.018)
Control mean	0.44	0.47	0.50	0.48	0.51
Equal effects (p-value)		0.054		0.705	
N	779,280	313,500	304,800	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 (p-value)		0.075		0.676	
N	1,300,500	519,660	516,600	770,400	470,100
B. Treated participants					
Completed x pandemic	0.004	-0.041*	0.021	-0.007	0.020
	(0.013)	(0.023)	(0.018)	(0.019)	(0.019)
Completed x during	-0.024	-0.027	-0.051**	-0.029	-0.011
	(0.015)	(0.026)	(0.022)	(0.021)	(0.022)
Completed x post	0.034**	0.047*	-0.003	0.026	0.041**
	(0.014)	(0.024)	(0.021)	(0.021)	(0.020)
Control mean	0.44	0.47	0.50	0.48	0.51
Equal effects (p-value)		0.118		0.618	
N	779,280	313,500	304,800	458,880	284,520

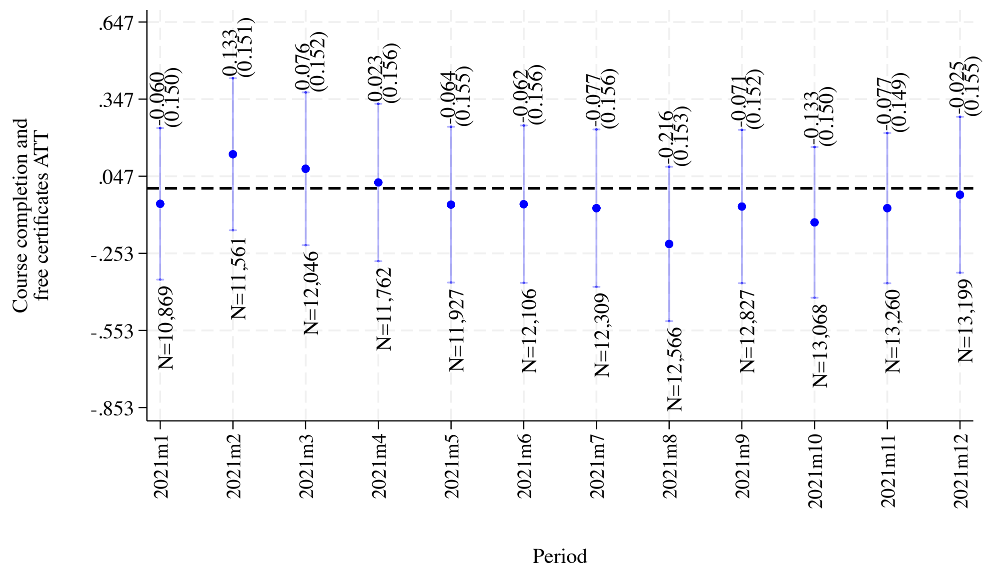
Note: This table reports the DiD estimates of equation 3. As we can only track platform activity for treated participants, Panels A and B report the ATT of certificates including and excluding control participants, respectively. Column 1 reports the estimates for all participants, columns 2 and 3 classify 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: Reduced Form and 2SLS Effects on Daily Wages

A. Reduced Form Effects



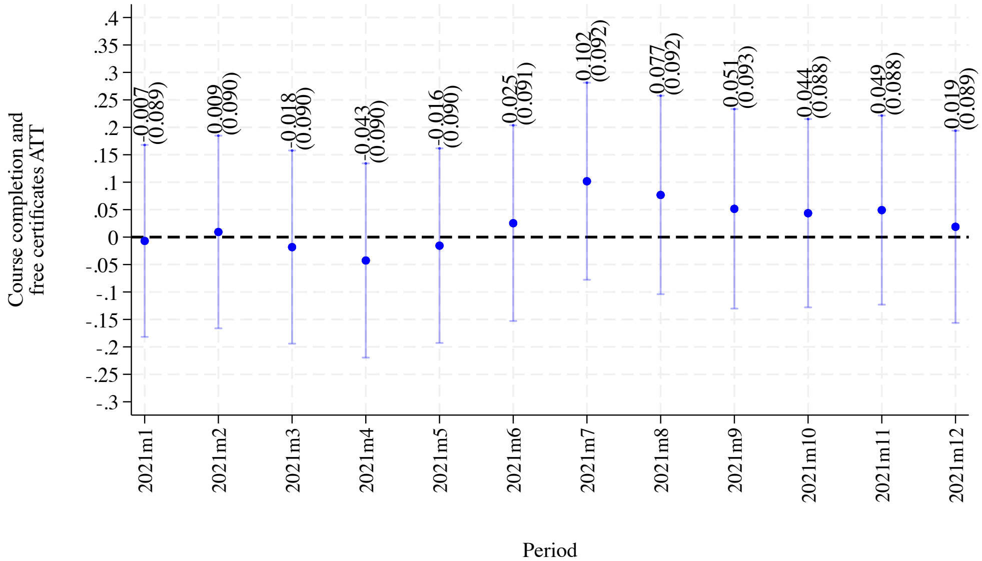
B. 2SLS Estimates



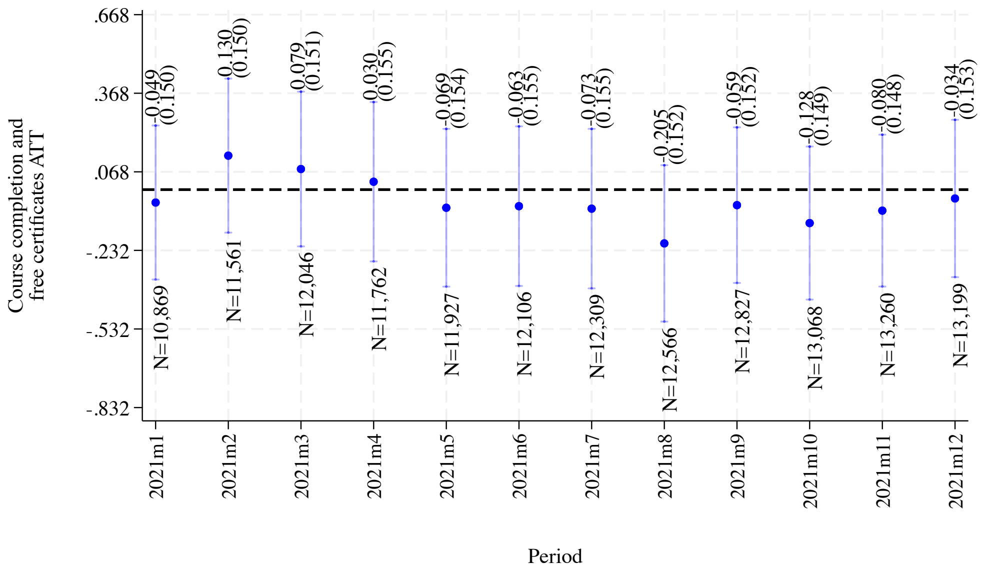
Note: This figure reports the treatment effects of free certificate eligibility and free certificates on daily wages for each period after the intervention. Panel A reports the estimates of equation 1, the effects of being eligible for free certificates. Panel B reports the estimates of equation 2a, 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 A.2: 2SLS Effects with Multiple Instruments

A. Formal employment



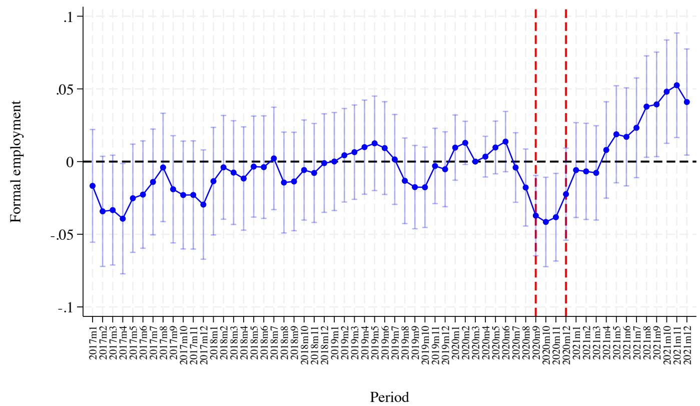
B. Daily wages



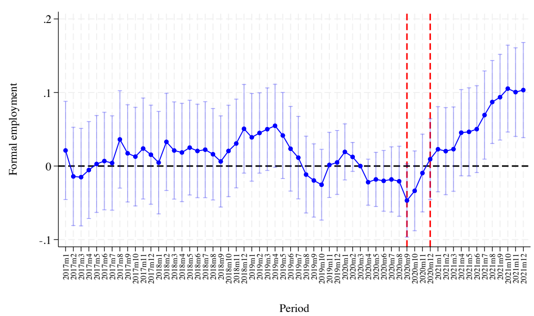
Note: This figure reports 2SLS effects of course completion on formal labor employment and wages using a multiple-instruments model with both rounds of the randomization as instruments for course completion. Panel A reports the estimates of equation 2a, the 2SLS estimates of completing at least one course 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.3: Event Study Effects on Formal Employment with Restricted Sample

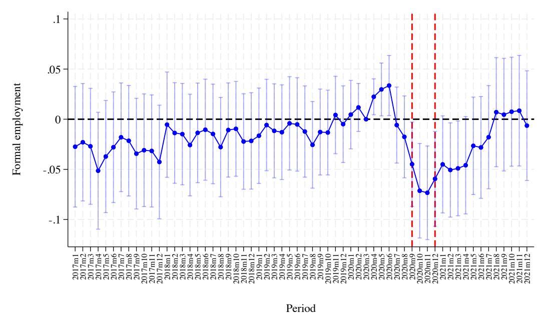
A. All participants



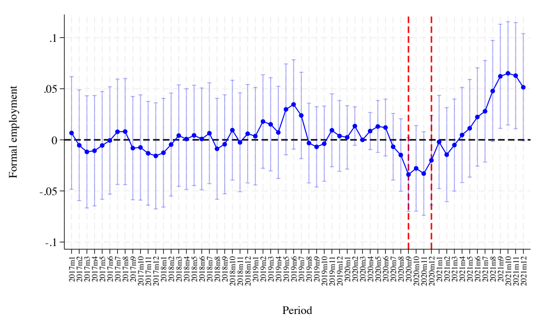
B. Low-income participants



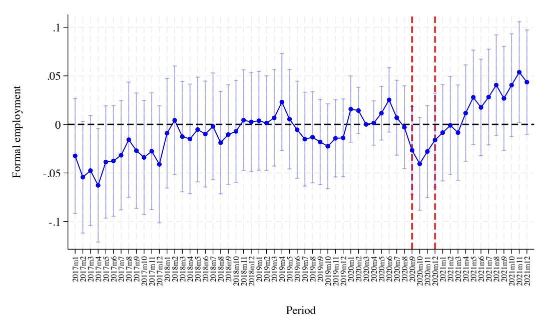
C. High-income participants



D. Women



E. Men



Note: This figure reports the event study estimates of equation 4 of MOOC certificates on formal employment 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.