# AI Training for Online Entrepreneurs*

## Evidence from a Large E-Commerce Experiment

Yizhou Jin and Zhengyun Sun

First Version: January 20, 2021
This Version: March 10, 2026

**Abstract**

We conducted a year-long experiment among millions of new sellers on an e-commerce platform. The treatment group received access to a free, algorithmically customized entrepreneur training program. With a 24% take-up rate, training raised seller revenue (1.7% ITT and 6.6% ATT) largely by inducing advertising adoption. Logit demand estimation on consideration-set data suggests that new sellers have 6.4% higher unobserved quality than incumbents, and that removing the program would reduce consumer surplus by 0.07%, despite new sellers making up only 0.25% of products in consideration sets. This is the largest-scale experiment and the first measurement of consumer welfare effects in the business training literature.

*Keywords:* business training, entrepreneurship, platform, consumer welfare
*JEL codes:* O12, L26, M53

*Jin (corresponding author): University of Toronto, yizhou.jin@utoronto.ca; Sun: Uber, zysun92@gmail.com. We thank Susan Athey, Jie Bai, Emily Breza, Long Chen, Shawn Cole, Jean-Etienne de Bettignies, Solène Delecourt, Avi Goldfarb, Asim Khwaja, Gabriel Kreindler, Michael Kremer, David McKenzie, Song Ma, Ariel Pakes, Gautam Rao, Yulu Tang, Yuan Tian, Joseph Stiglitz, Christopher Woodruff, and various seminar participants. We also thank the product development team at the unnamed e-commerce platform, and our data contacts and providers for making this project possible. The ability to publish is not contingent on results. Please email carolina_harvey@harvard.edu for questions about the data usage agreements. The trial was registered on the AEA Social Science Registry (AEARCTR-0006725). Human Subjects Approval was obtained from Harvard University (IRB19-0647).

# 1 Introduction

Much of today's entrepreneurship occurs on large digital platforms (Denes et al. 2022). In 2020, the last year these statistics are publicly available, Amazon marketplace saw 1.3 million new sellers, Alibaba's Taobao and TMall 6.8 million, and YouTube seven million new channels.[1] Platforms such as these offer entrepreneurs vast market access at low entry costs, but the resulting congestion has chocked the exploration of high-quality entrants (Bai et al. 2025). Jin and Sun (2026) show that, in 2019, only 0.6% of new sellers on a major e-commerce platform ever became viable businesses.

A widely adopted intervention is entrepreneur training. Nearly every major platform now operates one: Amazon Seller University, Alibaba's Taobao University, eBay's Seller Training Program, TikTok Shop Academy, Shopify Academy, and YouTube Creator Academy, among others. These programs historically shared a common template of curated tutorials on platform rules and tools for pricing, advertising, and customer service, delivered online to entrepreneurs at near-zero cost. Recently, several platforms introduced AI-powered services that analyze individual shop data and deliver personalized guidance.[2] However, whether such training can effectively promote entrant growth *at a large scale* is empirically untested.

In this paper, we experimentally evaluate an entrepreneur training program on a large e-commerce platform. In May 2019, we launched a randomized controlled trial among new sellers on the platform. Treatment group sellers received access to a customized training program: upon opt-in, a machine-learning algorithm analyzed each seller's product and operating statistics in real time and assigned relevant training materials sourced from public forums and tutorials. The program had negligible marginal cost and was offered for free. The trial lasted thirteen months, during which over two million new sellers gained access to the program, of whom 24% participated. We match the experimental variation with the platform's administrative data on sellers' product and operating statistics as well as consumers' search, browsing, and purchasing records, allowing us to construct seller-level and consideration-set-level panel datasets.

Training raised seller revenue by 1.7% (ITT) and 6.6% (ATT). The gain was driven almost entirely by advertising adoption. Pricing and product descriptions showed precise null effects. Customer service metrics improved, but organic traffic and conversion rates did not respond. Ad traffic instead surged in the first three months, when the vast majority of training occurred, and organic gains followed. This sequence suggests advertising-driven sampling built the track record needed for algorithmic visibility.

Treatment effects are largest among sellers just breaking into positive revenue (those at the 80th–90th percentile of the monthly revenue distribution) and become undetectable at the very top, indicating that the top-selling entrants likely would have succeeded regardless. Nonetheless, training participation is a strong predictor of success. Only 1.3% of new market entrants ever reach the top 10% of their sector's revenue distribution. Training participation in month 1 predicts this status as strongly as first-month log revenues,

[1] Statistics from Marketplace Pulse and Socialblade (only YouTube channels with over 5 subscribers are counted).

[2] Amazon Seller University and Shopify Academy still follow the curate-and-recommend model (video libraries and structured courses). TikTok Shop's Seller Assistant (launched 2024) analyzes each shop's activity to surface context-aware advice; Alibaba's AI Business Advisor generates tailored recommendations on pricing and inventory from seller data. There have been attempts to incorporate large-language-models into these training programs. eBay's AI Activate program grants sellers ChatGPT access, and Alibaba's AIGC suite generates and refines product images and videos. However, the unit economics remain relatively high so that such programs have not been offered at scale.

and its predictive power grows over time while early performance metrics fade. This pattern is consistent with participation screening for latent entrepreneurial ability before outcomes are observable. Does training benefit consumers? The entire business training literature has measured only supply-side outcomes; no prior study has estimated consumer welfare effects. The answer is not obvious: even if sellers earn more revenue, the resulting reallocation of consumer attention could displace higher-quality incumbents rather than expand the market. To address this, we estimate a logit demand model on consideration-set data, recovering sellers' latent product quality. Structural estimates suggest that new sellers have 6.4% higher unobserved quality than incumbents. The welfare gain operates through a reallocation channel: training makes high-quality new sellers more visible by helping them overcome the cold-start barrier. Removing the program would reduce consumer surplus by an estimated 0.07%, despite treated sellers comprising only 0.25% of products in consideration sets.

**Related Literature** This study contributes to a large experimental literature on business training. Recent meta-analyses demonstrate that training raises profits by 5–10%, overcoming earlier skepticism driven by underpowered studies (McKenzie and Woodruff 2014; McKenzie 2021).[3] However, most existing programs are offline, expensive to deliver, and difficult to scale or customize (Bloom et al. 2013; Bruhn et al. 2018). No prior study measures consumer welfare effects of entrepreneur training. Platform-based studies of seller guidance involve analytics dashboards or performance feedback rather than training, and none approaches the scale of this study or estimates consumer surplus. McKenzie and Puerto (2021) finds market expansion from training but does not estimate consumer surplus; the closest attempt is Atkin et al. (2017), which documents quality upgrading from export market access.

Since our experiment in 2019, the meaning of "AI" in this context has shifted, from the supervised-learning algorithms that match content to seller needs (as in our setting) to generative large language models that produce advice on demand. Otis et al. (2024) conduct the first RCT of this newer form, providing 640 Kenyan entrepreneurs with a GPT-4 business mentor, and find gains concentrated among high performers and losses among low performers, a pattern similar to ours. Together, these results suggest that training's dual role in developing skills and screening for latent ability persists across technologies, scales, and settings.

We address these gaps along three dimensions. First, ours is the first large-scale RCT of in-platform training, reaching over two million sellers at near-zero marginal cost ("experimenting at scale," Muralidharan and Niehaus 2017). Second, we show that advertising adoption is the high-return, easily nudgeable margin for platform entrants, paralleling Anderson et al. (2018)'s finding that marketing skills dominate finance skills offline and bringing a classic debate about advertising as an information barrier to entry (Shapiro 1982) to digital platforms. Third, logit demand estimation on consideration-set data provides, to our knowledge, the first estimate of how any business training intervention affects consumer welfare.

Section 2 describes the platform, the training program, and the experimental design. Section 3 presents treatment effects on seller outcomes, decomposes the mechanism through seller strategies and survey evidence, and examines heterogeneity including quantile treatment effects and the screening device. Section 4 estimates a logit demand model on consideration-set data to quantify consumer welfare effects. Section 5 concludes.

[3] Important innovations in program design include psychology-based personal initiative training (Campos et al. 2017), simplified rule-of-thumb accounting (Drexler et al. 2014), mentorship (Brooks et al. 2018), and peer networks (Cai and Szeidl 2018).

## 2 Background and Data

In this section, we introduce the setting of our study, our administrative data, and our experiment. Section 2.1 provides background information on the e-commerce platform, focusing on key platform rules for new sellers and descriptive evidence of new-business growth. Section 2.2 introduces the algorithmically customized entrepreneur training program. Lastly, Section 2.3 details our experimental design and the dissemination of the training program.

### 2.1 E-commerce: the Platform and the Sellers

Our research setting is a leading e-commerce platform in China. It hosts millions of active and independent sellers and hundreds of millions of consumers. In 2019, sales facilitated by the platform accounted for a majority of all e-commerce retail in the country.

Most sellers on the platform are third-party retailers. The platform does not directly sell to consumers in almost all sectors, with some notable exceptions such as fresh groceries and financial services. Our analysis excludes these sectors and features those in which the platform only intermediates—and does not compete with—sellers. In this setting, platform revenue primarily comes from in-platform advertising (approximately 80%), with the remaining share made up by sales commissions for large sellers and seller services such as return shipping insurance and advanced data analytics (Jin and Sun 2019).

**Platform matching and seller strategy** Almost all traffic to new sellers is intermediated by the platform: consumers arrive at the platform and search for a desired product; the platform then dissects the search query into keywords and ranks a set of related products and sellers. Sellers can boost their rankings organically or through in-platform advertising, both of which are covered by the training program. Organic traffic strategies aim to accumulate past sales and positive reviews through improved product listings, customer service, or discounts. In-platform advertising, primarily in the form of position auctions as analyzed by Edelman et al. (2007), Chen and He (2011), and Athey and Ellison (2011), is prevalent. The training program only covers basic introductions to the position auction mechanism and techniques for new sellers to get started.

Besides search, some established sellers can obtain a significant amount of traffic from recommendations: algorithms that the platform uses to directly recommend products to consumers based on their past consumption behavior. However, empirically, it is negligible for new sellers due to a well-known "cold-start" problem (Sun et al. 2021).

**Entry** Prospective sellers must undergo a rigorous verification process to register on the platform. A national ID or a business registry number is required and cross-referenced against the national digital identity database. The platform also collects a small deposit to further deter spam. Once registered, a seller is considered a **potential entrant**. In 2019, there were about 25,000 potential entrants per day (Jin and Sun 2026). One-third of them successfully listed at least one product and had non-zero traffic.[4] Further, about

[4]All new sellers who successfully listed products received nonzero traffic from the platform. However, listings must follow various platform rules. For example, detailed product descriptions and pricing must be provided; high-resolution and non-duplicative pictures were also required. In most cases, these rules were enforced with AI algorithms with the option to appeal to a platform representative.

half of these sellers (or 4,200 per day) ended up earning non-zero revenue. We call these new sellers **market entrants**. Nearly all new sellers (97%) are sole proprietors who view e-commerce as their primary occupation.

The cold-start problem is severe. Revenue distribution among new sellers is highly skewed: accounting for industry variation, a seller at the 95th percentile earns 70 times the revenue of the median seller. In aggregate, the approximately 700,000 new sellers in our sample combined accounted for roughly 1% of total platform sales during the sample period.

**Industry dynamics** Although it is easy to register as a seller on the platform, effective market entry remains difficult. During our research window, almost all sellers who posted products received some initial traffic from the platform in the first month, but only 50.2% completed a sale. Moreover, many sellers exited the platform after a month. 89.7% of sellers had visitors during the first month, but only 42.2% still did during the third month. For the average cohort of new sellers (with at least one product listing),[5] 0.05% and 1.7% of them ranked among the top 1% and 10% by traffic in their respective retail sectors in the first month of operation. At the end of the first year, growth among these new sellers becomes more divergent and concentrated at the top: while 0.08% of them are now placed in the top 1%, only 1.3% of are placed in the top 10%.

**Administrative data** Our analyses rely on the platform's administrative data. For each product, we observe listing and transaction prices and an indicator for whether the product is advertised through position auctions. We also observe the amount of traffic and revenues the product receives as well as its conversion rate and average customer rating. These statistics are typically aggregated to the seller level and then combined with other seller-level data, such as some basic characteristics (registration type, geographic locations, etc.), training records, ratings, refund frequencies, as well as customer service statistics.

We also obtain data on consumers' search, browsing, and purchasing behaviors, which is used to construct consideration sets. For each search session, we observe the corresponding consumer ID, the keywords associated with the search query, the set of products and sellers the consumer visited as well as the product or products she ends up buying, if any. By linking consumer and product IDs over time, we observe subsequent requests for returns, refunds, and repeat purchases.

## 2.2 An Algorithmically Customized Entrepreneur Training Program

We collaborated with the platform to develop and implement an algorithmically customized entrepreneur training program. We first sourced pre-existing training materials from official or high-quality public tutorials. We then selected a subset of materials in which the associated strategies and outcomes could be tracked by a quantitative key performance indicator ("KPI"). These materials were then parsed into a sequence of tasks using a proprietary machine-learning algorithm. Each task corresponded to a specific KPI threshold, often benchmarked to the sectoral average among incumbent sellers. Participants who did not meet the threshold were assigned the task with a suggested completion time frame based on difficulty level. Lastly, the platform developed an intuitive user interface that allowed participants to track their KPIs and view the detailed training materials. For example, as shown in Figure 2, one task asked sellers to post short-video product demos. The KPI used to identify sellers in need of this training was "the proportion of products with

[5] New-seller cohorts are created using the calendar month of sellers' first product listing.

short-video demo" for a seller. On the task detail page, sellers could find relevant tutorials and track the progress of their KPI. The KPI target used to judge task completion was shown and explained as the "sector average".

**Content of training** Training tasks covered four main areas: basic operations, ratings, marketing, and customer service. Table 1 provides representative examples; Appendix Table C3 provides the full list with classifications. The first type helped sellers effectively price and position their products. For example, some tasks introduced tools to include media in product descriptions and ways to offer quantity or repeat-purchase discounts.

Rating tasks introduced the consumer rating and review mechanisms. For example, several tasks introduced how seller rating is calculated from sub-components representing service, logistics, and product ratings. The corresponding KPIs were associated with improving each of the three components.

Marketing tasks taught sellers advertising techniques to attract visitors and increase site traffic. Sellers were nudged to choose keywords strategically and adopt automated bidding in the platform's pay-per-click position auctions. At a minimum, a new seller needed only to set a budget, however small it is, and the system would automatically analyze its product descriptions and labels to determine the keywords to bid on and the bid amount. If desired, the seller could launch a campaign that with tailored bidding strategies such as targeting certain consumer segments or search sessions at a certain time of day. It could also choose different keywords so as to highlight differentiation and avoid congestion.

The last group of tasks aimed to improve the quality and efficiency of customer service. Most salient was a series of introductions to an AI-based chatbot. They outlined simple procedures to set up automated responses as well as the appropriate use cases, such as responding to inquiries about business hours and delivery status.

Table 1: Examples of Training Tasks

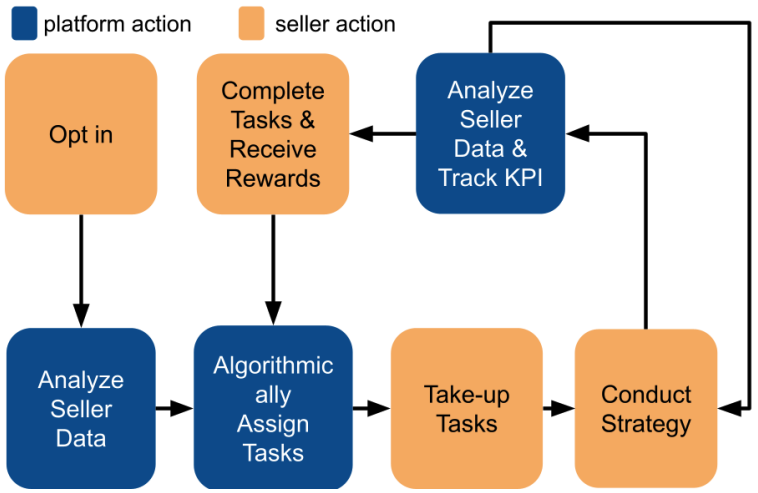
| **Task** | **KPI** | **Area** |
|---|---|---|
| Complete an order | payment received | basic |
| Offer product-level discounts | using discount tool on 1+ products | basic |
| Setup seller-wide coupons | use of "red-pocket" coupons | basic |
| Setup 7-day free return | free return status | basic |
| Optimize product titles | use of title-setting module | marketing |
| Set up advertising account | deposit funds into ads account | marketing |
| Set up automated ad campaign | keyword campaign and budget | marketing |
| Tips to improve delivery ratings | logistics component of seller rating | ratings |
| Tips to improve service ratings | service component of seller rating | ratings |
| Offer targeted coupons during inquiries | used discount tools in chat app | service |
| Set up chatbot automation | completed basic chatbot setup | service |

*Notes:* Representative examples of training tasks across four areas. Each task is associated with a measurable key performance indicator (KPI) used to trigger assignment and evaluate completion. See Appendix Table C3 for the full list of 35 tasks with classifications.

**Algorithmic customization and participation** Figure 1 illustrates the algorithmic pipeline. Tasks were assigned to each new seller based on a machine-learning algorithm. It used historical observational data to sequence all tasks available based on their predicted impact on new-seller revenues and survival. For each new seller participating in the training program, the algorithm then determined whether each task was relevant based on the seller's operations data, which resulted in a sequence of relevant tasks customized for that new seller.

This real-time and dynamic matching between sellers and tasks was made possible because each task is defined based on specific actions and quantifiable KPIs that were easily measurable by the platform, along with corresponding triggering and completion conditions that facilitated automation.

Figure 1: Training Program: Algorithmic Customization Pipeline



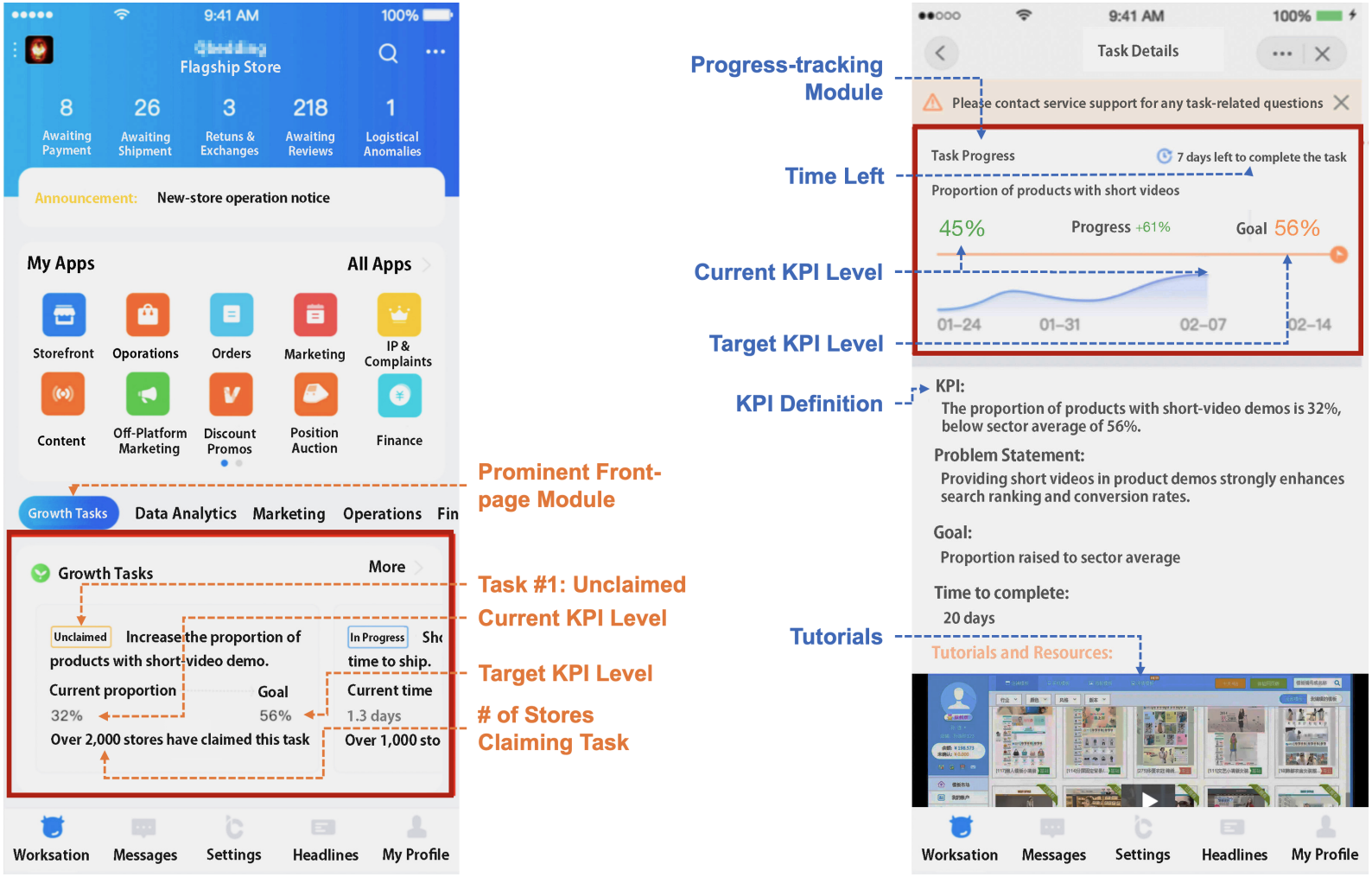

*Notes:* Flowchart illustrating the training program's algorithmic pipeline. Pre-existing training materials are mapped to measurable KPI thresholds, sequenced by a machine-learning algorithm based on predicted impact on new-seller revenue, and delivered to each participant based on their real-time operations data.

Sellers could choose to take up any number of tasks shown to them (Figure 2 panel (a)). Each task came with a suggested time frame for completion, detailed tutorials, and in some cases, live-streamed webinars to facilitate real-time interactions and Q&A (Figure 2 panel (b)). For each completed task, sellers earned a small reward in the form of a free trial to a supplementary service such as shipping-label printing. The monetary value of these rewards was estimated to be around $5 USD. However, sellers were unaware of the rewards before taking up their first tasks.

**Distribution, costs, and dependencies** The training program was launched on May 6, 2019 as part of our experiment outlined below. Over two million new sellers had accessed the training as of June 2020. The program was offered at no cost to new sellers. It was prominently featured on the platform's official and only seller-facing portal as a widget on the front page (Figure 2 panel (a)),[6] which guaranteed effective exposure to new sellers. It was also standalone from other platform operations: training participation and performance had no direct impact on search ranking or financing.

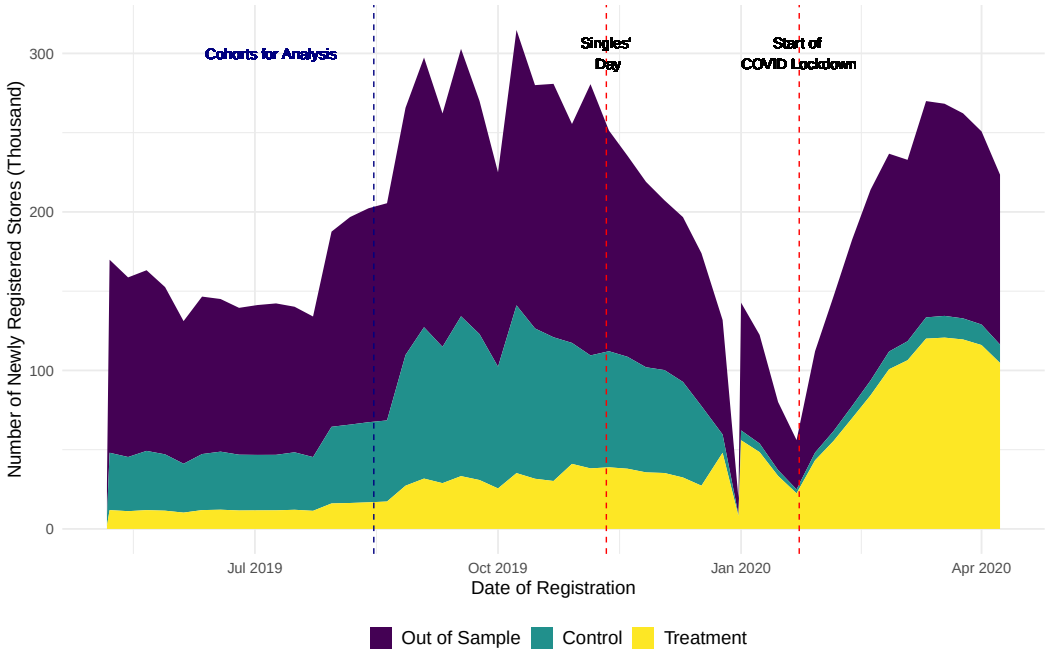
[6]The app was available on all major mobile and web operating systems.

Figure 2: Illustrations: Training Module and Tasks on the Seller's Portal App




(a) (b)

Figure 3: Experimental Sample and Treatment Assignment Over Time




*Notes:* This figure plots the training program roll-out time series and the number of affected sellers as a share of total new sellers entering the platform. Per company protocol, some sellers are randomly selected into our experiment and data sample. Among these sellers, we randomly select a treatment group. In order to construct a panel dataset, our empirical analysis focuses on new sellers registered before Aug 15, 2019 (left of blue line).

## 2.3 Experimental Design and Implementation

We designed and implemented the training intervention as a controlled experiment that randomized *access* to the training program among new sellers as they completed registration. In Table C5 columns (1) - (3), we verify that the treatment assignment was balanced. We were unable to randomize at the task level because tasks shared complex interdependencies and were continuously updated, revised, and occasionally deleted as our machine-learning algorithm was re-trained and based on user feedback.

Our empirical analysis focuses on new sellers registered between May 6 and August 15, 2019 (the first phase of the formal roll-out) whom we follow for nine months after registration. This restriction, shown in Figure 3 (left of the blue line), ensures a sufficiently long panel for each cohort. About 25,000 new sellers registered every day in 2019. Per company protocol, a portion of them were randomly selected into our experiment and data sample. We randomized among these new sellers, initially giving 25% training access and later expanding to 35% on October 29 and to 90% on December 26.[7] By June 2020, over two million new sellers had gained access to the program.

Our experimental design is implicitly supported by the Stable Unit Treatment Values Assumption (SUTVA). It may be violated when treated new sellers directly compete with control-group ones. However, this is exceedingly rare in reality. In Section 4.3, we construct a consideration-set dataset that over-stratifies

[7]To a seller, there is no difference between being in the control group and being out-of-sample.

on consumers who have visited at least one new seller during our research window. Even in this sample, new sellers account for only 0.2% of consideration-set members, and fewer than 1% of consideration sets contained more than one new seller. Therefore, the treatment status of one new seller can reasonably be assumed to have had no impact on other new sellers. Ad-market congestion is also unlikely to be a concern: only 7% of new sellers ever advertised (Table 6), whereas the majority of established sellers advertise,[8] so the marginal increase in new-seller ad adoption induced by training is negligible relative to the platform-wide advertising market (Jin and Sun 2026).

### 2.4 Training Take-up

44.6% of treatment-group sellers ("treated" sellers) browsed the training in the first month after registration; 24.1% took up at least one task ("participants"), and 12.6% completed at least one task ("finishers") within our research window (nine months after registration). No control-group sellers took up any tasks. Appendix A.1 details the timing of training takeup among treated sellers.

There is significant variation along the intensive margin of training participation: the median participant took up 3 tasks, but 29.8% only took up one task, and 9.6% of sellers took up more than 10 tasks.[9] 52.3% of participants are finishers, averaging 1.63 completed tasks, but the top 5.5% participants completed more than 6 tasks. Across types of tasks, the "basic operations" tasks had the highest average completion rate at 42.9%; marketing and customer services tasks had lower rates, at 39.7% and 22.3%, respectively.

There is advantageous selection into training take-up, as shown in columns 4 to 6 of table C5. Among treated (or training-eligible) sellers, training participants were more likely to have come from economically developed southern coastal provinces, have had formal business registration, and have posted products immediately after registration. This advantageous selection foreshadows a key finding in Section 3.4: training participation serves as a screening signal of latent entrepreneurial ability. In evaluating the impact of training, we thus report both the intent-to-treat effect (ITT) and the average-treatment-effect-on-the-treated (ATT).

## 3 Treatment Effects

### 3.1 Main Effects

We construct seller-level panel data and evaluate the impact of training on seller growth.

**Data cleaning** As described in Section 2.3, our panel dataset follows a cohort of new sellers for nine months after their registration. Treatment assignment appears balanced across a range of seller characteristics collected during the registration (Table C5). We present both ITT (intent-to-treat effects) and ATT (average treatment effect on the treated). We start with all registered new sellers—potential entrants—since our treatment assignment was done at the time of registration, and because this effect can be directly scaled

[8] On Amazon, nearly three-quarters of brands and sellers use Sponsored Products advertising (Jungle Scout *Amazon Advertising Report*, 2024). Given that our platform derives an even larger share of revenue from advertising, incumbent participation is likely at least as high.

[9] See Figure C2 for the full distribution.

to obtain the aggregate training impact. We then focus on new sellers that have earned nonzero revenues—market entrants—because it better aligns with how entry is traditionally measured, and because the effect on having revenues is minute. To account for the considerable heterogeneity across product categories, we develop a sector variable based on the sellers' product portfolio.[10]

**Impact on seller outcomes** We first evaluate the overall intent-to-treat (ITT) effect of the training with the new seller panel. For seller $i$ who registered on date $c$ with affiliated sector $s$, and for (relative) month $m$, we run the following specification :

$$Y_{imcs} = \beta Treatment_i + \alpha_m + \alpha_c + \alpha_s + \varepsilon_{imcs} \tag{1}$$

$Treatment_i$ is an indicator for having access to the training, $\alpha_m$, $\alpha_c$ and $\alpha_s$ are month, cohort and sector fixed effects. Standard errors are clustered by seller.

Table 2: Intent-to-Treat Effects Among All Registered New Sellers (Potential Entrants)

| | Log Revenue | Has Revenue | Log Traffic | Log Customers |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treatment | 0.017*** | 0.002** | 0.013** | 0.008*** |
| | (0.006) | (0.001) | (0.006) | (0.003) |
| Dep Var Mean | – | 0.19 | – | 0.57 |
| Observations | | 6,409,062 | | |
| Adjusted $R^2$ | 0.132 | 0.153 | 0.207 | 0.106 |

*Notes:* $N = 6,409,062$. Dependent variables are aggregated to the seller-month level. Log transformation is done after adding one. "Has revenues" is an indicator for earning positive revenues in a month. All regressions include cohort, month, and sector fixed effect as specified by equation 1. Dependent variable means are calculated with sellers in the control group across all months. Standard errors are clustered on the seller level. ** significant at 5% level and *** at 1%. Raw revenue and traffic numbers are omitted (labeled as "-") at the data provider's request.

Table 2 presents the estimates among all registered new sellers (potential entrants). Access to training led to a 1.7% increase in revenue (column 1) and a 0.8% increase in survival likelihood (column 2) as measured by a seller having positive revenue in a given month.[11] In aggregate, the approximately 178,000 treated sellers in our study cohort earned 1.8 million RMB in higher revenues due to training during our nine-month research window. Assuming comparable effects for subsequent cohorts, the program's two million treated sellers would have generated approximately 20 million RMB in additional revenue. With near-zero marginal cost of delivery, the program's return on investment is substantial. For comparison, offline management consulting interventions cost $177 per participant on average in developing countries (McKenzie 2021), while our program reached over two million sellers at negligible per-participant cost. The 1.7% ITT is smaller

[10]We classify sellers into sectors using the platform's proprietary algorithm which relies on NLP techniques and sellers' product labels. For sellers with products across multiple sectors, we take the top-selling product for each seller. For sellers without sectors due to a lack of product listing and sellers selling unclassified products together as an "unknown" category.

[11]Table C6 presents the treatment effect on revenues in levels with different winsorization thresholds and with inverse hyperbolic transformation (column 3). All estimated effects are positive.

than the 5–10% profit gains found in recent meta-analyses of offline training programs (McKenzie 2021), reflecting the low take-up rate; the 6.6% ATT falls squarely within this range. Both estimates apply to a base of over two million sellers at near-zero marginal cost, orders of magnitude larger than any prior training evaluation. This revenue increase was primarily driven by a positive effect on the number of visitors attracted: a 1.3% ITT (column 3), which led to an ITT of 0.8% on the number of purchasing customers (column 4).

Table 3: Intent-to-Treat Effects Among Market Entrants

| | Log Revenue | Has Revenue | Log Traffic | Log Customers | Conversion Rate | Star Rating | Refund Rate |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Treatment | 0.053*** | 0.006*** | 0.041*** | 0.029*** | −0.0001 | 0.008* | −0.0004 |
| | (0.016) | (0.002) | (0.014) | (0.009) | (0.0002) | (0.005) | (0.001) |
| Dep Var Mean | – | 0.4 | – | 1.57 | 0.03 | 0.57 | 0.21 |
| Adjusted $R^2$ | 0.077 | 0.095 | 0.107 | 0.065 | 0.043 | 0.121 | 0.040 |

*Notes:* This table replicates Table 2 on the sample of market entrants (registered new sellers with nonzero total revenue during our research window). $N = 1,771,371$. Conversion rate is the share of visitors (traffic) that made a purchase within a week, which is undefined for many potential entrants with no traffic. Star Rating scales between 0 to 5. Refund rate is calculated as total amount of refunds requested over the total payments made. Raw revenue and traffic numbers are omitted (labeled as "-") at the data provider's request.

Table 4: ATT Results on Main Outcomes: 2SLS Estimates with Training Take-up

| | Training Take-up | Has Revenues | Log Revenues | Log Traffic | Log Customers | Conversion Rate |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.257*** | | | | | |
| | (0.001) | | | | | |
| Take-up Training | | 0.006** | 0.066*** | 0.052** | 0.032*** | 0.0001 |
| | | (0.003) | (0.024) | (0.021) | (0.012) | (0.001) |
| Dep Var Mean | 0 | 0.19 | – | – | 0.57 | 0.04 |
| Adjusted $R^2$ | 0.232 | 0.153 | 0.133 | 0.208 | 0.106 | 0.048 |

*Notes:* This table reports the 2SLS (two-stage-least-square) estimates for the training ATT (average treatment effects on the treated). Column 1 reports the first-stage estimates, in which we regress the take-up indicator (for whether a treatment-group seller has taken up any task) on the treatment indicator. Columns 2 to 6 report the 2SLS estimates. All regressions include cohort, month, and sector fixed effects, with standard errors clustered at the seller level. Dependent variable means are calculated among control-group sellers. ** significant at 5% level and *** at 1% level. Raw revenue and traffic numbers are omitted (labeled as "-") at the data provider's request.

We now focus on market entrants and recalculate ITT for four reasons. First, this definition is more aligned with how entrants are measured offline. Second, training had no detectable impact on sellers starting to earn revenue (Table 2 column 2 primarily measures survival rate). Third, new sellers who had never made a sale were likely not serious about entry in the first place, let alone being impacted by training. Lastly, many measures, such as conversion rate, are undefined for non-market entrants.

As shown in Table 3, the ITT effects on revenue and traffic increase significantly among market entrants

to 5.3% and 4.1%, respectively. Training also had a significant impact on survival (1.5%). Conditioning on market entry is common in the training literature, where studies routinely restrict to firms with positive revenue, but because market-entrant status is itself a post-treatment outcome, these estimates embed compositional selection and should be interpreted descriptively. The IV results in Table 4 use the full registrant sample and are not subject to this concern. In contrast, training had no detectable impact on platform quality measures: conversion rate, star rating, and refund rate (more measures available in Table C9).

Next, we calculate the ATT (average treatment effects on the treated) on training participants, using treatment assignment as an instrument. Column 1 of Table 4 shows the first stage specification as equation 1: 25.7% of treated sellers took up at least one task, and participants exhibit advantageous selection on all performance measures (Table C8). The resulting ATT (6.6% on revenue) is comparable in magnitude to the ITT among the subsample of market entrants (5.3%, Table 3), consistent with advantageous selection among participants. See Table C7 for robustness check using alternative take-up measures.

**Summary of main effects** Table 5 summarizes the key estimates across the three specifications. Training raised traffic by 1.3% (ITT among potential entrants) to 5.2% (ATT), and revenue by 1.7% to 6.6%. There is no detectable impact on conversion rate, star ratings, or refund rate.

Table 5: Summary of Treatment Effects

| | ITT (Potential Entrants) | ITT (Market Entrants) | ATT |
|---|---|---|---|
| Traffic | +1.3%*** | +4.1%*** | +5.2%*** |
| Revenue | +1.7%*** | +5.3%*** | +6.6%*** |
| Has Revenue | +0.8%** | +1.5%*** | +3.2%** |
| Conversion Rate | n.s. | n.s. | n.s. |

*Notes:* Summary of intent-to-treat (ITT) and average treatment effect on the treated (ATT) estimates from Tables 2, 3, and 4. "n.s." denotes not significant. ** $p < 0.05$, *** $p < 0.01$.

## 3.2 Treatment Effect Dynamics

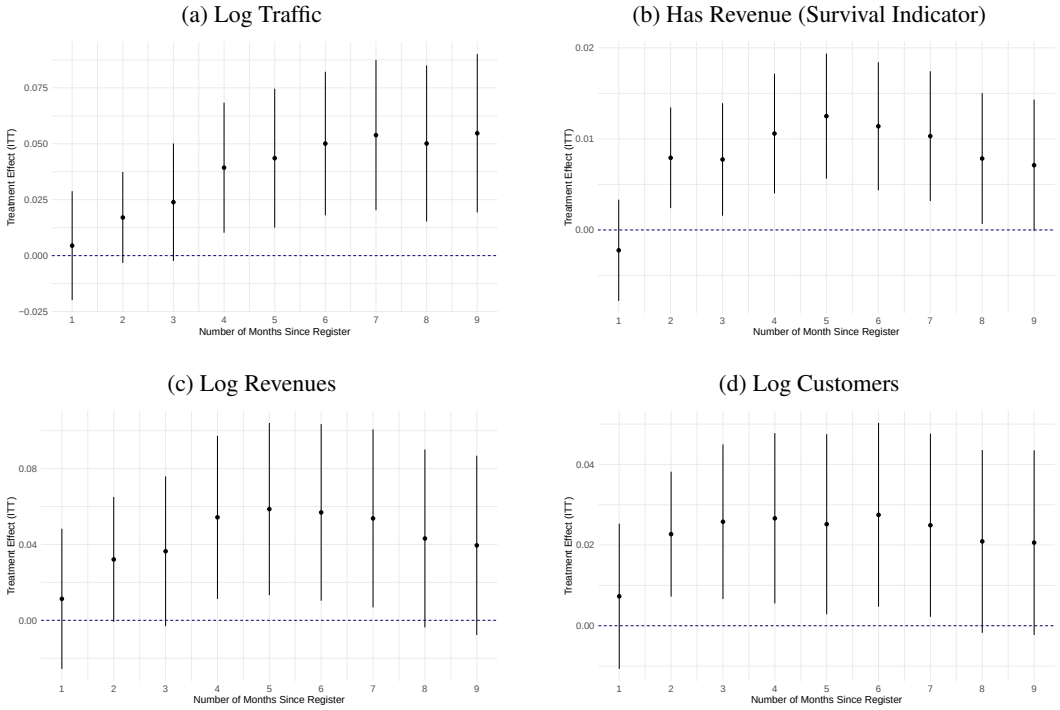
Figure 4 plots the month-by-month ITT coefficients $\{\beta^k\}_{k=1}^{9}$ from the panel specification

$$Y_{imcs} = \sum_{k=1}^{9} \beta^k Treatment_i M_{ik} + \alpha_m + \alpha_c + \alpha_s + \varepsilon_{imcs} \tag{2}$$

where $M_{ik}$ indicates month $k$ since entry. The treatment effect on traffic and revenues is positive and significant during months 2–4, consistent with the timing of training participation (Section 2.4), while conversion rate effects remain insignificant throughout.

The dynamics of *how* traffic increased are revealing. Figure 4(a) shows that the *ads* traffic effect peaks in months 2–4; Table 6 decomposes this effect by source and timing. Ads traffic increased first and sharply: the effect during months 1–3 (when 70% of tasks were taken up) was 7.60%, coinciding with the peak of training intensity. Organic traffic, by contrast, showed no contemporaneous response; it increased only in

Figure 4: Intent-to-Treat Effect Dynamics (Market Entrants)




*Notes:* The points in these figures represent the ITT estimates and the error bars the 95% confidence interval. We focus on *monthly* subsamples. The estimation method and variable definition are otherwise identical to those used to produce Table 3.

subsequent months ($t > 3$), with a smaller percentage change but a larger absolute contribution to total traffic given that organic traffic constitutes the vast majority of visits. This sequencing—paid traffic leading, organic traffic following—suggests that advertising-driven sampling generates the reviews, ratings, and sales history that eventually boost organic search ranking. Training helped sellers overcome the cold-start problem by purchasing consumer "sampling" upfront through ads, which then built the performance track record needed for organic visibility.

## 3.3 Mechanism: Seller Strategy and Advertising

We now focus on how training changed seller strategy, which mediates the effects discovered above on traffic and revenues. As shown in Table 6, of a wide range of strategies that the program trained—pricing, product offerings, marketing, ratings, and customer service—the most compelling change is an increase in advertising. The number of products sellers advertise through position auctions increased by 2.96%, with ads and organic traffic increasing by 4.36% and 0.90%, respectively (Table 6, pooled rows). The timing of these effects was revealed, albeit harder to detect. Since 70% of the tasks were taken up in the first three months of the training, we bisect our data into two periods: $t \leq 3$ and $t > 3$. The effect on ads traffic preceded the one on organic traffic. This suggests that SEO (search engine optimization) strategies targeting organic traffic were not immediately effective, instead, treated sellers were better at overcoming "cold-start" because they purchased more sampling upfront.

Table 6: Percentage ITT Effect on Seller Strategy (Potential Entrants)

| Variable | Treatment (%) | Dep Var Mean | Variable | Treatment (%) | Dep Var Mean |
|---|---|---|---|---|---|
| **Products** | | | **Marketing - Paid** | | |
| Post Products | 0.29 | 0.34 | # Ads Purchased | 2.96*** | 1.57 |
| | (0.29) | | | (1.00) | |
| Product Variety | 0.73 | 1.81 | Has Advertising | 2.25** | 0.07 |
| | (0.91) | | | (1.26) | |
| | | | Ads Traffic (t $\leq$ 3) | 7.60*** | - |
| **Ratings** | | | | (2.04) | |
| % of Good Reviews | 0.00 | 0.99 | Ads Traffic (t > 3) | 2.74* | - |
| | (0.02) | | | (1.49) | |
| Star Rating | 0.35 | 3.77 | Ads Traffic | 4.36*** | - |
| | (0.36) | | | (1.41) | |
| | | | % Ads Traffic | 2.25** | 0.03 |
| **Pricing** | | | | (1.13) | |
| Avg. Price Per Buyer | 0.50 | 80.72 | | | |
| | (0.59) | | **Marketing - SEO** | | |
| Avg. Price Per Product | 0.03 | 46.22 | Has Video Demo (t $\leq$ 3) | 1.97*** | 0.11 |
| | (0.59) | | | (0.72) | |
| | | | Has Title Change | -0.02 | 0.05 |
| **Service** | | | | (0.62) | |
| Active Time (min) | -0.41 | 19.17 | Organic Traffic (t $\leq$ 3) | 0.43 | - |
| | (0.67) | | | (0.36) | |
| Reply Time (min) | -0.29* | 389.02 | Organic Traffic (t > 3) | 1.14*** | - |
| | (0.15) | | | (0.29) | |
| Service Conversion | 1.40*** | 0.10 | Organic Traffic | 0.90*** | - |
| | (0.50) | | | (0.26) | |

*Notes:* This table presents the *percentage* ITT on seller strategy among potential entrants (all registered new sellers). We report the $\beta$ estimates in specification 1, in which "$t$" indexes months since seller registration. Price and Rating regressions do not include sellers with zero revenues. Standard errors are clustered on the seller level. All regressions include cohort, month, and main industry fixed effects. "Post products" is an indicator of having an active product listing during the month. "Avg. price per buyer" measures the transaction amount paid by each buyer, averaged across buyers of each seller and log-transformed. "Avg. price/product" measures the transaction amount divided by product quantity purchased, averaged across transactions of each seller and log-transformed. "Active time" is the number of minutes that a seller logged on to the customer service portal. "Reply time" is the average number of minutes a seller takes to respond to customer inquiries. Service conversion is the fraction of inquiring customers who make a purchase. "# of Product with Ads" counts the number of products that participated in the pay-per-click in-platform position auction. "Has Advertising" is an indicator for sellers using ads in the corresponding month. "% Ads Traffic" measures the share of traffic (or visitors) to the seller's site sourced from pay-per-click ads. "SEO" stands for search engine optimization. "Has Video Demo" and "Has Title Change" are indicators for sellers having done such SEO actions in a given month. * significant at 10% level, ** at 5% and *** at 1%=. Raw traffic numbers are omitted (labeled as "-") at the data provider's request.

The efficiency of customer service also increased, as measured by the percentage of serviced customers that made a purchase. However, this increase did not lead to a statistically significant change in (unconditional) conversion rate. This is because customer service was sparse among new sellers: 19 minutes per month on average and less than 1% of visitors serviced, with an average wait time of 6.5 hours for an inquiry response.

The pattern of selective response across channels is informative. Training covered advertising, SEO, customer service, pricing, and product strategy with comparable breadth (task completion rates of 39.7%, 42.9%, and 22.3% for marketing, basic operations, and customer service, respectively). Yet only advertising adoption responded with a corresponding increase in traffic and revenue. Pricing, product variety, category switching, and quality ratings were all unaffected (Tables C9 and C12, Appendix C).

Combining the effects on measurable strategies with their corresponding outcome metrics, advertising is the main mechanism through which training raised traffic and revenue. This result parallels Anderson et al.

(2018)'s finding that marketing skills dominate finance skills for small businesses. Jin and Sun (2026) analyzes the impact of this advertising shock on seller entry more carefully and provide a theoretical characterization of its relationship with organic ranking and platform-directed exploration.

**Survey evidence** An online survey of training participants ($n = 409$, May 2019) corroborates these revealed-preference findings. When asked about their biggest challenges in learning to operate an online store, 58.9% cited "too much information, hard to find the key points" and 50.1% cited "content is too scattered"—confirming that the primary friction is information filtering, not information scarcity. Sellers' evaluations of the algorithmically recommended tasks were largely positive: 77.5% rated task recommendations as accurate in matching their store's needs, and 87.3% found the training helpful overall.

Crucially, the survey reveals which task categories sellers found most helpful, broken down by engagement level (Table B2, Appendix B). Among sellers who completed only 1–2 tasks, store basics dominated. But as engagement increased, traffic and promotion tasks—including advertising—grew substantially in perceived helpfulness, rising from 46% to 47% of "most helpful" responses. This self-reported pattern mirrors Table 6: the most engaged sellers gravitated toward marketing training, and advertising was the channel through which training had its effect.[12]

### 3.4 Heterogeneity and Screening

**Quantile treatment effects** The average treatment effect masks substantial heterogeneity across the outcome distribution. We estimate quantile treatment effects (QTE) on first-month revenue at every percentile from the 80th to 99th. Since less than 20% of new sellers had ever earned any revenue, the QTE on revenues is mechanically zero below the 80th percentile. Among sellers with positive revenue, effects are largest at the lower end of the positive-revenue range—the 80th–90th percentiles—and fall to near zero at the very top.

Figure C5 presents the results across all nine months. In month 1, effects are small and imprecisely estimated. By months 2–3, effects of 0.03–0.05 log points emerge at the 80th–90th percentiles—roughly three to four times the ATE. The pattern peaks dramatically in month 4: QTE reaches 0.4–0.8 log points at the 80th percentile, declining steeply toward the very top of the distribution. This timing coincides with the peak of treatment effects on traffic (Figure 4). Effects moderate in months 5–6 and fade by months 7–9 as training intensity wanes. Throughout, effects are concentrated at the 80th–90th percentiles and near zero at the very top, indicating that training benefits sellers just entering the positive-revenue range rather than those already at the top who would have succeeded regardless.

Only 1.3% of new market entrants ever reach the top 10% of the revenue distribution in their sector. The QTE evidence shows that training's largest effects are at the 80th–90th percentiles of the monthly revenue distribution, while effects fade toward the very top, where sellers likely would have succeeded regardless (Figure C5, Appendix C). This raises a natural question: can training help the platform identify high-potential sellers earlier?

[12] A separate survey ($n = 566$, August 2019) collected demographic information. Training participants are mostly small and inexperienced: 71.9% have 1–2 employees and 74.3% have less than one year of e-commerce experience; 58.8% treat e-commerce as their main job. See Table B1 in Appendix B for details. A willingness-to-pay elicitation ($n = 676$) finds a median WTP of approximately zero, with a long right tail (Appendix B).

**Training Participation as a Screening Device** High-potential sellers are not identifiable from early performance: first-month revenue distributions overlap heavily between sellers who will eventually reach the top of the distribution and those who will not: first-month performance explains only 2.8% of variance in who reaches the top decile by months 4–9 (Table 7, column 1). The platform's ranking algorithm, which relies on historical conversion rates and traffic, cannot identify these high-potential sellers during the critical early months when they most need consumer exposure.

Moreover, there is no significant heterogeneity in treatment effects along *observable* dimensions such as gender, registration type, or preparedness (Tables C1 and C2, Appendix C). The meaningful heterogeneity is along *unobservable* dimensions—latent entrepreneurial ability that determines who will succeed.

Table 7: Training Participation Predicts Long-Run Top Sellers

| | *Dependent variable: indicator for top 10% of sector revenue* | | | |
|---|---|---|---|---|
| | Is SMB Months 4–9 | Is SMB Month 7 | Is SMB Month 8 | Is SMB Month 9 |
| | (1) | (2) | (3) | (4) |
| Training participation | 0.265*** | 0.380*** | 0.520*** | 0.496*** |
| | (0.038) | (0.063) | (0.062) | (0.062) |
| First-month log revenues | 0.250*** | 0.276*** | 0.268*** | 0.290*** |
| | (0.005) | (0.009) | (0.009) | (0.009) |
| First-month conversion rate | 0.756*** | 0.248* | 0.234 | 0.011 |
| | (0.075) | (0.149) | (0.151) | (0.161) |
| Observations | 158,766 | 158,766 | 158,766 | 158,766 |

*Notes:* OLS regressions among treated sellers. The dependent variable is an indicator for being an SMB (top 10% of revenue in the seller's sector), scaled by 100 so that coefficients are in percentage points. Column (1): SMB status during months 4–9. Columns (2)–(4): SMB status during the indicated month since entry. "Training participation" is an indicator for accepting at least one training task in month 1. "First-month log revenues" and "First-month conversion rate" are measured in month 1. All regressions include cohort and sector fixed effects. Standard errors in parentheses. * $p < 0.10$, *** $p < 0.01$.

Training participation provides an early signal of seller quality that performance data cannot yet reveal. Among treated sellers, those who voluntarily participate are drawn disproportionately from the portion of the distribution where treatment effects are positive. This is why the ATT (6.6%) substantially exceeds the ITT (1.7%): participants self-select from the upper tail. Training tasks are classified as "action-based" (list a product, issue coupons) or "effect-based" (achieve sales targets, reach traffic goals). The two types predict first-month outcomes—survival, traffic, and revenue—with similar magnitudes, suggesting that mere *participation* signals latent ability rather than that specific task *content* builds skills.[13]

Table 7 shows that participation predicts reaching the top decile with a coefficient of 0.265 percentage points ($p < 0.01$), comparable in magnitude to first-month log revenues (0.250 pp). Strikingly, the predictive power of participation *increases* in later months—rising to 0.520 pp by the ninth month since entry—while first-month conversion rate's predictive power fades to insignificance. This divergence confirms that

[13]Among treated sellers, accepting an action-based task predicts 1,372 additional page views and 2,904 additional RMB in first-month GMV ($p < 0.01$). Accepting an effect-based task predicts similar magnitudes: 1,194 page views and 2,911 RMB ($p < 0.01$). Completing (vs. merely accepting) action-based tasks does not predict meaningfully better outcomes, further suggesting that participation signals type rather than that task completion builds skill.

participation captures something distinct from early performance: latent entrepreneurial ability that takes months to manifest in outcomes.

Because the predictive power of participation persists after controlling for early performance and grows over time, it likely reflects selection on latent ability over and above the specific skills acquired through training. Incorporating training participation into the platform's traffic allocation could improve matching efficiency by directing more consumers to high-potential sellers during the information-poor early months after registration. Platforms could incorporate participation as a Bayesian prior in their exploration algorithm à la Jin and Sun (2026): early participation shifts the platform's belief about seller quality, generating additional exploration that self-corrects as actual performance data accumulates. This approach is incentive-compatible: if broader participation dilutes the screening signal, the platform can reduce the prior's weight to restore separation.

# 4 Welfare Impact of Training

Does training benefit consumers? The business training literature has largely only measured supply-side outcomes: profits, sales, productivity, and business practices (McKenzie and Woodruff 2014; McKenzie 2021). Our companion paper (Jin and Sun 2026) likewise focuses on entry and business success. No prior study has estimated whether training benefits the consumers these firms serve.

The answer is not obvious. Pricing is central to the division of surplus between consumers and sellers: higher seller revenue could reflect lower prices that expand the market, or higher prices that transfer surplus away from consumers. Moreover, both the available offerings and consumers' purchases feature vast quality heterogeneity that aggregate data cannot capture: even if consumers are buying more products at lower prices, they are not necessarily finding higher-quality matches. Training changes the composition of sellers that consumers encounter, as treated sellers invest in advertising and appear more frequently in search results. Whether this reallocation improves consumer welfare depends on whether the entering sellers offer higher quality than the incumbents they displace.

To answer this question, we estimate a logit demand model on consideration-set data. Consideration sets record the full set of sellers each consumer evaluated during a search session, allowing us to model consumer choices and recover sellers' latent product quality $\xi_j$. To quantify the program's consumer welfare impact, we equalize trained sellers' appearance in consideration sets with that of control sellers and measure the resulting welfare loss.

## 4.1 Consideration-Set Data

To quantify the welfare impact of training, we need consumer-level choice data. We construct consideration sets from search-session clickstream records: each set consists of all sellers a consumer visited during a single search session. Starting from a random subsample of new sellers in our experiment, we identify consumers who visited them and obtain their search keywords.[14] We then match a supplementary set of consumers who

[14]We focus on consumers who visited at least three sellers. Search sessions with one or two sellers have implausibly low purchasing rates, indicating data contamination from non-search channels largely irrelevant for new sellers.

only visited incumbents but had the same search intent, effort, and date.[15] This procedure is summarized in Figure C3.

The resulting sample contains 1.4 million consideration sets (Table 8), each consisting of 4.6 sellers on average. A total of 13,593 distinct keywords were searched across most popular sectors. Despite over-sampling, 82% of consideration sets had no new sellers. Consumers who visited new sellers are observably different: they search more sellers regardless of treatment status (Figure C4, Appendix C), illustrating how demand heterogeneity can bias quality inference using raw conversion rates.

Table 8: Summary Statistics: Consideration Set Sample

| Statistic | Mean | St. Dev. | Min | Pctl(25) | Pctl(99) | Max |
|---|---|---|---|---|---|---|
| Number of Sellers Browsed | 4.6 | 2.7 | 3 | 3 | 12 | 55 |
| % of New Sellers | 0.57 | 8.7 | 0 | 0 | 33 | 100 |
| Purchased (on the same day) | 0.180 | 0.390 | 0 | 0 | 0 | 1 |
| Purchased from New Sellers | 0.001 | 0.033 | 0 | 0 | 0 | 1 |
| Pay Amount (USD) | 4.8 | 21.2 | 0 | 0 | 753 | 2,050 |
| Avg. Seller Price Level (USD) | 34.4 | 86.3 | 0.0 | 8.8 | 582 | 9,674 |
| Avg. Seller Rating | 13.0 | 2.9 | 0 | 11.5 | 20 | 20 |

*Notes:* $N = 1,381,273$. Unit of analysis is a consideration set. Dollar amount converted to USD based on the average 2019 currency exchange rate of 1 RMB to 0.1448 USD.

## 4.2 A Structural Econometric Model

**Consumer Demand** Each consumer $i$ arrives at the platform with a predetermined consideration set $K_i$ of seller-product pairs $j \in K_i$. Consumer $i$ derives utility $U_{ij} = \delta_{ij} + \varepsilon_{ij}$ from each pair, where $\delta_{ij}$ captures the product's value and $\varepsilon_{ij}$ is idiosyncratic noise.

$$U_{ij} = \delta_{ij} + \varepsilon_{ij} = x_j\beta - \alpha p_j + \xi_j + \varepsilon_{ij} \tag{3}$$

$p_j$ is the price level of $j$, and $x_j$ a set of observable characteristics. $\xi_j$ is the residual, unobserved product value that is our main object of interest. We interpret $\xi_j$ as *product quality* in the vertical differentiation sense. Since the platform observe and utilize $p_j$ and $x_j$, the main source of friction our model captures comes from a mismatch between sellers' unobserved quality $\xi_j$ and their likelihood of appearing in consumers' consideration sets. $\varepsilon_{ij}$ is the i.i.d. consumer-seller idiosyncratic preferences that follows a type I extreme value distribution. We normalize the outside option of not purchasing to zero utility, which yields the logit choice probability:

$$P_{ij} = \frac{\exp(x_j\beta - \alpha p_j + \xi_j)}{1 + \sum_{k \in K_i} \exp(x_k\beta - \alpha p_k + \xi_k)} \tag{4}$$

[15] We limit the number of matched consumers per search query-search effort-date set at 50 for computational reasons, which implies over-sampling of sellers with fewer visitors and less popular keywords.

$P_{ij}$ is consumer $i$'s probability of purchasing from seller $j$. We later enrich the baseline model by adding consumer side heterogeneity, sector-specific fixed effect $\eta_s$ and sector-specific price coefficient $\alpha_s$.

In order to capture and simulate dynamic monthly path of seller outcomes so as to capture any potential compounding effects of early success, we model the dynamic evolution of seller traffic as primarily dependent on prior-periods conversion rates, resulting in the following model:

$$\mathscr{T}_{jt} = f(\mathscr{T}_{jt-1}, C_{jt-1}, C_{jt-2}) \tag{5}$$

$\mathscr{T}_{jt}$ is the traffic (number of visitors) of seller $j$ in month $t$[16], and $C_{j,t-1}$ the lag monthly conversion rate.

### 4.3 Identification and Estimation

We estimate the model using simulated maximum likelihood following Train (2009). To better fit the empirical setting, we make the following changes to the baseline model.

**Consumer Demand** We use the consumer-seller pair sample to estimate the demand parameters in particular sellers' $\xi_j$. Due to computation constraints, we sample a subset of sellers for actual estimation. The final sample is a seller-consumer matched pair dataset, where we explore seller-level variations. $x_j$ includes the number of products sellers offer and sellers' ratings. $p_j$ is the average price level that sellers charge.[17] To account for sector-level heterogeneity, we add a sector-specific intercept $\eta_s$ in the baseline model and enrich the baseline model by estimating the sector-specific price coefficient $\alpha_s$.

**Endogenous Strategies** One concern with the baseline model is that pricing level $p_j$ and strategies such as the number of products offered in $x_j$ could correlate with $\xi_j$, which would bias the estimates. To address this concern, we use a set of instruments to jointly determine the number of products posted $\text{prod}_{jt}$ and pricing level $p_{jt}$ with

$$\begin{bmatrix} p_{jt} \\ \text{prod}_{jt} \end{bmatrix} = \mathbf{Z}\beta^{fs} + \mu_j + \mu_t + \mu_s + \varepsilon_{jt} \tag{6}$$

$\mu_j$, $\mu_t$, and $\mu_s$ are seller, date, and sector fixed effects. The instruments $\mathbf{Z}$ exploit variation in the stringency of the platform's rule enforcement at the sector-month level. The platform frequently adjusts enforcement of rules governing product listings, pricing, and intellectual property. Stricter enforcement constrains sellers' pricing and product-posting strategies but does not directly affect consumer taste shocks.[18] More detailed estimation procedures are described in Appendix 4.1.

**Traffic Evolution** We use the new seller panel to estimate the traffic evolution function. We use the following specification to distinguish new entrants with no previous history and sellers with zero conversion

[16] We use log scale as the traffic distribution is very skewed.

[17] As mentioned before, ratings capture sellers' size more than quality. The price level seller charges is the quantity-weighted prices of all products sold by the sellers during the day. Such weighted prices reflect the relative popularity of products sellers' offerings as well as any sales or promotions.

[18] The instruments include the average frequency of rule violations and the share of sellers labeled as frequent rule violators in sector $s$ over the previous 30-day period, weighted by seller traffic. Rule violations come in different degrees of severity which we calculate separately.

rates in the previous periods from the rest:

$$T_{jts} = f(T_{j,t-1}, C_{j,t-k}) + g(I_{j,t-k}(t-k=0)) + h(I^c_{j,t-k}(C_{j,t-k}=0)) + \gamma_k + \gamma_t + e_{jts} \tag{7}$$

$C_{j,t-k}$ are lagged conversion rates in past periods $k = 1,2$, $I_{jt-k}(t-k=0)$ is an indicator for the initial two periods, and $I^c_{j,t-k}(C_{j,t-k}=0)$ are indicators for lagged conversion rates being exactly zero.[19] The specification also includes product category, calendar time and relative month fixed effect. In the baseline specification we start with linear functions for $f(\cdot)$, $g(\cdot)$ and $h(\cdot)$. In this setup, all the right-hand-side variables are determined in the previous periods.

## 4.4 Estimation Results

The main parameters of interest are $\xi_j$ for seller $j \in J$, which capture unobserved product quality. Table 9 reports their distribution. Focusing on sellers with at least one purchase record (28% of all sellers), new sellers have significantly higher $\xi_j$ than incumbents: 6.4% higher on average (pooled new-seller mean 0.267 vs. incumbent mean 0.251). The quality advantage is even more pronounced for treated sellers ($\xi = 0.296$) than control sellers ($\xi = 0.258$), consistent with training improving the products and services that consumers value.

Both treated and control new sellers have higher $\xi_j$ than incumbents, with no significant difference between the two groups among purchasing consumers. This means that the training does not directly raise unobserved quality; rather, both groups of new sellers offer genuinely higher value. The welfare gain from training thus operates purely through the *reallocation channel*: training makes higher-quality new sellers more visible by helping them overcome the cold-start barrier.

Table 9: Sellers' Characteristics and Estimated Type $\xi_j$

| Statistic | N | Mean | St. Dev. | Min | Max |
|---|---|---|---|---|---|
| All Sellers | 52,241 | 0.141 | 0.314 | −2.828 | 5.024 |
| **By Purchase Status** | | | | | |
| No Purchase | 37,617 | 0.098 | 0.051 | −2.828 | 0.101 |
| Has Purchase | 14,624 | 0.252 | 0.573 | −1.965 | 5.024 |
| **By Sellers' Type** | | | | | |
| Control | 3,366 | 0.136 | 0.296 | −1.445 | 4.939 |
| Treatment | 988 | 0.150 | 0.353 | −0.853 | 3.668 |
| Incumbent | 47,887 | 0.142 | 0.314 | −2.828 | 5.024 |
| **By Type Among Sellers with Purchase** | | | | | |
| Control | 788 | 0.258 | 0.593 | −1.370 | 4.939 |
| Treatment | 252 | 0.296 | 0.679 | −0.853 | 3.668 |
| Incumbent | 13,584 | 0.251 | 0.570 | −1.965 | 5.024 |

*Notes:* Distribution of estimated unobserved seller characteristics, $\xi_j$, on a sub-sample of new sellers and incumbents conditional on making at least one sale. See Appendix 4 for more details.

[19] Sellers have no past history of number of visitors and conversion rate in the initial periods and hence are subject to different matching rules.

### 4.5 Counterfactual: Removing the Training

We simulate a counterfactual that restricts training participants' appearance in consideration sets to match that of control-group sellers, holding constant the pool of active sellers, consumers' search behavior, seller strategies, and the platform's matching algorithm.

To conduct the analysis, we randomly sample sellers and their associated search sessions. For each seller, we obtain their strategies $x_j$ and estimated quality $\xi_j$, and calculate consumers' utility $V_j$. The counterfactual varies only the composition of consideration sets: we randomly remove training participants until their appearance rate matches that of control-group new sellers. The logit specification yields consumer surplus

$$CS = \frac{1}{\alpha} \sum_i \log \left[ 1 + \sum_{j \in K_i} \exp(V_j) \right]$$

and seller revenues $R_j = \sum_i \frac{\exp(V_j)}{1+\sum_{k \in K_i} \exp(V_k)} p_j$.

Table 10 reports the results. Without training, total revenue falls by 0.05% and consumer surplus by 0.07%. The mechanism is *market expansion*: trained new sellers, who have 6.4% higher unobserved quality than incumbents (Table 9), make consumers more likely to purchase within a search session by offering higher-quality alternatives. Despite comprising only 0.25% of products in consideration sets, their removal measurably reduces welfare. Treated sellers suffer a 7.7% drop in market share, as they become less visible and obtain a smaller share of a smaller pie, while control-group and incumbent sellers see moderate increases.

Our counterfactual removes trained sellers from consideration sets without modeling how the platform would reallocate the vacated slots or how consumers would adjust their search behavior. Given that new sellers, especially treated ones, offer substantially higher quality than the incumbents who would likely fill those slots (Table 9), we consider this a reasonable assumption in partial equilibrium.

Table 10: Welfare and Market Share Decomposition

| | **Revenue** | **CS** | **Market Share** | | |
|---|---|---|---|---|---|
| | | | Treatment | Control | Incumbent |
| Remove Training | −0.05 | −0.07 | −7.71 | 0.17 | 0.13 |
| | (0.05) | (0.03) | (0.63) | (0.11) | (0.02) |

*Notes:* Welfare and market share calculated with a random sample of 60,000 consumers. "CS" stands for consumer welfare. The row reports percentage changes from the baseline (where traffic assignment follows predicted traffic). Bootstrapped standard errors are in parentheses. See Appendix 4 for more details.

## 5 Conclusions

This paper experimentally evaluates a large-scale, algorithmically customized entrepreneur training program on a major e-commerce platform. Among more than two million new sellers randomly assigned to training access, those in the treatment group saw 1.7% higher revenue (ITT) and 6.6% higher revenue among participants (ATT). The mechanism is advertising adoption: despite equal training coverage across advertising,

SEO, pricing, and customer service, only advertising responded. Ad traffic increased sharply in the first three months, coinciding with the peak of training intensity, while organic traffic followed with a lag, consistent with advertising-driven sampling building the track record needed for organic visibility. Treatment effects are largest among sellers just breaking into positive revenue and fade toward zero at the very top. Training participation serves as an early screening signal of latent entrepreneurial ability. A logit demand model estimated on consideration-set data shows that new sellers have 6.4% higher unobserved quality than incumbents and that removing the program would reduce consumer surplus by 0.07%.

These findings carry several implications. First, information friction around advertising tools is a meaningful yet incomplete explanation for the entry difficulty described in Jin and Sun (2026). Training was effective precisely because it reduced this specific friction, nudging sellers to adopt readily available but underutilized advertising tools. But its effect was limited, indicating that there may be structural barriers beyond information friction. Second, training can serve as a screening device. Platforms can incorporate engagement into their ranking algorithms in an incentive-compatible way, for example by using training participation to update the platform's prior belief about seller quality within a Bayesian exploration algorithm. Third, we provide, to our knowledge, the first measurement of consumer welfare effects of any business training intervention, addressing a gap in a literature that has exclusively measured supply-side outcomes.

Several limitations warrant discussion. Our experiment was conducted on a single e-commerce platform in China, and external validity to other platforms, countries, or offline settings remains an open question. The training content was not separately randomized from the algorithmic matching, so we cannot disentangle the value of specific content from the value of algorithmic curation. The welfare estimates are based on short-term effects (nine months); longer-term dynamics, including potential reversion as in Davies et al. (2024) or compounding as trained sellers build reputations, remain unmeasured. The modest magnitude of the welfare effect (0.07%) reflects the small share of trained sellers in consideration sets (0.25%); scaling the program or combining it with more aggressive exploration of new-seller quality could amplify these gains.

For policy, our results suggest that platforms have both the incentive and the capability to facilitate entrant growth through low-cost, scalable training. The near-zero marginal cost of algorithmically customized training contrasts sharply with the high per-participant costs of offline business training and consulting interventions (Bloom et al. 2013; Bruhn et al. 2018; McKenzie 2021). Whether platforms should go further—for example, by directly subsidizing entrant advertising or incorporating training signals into ranking algorithms—is an important question that connects to broader debates about platform governance and competition policy (Chen and Tsai 2019; Farronato et al. 2023). In a companion paper (Jin and Sun 2026), we develop a theoretical framework for platform-mediated entry that characterizes optimal exploration and advertising policy, and show that diminishing returns make broad, shallow exploration optimal.

More broadly, this study bridges the development economics literature on entrepreneur training and the industrial organization literature on digital platforms. Platform-based entrepreneurship is now a global phenomenon, and understanding how platform governance shapes entry and competition is increasingly important. We hope this work inspires further cross-field research on the design of interventions that benefit both entrepreneurs and consumers in digitally mediated markets.

## References


Anderson, Stephen J et al. (2018). "Pathways to profits: The impact of marketing vs. finance skills on business performance". In: *Management Science* 64.12, pp. 5559–5583.

Athey, Susan and Glenn Ellison (2011). "Position auctions with consumer search". In: *The Quarterly Journal of Economics* 126.3, pp. 1213–1270.

Atkin, David et al. (2017). "Exporting and firm performance: Evidence from a randomized experiment". In: *The Quarterly Journal of Economics* 132.2, pp. 551–615.

Bai, Jie et al. (2025). "Stand Out from the Millions: Market Congestion and Information Friction on Global E-Commerce Platforms". Working Paper.

Bloom, Nicholas et al. (2013). "Does management matter? Evidence from India". In: *The Quarterly Journal of Economics* 128.1, pp. 1–51.

Brooks, Wyatt et al. (2018). "Mentors or teachers? Microenterprise training in Kenya". In: *American Economic Journal: Applied Economics* 10.4, pp. 196–221.

Bruhn, Miriam et al. (2018). "The impact of consulting services on small and medium enterprises: Evidence from a randomized trial in Mexico". In: *Journal of Political Economy* 126.2, pp. 635–687.

Cai, Jing and Adam Szeidl (2018). "Interfirm relationships and business performance". In: *The Quarterly Journal of Economics* 133.3, pp. 1229–1282.

Campos, Francisco et al. (2017). "Teaching personal initiative beats traditional training in boosting small business in West Africa". In: *Science* 357.6357, pp. 1287–1290.

Chen, Nan and Hsin-Tien Tsai (2019). "Steering via algorithmic recommendations". In: *Available at SSRN 3500407*.

Chen, Yongmin and Chuan He (2011). "Paid placement: Advertising and search on the internet". In: *The Economic Journal* 121.556, F309–F328.

Davies, Elwyn et al. (2024). "Can Zoom substitute for in-person business training? Experimental evidence from Mexico". In: *Journal of Development Economics* 166, p. 103168.

Denes, Matthew et al. (2022). *Entrepreneurship and the Platform Economy: Evidence from US Tax Returns*. Tech. rep. Working Paper.

Drexler, Alejandro et al. (2014). "Keeping it simple: Financial literacy and rules of thumb". In: *American Economic Journal: Applied Economics* 6.2, pp. 1–31.

Edelman, Benjamin et al. (2007). "Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords". In: *American economic review* 97.1, pp. 242–259.

Farronato, Chiara et al. (2023). *Self-preferencing at Amazon: evidence from search rankings*. Tech. rep. National Bureau of Economic Research.

Jin, Yizhou and Zhengyun Sun (2019). "Information Acquisition and the Return to Data". In: *Harvard University Working Paper*.

— (2026). "Entry and Advertising on Digital Platforms: Evidence from a Large E-Commerce Platform". Working paper.

McKenzie, David (2021). “Small business training to improve management practices in developing countries: re-assessing the evidence for ‘training doesn’t work’”. In: *Oxford Review of Economic Policy* 37.2, pp. 276–301.

McKenzie, David and Susana Puerto (2021). “Growing markets through business training for female entrepreneurs: a market-level randomized experiment in Kenya”. In: *American Economic Journal: Applied Economics* 13.2, pp. 297–332.

McKenzie, David and Christopher Woodruff (2014). “What are we learning from business training and entrepreneurship evaluations around the developing world?” In: *The World Bank Research Observer* 29.1, pp. 48–82.

Muralidharan, Karthik and Paul Niehaus (2017). “Experimentation at scale”. In: *Journal of Economic Perspectives* 31.4, pp. 103–24.

Otis, Nicholas G. et al. (2024). “The Uneven Impact of Generative AI on Entrepreneurial Performance”. Harvard Business School Working Paper No. 24-042.

Shapiro, Carl (1982). *Advertising as a Barrier to Entry?* US Federal Trade Commission, Bureau of Economics.

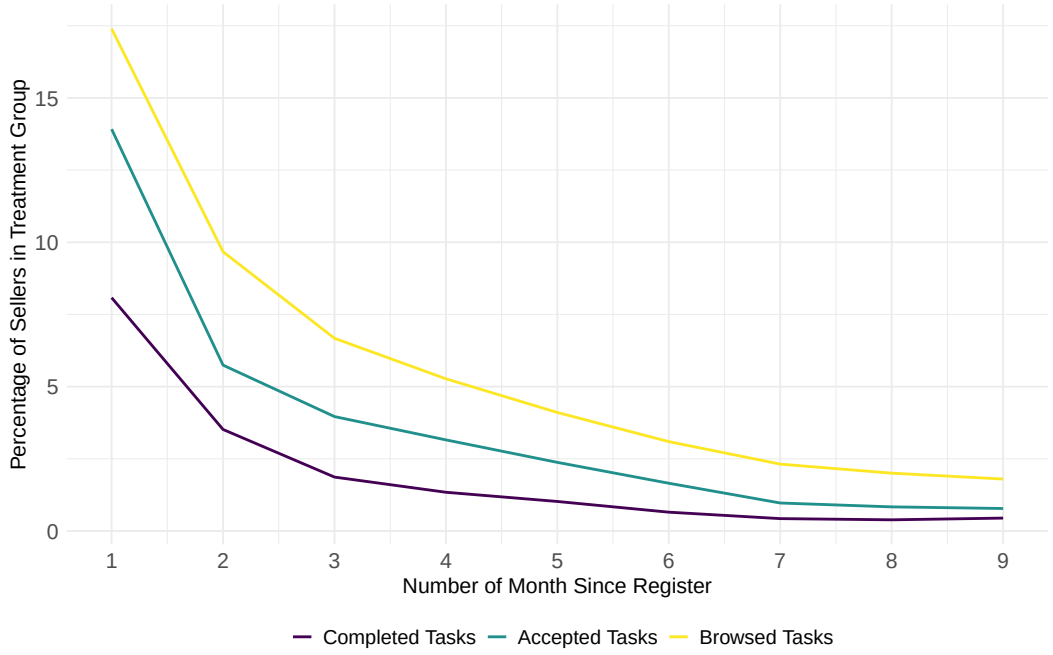
Sun, Tianshu et al. (2021). “The value of personal data in internet commerce: A high-stake field experiment on data regulation policy”. In: *Available at SSRN 3962157*.

Train, Kenneth E (2009). *Discrete choice methods with simulation*. Cambridge University Press.

# A Training Take-up Details [For Online Publication]

## A.1 Training Take-up Dynamics

Figure A1: Training Take-up Dynamics




Although sellers have access to the training for at least six months after the registration, the retention rate declined relatively quickly over time. 13.9% of treated sellers took up tasks during the first month, but only 4.0% of sellers continued to do so in the third month.[20] Conditional on taking up any tasks in the previous month, around 23% of sellers took up more tasks in the following month. Figure A1 shows the share of sellers who browsed, took up and completed the tasks in the subsequent months following the entry.

[20]Many sellers exit the platform after a month. 89.7% of sellers have visitors during the first month, but only 42.2% still do during the third month. The rates are similar for sellers in the treatment and control group.

# B Seller Survey Details[For Online Publication]

**Seller Demographics** An online survey ($n = 566$, August 2019) was stratified by training engagement, oversampling more engaged sellers. Since most respondents are training participants, they may not form a representative sample. These respondents are likely to be more active and have higher sales, with higher than average ownership of manufacturing factories (32.5%) and offline stores (19.2%).

The survey shows that even among training participants, sellers differ in terms of their background, experience, education, and financial resources. However, while the vast majority is small and inexperienced, a substantial share is reasonably educated and expresses clear interest in e-commerce. Results show that 71.9% have 1 or 2 employees, 74.3% have no or less than one year of e-commerce experience, and 68.2% have completed at least high school education. About 58.8% report that running the e-commerce store is their main job and 48.2% have invested more than 10,000 RMB ($1,430) into their online businesses.

Table B1: Summary of Sellers' Survey

| Category | Fraction | Category | Fraction |
|---|---|---|---|
| **Respondent Chars** | | **Business Chars** | |
| *Education* | | *Sources of Supply* | |
| Primary | 2.8% | Own factory | 32.5% |
| Middle School | 29.0% | Offline wholesale markets | 19.2% |
| High School | 23.7% | Online whotesale markets | 21.7% |
| Some College | 28.0% | Distribution/brand subsidary | 19.5% |
| Bachelors | 15.4% | Others | 7.1% |
| Master's and Above | 0.8% | | |
| Professional Degrees (e.g. MBA) | 0.3% | *Number of Employees (inc. owners)* | |
| | | 1 - 2 persons | 71.9% |
| *Exp in Retail* | | 3 - 5 persons | 21.8% |
| None | 36.7% | 6 - 10 persons | 3.9% |
| Less than a year | 25.6% | >10 persons | 2.4% |
| 1 to 3 years | 17.2% | | |
| More than 3 years | 20.5% | *Total investments* | |
| | | <5k RMB | 32.3% |
| *Exp in E-commerce* | | 5k - 10k RMB | 19.5% |
| None | 36.3% | 10k - 50k RMB | 25.2% |
| Less than a year | 38.0% | 50k - 100k RMB | 9.3% |
| 1 to 3 years | 16.5% | 100k - 200k RMB | 5.0% |
| More than 3 years | 9.2% | >200k RMB | 8.7% |
| *Goal* | | | |
| No specific goal | 3.1% | | |
| As part-time job | 19.2% | | |
| As main job | 58.8% | | |
| Expand offline business online | 18.9% | | |

*Notes:* Online survey implemented with users assigned to treatment group for the training intervention in August 2019. Separate messages were sent out based on sellers' engagement with the training defined by number of tasks accepted and whether or not sellers have browsed contents of the training. Survey response rates are higher among sellers that were more engaged in the training. All fractions shown adjusted for the sampling and response rate differential.

Table B2: Survey: Most and Least Helpful Task Categories by Engagement Level (%)

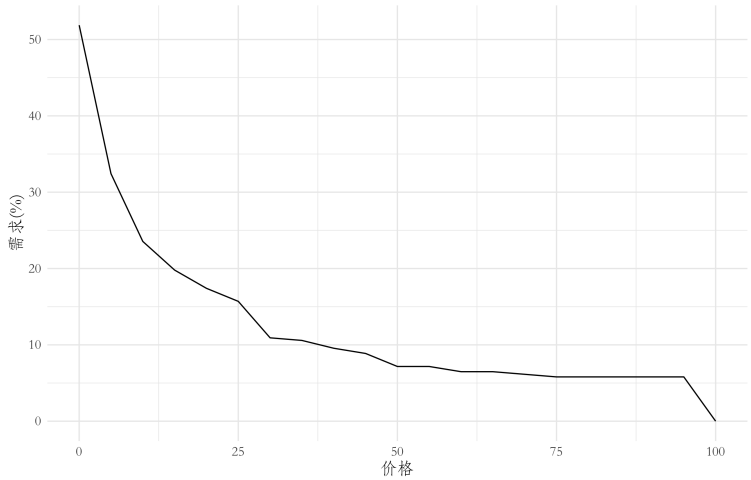
| | Most Helpful | | | Least Helpful | | |
|---|---|---|---|---|---|---|
| Task Category | 1–2 tasks | 3–4 tasks | 5+ tasks | 1–2 tasks | 3–4 tasks | 5+ tasks |
| Store Basics | 33.9 | 38.2 | 33.3 | 30.0 | 26.1 | 33.8 |
| Traffic & Promotion | 46.2 | 43.5 | 46.9 | 31.2 | 31.1 | 40.0 |
| Purchase Conversion | 9.4 | 11.8 | 7.4 | 9.1 | 12.2 | 8.7 |
| Customer Service | 4.6 | 2.7 | 8.6 | 13.0 | 15.6 | 8.7 |
| User Operations | 5.8 | 3.8 | 3.7 | 16.6 | 15.0 | 8.7 |
| *N* | 519 | 186 | 81 | 494 | 180 | 80 |

*Notes:* Online survey of training participants (May 2019). Respondents identified the most and least helpful task categories. Columns show proportions by engagement level (number of tasks completed). Store Basics: deposit, listing, decoration. Traffic & Promotion: free traffic, paid advertising, content marketing. Purchase Conversion: pre-sales, discounts. Customer Service: CS tools, logistics. User Operations: repeat customers, fan engagement.

### Most and Least Helpful Task Categories

**Willingness to Pay for Training** A separate survey wave ($n = 676$) elicited sellers' willingness to pay (WTP) for the training using two methods: an open-ended Van Westendorp price sensitivity meter and a discrete choice list with ascending price points (5–100 RMB/month). Figure B1 plots the demand curves. The median WTP is approximately zero—most sellers would not pay for training at any positive price—but there is a long right tail: roughly 20% would pay 20 RMB/month and 5–8% would pay up to 75 RMB/month. WTP is higher among sellers who treat e-commerce as their main business and those with greater prior investment (Figure B2), consistent with more committed sellers placing higher value on the program.

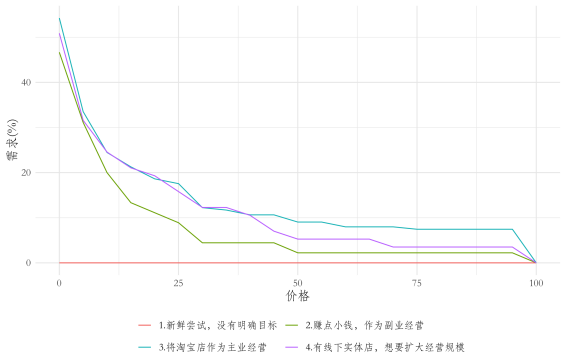
Figure B1: Willingness to Pay for Training: Demand Curve



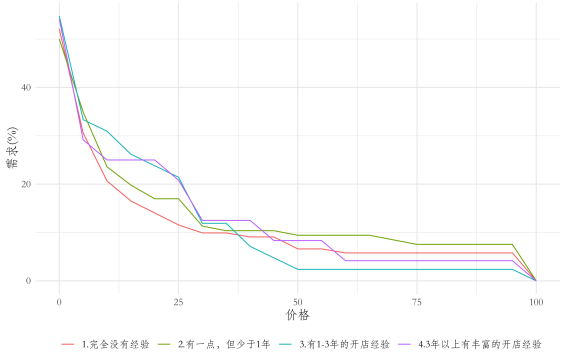

*Notes:* Demand curve (complementary CDF) from the discrete choice elicitation method. X-axis: monthly price in RMB; Y-axis: share of respondents willing to pay at least that price.

Figure B2: WTP by Business Goal and E-Commerce Experience

(a) By Business Goal




(b) By E-Commerce Experience




*Notes:* Demand curves from the discrete choice elicitation, split by seller characteristics. I
business goal (trying new things, side income, main business, expand offline store). Panel
(none, <1 year, 1–3 years, 3+ years).

# C Additional Tables and Figures [For Online Publication]

**Heterogeneous Treatment Effects** Is the training particularly effective for certain types of new sellers? We examine the heterogeneous impact of the training by sellers' characteristics in the baseline. Because the content of the training mainly targets basic operations and marketing, we expect sellers with limited previous exposure to e-commerce to benefit more from the training, as it helps to close their knowledge gap. We characterize sellers from the following dimensions: registration type, gender (if registered as individuals), location, whether any products were posted on the first day and whether the store is registered as a B2C store on the first day[21]. Since no sellers took up tasks on the first day of entry, we consider listing products and sellers' registration type on the first day as part of pre-treatment characteristics. We then estimate the following specification:

$$Y_{imcs} = \beta Treatment_i + \gamma Type_i + \delta Treat_i \times Type_i + \alpha_m + \alpha_c + \alpha_s + \varepsilon_{imcs} \tag{8}$$

where as before $Treat_i$ denotes assignment to treatment or control group and $Type_i$ specify whether sellers have characteristics aforementioned. $\delta$ captures the heterogeneous treatment effects on different types of sellers. Table C1 summarizes the results on log revenues by seller types. Overall, estimated $\beta$ in these specifications have similar magnitudes as estimates using equation 1, but there are no differential treatment effects by seller types, gender, actions on the first day of entry on revenues. The slightly surprising result is that there is no differential treatment effect for sellers with different preparedness levels. Compared to the rest, sellers who post products on the first day of entry could be better prepared or more experienced. Hence, these sellers might find the basic part of the training less useful, yet we do not find such results[22].

To evaluate the impact of the offline business environment, we group the sellers based on their registered locations[23]. Table C2 presents the results on heterogeneous treatment effect by seller locations on log monthly revenues. There are significant variations in average performance for sellers from different parts of the country. Sellers from the southern coastal provinces significantly outperform the rest, while those coming from the less-developed western part lagged. Training is less helpful for sellers from less developed regions, as these sellers are less likely to take-up the training (see Table C5)[24].

[21] To register as a B2C store, potential sellers must obtain formal approval from the platform. The minimum requirements include having a brand name and a formally registered firm. 97% of sellers in the sample are registered as C2C stores. C2C stores can be converted to B2C stores later on. Among all sellers that eventually become B2C stores, 66.7% of them converted later on, and sellers in the treatment group are more like to convert.

[22] We also do not find training to be more useful for sellers who post products after the first day of entry.

[23] The location information on ID cards for individual sellers may not reflect where the sellers reside at the moment because the location indicates ID card holder's birthplace rather than current residence. The internal migration patterns imply that we are under-counting sellers living in the coastal provinces, as these provinces are major destinations of migration. Similarly, the firm's registered locations might not be the same as where the firms operate, but in this case, the direction of the bias is unclear.

[24] Training participants are significantly less likely to be sellers from the western provinces.

Table C1: Treatment Effect Heterogeneity on Sellers' Basic Types

| | Registration Type | | Post Products | | B2C Sellers |
|---|---|---|---|---|---|
| | Female | Firm | First Day | Later Days | |
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.016* | 0.009 | 0.014** | 0.011* | 0.016** |
| | (0.008) | (0.006) | (0.007) | (0.006) | (0.006) |
| Seller Type | −0.494*** | 1.504*** | −0.130*** | 1.265*** | 3.842*** |
| | (0.006) | (0.009) | (0.008) | (0.008) | (0.050) |
| Treatment × Seller Type | 0.003 | 0.015 | 0.013 | 0.014 | 0.077 |
| | (0.012) | (0.017) | (0.015) | (0.013) | (0.098) |
| Observations | | | 6,409,062 | | |
| Adjusted $R^2$ | 0.138 | 0.174 | 0.132 | 0.163 | 0.147 |

*Notes:* Standard errors clustered by seller. All regressions include month, cohort and initial sector fixed effect. Dependent variable is monthly revenues in log scale after adding one to base level. The interaction variables are indicators for whether or not sellers are females, are registered as firms, post products on the very first day of entry or during some later days and lastly whether or not sellers register as B2C sellers. * significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C2: Treatment Effect Heterogeneity by Sellers' Registered Location

| | Beijing Vicinity | Resource-Oriented | Northeast | Coastal South | Central | West |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.015** | 0.018*** | 0.019*** | 0.014* | 0.015** | 0.019*** |
| | (0.007) | (0.006) | (0.006) | (0.007) | (0.007) | (0.007) |
| Location | −0.039*** | −0.371*** | −0.273*** | 0.470*** | −0.128*** | −0.642*** |
| | (0.008) | (0.013) | (0.012) | (0.006) | (0.007) | (0.007) |
| Treatment × Location | 0.011 | −0.042 | −0.048* | 0.006 | 0.008 | −0.013 |
| | (0.017) | (0.026) | (0.025) | (0.013) | (0.014) | (0.014) |
| Observations | | | 6,409,062 | | | |
| Adjusted $R^2$ | 0.132 | 0.132 | 0.132 | 0.138 | 0.132 | 0.136 |

*Notes:* Standard errors clustered by seller. All regressions include month, entry date and main industry fixed effect. Dependent variable is monthly revenues in log scale. Indicators are sellers registration locations clustered into different regions. Beijing Vicinity includes Beijing, Tianjin, Heibei and Shandong; resource-oriented provinces include Shanxi, Neimenggu, Gansu and Ningxia; northeastern provinces are Heilongjiang, Jilin and Liaoning; coastal southern provinces are Jiangsu, Shanghai, Zhengjiang, Fujian, Guangdong and Hainan; central provinces are Anhui, Jiangxi, Henan, Hubei and Hunan; western provinces are Tibet, Xinjiang, Yunnan, Guangxi, Sichuan, Chongqing, Guizhou, Shaanxi and Qinghai. * significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C3: Example of Tasks and Their Classifications

| **Task** | **Indicator** | **Area of Focus** | **Function** | **Type** |
|---|---|---|---|---|
| Acquire customers' reviews | reviews | ratings | knowledge | outcome |
| Acquire free traffic | visitors from search channel | marketing | knowledge | outcome |
| Choose proper promotion products | payment received | basic | knowledge | outcome |
| Complete pricing and inventory information | payment received | basic | knowledge | outcome |
| Complete an order | payment received | basic | knowledge | outcome |
| Expand base of followers | followers | customers | knowledge | outcome |
| Improve "add to shopping cart" | add to cart | marketing | knowledge | outcome |
| Improve buyer review section | reviews | basic | reminder | outcome |
| Improve conversion rate: inqury | conversion | service | knowledge | outcome |
| Improve conversion rate: make payment | conversion | marketing | knowledge | outcome |
| Improve fans' engagement | followers | customers | reminder | outcome |
| Improve payments from returning customers | payments received | customers | knowledge | outcome |
| Engage with customers via microblog | followers' activities | customers | reminder | action |
| Improve ratings on customer service | ratings | ratings | knowledge | outcome |
| Improve ratings on product quality | ratings | ratings | knowledge | outcome |
| Decorate store front-page on app | decoration | basic | reminder | action |
| Improve tag/bookmark rates | bookmarked | marketing | knowledge | outcome |
| Improve ratings on delivery | ratings | service | knowledge | outcome |
| Optimize products' titles | traffic | marketing | knowledge | outcome |
| Participate in official sales events | sign-up | marketing | knowledge | action |
| Participate in product position auctions | paid ads | marketing | knowledge | action |
| Pay security deposits | deposits | basic | reminder | action |
| Post products on store page | number of products | basic | reminder | action |
| Set up automated position auction campaign | paid ads | marketing | knowledge | action |
| Setup bonus after purchase | bonus | basic | reminder | action |
| Setup free return and refund | return policy | basic | reminder | action |
| Setup free trial / offer free samples | free trial | basic | reminder | action |
| Setup store coupons and discount | coupons | basic | reminder | action |
| Shorten average time to delivery | delivery time | service | knowledge | outcome |
| Shorten response time to customer inquires | response time | service | knowledge | outcome |
| Subscribe to storefront customization | "hip-store" subscription | basic | reminder | action |
| Upload videos for product descriptions | videos | basic | reminder | action |
| Upsell to increase per-consumer spending | avg. order size | marketing | knowledge | outcome |

*Notes:* Listed tasks are a subset of all tasks offered to the sellers. Over time service providers also created more tasks and the platform invested in streamlining and regularizing the tasks offered. Tasks are ordered in sequence of priorities. Each task is triggered by a particular indicator. For the outcome based tasks, comparisons are made with other sellers in the same industry. Tasks are classified based on main area of focus, the functions they served and how they are evaluated.

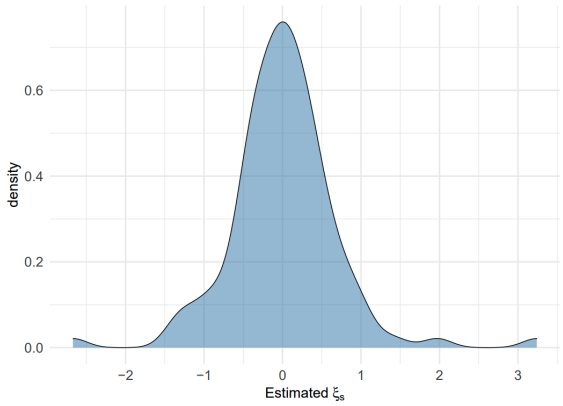
Table C4: Prediction Precision: Traffic Matching Function

| Lagged Traffic Degree | Lagged Conv. Rate Deg | FE | $R^2$ | RMSE | MAE |
|---|---|---|---|---|---|
| 1 | 1 | Y | 0.63 | 1.63 | 1.18 |
| 1 | 1 | N | 0.6 | 1.7 | 1.23 |
| 2 | 1 | N | 0.6 | 1.69 | 1.22 |
| 2 | 1 | N | 0.61 | 1.68 | 1.21 |
| 3 | 2 | N | 0.6 | 1.69 | 1.22 |

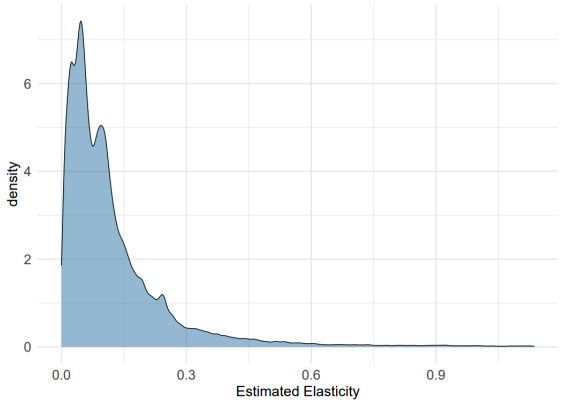
*Notes:* Table shows measure of prediction's precision with different specifications on current period traffic. Precision calculated on out-of-sample data.

Figure C1: Cross-Sector Differences in Structural Parameters

(a) $\xi_s$




(b) $\alpha_s$




*Notes:* These figures plot the distribution of two key latent parameters from our model estima
sectors. $\xi_s$ measures sector-level residual in unobserved seller quality, while $\alpha_s$ measures t
average price elasticity is $-0.22$, which is smaller than typical estimates in the literature be
consideration sets that already restrict the price range.

Table C5: Balance Check and Selective Participation
(at time of registration and treatment assignment)

| | Full Sample | | | Sellers in Treatment Group | | |
|---|---|---|---|---|---|---|
| | Treatment | Control | (1) - (2) Difference | Participants | Non Participants | (4) - (5) Difference |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Is Firm | 0.268 | 0.264 | 0.004*** | 0.275 | 0.266 | 0.008*** |
| | (0.443) | (0.441) | (3.33) | (0.446) | (0.442) | (3.18) |
| Female Owner | 0.455 | 0.453 | 0.002 | 0.442 | 0.458 | -0.16*** |
| (among individual sellers) | (0.497) | (0.497) | (1.27) | (0.497) | (0.498) | (4.61) |
| Region: Coastal South | 0.435 | 0.434 | 0.001 | 0.518 | 0.413 | 0.105*** |
| | (0.496) | (0.496) | (0.28) | (0.500) | (0.492) | (3.573) |
| Region: West | 0.118 | 0.118 | 0.0001 | 0.089 | 0.126 | -0.038*** |
| | (0.323) | (0.323) | (0.34) | (0.284) | (0.332) | (21.41) |
| List Products on the 1st Day | 0.213 | 0.212 | 0.0005 | 0.245 | 0.204 | 0.04*** |
| | (0.409) | (0.409) | (0.44) | (0.430) | (0.403) | (16.04) |
| # Listed Products | 1.539 | 1.535 | 0.004 | 2.307 | 1.342 | 0.965*** |
| | (2.166) | (2.164) | (0.65) | (1.988) | (2.167) | (80.81) |
| Traffic | - | - | 0.006 | - | - | 1.527*** |
| | (2.289) | (2.283) | (0.91) | (2.259) | (2.191) | (115.35) |
| Conversion Rate | 0.051 | 0.051 | 0.0002 | 0.055 | 0.050 | 0.006*** |
| | (0.161) | (0.162) | (0.52) | (0.120) | (0.171) | (7.34) |
| | (3.312) | (3.304) | (1.144) | (3.726) | (3.048) | (100.66) |
| Revenues | - | - | 0.01 | - | - | 2.134*** |
| | (3.312) | (3.304) | (1.144) | (3.726) | (3.048) | (100.66) |
| Observations | 177,026 | 535,092 | 712,118 | 36,189 | 140,837 | 177,026 |

*Notes:* The timing of the assignment limits the available baseline information to variables collected during the registration and actions taken on the first day of entry, most of which is listed here. Columns 1, 2, 4, and 5 present means and standard deviations (in parentheses). Columns 3 shows the difference in means across the treatment and control group (training participants and non-participants) in the full sample with the corresponding t-statistics in parentheses. Columns 6 shows the comparison between training participants and non-participants. For the second row (female owner) the value in parentheses shows the estimate with a further sample restriction that includes only those not registered as firms. Participation is defined as having taken up any tasks during the nine-month period. Traffic, conversion rate, revenues and number of product posted are for the first month outcomes. * significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C6: Intent-to-Treat Effects on Revenues (Robustness Check on Revenue Definition)

| | | | | Raw / Winsorized | | |
|---|---|---|---|---|---|---|
| | Indicator | Log | I.H.S. | 99th | 99.5th | 99.9th |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.002** | 0.017*** | 0.018*** | 25.436** | 26.175 | 85.084 |
| | (0.001) | (0.006) | (0.007) | (12.478) | (22.308) | (54.321) |
| Dep Var Mean | 0.19 | – | – | – | – | – |
| Observations | | | 6,409,062 | | | |
| Adjusted $R^2$ | 0.152 | 0.132 | 0.134 | 0.027 | 0.024 | 0.019 |

*Notes:* Dependent variables are total revenues in the seller sample. All regressions include cohort, initial sector and relative month fixed effect. In column 3, the revenue is transformed with inverse hyperbolic sine. We calculate thresholds for winsorization by the relative month since registration. Dependent variable means calculated with sellers in the control group.* significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C7: ATT Results with Alternative Take-up Definition: the Number of Tasks Taken-up

| | Any Revenues | Revenues | # Visitors | # Buyers | Conversion Rate |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| # Tasks | 0.028*** | 0.284*** | 0.208*** | 0.139*** | 0.002 |
| | (0.011) | (0.088) | (0.077) | (0.043) | (0.002) |
| Dep Var Mean | 0.19 | – | – | 0.57 | 0.04 |
| Observations | 6,409,062 | 6,409,062 | 6,409,062 | 6,409,062 | 2,593,762 |
| $R^2$ | 0.155 | 0.136 | 0.209 | 0.109 | 0.049 |

*Notes:* The first stage specification is the following:

$$Tasks_{it} = \sum_{k=1}^{9} \gamma_k Treatment_i \mathbf{I}_{t=k} + \alpha_t + \alpha_s + \alpha_c + \varepsilon_{it}$$

where $Tasks_{it}$ is number of tasks taken-up for seller $i$ during month $i$. $\mathbf{I}_{t=k}$ is an indicator for being in the $k$th month since registration. All specifications are two-stage least square results using treatment assignment interacting with dummies for relative month as the instrument. Number of visitors, number of buyers and revenues (total payments received) are monthly total in log after adding one to the level. All regressions include cohort, relative month and initial industry fixed effect. Standard errors clustered at seller level. Dependent variable means calculated with sellers in the control group. * significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C8: OLS Results on Main Outcomes

| | Log Revenues | Any Revenues | Revenues | Log # Visitors | | Log # Buyers | | Conversion |
|---|---|---|---|---|---|---|---|---|
| | | | | Full | Have Visitors | Full | Have Visitors | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Take-up Tasks | 1.847*** | 0.223** | 0.771*** | 1.735** | 0.703*** | 0.825*** | 0.686*** | 0.005*** |
| | (0.017) | (0.002) | (0.019) | (0.014) | (0.015) | (0.009) | (0.012) | (0.0004) |
| Dep Var Mean | – | 0.19 | – | – | – | 0.57 | 1.4 | 0.04 |
| Sample | Full | Full | Earn Rev. | Full | Have Vis. | Full | Have Vis. | Have Vis. |
| Observations | 1,593,234 | 1,593,234 | 314,376 | 1,593,234 | 646,894 | 1,593,234 | 646,894 | 646,894 |
| Adjusted $R^2$ | 0.193 | 0.204 | 0.109 | 0.274 | 0.128 | 0.159 | 0.097 | 0.047 |

*Notes:* Sample restricted to sellers with access to training. Main explanatory variable is an indicator of take up of at least during the sample period. Traffic (number of visitors), number of buyers and revenues (total payments received) are monthly total in log after adding one to the level. All regressions include cohort, relative month and initial industry fixed effect. Standard errors clustered at seller level. Dependent variable means calculated with sellers in the control group. * significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C9: Treatment Effect on Sellers' Ratings and Refunds

| Variable | Treatment | Dep Var Mean | Variable | Treatment | Dep Var Mean |
|---|---|---|---|---|---|
| **Ratings** | | | **Refunds and Reviews** | | |
| Products | 0.004 | 1.25 | % Refund (Amount) | -0.0004 | 0.21 |
| | (0.004) | | | (0.001) | |
| Service | 0.004 | 1.26 | % Complaints | -0.003 | 0.05 |
| | (0.004) | | | (0.006) | |
| Logistics | 0.005 | 1.26 | Rule Violations | 0.0002 | 0.22 |
| | (0.004) | | | (0.001) | |
| | | | % Good Reviews | -0.00001 | 0.99 |
| | | | | (0.0002) | |

*Notes:* Table presents estimated coefficients $\beta$ on treatment assignment dummy with specification 1. Standard errors clustered by seller. All regressions include month, entry date and main industry fixed effect. Ratings are customer ratings variables that the platform calculates and assignes to sellers based on customers' reviews and ratings. The ratings scale between 0 to 5, on the dimensions of accuracy of product descriptions, quality of customer service and logistics. % refunds calculated as total refunds requested over total payments made. % complaints defined as number of complaints over total number of orders. Rule violations is the frequency that sellers violate the platform's rules, see more details in appendix 4, descriptions of the instruments. Share of good reviews is the share of good reviews out of all the reviews that sellers get. Vast majority of the reviews are positive. * significant at 10% level, ** significant at 5% level and *** significant at 1% level.

Table C10: Treatment Effect on Service Quality Ratings (DSR)

| | Descriptions (1) | Service (2) | Shipping (3) | Logistics (4) | Star Rating (5) |
|---|---|---|---|---|---|
| Treatment | 0.004 | 0.004 | 0.005 | 0.005 | 0.008* |
| | (0.004) | (0.004) | (0.004) | (0.004) | (0.005) |
| Cohort FE | X | X | X | X | X |
| Industry FE | X | X | X | X | X |
| Dep Var Mean | 1.25 | 1.26 | 1.26 | 1.26 | 0.57 |
| Observations | | | 6,409,062 | | |
| Adjusted $R^2$ | 0.181 | 0.181 | 0.181 | 0.181 | 0.121 |

*Notes:* Table presents estimated ITT coefficients on treatment assignment with specification 1. Dependent variables are the platform's Detailed Service Ratings (DSR) on four dimensions (descriptions, service, shipping, logistics) and the overall seller star rating. Standard errors clustered by seller. All regressions include cohort and initial sector fixed effects. * significant at 10% level, ** at 5% and *** at 1%.

Table C11: Treatment Effect on Conversion Rates and Consumer Engagement

| | CR All | CR WL | CR App | Stay Time | Use CS |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Treatment | 0.0001 | −0.0002 | −0.0002 | 0.003 | 0.001*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.003) | (0.001) |
| Cohort FE | X | X | X | X | X |
| Industry FE | X | X | X | X | X |
| Dep Var Mean | 0.04 | 0.04 | 0.03 | 1.01 | 0.1 |
| Observations | 2,593,762 | 2,593,762 | 2,593,762 | 6,409,062 | 2,685,771 |
| Adjusted $R^2$ | 0.048 | 0.043 | 0.041 | 0.251 | 0.083 |

*Notes:* Table presents estimated ITT coefficients on treatment assignment with specification 1. "CR All" is the overall conversion rate; "CR WL" and "CR App" are conversion rates on the web and app channels, respectively. "Stay Time" is log minutes spent on the seller's page. "Use CS" is an indicator for the buyer initiating customer service contact. Standard errors clustered by seller. All regressions include cohort and initial sector fixed effects. * significant at 10% level, ** at 5% and *** at 1%.

Table C12: Treatment Effect on Pricing, Category Switching, and Other Outcomes

| | **Pricing** | | **Category Switching** | | | |
|---|---|---|---|---|---|---|
| | Per Buyer | Per Product | Switch | Add | Remove | Any Change |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Treatment | 0.005 | 0.0003 | −0.00001 | −0.0001 | 0.00002 | −0.0003 |
| | (0.006) | (0.006) | (0.0003) | (0.0001) | (0.00002) | (0.0004) |
| Dep Var Mean | 4.39 | 3.83 | 0.05 | 0.02 | 0 | 0.09 |
| N | 1,253,284 | 1,253,284 | 6,409,062 | 6,409,062 | 6,409,062 | 6,409,062 |

*Notes:* Table presents estimated ITT coefficients on treatment assignment with specification 1. Columns 1–2: average transaction price per buyer and per product (log), restricted to sellers with nonzero revenues. Columns 3–6: indicators for category switching, adding, removing, or any change in product categories. Standard errors clustered by seller. All regressions include cohort and initial sector fixed effects. * significant at 10% level, ** at 5% and *** at 1%.

Table C13: Buyer Characteristics by Seller Treatment Status

| | Size of Set | Buyer Rating | Spending Last Week | Page View Last Week |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Control | 1.268*** | 0.028*** | 0.198*** | 0.021*** |
| | (0.006) | (0.005) | (0.006) | (0.003) |
| Treatment | 1.392*** | 0.060*** | 0.232*** | 0.014*** |
| | (0.007) | (0.005) | (0.006) | (0.003) |
| Control − Treatment | −0.12*** | −0.032*** | −0.034*** | 0.0067* |
| | (0.0086) | (0.0063) | (0.0077) | (0.0036) |
| Observations | | | 2,676,501 | |
| Adjusted $R^2$ | 0.440 | 0.146 | 0.152 | 0.132 |

*Notes:* Each column regresses a buyer characteristic on indicators for visiting a control or treatment new seller, with sector and cohort fixed effects. "Size of Set" is the number of sellers in the buyer's consideration set. "Buyer Rating" is the platform's buyer quality rating. "Spending Last Week" and "Page View Last Week" are the buyer's activity in the week prior to the search session. The bottom row reports the difference between control and treatment coefficients with the *t*-test in parentheses. * significant at 10% level, ** at 5% and *** at 1%.

Figure C2: Training Participation: Tasks Take-up and Completion

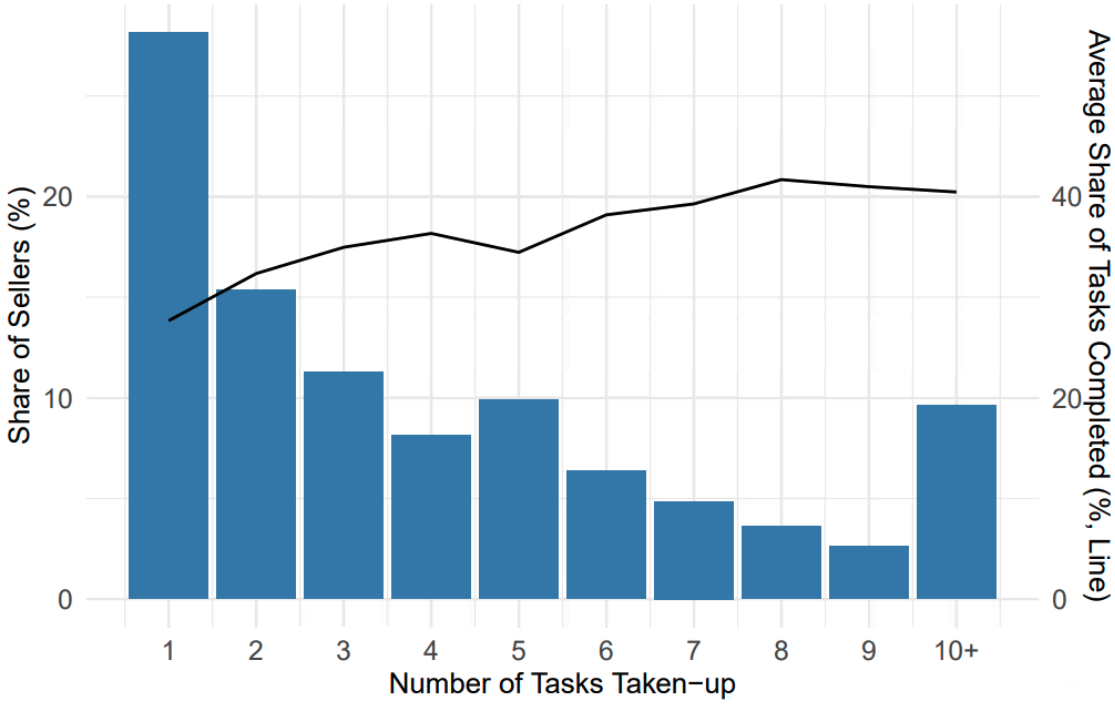



Figure C3: Procedures of Constructing the Consumer Sample

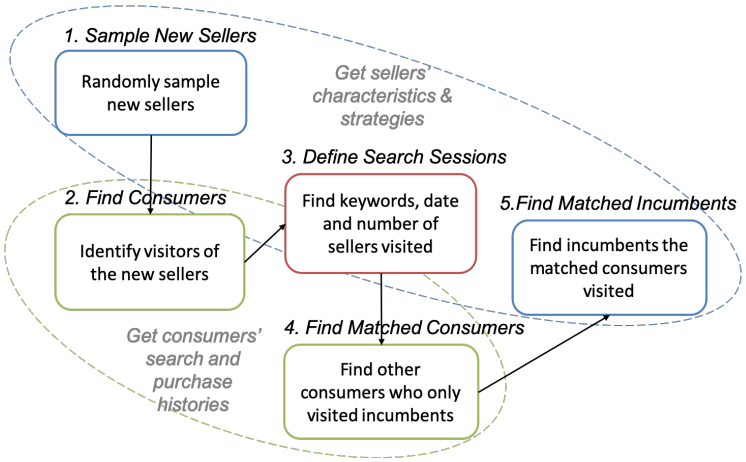
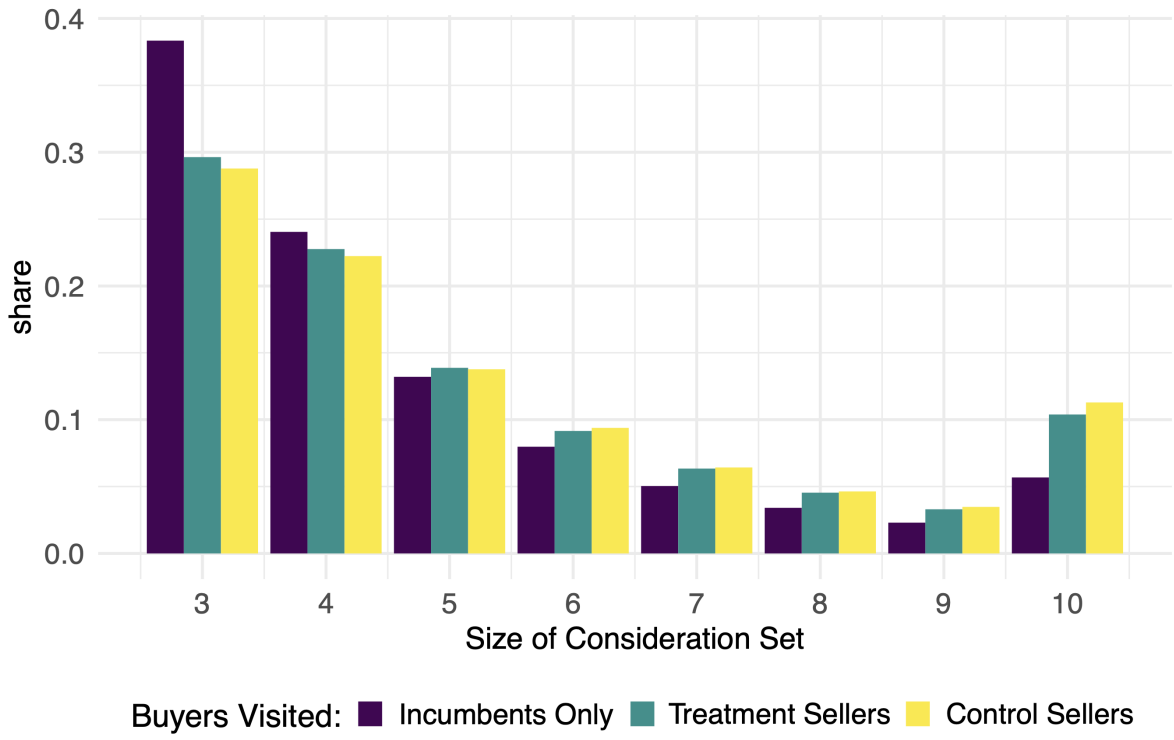
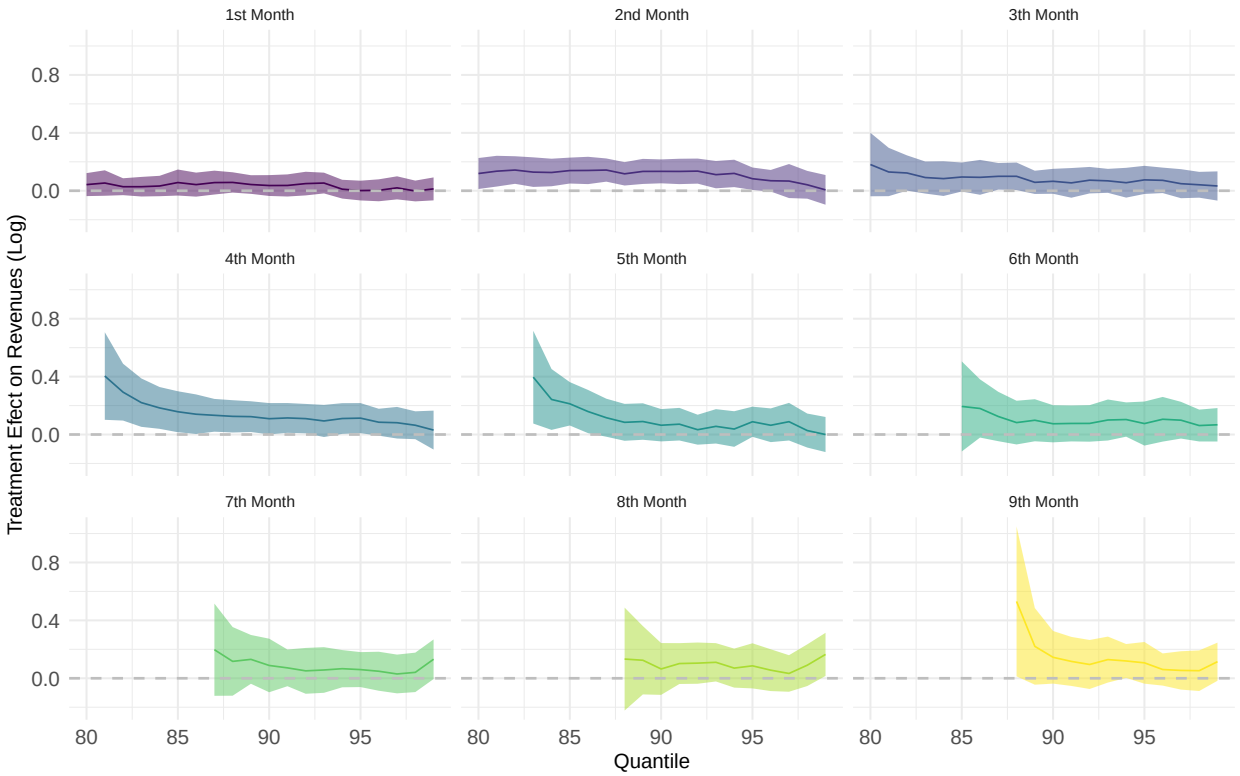


Figure C4: Distribution of the Size of Consideration Sets by Seller Type




*Notes:* Density histograms of the size of consideration sets (number of sellers considered) are plotted by their composition. "Incumbent Only" includes sets that consist only of incumbent sellers. "Control Sellers" and "Treatment Sellers" include sets with at least one control-group or treatment-group seller. We focus on sizes three and above because those that contain one or two sellers have implausibly low purchasing rates, indicating data contamination from non-search channels that are largely irrelevant for new sellers.

Figure C5: Quantile Treatment Effects on Revenue by Month




*Notes:* Quantile treatment effects (QTE) on log revenue, estimated at every percentile from the 80th to 99th on the full balanced sample. Each panel corresponds to a month since seller registration (months 1–9). Solid lines are point estimates; shaded regions show 95% confidence intervals. The dashed horizontal line marks zero. Fewer than 20% of new sellers ever earn positive revenue, so QTE below the 80th percentile is mechanically zero. Effects are concentrated at the 80th–90th percentiles and peak in month 4, consistent with the timing of treatment effects on traffic (Figure 4).