Received: 4 October 2019 | Revised: 18 August 2020 | Accepted: 18 December 2020

DOI: 10.1111/jems.12420

SPECIAL ISSUE

Journal of Economics & Management Strategy WILEY

Evaluating entrepreneurship training: How important are field experiments for estimating impacts?

Robert W. Fairlie[1,2]

[1]Department of Economics, University of California, Santa Cruz, California, USA

[2]NBER, Cambridge, Massachusetts, USA

Correspondence
Robert W. Fairlie, Department of Economics, Engineering 2 Bldg, University of California, Santa Cruz, CA 95064, USA.
Email: rfairlie@ucsc.edu

Abstract

Governments and donors around the world spend billions of dollars subsidizing entrepreneurship training programs. Unbiased evaluation estimates are paramount to understanding whether subsidies and investments in these programs are justified. The goal of this paper is to compare nonexperimental to experimental methods for evaluating the effectiveness of entrepreneurship training programs that provide training in management, marketing, accounting, legal, and other aspects of starting and running businesses. For the comparison, I take advantage of an unprecedented setting in which experimental and nonexperimental estimates are derived from the same underlying population of study participants. The Growing America through Entrepreneurship field experiment provides a uniquely large sample allowing for a separate nonexperimental analysis using only the control group. Experimental estimates indicate null effects of entrepreneurship training on business outcomes: business ownership, sales, and employment. In contrast, nonexperimental estimates using an extremely rich set of controls, including typically unobservable characteristics, such as previous family business experience, credit problems, and personality traits, find large, positive, and statistically significant effects. The nonexperimental estimates range from 21 to 22 percentage points (pp) for business ownership, $1300–2000 for average monthly sales, and 4–6 pp for hiring any employees at 1.5- and 5-year follow-ups. Nearest-neighbor and propensity-score matching models using detailed individual characteristics provide similarly large, positive, and statistically significant effects of entrepreneurship training on business outcomes. The findings have implications for choosing evaluation methods of government programs and provide evidence of positive selection bias which has more general implications on the scalability of entrepreneurship training programs to broader populations.

1 | INTRODUCTION

Expenditures on subsidizing entrepreneurship training are large and growing rapidly around the world (OECD, 2017). In the United States alone, there exist more than 1000 Small Business Administration (SBA)-subsidized Small Business Development Centers (SBDCs) and at least 800 other not-for-profit programs providing entrepreneurship training,

which includes counseling, consulting, and classroom training in management, marketing, accounting, legal, and other aspects of starting and running businesses.[1] Justifying these expenditures, administrators often document how many participants are served, how many businesses are created, and/or how many employees those businesses hired.[2] Interestingly, however, the few randomized control trial (RCT) evaluations of entrepreneurship training programs to date show mostly null effects (Benus & Michaelides, 2010; Davis et al., 2013; Fairlie et al., 2015).[3] Even a very expensive training program targeted at youth (12,000 euros) had mixed success when evaluated with an RCT (Astebro & Hoos, 2016).

Are entrepreneurship training programs susceptible to selection bias because training is opted into and not exogenously determined? This paper provides a novel comparison of experimental and nonexperimental methods in evaluating the effectiveness of entrepreneurship training and development programs. Estimates might differ due to either positive or negative selection bias or any other unobservable characteristic of the individual correlated with training receipt and business outcomes. Positive bias might occur, for example, if the individuals most willing to invest in training programs are the ones that benefit the most from these programs. Highly motivated individuals might have also started successful businesses without the help of training programs. On the other hand, selection might be negative if the marginal, most in-need individuals are the ones who sign up for entrepreneurship training and development programs. In either case, estimates of the effectiveness of training programs for these selected individuals might be very different than estimates of the overall effectiveness of programs. Improving our understanding of the potential bias, either positive or negative, from nonexperimental estimates of the effectiveness of entrepreneurship training programs is of paramount importance for policy as enthusiasm and expenditures on these programs continue to rise around the world.

To compare experimental and nonexperimental estimates, I take advantage of an unprecedented setting in which experimental and nonexperimental estimates are derived from the same underlying sample of study participants. To the best of my knowledge, this approach has not been previously taken to estimate the potential bias from nonexperimental estimates of entrepreneurship training and development programs. I analyze the largest RCT on entrepreneurship training ever conducted: GATE. The uniquely large size of this field experiment permits a separate nonexperimental analysis with only the control group, which draws from the same underlying participant population that applied for the program and consists of roughly 2000 observations. Using this sample, I estimate training effects to compare to the experimental estimates using the full sample. Another advantage of the study is that the surveys conducted for GATE provide extensive information on study participants, including detailed demographic, baseline income and business ownership, and personality characteristics. The availability of detailed controls removes the concern that the nonexperimental evaluation techniques used, such as ordinary least squares (OLS), nearest-neighbor matching, and propensity-score matching, are missing some key individual characteristics to use as controls or for matching observations. In addition to estimating the "bias" from these techniques the comparison provides some information on whether individuals positively or negatively select into entrepreneurship training, which is an interesting and important question in its own right. Information on the direction of selection is useful for policymakers considering broadening the reach of existing entrepreneurship training programs.

GATE was a longitudinal field experiment conducted by the US Department of Labor (DOL) and the SBA in which free entrepreneurship training was randomly offered to individuals interested in starting or improving a business. More than 4000 individuals applied for a limited number of slots at 14 different SBDCs and nonprofit community-based organizations (CBOs) located across seven sites in three states. SBDCs and CBOs are the predominant providers of entrepreneurship training services in the US market. Subjects assigned to the treatment group were offered an array of best-practice training services whereas subjects assigned to the control group were not offered any free services. Follow-up surveys at 6, 18, and 60 months after treatment assignment yield a rich set of outcome measures. The 60-month follow-up provides rare measures of long-run outcomes.

[1]SBDCs exist in all 50 states, and are administered and funded through partnerships between the SBA and public colleges and not-for-profits. See http://www.sba.gov/content/small-business-development-centers-sbdcs for a directory of SBDCs, Aspen Institute (2012) for information on other nonprofit programs, and European Commission (2016) for a description of programs in the European Union.

[2]For example, Chrisman's (2004, 2017) ongoing annual evaluations of national SBDC contributions focus on the "difference between the growth rate of clients and weighted average growth rates for all businesses was used to estimate the incremental or marginal changes in sales and employment growth of the sample." See also Gu et al. (2008) for a discussion.

[3]One exception is a small earlier demonstration experiment conducted in Washington and Massachusetts in 1992 (Benus et al., 1994). The study found positive impacts on self-employment, total earnings, and job creation from a training assistance program that allowed for concurrent unemployment insurance (UI) benefit payments and a lump-sum benefit payment. Another exception is that the second round of experimental Growing America through Entrepreneurship (GATE) programs targeting rural dislocated workers in North Carolina found training to increase self-employment roughly 3 years later (Davis et al., 2013). The other experimental GATE II program targeted older dislocated workers in Virginia and found no increase in self-employment 2 years later. For research using nonrandomized approaches to identifying effects of self-employment training programs, see, for example, Kosanovich and Fleck (2001) and Rodriguez-Planas (2010).

Although business ownership rates increased 20–25 percentage points (pp) among recipients of entrepreneurship training from the GATE program, estimates of positive effects on business ownership rates and other outcomes disappear once the control group is used. Experimental estimates of average treatment effects suggest that entrepreneurship training has null impacts on business ownership, sales, and employment. The only exception is that entrepreneurship training increases the likelihood of business ownership in the short run (by 13 pp 6 months after random assignment). However, by the 18-month follow-up survey wave the effect disappears. Using the same underlying sample of study participants but drawing from the control group, nonexperimental estimates are different. I find that receipt of entrepreneurship training has large, positive, and statistically significant "effects" on business ownership, sales, and employment. For example, at the 1.5-year follow-up the nonexperimental estimates indicate increases of 21 pp for business ownership, $1500 for average monthly sales, and 4 pp for hiring any employees. Nearest-neighbor and propensity-score matching models using detailed characteristics provide similarly large, positive, and statistically significant effects of entrepreneurship training on business outcomes. Additionally, the experimental versus nonexperimental estimate comparison draws a similar conclusion when focusing on one-on-one counseling services within entrepreneurship training and whether the applicant was self-employed or not self-employed at baseline. These findings indicate that positive selection into entrepreneurship training is an important concern and needs to be addressed in evaluations of programs that provide entrepreneurship training and related services.

The rest of the paper proceeds as follows. Section 2 provides more details on GATE, including its research design and implementation, and a treatment/control group balance check. Section 3 presents the nature of the training services received by subjects. Section 4 presents experimental estimates. Section 5 presents nonexperimental estimates, and results for one-on-one counseling services and by baseline self-employment. Section 6 concludes.

2 | EXPERIMENTAL DESIGN AND STUDY PARTICIPANTS

The field experiment used for the analysis is the GATE project. The experimental evaluation was designed and implemented by the US DOL in partnership with the US SBA. The goal of GATE was to assist would-be-entrepreneurs to create and/or grow businesses.[4] The evaluation was designed to capture existing representative training providers and recipients. The intake/treatment phase of the evaluation ran from September 2003 to July 2005 in seven sites that captured both urban and rural areas. Follow-up surveys were conducted at 6-, 18-, and 60-month postrandom assignment for each participant.

Individuals entered the study by completing an application process for a standard offer of free training from one of 14 established providers.[5] The application process started with an orientation meeting at one of 21 One-Stop Career Centers in the seven sites. Anyone attending the orientation meeting could then apply by completing and mailing a form with questions on demographics, work and business experience, and the individual's current business or new business idea. Program coordinators randomized applicants to treatment or control with equal probability. Program administrators for each training provider then offered treatment applicants a standard array of free-training services, and told control applicants that the GATE program did not have the capacity to offer them services and did not offer referrals to any other services. Individuals in both treatment and control groups were notified that they would be mailed follow-up surveys.

GATE is the largest-ever randomized evaluation of entrepreneurship training and assistance, with 4197 individuals randomized at baseline. The uniquely large size of the field experiment is essential for creating a large enough sample size for the control group to explore nonexperimental estimates of the effects of entrepreneurship training in addition to experimental estimates.[6]

2.1 | Entrepreneurship training services

GATE training providers were chosen with representativeness of the subsidized training market in mind. Fourteen different, established organizations from seven different sites participated in the GATE study, with a mix of SBA-

[4] See http://www.doleta.gov/projectgate/for more information.

[5] Training providers marketed GATE through several channels: online, on-site electronic kiosks, merchandising, paper materials, direct mail (insert with UI checks), mass media, and referral networks with CBOs.

[6] In contrast, the RCT evaluation of the GATE II program in North Carolina (dislocated rural workers) yielded only 294 observations in the control group (196 with answers to the training receipt question) and the evaluation of the Virginia program yielded only 191 observations in the control group (158 with answers to the training receipt question).

funded SBDCs and nonprofit CBOs, and urban and rural locations that are characteristic of the market. The 14 participating providers deliver services in and around Philadelphia; Pittsburgh; Minneapolis/St. Paul; Duluth, Minnesota; Virginia, Minnesota; Portland, Maine; Lewiston, Maine; and Bangor, Maine (see Bellotti, 2006 for more details). SBDCs and CBOs offer similar services, as detailed below, but differ somewhat in their stated goals. The SBDCs tend to emphasize helping small business owners grow (or start) their businesses to contribute to the local economy, whereas the CBOs tend to emphasize small business ownership as a path to self-sufficiency. Both types of organizations employ experienced business consultants to deliver one-on-one and group trainings.

GATE training was customized for the individual from an array of services offered by the provider, as is typical in the subsidized market.[7] Training began with a one-on-one assessment to produce a service plan that typically combined one-on-one with selected group services. The majority of treatment group individuals then received one-on-one counseling/consulting that was customized to the individual's experience, capability, circumstances, and opportunities. The largest component of training, however, was classroom/group training(s). These targeted a variety of general and specialized topics at different experience levels. Introductory courses/seminars/workshops covered subjects, such as management practices, legal structure, business plans, and marketing. Intermediate and advanced group trainings covered subjects, including managing growth, obtaining financing, legal risks, and personnel issues. More specialized group trainings covered, for example, accounting, information technology, and web-based businesses. Benus et al. (2009) estimate that the total cost of providing training to GATE recipients is $1321 per person.

2.2 | Representation of study participant sample

GATE was designed to estimate treatment effects on recipients who are representative of those served by subsidized training providers. GATE services, like, most subsidized training programs in the United States, were marketed to any individual interested in starting or growing a business.[8] Although there are no sources of nationally representative data on the characteristics of training entrants, a comparison to the characteristics of self-employment entrants is useful as a very rough approximation. To examine the characteristics of self-employment entrants I create panel data from the Current Population Survey (CPS) for similar years. Although the CPS data are usually used as cross-sectional data, panel data can be created from the underlying data files to measure business starts by individuals (Fairlie, 2013; Fairlie & Chatterji, 2013). Table A1 reports the average characteristics of GATE participants and self-employment entrants from the CPS. GATE participants do not differ substantially from self-employment entrants on observable characteristics. For example, 47% of GATE participants is female compared with 42% of self-employment entrants. GATE participants are more likely to be African-American, but less likely to be Latino.[9] The percentage of white and Asian races is similar. Average schooling levels are also reasonably similar with 14 years of schooling among GATE participants and 13 years of schooling among self-employment entrants. Average age and income are roughly similar (although GATE participants are less likely to come from the highest income group). The biggest difference is that 55% of GATE participants is not employed at the time of application, which is expected because participants are seeking entrepreneurship training services. But, self-employment entrants also have a high percentage coming from unemployment (21%) which is consistent with the findings from previous studies showing a strong relationship between unemployment and self-employment entry (Fairlie, 2013; Krashinsky, 2005; Parker, 2018).

In all, the available data suggests that GATE obtained a representative sample of subsidized training recipients. In the choice of sites for training service providers, one of the main goals was to create a representative sample (Bellotti, 2006). Although self-employment entrants are not a perfect comparison group they provide at least some assurance that GATE is capturing a wide range of interested entrants.

[7]For example, sba.gov describes SBDCs as providing "... extensive, one-on-one, long-term professional business advising, low-cost training and other specialized services" (http://www.sba.gov/content/small-business-development-centers-sbdcs).

[8]Some assistance programs target recipients of social insurance. Demonstration programs in Washington and Massachusetts starting in 1989, and Self-Employment Assistance programs in several states starting in 1993, target UI recipients. These programs provide concurrent UI benefits or lump-sum payments, and exempt search requirements for wage/salary jobs to help start self-employment activities (Weigensberg et al., 2017). Similar programs exist in Europe (e.g., Baumgartner and Caliendo 2008). The Self-Employment Investment Demonstration, implemented from 1988 to 1992 in five states, targeted Aid to Families with Dependent Children (AFDC) recipients.

[9]An extensive literature documents and explores the causes and consequences of low rates of business entry and success among African-Americans (see, e.g., Fairlie & Robb, 2007, 2008). Interestingly, there is no evidence that blacks have less preference for becoming self-employed and in fact might have a stronger preference (Koellinger & Minniti, 2006; Walstad & Kourilsky, 1998).

2.3 | Balance check

Table 1 checks for treatment versus control balance on baseline characteristics at both baseline, and each of the three follow-ups (sample composition changes over time due to attrition). Random assignment was not stratified by site, but the top rows show that each site produced roughly 50–50 assignments nevertheless (Columns 1–3). Among the numerous baseline characteristics measured in the application, only one, age, is statistically different between treatment and control. One would expect to find one or two significant differences by chance, and the magnitude of the age difference is small (<1 year). In any case, when estimating treatment effects results are presented both without covariates as well as with controls for a large set of detailed baseline characteristics.

Table 1 also compares treatment and control completion rates and baseline characteristics for each of the three follow-up surveys. The bottom row shows that control group members are more likely to attrit: the completion rate differs by 4–5 pp, on a base of 56–80%, for each follow-up wave. However, despite differential attrition rates overall, we do not find differences in the observable composition of the treatment versus control groups, based on characteristics observed in the baseline.[10] The number of significant differences is about what one would expect to find by chance, and the magnitude of these differences is small.

2.4 | Beyond the balance check

The analysis also relies on the use of the control group as a separate sample to estimate the effectiveness of entrepreneurship training using nonexperimental methods. The balance check discussed above thus provides another test of whether there are differences between these two samples. Experimental estimates use the full sample of treatment and control observations, and nonexperimental estimates use the sample of control observations. Both samples are derived from the same underlying population of study participants who were initially interested in applying for and receiving entrepreneurship training through the program. The only difference between the two groups is random assignment and as the results from Table 1 indicate the two groups are balanced on observables. This is the advantage of using the control group for the nonexperimental analysis instead of a separate sample of individuals not from the original experimental study. That group would likely differ on observables, unobservables, context, and many other characteristics.

2.5 | Econometric model

The base specification for nonexperimental estimates of entrepreneurship training effects is a standard regression of business outcomes on training *receipt*. Thus, the comparison experimental specification for estimating average treatment effects on business outcomes estimates the effects of *receiving* entrepreneurship training (i.e., local average treatment effects [LATEs]) instead of estimating the effects of being *offered* free entrepreneurship training (i.e., "intent-to-treat" effects [ITT]). LATE estimates are needed for the comparison to nonexperimental estimates because they adjust for treatment and control differences in the take-up of actual training. Nonexperimental estimates by necessity focus on the effects of actual training received.

I estimate the first-stage OLS regressions of the form

$$E_{it} = \omega + \gamma X_{\mathrm{ib}} + \pi T_{\mathrm{ib}} + u_{it}, \tag{2.1}$$

where E_{it} measures whether individual i had obtained any training by follow-up survey t,[11] X_{ib} is a vector of baseline covariates (indexed by b for "baseline") reported in Table 1, and $T_{\mathrm{ib}} = 1$ if i was assigned to the treatment group. The second-stage regression for an outcome of interest y_{it}, measured for individual i at time t, is then

$$y_{it} = \alpha + \beta X_{ib} + \Delta \hat{E}_{\mathrm{ib}} + \varepsilon it, \tag{2.2}$$

[10] See Fairlie et al. (2015) for a detailed analysis of the effects of attrition. The experimental results are not sensitive to differential attrition.

[11] Estimates of the first-stage relationships between treatment assignment and intensive margins of training receipt reveal similar results (see Fairlie et al., 2015). Note that there is only one instrument and hence separate effects on extensive and intensive margins cannot be identified.

TABLE 1 Treatment/control comparison of characteristics for GATE experiment

	Baseline			Follow-up Wave 1			Follow-up Wave 2			Follow-up Wave 3		
	Treatment group (1)	Control group (2)	*p* Value for treat–control (3)	Treatment group (4)	Control group (5)	*p* Value for treat–control (6)	Treatment group (7)	Control group (8)	*p* Value for treat–control (9)	Treatment group (10)	Control group (11)	*p* Value for treat–control (12)
Philadelphia	28.7%	27.5%	0.40	26.8%	25.6%	0.43	25.1%	24.0%	0.49	23.1%	22.0%	0.53
Pittsburgh	13.8%	14.6%	0.43	13.7%	14.3%	0.58	14.0%	14.2%	0.82	14.5%	14.4%	0.92
Minneapolis–St. Paul	39.8%	39.0%	0.58	41.1%	39.1%	0.24	42.3%	40.4%	0.29	43.9%	42.0%	0.35
Duluth	4.6%	5.0%	0.54	4.6%	5.1%	0.51	4.7%	5.1%	0.60	5.0%	4.9%	0.99
Maine	13.1%	13.9%	0.48	13.9%	15.9%	0.09	14.0%	16.3%	0.08	13.6%	16.7%	0.03
Female	47.2%	45.7%	0.32	48.5%	46.4%	0.22	48.8%	46.9%	0.31	48.1%	47.1%	0.62
Black	30.5%	30.6%	0.91	29.1%	29.8%	0.65	27.6%	28.3%	0.69	25.3%	26.0%	0.70
Latino	6.2%	5.1%	0.12	6.3%	4.9%	0.09	6.4%	5.1%	0.12	6.4%	5.2%	0.19
Asian	4.6%	4.5%	0.86	3.8%	3.3%	0.42	3.3%	2.9%	0.52	3.1%	2.8%	0.71
Other	7.9%	8.1%	0.80	7.7%	7.6%	0.91	7.4%	7.0%	0.64	7.4%	6.6%	0.47
Not US born	10.0%	10.2%	0.83	8.9%	9.2%	0.81	8.3%	8.7%	0.67	7.1%	8.1%	0.34
Age	42.08	42.77	0.03	42.73	43.42	0.04	43.16	43.81	0.07	43.91	44.16	0.54
Married	48.1%	48.4%	0.81	49.4%	48.6%	0.64	50.2%	49.0%	0.54	51.4%	49.6%	0.38
Has children	46.7%	46.1%	0.68	45.4%	45.1%	0.88	45.4%	44.6%	0.69	44.0%	42.8%	0.58
Highest grade completed	14.39	14.52	0.07	14.53	14.61	0.28	14.59	14.66	0.38	14.75	14.78	0.77
HH income $25,000–49,999	32.6%	33.7%	0.46	33.0%	34.0%	0.56	32.9%	33.4%	0.77	31.9%	34.5%	0.18
HH income $50,000–74,999	17.9%	17.2%	0.55	18.5%	17.5%	0.45	19.2%	17.8%	0.31	20.1%	17.2%	0.06
HH income $75,000–99,999	6.9%	7.2%	0.70	7.1%	7.2%	0.91	7.4%	7.3%	0.92	8.1%	7.4%	0.53
HH income $100,000+	6.3%	7.0%	0.31	6.9%	7.4%	0.56	7.5%	8.0%	0.59	8.8%	8.9%	0.96
Self-emp. at application	18.3%	19.5%	0.33	19.3%	20.4%	0.41	19.8%	21.2%	0.34	20.3%	21.5%	0.48

TABLE 1 (Continued)

	Baseline			Follow-up Wave 1			Follow-up Wave 2			Follow-up Wave 3		
	Treatment group (1)	Control group (2)	*p* Value for treat–control (3)	Treatment group (4)	Control group (5)	*p* Value for treat–control (6)	Treatment group (7)	Control group (8)	*p* Value for treat–control (9)	Treatment group (10)	Control group (11)	*p* Value for treat–control (12)
Has a health problem	8.7%	8.3%	0.63	9.0%	8.9%	0.90	9.1%	8.9%	0.85	8.9%	8.4%	0.69
Has relatives or friends who have been previously SE	70.3%	70.4%	0.93	71.7%	72.0%	0.85	72.9%	72.5%	0.81	73.6%	73.1%	0.78
Ever worked for relatives or friends who are SE	31.7%	32.0%	0.81	31.7%	31.8%	0.96	31.6%	31.7%	0.97	30.9%	31.5%	0.77
Has a bad credit history	45.4%	43.9%	0.34	43.3%	43.2%	0.94	41.8%	41.5%	0.87	38.9%	39.4%	0.79
Currently receiving UI benefits	39.9%	38.1%	0.24	41.1%	39.7%	0.40	42.1%	39.3%	0.12	43.0%	41.1%	0.35
Has health insurance from current employer	16.8%	18.1%	0.26	16.6%	17.5%	0.48	16.6%	17.6%	0.46	16.8%	17.1%	0.84
Autonomy index	1.7%	−1.7%	0.27	−1.1%	−1.9%	0.81	−0.7%	−1.7%	0.79	−2.0%	−4.9%	0.49
Risk tolerance index	−0.2%	0.2%	0.87	2.6%	−1.1%	0.27	1.3%	−2.0%	0.34	−0.7%	−4.4%	0.35
Unemployed at application	55.3%	55.4%	0.92	55.0%	55.5%	0.78	55.5%	54.6%	0.63	55.8%	55.4%	0.85
F-test for all variables			0.56			0.53			0.69			0.80
Sample size	2094	2103		1758	1691		1563	1475		1274	1176	
Percent of baseline sample	100.0%	100.0%		84.0%	80.4%	0.003	74.6%	70.1%	0.001	60.8%	55.9%	0.001

Notes: (1) All reported characteristics are measured at the time of application, before random assignment. (2) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (3) The autonomy index is created from standardizing responses on a scale of 1–5 to whether the statement "I enjoy working independently" is true about themselves. The risk aversion index is created from combining standardized responses to "I'm only willing to take a risk if I am sure everything will work out" and "I am not prepared to risk my savings for my business."

Abbreviations: GATE, Growing America through Entrepreneurship; HH, household; SE, self-employed; UI, unemployment insurance.

where $\hat{E_{ib}}$ is the predicted likelihood of training receipt, and u_{it} and ε_{it} are the error terms. Δ provides an estimate of the LATE. The LATE estimates are essentially scaled-up values of the ITT estimates (which are estimated from the regression of y_{it} on T_{ib}).

The main nonexperimental regressions simply replace predicted training receipt, $\hat{E_{ib}}$, with actual training receipt, E_{it}, in (2.2). These regressions are estimated with only the control group sample.

3 | ENTREPRENEURSHIP TRAINING RECEIPT

Before turning to experimental and nonexperimental estimates of entrepreneurship training I examine training receipt. Given that not all of the treatment groups might have received training and the control group was not restricted from obtaining training elsewhere, it is important for the experimental analysis to first examine whether and how the GATE treatment changed the use of training services. The training receipt differential between the treatment and control groups needs to be large enough to be able to detect treatment effects. Additionally, to estimate nonexperimental training effects for the control group it is important to verify that at least some of the control group received training services on their own creating variation in the training receipt variable.

Table 2 reports entrepreneurship training receipt by the control and treatment groups. Both the treatment and control groups received entrepreneurship training. Among the control group, 44% received any training by Wave 1 (6 months following random assignment). Among the treatment group, 81% received training which is 37 pp higher than the control group.[12] Cumulating across waves, 65% of the control group received any training by 5 years after random assignment. Among the treatment group, 89% received any training. The results indicate that: (i) a relatively large share of the control group received at least some entrepreneurship training, and (ii) there is a large treatment/control group difference in training receipt. The first result is important for the nonexperimental control group analysis, and the second result is important for the experimental analysis.

Turning to hours of training, it is important to distinguish between mean hours and hours per recipient. For the control group, each of the training recipients received, on average, 15 h of training in the first 6 months, and 29 h of training over the 5-year period. Training per recipient was higher in the first 6 months for the treatment group (19 h), but was similar over the full study period, 31 h. For the experimental design, it is important to focus on mean hours unconditional on receipt instead of mean hours per recipient to remove selection concerns. The control group received an average of 6.6 h of training by Wave 1. The treatment group received more than twice the number of hours of training by Wave 1: 15.6 versus 6.6. By the 5-year follow-up mean hour of training receipt for the treatment group was 28 h which was 8.5 more hours than the control group (Wave 10).[13]

Returning to Table 2, note also that the cumulative differences in training hours are driven by the Wave 1 effects; this is unsurprising, given that the sample is comprised of people with a demand for training at the time of random assignment. We would only expect to find differences in training obtained at later horizons if there is a strong complementarity between training obtained in the short run (between random assignment and Wave 1 follow-up) and training obtained later (between Waves 1 and 2, and/or between Waves 2 and 3).

Table 2 also reports evidence on how treatment assignment affects the type/quality of training received. The results suggest that an individual in the treatment group is more likely to obtain customized training; for example, 30 pp more likely to receive one-on-one assistance. Conditional on receiving any training the percentage of the control and treatment groups receiving one-on-one training services, the difference is smaller (44% of control group recipients and 65% of treatment recipients). And, among recipients of customized training the average hours received by the control and treatment groups are similar.

Table A3 reports self-reports by study participants on whether and what parts of training services they found to be helpful. All of the reported estimates condition on training receipt. The results suggest that an individual in the treatment group receiving training is 16 pp more likely to receive help refining their business idea. Among treatment group respondents 52% of

[12]Examining who receives entrepreneurship training, I find only a few characteristics that predict the take-up of training by each follow-up wave. Focusing on the main effects I find some evidence that African-Americans and the more educated are more likely to receive training (see Table A2). Examining differential take-up between the treatment and control groups, I find only a few significant differences. *F*-tests for differential take-up for all covariates do not reject equality in any of the three follow-up waves.

[13]The 8.5 h differential only measures training time and not any "homework or other multiplier." The multiplier could be substantial. In evaluating the effectiveness of a standard 5-credit college course, for example, treatment effects would work through much more than the 30 or so hours of instructional time per term. Similarly, the impact of preventative medical care works through more than the few hours (or minutes) of office visits. Finally, many board members and advisors of small companies only provide a handful of hours of advice or training each quarter. The provision of key bits of information, and how recipients then apply that information to the businesses is the key factor (Fairlie et al., 2015).

recipients reported it "very useful" and 34% "somewhat useful." The comparable proportions for those in the control group who obtained any training are 36% and 41%. Table A3 also reports treatment and control group responses to questions about whether training helped with 12 specific aspects of business planning and operation. (The sample is again comprised of training recipients only.) The treatment group is more likely to respond that training was helpful for each of the 12 aspects.

In all, the evidence suggests that the experiment produced training in substantially greater quantity and quality for treatment relative to control individuals. Hence any null experimental effects are likely due to training that is ineffective per se, rather than to a low-powered identification strategy (Fairlie et al., 2015). The evidence also suggests that, if anything, the control group who received training received lower-quality training than the treatment group, and thus any larger estimated effects of entrepreneurship training using nonexperimental methods with the control group sample are not simply due to higher-quality training.

Turning to the analysis of the comparison between the experimental and nonexperimental estimates it is important to highlight similarities and differences in training. First, although the evidence does not show major differences along these lines it is possible that the treatment group received training at a different time than that of the control sample. Table 2 only shows training receipt up to the point of time of each follow-up survey. The control sample might have waited longer, for example, lessening potential impacts as measured on the follow-up survey. On the other hand, the control sample might have been in a better position to benefit from training by waiting longer. Second, there is some suggestive evidence showing that the control sample received training that was not as high quality as what the treatment group received. The only evidence on this point is through self-reported surveys which might be subject to social desirability bias. In any case, the quality of training might differ somewhat between the two samples. Third, although the treatment groups attended training programs that already existed and were part of the local market for training services, the control sample might have received training from programs that were more likely to target benefits to participants who could benefit the most from training. These programs might have discouraged other individuals from pursuing training and entrepreneurship. Overall, there might be differences in the timing, quality, and targeting of training between what the treatment and control samples received that could affect the comparison of estimates. The direction of any bias from these training differences in the comparison between experimental and nonexperimental estimates, however, is not clear and might be small.

There also might be differences in terms of selection into training between the treatment and control samples. The control sample was rejected from receiving training services through the study and had to go out and find training on their own. This is very different than for the treatment group that received training services directly through the program. The selection into types and quality of entrepreneurship training could differ between the two samples which are part of the selection bias that the comparison of the experimental and nonexperimental estimates is trying to capture.

4 | EXPERIMENTAL ESTIMATES OF ENTREPRENEURSHIP TRAINING

In this section, I first estimate experimental effects using the full study participant sample. I focus on estimating the effects of entrepreneurship training on business ownership and other business outcomes. Table 3 reports experimental LATE estimates of Equation (2.2).[14] Results with and without controls for baseline covariates are similar, and the discussions below focus on the former (Column 2). The average impact of entrepreneurship training on business ownership at Wave 1 (the 6-month follow-up) is positive and significant: 13.4 pp on a base (control group mean) of 35.9 pp. At the 18-month follow-up, the treatment effect point estimate is smaller and no longer statistically significant. Sixty months after random assignment, the treatment and control groups have nearly identical levels of business ownership. These results are not driven by changes in sample composition: I get similar results after restricting the sample to Wave 3 respondents. In all, the results in this first panel show only positive short-term effects of entrepreneurship training on business ownership that dies out quickly over time.[15]

I turn to experimental estimates of impacts on business outcomes. Examining the average treatment effects on business sales and hiring employees (Table 3), I find no significant effects at any horizon, suggesting that

[14]ITT estimates are reported in Table A4. As expected given the noncompliance rates, the point estimates are generally scaled down by a factor of 2–3 relative to the LATE estimates. None of the statistical inferences changes. Because most of the LATE estimates are close to zero the "scaled down" ITT estimates also tend to be close to zero.

[15]The results are not due to the influence of side or casual businesses, or disguised unemployment (Carter & Sutch, 1994). Defining business ownership with 30 or more hours worked per week, I find lower rates of business ownership, but similar treatment–control differences. I also restrict business ownership to only include businesses reporting positive sales at each survey wave to remove nonserious self-employment activities. Again, I find similar results.

the marginal businesses had low levels of sales and generally did not hire employees. Note that these results do not condition on business ownership, and thus capture the treatment's overall impact on sales and hiring employees.

I focus on these three common measures of business success, however, estimates from the experiment show quite resounding evidence of null effects across numerous additional outcomes (see Fairlie et al., 2015). For example, when switching from the extensive margin of employment to using the average number of employees as the employment outcome the results are the same: no effects on any time horizon. The null results are also not sensitive to alternative measures of outcomes, outliers, and sample restrictions.

5 | NONEXPERIMENTAL ESTIMATES OF ENTREPRENEURSHIP TRAINING

In this section, I estimate nonexperimental regressions for the effects of entrepreneurship training using the control group sample. I start by estimating a straightforward OLS regression in which the dependent variable is business ownership.[16] Specifications with no controls, main demographic controls (i.e., those found in data sets, such as the CPS and ACS), and a rich set of detailed controls are reported in Table 4. All of these variables are measured at baseline. For business ownership, the nonexperimental entrepreneurship training effect estimates are large, positive, and statistically significant at the 6-, 18-, and 60-month follow-up waves. For the specification using the extensive set of controls, the estimates indicate that entrepreneurship training increases business ownership rates by 21 pp (base of 38 pp) at the 5-year follow-up. In contrast, the experimental results reveal a point estimate of 1 pp that is not statistically significant. At the 1.5-year follow-up, the nonexperimental estimates indicate an equally large and significant effect (22 pp) compared with a null experimental finding. The only follow-up wave in which the estimates are more similar is the first one at 6 months. The nonexperimental estimates indicate a short-run effect of 15.3 pp compared with an experimental estimate of 13.3 pp. This finding might just be the result of study participants initially trying self-employment in the first few months after the random assignment is given their interest in entrepreneurship training and the follow-up survey only being 6 months later.

The large, positive, and statistically significant estimates of entrepreneurship training effects are found with and without including controls. Specifications 2, 5, and 8 include the standard set of demographic and education controls included in most large data sets, such as the CPS and ACS. The point estimates on training become smaller after the inclusion of this standard set of controls relative to the specification without controls, implying that they capture at least some heterogeneity.

The baseline survey data collected as part of the GATE experiment provide a unique set of detailed characteristics of individuals. Typically unobservable characteristics, such as baseline household income, whether self-employed at the time of application, previous family business/work experience (has relatives or friends who have been previously self-employed; has ever worked for relatives or friends who are self-employed), health-related information (has health insurance from current employer; has health problems) credit history problems, currently receiving UI benefits, and personality characteristics (autonomy index; risk tolerance index) are included. Previous research indicates that these characteristics along with the demographic characteristics are important determinants of business ownership and outcomes.[17] Even after controlling for this rich set of typically unobservable characteristics the nonexperimental regressions indicate large, positive, and statistically significant training effects.

The comparison of findings is similar for business sales. Table 4 also reports estimates for the same three follow-up waves where sales are the dependent variable. The nonexperimental estimates indicate that training increases monthly sales by $2035 at the 5-year follow-up. Monthly sales are $1285 higher at the 1.5-year follow-up. In contrast, experimentally there is no estimated effect of training on business sales, with precise estimates of null treatment effects.[18]

[16] Estimates of marginal effects from probit and logit models are similar.

[17] See, for example, evidence on the importance of gender (Koellinger et al., 2013), race (Fairlie & Robb, 2007), age (Zissimopoulos & Karoly, 2007), marital status and children (Hundley, 2000; Parker, 2008), education (van der Sluis et al., 2008), family business experience (Dunn & Holtz-Eakin, 2000; Fairlie & Robb, 2007), liquidity constraints (Evans & Jovanovic, 1989), health insurance (Fairlie et al., 2011), unemployment, and personality traits (Caliendo et al., 2014; Kerr et al., 2017). Also, see Parker (2018) and Simoes et al. (2016) for reviews of the literature.

[18] Estimates from log specifications provide similar results. Monthly sales are 0.45 log points higher with entrepreneurship training receipt at the 5-year follow-up, and are 0.55 log points higher at the 1.5-year follow-up.

TABLE 2 Treatment and control groups receipt of entrepreneurship training

	RA to Wave 1 (6-month period)		Waves 1–2 (12-month period)		Year before Wave 3 (12-month period)		Cumulative to Wave 2		Cumulative to Wave 3	
	Percent receiving	Mean hours	Percent receiving	Mean hours	Percent receiving	Mean hours	Percent receiving	Mean hours	Percent receiving	Mean hours
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment group										
Any entrepreneurship training	81.2%	15.6	41.5%	7.3	26.1%	4.6	86.4%	22.9	88.7%	27.6
Attended classes, workshops, or seminars	66.8%	13.8	35.0%	6.6	22.1%	4.0	74.1%	20.4	77.0%	24.4
Received one-on-one counseling or technical assistance	52.5%	1.8	18.0%	0.8	10.0%	0.6	58.8%	2.6	63.7%	3.3
Control group										
Any entrepreneurship training	44.0%	6.6	37.9%	6.7	28.7%	5.7	57.3%	13.3	65.4%	19.1
Attended classes, workshops, or seminars	37.7%	5.8	32.7%	6.1	25.1%	5.2	50.6%	11.9	58.7%	17.1
Received one-on-one counseling or technical assistance	19.2%	0.9	13.8%	0.7	10.3%	0.6	27.9%	1.5	34.4%	2.1
Treatment–control difference and (standard error)										
Any entrepreneurship training	0.372	8.99	0.036	0.63	−0.026	−1.10	0.291	9.62	0.232	8.52
	(0.015)	(0.72)	(0.018)	(0.73)	(0.018)	(0.71)	(0.016)	(1.15)	(0.017)	(1.53)
Attended classes, workshops, or seminars	0.290	7.97	0.024	0.50	−0.030	−1.15	0.235	8.47	0.183	7.31
	(0.016)	(0.68)	(0.017)	(0.68)	(0.017)	(0.68)	(0.017)	(1.08)	(0.019)	(1.38)
Received one-on-one counseling or technical assistance	0.333	0.99	0.043	0.10	−0.003	0.05	0.310	1.09	0.293	1.14
	(0.015)	(0.15)	(0.013)	(0.11)	(0.012)	(0.17)	(0.017)	(0.21)	(0.019)	(0.30)

Notes: (1) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application.

Abbreviation: RA, random assignment.

TABLE 3 Experimental estimates of entrepreneurship training effects on business outcomes

	Entrepreneurship training coefficient			
	Without controls	With controls	Control man	*N*
Dependent variable	(1)	(2)	(3)	(4)
Business owner at W1 survey date	0.1252	0.1337	0.3592	3443
	(0.0446)	(0.0395)		
Business owner at W2 survey date	0.0742	0.0691	0.4091	3032
	(0.0616)	(0.0570)		
Business owner at W3 survey date	0.0406	0.0105	0.3794	2446
	(0.0844)	(0.0810)		
Monthly business sales at W1 survey date	−1081.7	−940.2	1828.8	3210
	(751.0)	(733.6)		
Monthly business sales at W2 survey date	−606.0	−441.1	2132.7	2794
	(1153.9)	(1115.1)		
Monthly business sales at W3 survey date	−2097.7	−2552.2	2909.2	2323
	(2280.4)	(2288.5)		
Has any employees at W1 survey date	0.0353	0.0363	0.0722	3438
	(0.0248)	(0.0245)		
Has any employees at W2 survey date	0.0133	0.0065	0.0939	3023
	(0.0368)	(0.0362)		
Has any employees at W3 survey date	−0.0736	−0.0871	0.1104	2436
	(0.0525)	(0.0534)		

Notes: (1) The first stage in the IV (LATE) model regresses receipt of entrepreneurship training on treatment. The second stage regresses the listed outcome on predicted receipt of entrepreneurship training. (2) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (3) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

Abbreviation: LATE, local average treatment effect.

Employment by the business also follows the same pattern. Table 5 reports estimates. Nonexperimental estimates indicate that training has large, positive, and statistically significant effects on having any employees. At the 1.5-year follow-up the increase is 4.4 pp, and at the 5-year follow-up the increase is 6.4 pp. For comparison, in both cases the experimental estimates indicate no effect on having employees.

Although the nonexperimental regressions attempt to control for the long list of detailed individual characteristics there might be unobservable differences between those who choose to obtain entrepreneurial training and those who do not. One method of investigating this issue is to compare the average observable characteristics of the two groups. Large differences in these observable characteristics potentially indicate differences in unobservable characteristics that are correlated with both entrepreneurship training receipt and business outcomes. Table 5 recreates the balance check from randomization into treatment and control groups reported in Table 1, but in this case only the control group sample is included and the groups are distinguished by receipt or no receipt of entrepreneurship training. Several of the variables show a statistical difference between the training receipt and no training receipt groups. For example, the training receipt group has higher baseline income, self-employment probability, age, and risk tolerance, but none of these differences is very large in magnitude. The finding is disconcerting because even small observed differences between groups may lead to substantially biased nonexperimental estimates of entrepreneurship training effects on business outcomes.

TABLE 4 Nonexperimental regression estimates of entrepreneurship training effects on business outcomes

	Entrepreneurship training coefficient			
	Without controls	Basic controls	Detailed controls	*N*
Dependent variable	(1)	(2)	(3)	(4)
Business owner at W1 survey date	0.196***	0.178***	0.153***	1,685
	(0.023)	(0.023)	(0.022)	
Business owner at W2 survey date	0.258***	0.238***	0.221***	1462
	(0.025)	(0.025)	(0.025)	
Business owner at W3 survey date	0.228***	0.222***	0.210***	1162
	(0.029)	(0.030)	(0.031)	
Monthly business sales at W1 survey date	836.0*	888.0*	561.2	1575
	(466.300)	(480.426)	(498.920)	
Monthly business sales at W2 survey date	1248.0**	1418.8***	1284.6**	1337
	(492.490)	(506.489)	(533.815)	
Monthly business sales at W3 survey date	2058.1**	2508.8***	2034.5**	1101
	(877.957)	(910.799)	(962.796)	
Has any employees at W1 survey date	0.035***	0.039***	0.030**	1685
	(0.013)	(0.013)	(0.013)	
Has any employees at W2 survey date	0.050***	0.050***	0.044***	1457
	(0.015)	(0.016)	(0.016)	
Has any employees at W3 survey date	0.068***	0.064***	0.064***	1158
	(0.019)	(0.020)	(0.021)	

Notes: (1) The sample includes only control group observations. Each cell represents an ordinary least-square (OLS) regression of the listed outcome on receipt of entrepreneurship training. (2) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (3) Covariates include basic set (program sites, female, race, immigrant, age, married, children, and education level) and detailed set (+household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance).

*denote statistical significance at the 0.10.

**denote statistical significance at the 0.05.

***denote statistical significance at the 0.01.

5.1 | Matching etimator estimates

To further control for unobserved heterogeneity, I estimate the relationship using matching estimators. These estimators attempt to compare the outcomes of individuals who are as similar as possible with the exception of treatment status (i.e., receipt or no receipt of training among the control group). I first estimate the entrepreneurship training effect using nearest-neighbor matching. In nearest-neighbor matching the goal is to first create a similarity measure which determines how "close" two observations are to each other. Then the "nearest-neighbor" observations are identified from the calculated distances between pairs of observations based on the specified set of covariates, and are matched for the comparison of mean outcomes between the two groups.[19]

Table 6 reports estimates for the three business outcome measures and three follow-up waves. Estimates are provided for two sets of matches—the standard set of background characteristics and the expanded set of detailed demographic, income, health, and previous business/work experience characteristics. Expanded sets of characteristics for matching the increase the enforced similarity between groups, but increase estimate variance compared with more limited sets of characteristics for matching. Nearest-neighbor matching estimates indicate large effects on business

[19]The nearest-neighbor matches use the Mahalanobis distance between the vectors of covariates of observations, which accounts for correlation between covariates and having different measurement scales.

TABLE 5 Training receipt/no receipt comparison of control group

	Follow-up Wave 1			Follow-up Wave 2			Follow-up Wave 3		
	Receipt group (4)	No receipt group (5)	*p* Value for R–NR (6)	Receipt group (7)	No receipt group (8)	p Value for R–NR (9)	Receipt group (10)	No receipt group (11)	*p* Value for R–NR (12)
Philadelphia	23.2%	27.4%	0.05	22.8%	25.8%	0.19	22.3%	21.6%	0.78
Pittsburgh	11.5%	16.6%	0.00	12.5%	16.8%	0.02	12.6%	18.1%	0.02
Minneapolis–St. Paul	41.1%	37.5%	0.13	42.0%	38.1%	0.13	42.3%	41.2%	0.73
Duluth	4.5%	5.5%	0.32	4.8%	5.4%	0.57	5.4%	4.2%	0.37
Maine	19.8%	12.9%	0.00	17.9%	13.9%	0.04	17.5%	14.9%	0.25
Female	48.3%	44.7%	0.15	48.7%	44.5%	0.11	49.1%	43.2%	0.05
Black	29.1%	30.3%	0.59	28.0%	28.9%	0.74	27.2%	24.1%	0.26
Latino	5.3%	4.7%	0.58	6.0%	4.0%	0.08	6.2%	3.5%	0.03
Asian	2.4%	3.9%	0.08	2.6%	3.4%	0.42	2.6%	3.2%	0.56
Other	7.4%	7.9%	0.74	7.2%	6.9%	0.84	7.2%	5.7%	0.32
Not US born	10.2%	8.4%	0.19	9.6%	7.7%	0.21	9.3%	6.0%	0.03
Age	44.29	42.75	0.00	44.53	42.81	0.00	44.72	43.01	0.01
Married	52.4%	45.7%	0.01	51.6%	45.5%	0.02	52.6%	43.6%	0.00
Has children	47.0%	43.7%	0.18	45.5%	43.7%	0.48	44.9%	39.2%	0.06
Highest grade completed	14.93	14.36	0.00	14.96	14.26	0.00	15.04	14.29	0.00
HH income $25,000–49,999	33.4%	34.3%	0.70	32.9%	33.8%	0.72	34.3%	34.6%	0.91
HH income $50,000–74,999	17.5%	17.5%	0.99	17.4%	18.3%	0.68	17.2%	17.2%	0.99
HH income $75,000–99,999	8.1%	6.3%	0.14	8.3%	5.6%	0.04	8.9%	4.2%	0.00
HH income $100,000+	9.5%	5.8%	0.01	9.4%	6.1%	0.02	9.8%	7.2%	0.13
Self-emp. at application	22.9%	18.5%	0.03	23.7%	18.0%	0.01	23.0%	18.8%	0.09
Has a health problem	9.2%	8.7%	0.71	8.9%	8.7%	0.93	8.2%	8.6%	0.82
Has relatives or friends who have been previously SE	75.3%	69.3%	0.01	74.1%	69.9%	0.08	75.1%	69.0%	0.03
Ever worked for relatives or friends who are SE	32.9%	30.9%	0.39	32.8%	29.9%	0.24	33.2%	28.0%	0.07
Has a bad credit history	39.9%	45.8%	0.02	41.2%	42.1%	0.73	39.8%	39.0%	0.79
Currently receiving UI benefits	39.0%	40.2%	0.64	38.9%	39.6%	0.78	39.7%	43.4%	0.23
Has health insurance from current employer	15.5%	19.1%	0.05	15.4%	20.7%	0.01	15.7%	20.1%	0.07
Autonomy index	1.6%	−4.6%	0.21	−3.0%	0.1%	0.55	−6.3%	−2.5%	0.55
Risk tolerance index	4.3%	−5.3%	0.05	2.8%	−7.9%	0.04	2.3%	−16.4%	0.00

TABLE 5 (Continued)

	Follow-up Wave 1			Follow-up Wave 2			Follow-up Wave 3		
	Receipt group (4)	No receipt group (5)	*p* Value for R–NR (6)	Receipt group (7)	No receipt group (8)	p Value for R–NR (9)	Receipt group (10)	No receipt group (11)	*p* Value for R–NR (12)
F-test for all variables			0.53			0.69			0.80
Sample size	742	944		838	625		762	403	

Notes: (1) The sample includes only control group observations. (2) All reported characteristics are measured at the time of application, before random assignment. (3) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (4) The autonomy index is created from standardizing responses on a scale of 1–5 to whether the statement "I enjoy working independently" is true about themselves. The risk aversion index is created from combining standardized responses to "I'm only willing to take a risk if I am sure everything will work out" and "I am not prepared to risk my savings for my business."

Abbreviations: HH, household; SE, self-employed; UI, unemployment insurance.

TABLE 6 Nearest-neighbor matching estimates of entrepreneurship training effects on business outcomes

	Entrepreneurship training coefficient		
	Basic controls	Detailed controls	*N*
Dependent variable	(1)	(2)	(3)
Business owner at W1 survey date	0.181***	0.143***	1685
	(0.029)	(0.027)	
Business owner at W2 survey date	0.243***	0.221***	1462
	(0.029)	(0.029)	
Business owner at W3 survey date	0.233***	0.192***	1162
	(0.032)	(0.035)	
Monthly business sales at W1 survey date	696.204	516.117	1575
	(708.477)	(536.663)	
Monthly business sales at W2 survey date	1188.312***	813.636	1337
	(450.127)	(598.017)	
Monthly business sales at W3 survey date	2024.934***	1996.344*	1101
	(662.356)	(1120.232)	
Has any employees at W1 survey date	0.045***	0.031*	1685
	(0.016)	(0.017)	
Has any employees at W2 survey date	0.041**	0.047***	1457
	(0.018)	(0.018)	
Has any employees at W3 survey date	0.067***	0.067***	1158
	(0.018)	(0.019)	

Notes: (1) The sample only includes control group observations. (2) Nearest-neighbor matching compares observations in the receipt and no receipt groups based on the closest values of the set of covariates using the Mahalanobis distance. (3) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (4) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

*denote statistical significance at the 0.10.

**denote statistical significance at the 0.05.

***denote statistical significance at the 0.01.

TABLE 7 Propensity-score matching estimates of entrepreneurship training effects on business outcomes

	Entrepreneurship training coefficient		
	Basic controls	Detailed controls	*N*
Dependent variable	(1)	(2)	(3)
Business owner at W1 survey date	0.219***	0.163***	1685
	(0.029)	(0.027)	
Business owner at W2 survey date	0.249***	0.242***	1462
	(0.029)	(0.029)	
Business owner at W3 survey date	0.224***	0.174***	1162
	(0.034)	(0.037)	
Monthly business sales at W1 survey date	547.596	1182.939**	1575
	(514.012)	(560.219)	
Monthly business sales at W2 survey date	1684.058***	1006.384**	1337
	(566.036)	(457.623)	
Monthly business sales at W3 survey date	1987.190*	2191.743***	1101
	(1028.732)	(735.579)	
Has any employees at W1 survey date	0.033**	0.040**	1685
	(0.015)	(0.016)	
Has any employees at W2 survey date	0.050***	0.053***	1457
	(0.018)	(0.018)	
Has any employees at W3 survey date	0.059***	0.059***	1158
	(0.020)	(0.020)	

Notes: (1) The sample only includes control group observations. (2) Propensity-score matching compares observations in the receipt and no receipt groups based on the closest values of predicted probabilities of receipt from a first-stage logit regression. (3) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (4) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

*denote statistical significance at the 0.10.

**denote statistical significance at the 0.05.

***denote statistical significance at the 0.01.

ownership. At the 1.5- and 5-year follow-up waves business ownership is estimated to be 20–24 pp higher among individuals receiving entrepreneurship training than not receiving training in the control group. The estimates are less clean for monthly business sales, but half the reported estimates show large, positive, and statistically significant effect estimates of entrepreneurship training. All of the point estimates are large and positive. Business employment is also estimated to be much higher with the receipt of entrepreneurship training. The likelihood of having employees increases by 4–5 pp 1.5 years later, and 7 pp 5 years later.

The estimates are robust to several alternative specifications. First, I remove the continuous and somewhat continuous variables, such as age, highest grade completed, the risk tolerance index, and the autonomy index. The estimates are similar. I continue to find large, positive, and statistically significant effects on outcomes. Next, I allow for multiple nearest-neighbor matches instead of only one match. Allowing for multiple matches results in very similar estimates. Finally, I match exactly on baseline self-employment given its importance in determining future self-employment. I continue to include all of the other characteristics in the nearest-neighbor matches. Again, the estimates are robust to this alternative specification of the model.

Table 7 reports estimates from an alternative method of matching, propensity-score matching. In this case, individuals are matched on predicted probabilities of treatment, referred to as propensity scores.[20] Predicted probabilities are calculated in a first-stage logit regression of training receipt within the control group. The propensity score estimates also indicate large positive effects on all business outcomes. The estimates for business ownership in the 1.5- and 5-year follow-up waves range from 17 to 25 pp. The estimates for monthly business sales are $1000–$2200. Employment levels are estimated to be 5–6 pp higher among training recipients in the control sample.

These estimates of the effect of entrepreneurship training on business ownership, sales, and employees using the control sample are large, positive, and statistically significant, contrasting sharply with the estimates of null effects found in the experiment. This discrepancy raises concerns about positive selection into entrepreneurship training resulting in an overstatement of the nonexperimental estimates of the effects of training on business outcomes. Furthermore, even a rich set of controls—for baseline household income level, self-employment status, health problems, work experience in a family business, credit history, UI receipt, employer-provided health insurance, personality traits, and standard demographic controls—cannot purge nonexperimental treatment effect estimates of strong upward selection bias. Techniques, such as nearest-neighbor matching and propensity-score matching, to address selection also do not change the conclusion.

The finding contributes to a broader literature on comparing experimental to nonexperimental methods. The seminal study in labor economics by LaLonde (1986) compares experimental estimates of the effects of a job training program (i.e., the National Supported Work [NSW] Demonstration) to commonly used nonexperimental methods at the time and finds that these techniques do not perform well. Using the same data Dehejia and Wahba (2002), however, show that propensity-score matching methods perform reasonably well. But, follow-up work by Smith and Todd (2005) also using the same data indicate that estimates from propensity-score matching are highly sensitive to both the set of variables included in the scores and the analysis sample. They conclude that "while propensity-score matching is a potentially useful econometric tool, it does not represent a general solution to the evaluation problem." Additional research highlights problems with nonexperimental methods using evaluations of mandatory work problems (Michalopoulos et al., 2004).

The comparison of experimental to nonexperimental estimates from GATE adds to this literature. I have access to an extensive set of variables predicting entrepreneurship. The baseline survey was designed to capture many of the determinants of entrepreneurship found in previous studies. I have tried numerous permutations of these variables with little effect on the results. I have also tried different matching techniques with little difference in results. And, finally, I consistently find the same large positive estimates of effects across different dependent variables and even with three different follow-up survey waves (which are rare to have in RCTs). Although these are reasonable things to try (Caliendo & Kopeinig, 2008) perhaps one problem is that many of the determinants of entrepreneurship are not that well known or measurable. There are many idiosyncratic reasons that people start businesses. There is no way to measure these impetuses. For example, we cannot measure the exogenous arrival of innovative ideas or the family circumstances leading to when it is a good time to start a business. In other settings propensity-score matching methods might perform better.

5.2 | Comparing estimates of the impact of one-on-one counseling services

One-on-one counseling services relative to classroom instruction might provide the largest impacts on business outcomes within broader entrepreneurship training assistance. I compare experimental and nonexperimental estimates of individualized one-on-one counseling effects. For the treatment group, 64% received individualized counseling services compared with 34% of the control group (Table 2). For any training services, as noted above, 81% of the treatment group and 44% of the control group received services. The shift in the treatment–control differential suggests that the LATE (IV) treatment estimates for one-on-one counseling will be rescaled versions of the ones for any entrepreneurship training services according to the equations in Section 2.4.

Table 8 reports experimental and nonexperimental estimates of the impact of one-on-one counseling services on business outcomes. I compare LATE estimates from the full experimental sample, and three nonexperimental methods:

[20] A logit model is estimated to recover the predicted probabilities of training receipt. Estimates are similar using a probit model to estimate the predicted probabilities. I also verify that there is common support for the propensity-score match.

TABLE 8 Estimates of one-on-one counseling training effects on business outcomes

	One-on-one counseling training coefficient			
	Experimental (LATE)	**OLS**	**Nearest neighbor**	**Propensity score**
Dependent variable	**(1)**	**(2)**	**(3)**	**(4)**
Sample	Full	Control	Control	Control
Business owner at W1 survey date	0.149***	0.145***	0.203***	0.203***
	(0.043)	(0.028)	(0.036)	(0.036)
Business owner at W2 survey date	0.065	0.203***	0.238***	0.238***
	(0.053)	(0.028)	(0.035)	(0.035)
Business owner at W3 survey date	0.008	0.199***	0.230***	0.230***
	(0.064)	(0.030)	(0.037)	(0.037)
Monthly business sales at W1 survey date	−1065.084	662.204	1755.178**	1755.178**
	(835.542)	(637.185)	(885.688)	(885.688)
Monthly business sales at W2 survey date	−420.809	683.900	1171.716**	1171.716**
	(1075.931)	(597.858)	(563.184)	(563.184)
Monthly business sales at W3 survey date	−2031.623	1756.910*	1675.750	1675.750
	(1813.822)	(953.078)	(1032.057)	(1032.057)
Has any employees at W1 survey date	0.042	0.013	0.040**	0.040**
	(0.027)	(0.017)	(0.020)	(0.020)
Has any employees at W2 survey date	0.006	0.062***	0.070***	0.070***
	(0.034)	(0.018)	(0.023)	(0.023)
Has any employees at W3 survey date	−0.073*	0.059***	0.072***	0.072***
	(0.043)	(0.021)	(0.024)	(0.024)

Notes: (1) The sample used in Specifications 2–4 includes only control group observations. (2) Nearest-neighbor matching compares observations in the receipt and no receipt groups based on the closest values of the set of covariates using the Mahalanobis distance. (3) Propensity-score matching compares observations in the receipt and no receipt groups based on the closest values of predicted probabilities of receipt from a first-stage logit regression. (4) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (5) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

Abbreviations: LATE, local average treatment effect; OLS, ordinary least-square.

*denote statistical significance at the 0.10.

**denote statistical significance at the 0.05.

***denote statistical significance at the 0.01.

OLS, nearest-neighbor, and propensity-score matching. The experimental estimates show null effects for all outcomes (other than 6-month impacts on self-employment) which is expected because the switch from the receipt of any entrepreneurship training services to only one-on-one counseling services essentially just scales up the estimates. And, scaling-up zero impact estimates results in zero impact estimates. For self-employment at all follow-up waves, the nonexperimental estimates indicate large effects. The receipt of one-on-one counseling is associated with a 20–24 pp higher business ownership rate at the 1.5- and 5-year follow-up surveys. The estimates for business sales vary more, but generally indicate large positive effects from individualized counseling. Finally, nonexperimental estimates of the effects of one-on-one counseling on hiring employees consistently show large, positive effects.

Overall, changing the focus from any entrepreneurship training services to only one-on-one counseling services, which are likely to have the largest impact, does not change the conclusions. Experimental estimates for one-on-one counseling show null effects whereas nonexperimental estimates indicate large, positive effects.

5.3 | Comparing estimates based on baseline self-employment

As noted above, roughly 20% of both the treatment and control groups is self-employed at the time of random assignment. The GATE program was open to both individuals who wanted to start a business and to individuals who wanted to grow their pre-existing business. The objective corresponds with the general objectives of SBDCs which is to provide "assistance to *current* and *prospective* small business owners." The bias in nonexperimental estimates might differ between non-self-employed and self-employed at baseline because of potential differences in selection into who seeks training among the two groups.

To investigate this question, Tables 9 and 10 report experimental and nonexperimental estimates of the impact of entrepreneurship training on self-employment, sales, and employees at the three follow-up waves for the baseline not self-employed and self-employed, respectively. For those interested in starting a business the estimates do not differ substantially from the full sample. Experimental estimates indicate null effects on business ownership (other than the

TABLE 9 Estimates of entrepreneurship training effects on business outcomes—not self-employed at application

	Entrepreneurship training coefficient					
	Experimental (LATE)	OLS	Nearest neighbor	Propensity score		
Dependent variable	(1)	(2)	(3)	(4)	*N*	*N*
Sample	Full	Control	Control	Control	Full	Control
Business owner at W1 survey date	0.160***	0.137***	0.139***	0.171***	2566	1241
	(0.044)	(0.025)	(0.030)	(0.033)		
Business owner at W2 survey date	0.077	0.226***	0.220***	0.216***	2226	1065
	(0.062)	(0.029)	(0.032)	(0.030)		
Business owner at W3 survey date	−0.015	0.185***	0.207***	0.208***	1790	845
	(0.088)	(0.034)	(0.034)	(0.036)		
Monthly business sales at W1 survey date	−428.510	−208.166	−109.798	1242.956	2445	1187
	(674.895)	(466.608)	(492.714)	(769.061)		
Monthly business sales at W2 survey date	24.438	1066.353*	1023.739*	1154.469	2090	993
	(1219.336)	(594.452)	(534.969)	(737.935)		
Monthly business sales at W3 survey date	−1180.045	706.537	943.842	1377.236	1711	801
	(2194.538)	(975.723)	(745.844)	(1159.634)		
Has any employees at W1 survey date	0.027	0.011	0.022	0.017	2562	1241
	(0.023)	(0.013)	(0.016)	(0.016)		
Has any employees at W2 survey date	0.021	0.046***	0.036**	0.065***	2222	1063
	(0.036)	(0.016)	(0.018)	(0.019)		
Has any employees at W3 survey date	−0.034	0.064***	0.079***	0.064***	1786	842
	(0.054)	(0.022)	(0.018)	(0.021)		

Notes: (1) Only participants who are not self-employed at the time of application are included. The sample used in Specifications 2–4 includes only control group observations. (2) Nearest-neighbor matching compares observations in the receipt and no receipt groups based on the closest values of the set of covariates using the Mahalanobis distance. (3) Propensity-score matching compares observations in the receipt and no receipt groups based on the closest values of predicted probabilities of receipt from a first-stage logit regression. (4) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (5) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

Abbreviations: LATE, local average treatment effect; OLS, ordinary least-square.

*denote statistical significance at the 0.10.

**denote statistical significance at the 0.05.

***denote statistical significance at the 0.01.

TABLE 10 Estimates of entrepreneurship training effects on business outcomes—self-employed at application

	Entrepreneurship training coefficient					
	Experimental (LATE)	OLS	Nearest neighbor	Propensity score		
Dependent variable	(1)	(2)	(3)	(4)	*N*	*N*
Sample	Full	Control	Control	Control	Full	Control
Business owner at W1 survey date	0.077	0.181***	0.227***	0.183***	640	322
	(0.086)	(0.048)	(0.049)	(0.048)		
Business owner at W2 survey date	0.058	0.182***	0.191***	0.140**	580	290
	(0.131)	(0.059)	(0.060)	(0.063)		
Business owner at W3 survey date	0.122	0.307***	0.296***	0.285***	476	233
	(0.196)	(0.070)	(0.080)	(0.088)		
Monthly business sales at W1 survey date	−2596.191	3909.813**	2698.961	3292.096	544	274
	(2720.277)	(1791.482)	(2074.084)	(2383.883)		
Monthly business sales at W2 survey date	−1754.785	2468.448*	2093.737**	1561.131*	493	243
	(2870.949)	(1275.841)	(1026.019)	(947.648)		
Monthly business sales at W3 survey date	−6760.638	7409.598**	7359.209***	5916.062***	444	219
	(8326.640)	(2860.896)	(2394.290)	(1524.700)		
Has any employees at W1 survey date	0.127	0.123***	0.140***	0.113**	639	322
	(0.081)	(0.043)	(0.051)	(0.045)		
Has any employees at W2 survey date	−0.015	0.035	0.093*	0.069	575	287
	(0.113)	(0.051)	(0.050)	(0.054)		
Has any employees at W3 survey date	−0.317*	0.064	0.077	0.132**	471	232
	(0.171)	(0.063)	(0.060)	(0.052)		

Notes: (1) Only participants who are self-employed at the time of application are included. The sample used in Specifications 2–4 includes only control group observations. (2) Nearest-neighbor matching compares observations in the receipt and no receipt groups based on the closest values of the set of covariates using the Mahalanobis distance. (3) Propensity-score matching compares observations in the receipt and no receipt groups based on the closest values of predicted probabilities of receipt from a first-stage logit regression. (4) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (5) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

Abbreviations: LATE, local average treatment effect; OLS, ordinary least-square.

*denote statistical significance at the 0.10.

**denote statistical significance at the 0.05.

***denote statistical significance at the 0.01.

6-month wave) and on sales and employees. The nonexperimental estimates consistently show large, positive effects on business ownership and employees. For sales, the point estimates become less precise, but are large and positive in all cases, and marginally significant in a couple of cases.

Table 10 reports estimates for the smaller sample of self-employed individuals who were self-employed at the time of application to the field experiment. The comparison between experimental and nonexperimental estimates shows even larger differences. Experimentally, I find null effects for all outcomes. Nonexperimentally, I find very large positive estimates of entrepreneurship training on business ownership. At the 5-year follow-up the three nonexperimental techniques each show a roughly 30 pp effect on business ownership. The estimated training effects on monthly sales after 5 years are also large in magnitude indicating 6–7.5 thousand dollars. Estimated effects on employees are less precise, but consistently show positive point estimates and several statistically significant effects.

Overall, even with smaller sample sizes and less precise estimates the results indicate major differences between experimental and nonexperimental estimates regardless of the goal of training receipt. For both those who want to start a

business and those who want to grow a pre-existing business nonexperimental estimates indicate that entrepreneurship training substantially improves business outcomes, whereas experimental estimates indicate null effects.

6 | CONCLUSION

Governments and philanthropic organizations around the world devote substantial resources to entrepreneurship training even though there is little causal evidence on its effectiveness. Instead, reporting the total number of clients served, making simple pre–post comparisons, or nonexperimental estimates of positive effects are cited as justification of the "success" of programs. This paper provides the first evidence in the literature on whether nonexperimental estimates of the effects of entrepreneurship training are similar to experimental estimates for the same underlying population. Additionally, the study provides the first estimates of whether there is positive, negative, or no selection into entrepreneurship training.

Experimental estimates of entrepreneurship training indicate null effects on business ownership, sales, and employees, and many additional business outcomes. The findings are robust to alternative specifications, outliers, and differential attrition. In contrast, for the same underlying group of study participants but focusing on the control group subsample, I find large, positive nonexperimental estimates of the effects of entrepreneurship training receipt on business ownership, sales, and employment. The largest-ever field experiment on entrepreneurship training, GATE, makes it possible to isolate the control group sample to conduct this nonexperimental analysis. Another feature of the study design is that in conducting the GATE experiment a very detailed set of baseline demographic characteristics, previous business, work and unemployment experience, and health insurance was collected. Even information on family business experience, credit history, and personality characteristics, such as risk tolerance and preference for autonomy, was collected. In most cases, this type of information is not available to researchers evaluating the effectiveness of training programs. Finally, information on the receipt of entrepreneurship training services was collected for the control group in addition to the treatment group in the experiment.

The differential estimates do not appear to be due to the control group receiving more or superior training than the treatment group. If anything the detailed evidence from the follow-up surveys on use of services, types of services, and satisfaction with services suggests that the control group appears to have, if anything, received less and lower-quality services, which would create a downward bias on nonexperimental estimates relative experimental estimates.[21] Another concern is that the participants in the GATE study are different than individuals typically seeking entrepreneurship training. GATE was implemented with this concern in mind, however, by being careful to use a broad range of training providers in several locations, and advertise to a wide range of potentially interested individuals. More research from other programs and locations would be useful, but the findings here raise general concerns about estimating training effects. Are these results generalizable to other settings? They might be especially in cases where would-be-entrepreneurs or entrepreneurs are investing time and money into improving their chances of success. It is quite possible that in a wide range of these cases there is some type of positive selection bias.

The simplest evaluations of programs essentially perform a before versus after calculation. Using this "evaluation" approach, I can estimate what GATE training services produced for the treatment group that received entrepreneurship training.[22] At baseline, 21% of the treatment group receiving training services was self-employed. At the follow-up surveys from 41 to 46% of the group were self-employed. Thus, GATE training services "created" a new business owner for every 4–5 participants (20–25 pp increase). Focusing on employees, however, GATE entrepreneurship training apparently "created" no jobs—roughly 10% of participants had a business with any employees entering the program and roughly 10% of participants had a business with any employees at each of the follow-up waves.[23] Revenues are not available for business owners at baseline so a before/after comparison cannot be made.

The results presented here speak to the importance of choosing evaluation methods for entrepreneurship training and potentially in other settings in which there could be selection bias, such as government-sponsored education, loan, and R&D grant programs. Although RCTs are expensive and difficult to implement they might be necessary given the evidence provided here of a large upward bias in nonexperimental estimates of the effectiveness of

[21] Additionally, any Hawthorne effects should similarly bias experimental estimates upward relative to nonexperimental estimates.

[22] E.g., the New York Senate in 2011 justified extending the self-employment assistance program by stating that the program has been successful in developing small businesses among unemployment insurance recipients and has generated over 4,000 jobs and $16 million in state tax revenue (at no cost to the state).

[23] Similarly, the average number of employees does not increase.

entrepreneurship training.[24] Positive selection might be especially strong in entrepreneurship training receipt suggesting that those most able to benefit from the management, marketing, accounting, legal, and other training provided are the ones most likely to obtain training. This has broad implications for policy. On the one hand, program administrators may want the individuals who are most likely to benefit to be the same individuals who receive training. But, on the other hand, if governments and philanthropists want to scale up programs they will need to temper expectations about large, positive effects (Huber et al., 2012; Oosterbeek et al., 2010; Van Praag, 2005). Although the null effects found for the GATE program make these two points moot, other entrepreneurship training programs perhaps with additional features or higher and longer intensity of services may have positive effects that policymakers want to replicate and provide at a larger scale.

ORCID


Robert W. Fairlie http://orcid.org/0000-0002-8634-9785

REFERENCES

Aspen Institute. (2012). *The state of business development services*. FIELD at the Aspen Institute. https://www.aspeninstitute.org/publications/state-business-development-services

Astebro, T. B., & Hoos, F. (2016). *The effects of a training program to encourage social entrepreneurship: Field-experimental evidence* (HEC Paris Research Paper No. SPE-2016-1128). January 13, 2016.

Baumgartner, H. J., & Caliendo, M. (2008). Turning unemployment into self-employment: Effectiveness of two start-up programmes. *Oxford Bulletin of Economics and Statistics*, *70*(3), 347–373.

Bellotti, J. (2006). *Growing America through entrepreneurship: Interim report* (US Department of Labor Report).

Benus, J., Johnson, T. R., Wood, M., Grover, N., & Shen, T. (1994). *Self-employment programs: A new reemployment strategy (Final impact analysis of the Washington and Massachusetts self-employment demonstrations)* (US Department of Labor Report).

Benus, J. M., & Michaelides, M. (2010). *Are self-employment training programs effective? Evidence from Project GATE* (MPRA Paper 20880).

Benus, J., Shen, T., Zhang, S., Chan, M., & Hansen, B. (2009). *Growing America through entrepreneurship: Final evaluation of project GATE* (US Department of Labor Report).

Caliendo, M., Fossen, F., & Kritikos, A. (2014). Personality characteristics and the decision to become and stay self-employed. *Small Business Economics*, *42*, 787–814.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72.

Carter, S. B., & Sutch, R. (1994). *Self-employment in the age of big business: Toward an appreciation of an American Labor Market Institution* (UC Riverside Working Paper).

Chrisman, J. J. (2004). *Economic impact of Small Business Development Center counseling activities in the United States: 2000–2002*. Mississippi State University.

Chrisman, J. J. (2017). *Economic impact of Small Business Development Center counseling activities in the United States: 2014–2015*. Mississippi State University.

Davis, S., Michaelides, M., Poe-Yamagata, E., & Davis, A. (2013). *Evaluation of the GATE II grants: Is self-employment training effective for rural and older dislocated workers*? (Final Report to Employment and Training Administration). US Department of Labor.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, *84*(1), 151–161.

Dunn, T., & Holtz-Eakin, D. (2000). Financial capital, human capital, and the transition to self-employment: Evidence from intergenerational links. *Journal of Labor Economics*, *18*, 282–305.

Evans, D., & Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy*, *97*(4), 808–827.

European Commission. (2016). *European employment observatory review: Self-employment in Europe* (European Commission).

Fairlie, R. (2013). Entrepreneurship, economic conditions, and the great recession. *Journal of Economics & Management Strategy*, *22*(2), 207–231.

Fairlie, R. W., & Chatterji, A. K. (2013). High-technology entrepreneurship in Silicon Valley. *Journal of Economics & Management Strategy*, *22*(2), 365–389.

Fairlie, R. W., Kapur, K., & Gates, S. (2011). Is employer-based health insurance a barrier to entrepreneurship? *Journal of Health Economics*, *30*(1), 146–162.

Fairlie, R. W., Karlan, D., & Zinman, J. (2015). Behind the GATE experiment: Evidence on effects of and rationales for subsidized entrepreneurship training. *American Economic Journal: Economic Policy*, *7*(2), 125–161.

[24]Alternative empirical approaches that are useful include using natural experiments, regression discontinuity designs, and other methods of reducing observable and unobservable differences between groups (see, e.g., Lyons & Zhang, 2018).

Fairlie, R. W., & Robb, A. M. (2007). Why are black-owned businesses less successful than white-owned businesses? The role of families, inheritances, and business human capital. *Journal of Labor Economics*, *25*(2), 289–323.

Fairlie, R. W., & Robb, A. M. (2008). *Race and entrepreneurial success: Black-, asian-, and white-owned businesses in the United States*. Cambridge: MIT Press.

Gu, Q., Karoly, L. A., & Zissimopoulos, J. M. (2008). *Small business assistance programs in the United States: An analysis of what they are, how well they perform, and how we can learn more about them* (RAND Working Paper No. WR- 603-EMKF).

Huber, L. R., Sloof, R., & vanPraag, M. (2012). *The effect of early entrepreneurship education: Evidence from a randomized field experiment* (Working Paper). April 18.

Hundley, G. (2000). Male/female earnings differences in self-employment: The effects of marriage, children, and the household division of labor. *ILR Review*, *54*(1), 95–114.

Kerr, S. P., Kerr, W. R., & Xu, T. (2017). *Personality traits of entrepreneurs: A review of recent literature* (Vol. w24097). National Bureau of Economic Research.

Koellinger, P., & Minniti, M. (2006). Not for lack of trying: American entrepreneurship in black and white. *Small Business Economics*, *27*(1), 59–79.

Koellinger, P., Minniti, M., & Schade, C. (2013). Gender differences in entrepreneurial propensity. *Oxford Bulletin of Economics and Statistics*, *75*, 213–234.

Kosanovich, W. T., & Fleck, H. (2001). *Final report: Comprehensive assessment of self-employment assistance programs* (US Department of Labor Report).

Krashinsky, H. A. (2005). *Self-employment for less-educated men*. Mimeo: University of Toronto.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, *76*(4), 604–620.

Lyons, E., & Zhang, L. (2018). Who does (not) benefit from entrepreneurship programs? *Strategic Management Journal*, *39*(1), 85–112.

Michalopoulos, C., Bloom, H. S., & Hill, C. J. (2004). Can propensity-score methods match the findings from a random assignment evaluation of mandatory welfare-to-work programs? *Review of Economics and Statistics*, *86*(1), 156–179.

Oosterbeek, H., vanPraag, M., & Ijsselstein, A. (2010). The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, *54*, 442–454.

OECD. (2017). *SME and entrepreneurship policy series*. Paris: Organisation for Economic Co-operation and Development Press.

Parker, S. C. (2008). Entrepreneurship among married couples in the United States: A simultaneous probit approach. *Labour Economics*, *15*, 459–481.

Parker, S. C. (2018). *The economics of entrepreneurship*. Cambridge University Press.

Van Praag, M. (2005). *Successful entrepreneurship: Confronting economic theory with empirical evidence*. Edward Elgar.

Rodriguez-Planas, N. (2010). Channels through which public employment services and small business assistance programmes work. *Oxford Bulletin of Economics and Statistics*, *72*(4), 458–485.

Simoes, N., Crespo, N., & Moreira, S. B. (2016). Individual determinants of self-employment entry: What do we really know? *Journal of Economic Surveys*, *30*(4), 783–806.

Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: A review of the empirical literature. *Journal of Economic Surveys*, *22*, 795–841.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1–2), 305–353.

Walstad, W. B., & Kourilsky, M. L. (1998). Entrepreneurial attitudes and knowledge of black youth *Entrepreneurship Theory and Practice*, *23*(2), 5–18.

Weigensberg, E., Needels, K., Gould-Werth, A., Patnaik, A., & Lee, J. (2017). *A study of the self-employment assistance program: Helping unemployed workers pursue self-employment*. Mathematica Policy Research.

Zissimopoulos, J. M., & Karoly, L. A. (2007). Transitions to self-employment at older ages: The role of wealth, health, health insurance and other factors. *Labour Economics*, *14*(2), 269–295.

How to cite this article: Fairlie, R. W. (2023). Evaluating entrepreneurship training: How important are field experiments for estimating impacts? *Journal of Economics & Management Strategy*, 32, 607–635. https://doi.org/10.1111/jems.12420

APPENDIX A

TABLE A1 Comparison of GATE participants to self-employment entrants

	GATE participants	Self-employment entrants
	(1)	(2)
Female (%)	46.5	41.9
Black (%)	30.6	9.2
Latino (%)	5.6	14.6
Asian (%)	4.6	4.9
Other (%)	8.0	2.7
Not US born (%)	10.1	20.0
Age	42.42	42.71
Married (%)	48.3	64.3
Highest grade completed	14.45	13.14
HH income $25,000–49,999 (%)	33.2	29.8
HH income $50,000–74,999 (%)	17.6	18.2
HH income $75,000–99,999 (%)	7.0	10.7
HH income $100,000+ (%)	6.6	16.1
Unemployed (baseline/time 0) (%)	55.3	20.7
Sample Size	4197	6086

Notes: (1) For GATE experiment participants, all reported characteristics are measured at the time of application, before random assignment. (2) Self-employment entrants are derived from the Current Population Survey (CPS) microdata for similar years. The sample consists of individuals who are not self-employed in the initial survey month of the 2-month pair, but are self-employed in the second survey month.

Abbreviations: GATE, Growing America through Entrepreneurship; HH, household.

TABLE A2 Regressions for probability of receiving entrepreneurship training

	W1	W2	W3
	(1)	(2)	(3)
Female	0.0242	0.0410	0.0413
	(0.0259)	(0.0276)	(0.0295)
Black	0.0843	0.0694	0.1123
	(0.0365)	(0.0403)	(0.0446)
Latino	0.0688	0.0878	0.1027
	(0.0628)	(0.0643)	(0.0642)
Asian	−0.1439	−0.0679	−0.0730
	(0.0846)	(0.0988)	(0.1129)
Other	0.0072	0.0260	0.1006
	(0.0487)	(0.0536)	(0.0550)
Not US born	0.0706	0.0498	0.0719
	(0.0523)	(0.0564)	(0.0582)

TABLE A2 (Continued)

	W1 (1)	W2 (2)	W3 (3)
Age	0.0100	0.0056	0.0017
	(0.0098)	(0.0105)	(0.0113)
Age squared	−0.0001	0.0000	0.0000
	(0.0001)	(0.0001)	(0.0001)
Married	0.0202	0.0308	0.0485
	(0.0306)	(0.0334)	(0.0365)
Has children	0.0267	0.0154	0.0327
	(0.0286)	(0.0311)	(0.0331)
Highest grade completed	0.0213	0.0393	0.0446
	(0.0112)	(0.0123)	(0.0135)
College education	0.0136	−0.0304	−0.0652
	(0.0498)	(0.0540)	(0.0570)
HH income $25,000–49,999	0.0232	0.0250	0.0709
	(0.0326)	(0.0359)	(0.0390)
HH income $50,000–74,999	0.0185	0.0199	0.0705
	(0.0414)	(0.0456)	(0.0498)
HH income $75,000–99,999	0.0808	0.1354	0.2253
	(0.0577)	(0.0594)	(0.0599)
HH income $100,000+	0.0917	0.0795	0.0980
	(0.0565)	(0.0603)	(0.0651)
Wage/salary work	−0.0133	−0.0089	−0.0311
	(0.0380)	(0.0425)	(0.0453)
Self-employed with no employees	0.0795	0.0867	0.0274
	(0.0460)	(0.0457)	(0.0493)
Self-employed with employees	0.0118	0.0315	0.0014
	(0.0455)	(0.0472)	(0.0503)
Has a health problem	0.0037	−0.0063	−0.0032
	(0.0470)	(0.0520)	(0.0586)
Has relatives or friends who have been previously SE	0.0406	0.0216	0.0270
	(0.0319)	(0.0341)	(0.0370)
Ever worked for relatives or friends who are SE	0.0078	0.0248	0.0410
	(0.0306)	(0.0329)	(0.0348)
Has a bad credit history	−0.0395	0.0277	0.0369
	(0.0293)	(0.0318)	(0.0346)
Currently receiving UI benefits	−0.0386	−0.0297	−0.0610
	(0.0292)	(0.0312)	(0.0330)

(Continues)

TABLE A2 (Continued)

	W1 (1)	W2 (2)	W3 (3)
Has health insurance from current employer	−0.0586	−0.0900	−0.0707
	(0.0385)	(0.0426)	(0.0462)
Autonomy index	0.0066	−0.0156	−0.0078
	(0.0119)	(0.0131)	(0.0139)
Risk tolerance index	0.0174	0.0176	0.0404
	(0.0127)	(0.0140)	(0.0150)
Managerial experience	0.0142	0.0455	0.0554
	(0.0279)	(0.0299)	(0.0327)
Treatment	0.1198	0.2980	0.4986
	(0.3111)	(0.3319)	(0.3616)
Female ∗ treatment	−0.0216	−0.0557	−0.0635
	(0.0327)	(0.0333)	(0.0352)
Black ∗ treatment	−0.0726	−0.0443	−0.0424
	(0.0475)	(0.0500)	(0.0546)
Latino ∗ treatment	−0.0673	−0.0716	−0.0872
	(0.0765)	(0.0754)	(0.0770)
Asian ∗ treatment	0.1176	0.0108	0.0860
	(0.1081)	(0.1202)	(0.1342)
Other ∗ treatment	0.0194	0.0126	−0.0407
	(0.0612)	(0.0631)	(0.0629)
Not US born ∗ treatment	−0.1275	−0.0941	−0.1264
	(0.0683)	(0.0703)	(0.0763)
Age ∗ treatment	0.0090	0.0069	0.0040
	(0.0120)	(0.0125)	(0.0134)
Age squared ∗ treatment	−0.0001	−0.0001	−0.0001
	(0.0001)	(0.0001)	(0.0002)
Married ∗ treatment	0.0040	−0.0086	−0.0441
	(0.0385)	(0.0401)	(0.0429)
Has children ∗ treatment	−0.0196	0.0039	0.0161
	(0.0358)	(0.0369)	(0.0390)
Highest grade completed ∗ treatment	0.0109	0.0003	−0.0124
	(0.0144)	(0.0150)	(0.0163)
College education ∗ treatment	−0.0352	−0.0327	0.0356
	(0.0624)	(0.0642)	(0.0678)
HH income $25,000–49,999 ∗ treatment	−0.0554	−0.0262	−0.0829
	(0.0424)	(0.0443)	(0.0472)

TABLE A2 (Continued)

	W1 (1)	W2 (2)	W3 (3)
HH income $50,000–74,999 * treatment	−0.0216	−0.0321	−0.0875
	(0.0518)	(0.0547)	(0.0581)
HH income $75,000–99,999 * treatment	−0.0813	−0.1646	−0.2571
	(0.0708)	(0.0720)	(0.0722)
HH income $100,000 + * treatment	−0.0573	−0.0680	−0.0968
	(0.0677)	(0.0708)	(0.0751)
Wage/salary work * treatment	−0.0189	−0.0173	0.0127
	(0.0482)	(0.0508)	(0.0537)
Self-employed with no employees * treatment	−0.0613	−0.0610	−0.0432
	(0.0557)	(0.0539)	(0.0584)
Self-employed with employees * treatment	0.0488	0.0310	0.0283
	(0.0545)	(0.0533)	(0.0569)
Has a health problem * treatment	0.0285	0.0052	0.0059
	(0.0585)	(0.0617)	(0.0682)
Has relatives or friends who have been previously SE * treatment	−0.0264	−0.0083	−0.0123
	(0.0404)	(0.0414)	(0.0447)
Ever worked for relatives or friends who are SE * treatment	−0.0099	−0.0086	−0.0060
	(0.0377)	(0.0386)	(0.0402)
Has a bad credit history * treatment	0.0457	0.0010	−0.0225
	(0.0367)	(0.0377)	(0.0404)
Currently receiving UI benefits * treatment	0.0552	0.0533	0.0827
	(0.0373)	(0.0380)	(0.0399)
Has health insurance from current employer * treatment	0.1029	0.1439	0.1085
	(0.0484)	(0.0503)	(0.0537)
Autonomy index * treatment	−0.0046	0.0166	0.0137
	(0.0155)	(0.0163)	(0.0175)
Risk tolerance index * treatment	−0.0091	−0.0094	−0.0325
	(0.0164)	(0.0169)	(0.0180)
Managerial experience * treatment	0.0302	−0.0100	−0.0212
	(0.0351)	(0.0362)	(0.0392)

Notes: (1) All reported characteristics are measured at the time of application, before random assignment. (2) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application.

Abbreviations: HH, household; SE, self-employed; UI, unemployment insurance.

TABLE A3 Self-reported amount that entrepreneurship training helped recipients in various ways

How would you rate the overall usefulness of the services you have received?	Very useful (%)	Somewhat useful (%)	Not very useful (%)	Not at all useful (%)
Treatment group	51.7	33.7	8.5	6.1
Control group	35.8	40.8	10.8	12.7

	Treatment group			Control group		
GATE services	A lot (%)	Somewhat (%)	Not at all (%)	A lot (%)	Somewhat (%)	Not at all (%)
Helped with applying for loans	12.6	21.5	65.9	5.9	17.2	76.8
Helped with deciding whether to pursue self. emp.	39.5	33.1	27.4	23.6	30.0	46.4
Helped with refining the business idea	34.1	37.2	28.8	23.0	32.3	44.7
Helped with credit issues	16.4	25.8	57.7	10.9	17.3	71.7
Helped with developing a marketing strategy	31.4	37.4	31.2	19.6	31.6	48.8
Helped with legal issues	19.3	35.5	45.2	11.3	28.2	60.6
Helped with accounting issues	23.7	35.9	40.4	12.1	26.9	61.0
Helped with hiring and dealing with employees	12.7	24.7	62.6	7.3	18.1	74.5
Helped with networking	28.7	37.9	33.4	23.1	31.2	45.7
Helped with using computers and technology	13.3	26.5	60.2	12.1	22.2	65.7
Helped with dealing with clients	16.7	35.1	48.2	11.3	30.4	58.3
Helped with providing psychological support	16.6	31.0	52.4	13.1	23.8	63.1

Notes: (1) Sample includes treatment and control group participants who received any entrepreneurship training by Wave 1 follow-up survey (6 months). (2) Evaluation of services was asked at W1.

Abbreviation: GATE, Growing America through Entrepreneurship.

TABLE A4 ITT estimates compare to Table 3

	Intent-to-treat estimates		
	No covariates	Covariates	*N*
Dependent variable	**(1)**	**(2)**	**(3)**
Business owner at W1 survey date	0.0464	0.0517	3443
	(0.0166)	(0.0153)	
Business owner at W2 survey date	0.0216	0.0208	3032
	(0.0179)	(0.0172)	
Business owner at W3 survey date	0.0095	0.0025	2446
	(0.0197)	(0.0194)	
Monthly business sales at W1 survey date	−406	−369	3210
	(282)	(288)	
Monthly business sales at W2 survey date	−186	−140	2794
	(353)	(353)	
Monthly business sales at W3 survey date	−495	−620	2323
	(539)	(556)	
Has any employees at W1 survey date	0.0131	0.0140	3438
	(0.0092)	(0.0095)	
Has any employees at W2 survey date	0.0039	0.0020	3023
	(0.0107)	(0.0110)	
Has any employees at W3 survey date	−0.0172	−0.0209	2436
	(0.0123)	(0.0128)	

Notes: (1) The ITT model regresses the listed outcome on treatment. (2) Waves 1, 2, and 3 surveys are conducted at 6, 18, and 60 months after the time of application. (3) Covariates include program sites, female, race, immigrant, age, married, children, education level, household income, self-employed at application, health problems, worked in family business, bad credit history, unemployment compensation, employer-provided health insurance, autonomy, and risk tolerance.

Abbreviation: ITT, intent-to-treat.