Can Firms Use Self-Selection to Improve the Efficacy of Human Capital Investments? Evidence from a Field Experiment

Abstract

In a field experiment, we find a mandatory mentorship program raises worker productivity while a voluntary version of the program does not. A significant reason for this difference is that the lowest productivity employees do not participate when the program is voluntary despite having the largest treatment benefits. Even though our study firm has high-powered incentives, it cannot rely on self-selection to allocate workers to the program. As such, we estimate large differences in the returns to voluntary and mandatory programs. Our findings have implications for human capital investment, resource allocation, experimental design, productivity dispersion, and inequality.

1 Introduction

Stein (2003) begins his handbook chapter by stating, "A fundamental question in corporate finance is this: to what extent does capital get allocated to the right investment projects?" For modern, labor intensive firms with few physical assets, this question becomes: to what extent do human capital investments get allocated to the right workers? Prior work has shown that human capital is an increasingly important driver of firm value (Zingales, 2000; Edmans, 2011; Bloom, Sadun, and Van Reenen, 2016; Nishesh, Ouimet, and Simintzi, 2022), but relatively little work has examined the implications of how firms allocate investments aimed at developing human capital, which totaled \$100 billion in 2022 for U.S. firms (Statista, 2022).[1] We provide novel field experimental evidence on a ubiquitous decision: whether to make human capital investment programs mandatory or voluntary. Whether a mandatory or voluntary program is more effective—and improves resource allocations—hinges on workers' self-selection. If those who would benefit most opt into such programs, voluntary programs provide a cost-effective, targeted investment. If, instead, workers volunteer at random (or worse, negatively with program gains), then mandates may be needed to ensure those with the largest benefits participate in programs. We investigate the implications of this voluntary versus mandatory design choice by conducting a field experiment on mentorship.

We implemented the experiment in a U.S.-based inbound sales call center. Workers at the firm answer incoming calls to sell digital products, like television and internet subscriptions. They are strongly incentivized to increase sales, as commissions comprise over a third of the median employee's compensation. This setting is well-suited to

[1]This spending represents a potentially significant source of intangible capital, which has itself been recognized as increasingly important for U.S. firms (Eisfeldt and Papanikolaou, 2014; Crouzet et al., 2022; Crouzet and Eberly, 2023). While most of the literature has focused on whether firms can rationalize these investments (Becker, 1975; Acemoglu and Pischke, 1999; Fudenberg and Rayo, 2019), far less work examines how program features influence investment returns.

evaluate the design of workplace programs for at least four reasons: (i) sales agents work independently of each other, (ii) we have individual daily sales performance data for all workers, (iii) inbound calls are randomly assigned to sales agents, and (iv) the firm regularly hires new sales agents in cohorts that train together, allowing the program to be administered under different conditions for similar groups.

Our experiment entailed two levels of randomization: the first (high level) is at the cohort level for training classes of new hires, and the second (low level) is at the worker level within a cohort. At the high level, new hire cohorts were randomized into one of two groups, labeled the Voluntary-Condition and the Mandatory-Condition. For Voluntary-Condition cohorts, on the first day of training, the firm's staff asked each new hire to write a private message indicating whether they wanted to participate in the mentorship program. For those who opted in, the lower-level treatment involved randomly assigning sales agents to have a mentor or not. Agents who opted out of participation were not assigned a mentor. For Mandatory-Condition cohorts, the lower-level treatment involved randomly assigning agents to have a mentor or not, without first prompting workers to opt in or out. All mentor assignments were randomly drawn from a pool of established, non-supervisory sales agents who volunteered to be mentors in the program. Matched mentor-protégé pairs were asked to meet for 30 minutes each week for four weeks and to record what they discussed on a worksheet.

We first show that mentorship did not affect the productivity of agents who opted into the program in the Voluntary-Condition. Treatment effects on sales revenue and revenue-per-call for workers who opt into the program are not significantly different from zero. This result could be due to the ineffectiveness of the mentorship program, or it could be driven by *who* selects into the program. Specifically, if employees who benefit the least from the program are the most likely to select in (and if those who benefit the most are the least likely to select in), then the misallocation is due to heterogeneous treatment effects and self-selection. This may be the case in our setting, as a significant share of new hires—about 17%—opted out of program participation.

Our next set of results quantify whether heterogeneous treatment effects and self-selection are large enough in practice to be an important factor for firms' investments in human capital. We begin by estimating the overall effectiveness of the mentorship program (absent self-selection) from the agents randomly assigned a mentor in the Mandatory-Condition. Here we find a large positive effect of the program on productivity. Specifically, daily sales revenue and revenue-per-call increased by 19% and 12%, respectively, over workers' first two months of tenure. Approximately 45% of the initial treatment effects persist through agents' first six months of tenure (well after the program ends), suggesting that the program augmented workers' skills. The positive treatment effects of the mandatory program indicate that general program ineffectiveness is not likely the reason for the negligible treatment effects in the voluntary program. The large difference in treatment effects between the voluntary and mandatory conditions may be explained by heterogeneous treatment effects and self-selection, but other factors—like program framing—could also give rise to these results.

To estimate the impact that self-selection has on the differences in treatment effects between the two programs, we investigate which agents opt out of the voluntary program. We find substantial evidence of self-selection on productivity, where the weakest workers opt out of the program, by comparing the productivity of agents who opted out to those who opted in but were not randomly assigned a mentor. Workers who opt out have average daily revenue and revenue-per-call that are, respectively, 31% and 23% lower than those who opted into the program but were not treated. At first glance, this is surprising because workers at this firm have high powered incentives through commission pay to improve their productivity. To investigate what factors explain the opt out decision, we examined demographics, personality characteristics, and data in the firm's personnel files. One factor, in particular, stands out: workers with low pre-hire assessments assigned during the recruiting process are more likely to opt out of program participation. Workers' demographics, work history, and personality characteristics are not strong predictors of the opt-out decision.[2]

Finally, we quantify heterogeneous treatment effects by estimating how treatment effects vary based on workers' underlying propensity to opt out of the program. When we impute opt-out propensity scores using data from the Voluntary-Condition and examine treatment effect heterogeneity in the Mandatory-Condition, we find that workers who are the least likely to participate in the program have the highest gains. The Voluntary-Condition leaves these workers untreated. As a result, ROI calculations from the Voluntary-Condition alone would yield a false-negative regarding whether the program should be adopted at scale.

To highlight the consequences of such false-negative assessments, suppose that a manager were to conduct a pilot experiment to decide whether or not to invest in human capital development programs. Many experiments in firms tend to recruit volunteers rather than making changes mandatory (List, 2022), suggesting that the Voluntary-Condition in our experiment mimics most experiments that would be conducted in other settings. Here we quantify how much the firm would lose by not adopting the mandatory version of the program due to the false-negative in a Voluntary-Condition

[2] We do not evaluate the extent to which experimental workers' beliefs about the efficacy of the program predicted their opt-out decisions, as we consciously chose not to elicit subjects' ex-ante beliefs about treatment effects because this prompt could have changed sorting into participation. To get at belief-related mechanisms behind the opt-out decision, we rely on evidence from a nationally representative survey of workers that we conducted. In the survey, we find wide variation in firms' practices regarding whether programs are mandatory or voluntary, with substantial rates of non-participation in voluntary programs. Workers cite time constraints and inconvenience as the primary reasons for non-participation; these issues are less applicable for the experiment, as the program was conducted during work hours. The next most common reasons highlight skepticism about personal benefits and the desire to avoid interacting with coworkers or bosses. Intimidation around interacting with more productive coworkers is a theme that emerged in interviews during our prior work in this setting (Sandvik et al., 2020), although we are unable to detect evidence on this factor through variation in personality characteristics of experimental subjects.

only pilot. A conservative back-of-the-envelope calculation suggests that randomizing mentorship among workers who opted out of the program would have raised aggregate treatment gains on revenue and revenue-per-call in the Voluntary-Condition by 6% and 5% of baseline productivity, respectively. If we suppose that the true treatment effect in the Voluntary-Condition is zero, then treatment effect heterogeneity based on selection into the program explains approximately one-third of the difference in treatment effects between the Voluntary- and Mandatory-Conditions.

Ultimately, the firm gained an additional $536,000 in revenues from treating 127 agents in the Mandatory-Condition over a six-month post-treatment horizon. The total costs, including overhead costs, associated with the mandatory program were $97,000. As such, the firm realized a $439,000 return on a $97,000 investment by implementing the mandatory mentorship program. Had the voluntary program been mandatory instead, we estimate that the firm would have gained an additional $201,000, assuming that one-third of the full treatment effect carries over from the Mandatory-Condition.[3]

Our findings indicate that self-selection and treatment effect heterogeneity explain a significant portion, but not all, of the difference in effectiveness between the two programs. While there are several possible alternative explanations for the differences in treatment effects between the two conditions, we find little evidence for any of them. In particular, we show that, across both conditions, the mentorship program had no impact on retention, which, therefore, cannot explain the observed sales revenue treatment effects. Furthermore, our findings do not appear to be driven by information leakages, discouragement or encouragement from treatment status, other violations of the Stable Unit Treatment Value Assumption (SUTVA), or crowding out of organic mentorship. In particular, we expected some form of framing effect—like a signal

[3] The long-term benefits could possibly be even greater if managers can effectively refine allocation procedures relative to random treatment assignment (Li et al., 2020; Johnson et al., 2022).

that a mandatory program means workers should engage with treatment—would explain part of the remaining difference in treatment effects. However, our proxies for engagement—meeting completion rates and worksheet contents—explain little of the remaining disparities between conditions.

Our findings contribute to the literature in several ways. First, we contribute to the growing literature that seeks to explain the reasons for disparities in performance and wages across workers. Prior research has shown that management practices, managerial talent, capital formation, labor market concentration, and firm size are important drivers of differences in productivity, innovation, and compensation across firms and workers (Bertrand and Schoar, 2003; Bloom and Van Reenen, 2007; Larrain, 2015; Lazear, Shaw, and Stanton, 2015; Mueller, Ouimet, and Simintzi, 2017b; Custódio, Ferreira, and Matos, 2019; Benmelech, Bergman, and Kim, 2022). Other work directly examines the allocation implications of different work and personnel management practices (Benson et al., 2019, 2024). Our experiment provides an additional explanation for these differences: workers who are the most likely to benefit from workplace programs and other human capital development resources may be the least likely to take advantage of them. This is somewhat surprising given workers' high-powered incentives at this firm (Lazear, 2000; Lazear et al., 2016). As such, our findings shed light on why wage inequality may persist despite the presence of performance-based pay and the efforts made by managers and regulators to reduce such disparities (Pan et al., 2022).

Another contribution of our paper is that we experimentally identify a large positive intent-to-treat effect of mentorship programs in a workplace setting. While a positive treatment effect might be expected because of the widespread prevalence of mentorship programs (Gutner, 2009), causal evidence is thin due to nonrandom selection and participation in most mentorship settings (Allen et al., 2017). As such, our evidence adds to a growing literature on the efficacy of mentorship (Lyle and Smith, 2014; Porter and Serra, 2020; Ginther et al., 2020). Furthermore, our use of a randomized controlled trial showcases the importance of using experiments to identify effects and mechanisms that cannot be teased out of observational data (for other examples of experiments in financial economics, see Iyer et al. (2008); Custodio et al. (2020, 2022); Kim et al. (2022)).

Our findings underscore that human capital development investments have design features that make them distinct from physical capital investments, which have been widely studied (Panousi and Papanikolaou, 2012; Kogan et al., 2017; Campello et al., 2024). Specifically, we demonstrate that the choice between making a workplace program voluntary versus mandatory is not trivial, as the net present value of a program depends crucially on how workers participate (self-selection) and heterogeneous treatment effects. Our results suggest that employers may missallocate program resources if the individuals who opt out are actually those who would benefit the most from participation. Experiments such as ours can help to improve allocation decisions, as sorting by the best or most eager workers into voluntary programs may obfuscate ineffective investments.

Finally, our findings contribute to the nascent literature in finance and accounting on human capital management (Agrawal, Hacamo, and Hu, 2021; Hacamo and Kleiner, 2022; Liu, Makridis, Ouimet, and Simintzi, 2023). Investors are increasingly interested in understanding the efforts that firms make to invest in their employees (Investors, 2018), and recent regulation encourages firms to disclose "human capital measures and objectives that address the attraction, development, and retention of personnel..." (SEC, 2020). Related to this, Bourveau et al. (2022) show that firms increasingly disclose how they recruit, engage, educate, and retain their employees. Workplace programs, such as mentorship programs, can be an effective way for firms to invest in human capital development. However, our results suggest that organizations cannot always rely on workers with the largest treatment gains to sort into voluntary programs.

2 Firm Setting

The mentoring program occurred in an inbound sales call center from January to December 2019, with data collection on protégé performance continuing after the conclusion of the mentorship program. The firm markets and sells the services of several companies/brands, most of which are television, phone, and internet providers. Sales agents answer incoming calls from potential customers and sell digital services with the goal of closing sales and up-selling premium service packages. Firm insiders report that learning the sales process (e.g., how to run credit checks for equipment lease compliance or determine whether callers qualify for regional sales promotions) and how to up-sell can be challenging for new hires.

When hired, sales agents begin a two-week training program, where they learn the sales process through lectures and by listening to other agents' live calls. Once agents complete their two-week training, they are allocated to a team and begin responding to sales calls. Teams are typically comprised of 10–15 individuals, overseen by a (direct) sales manager, who is responsible for monitoring performance and troubleshooting issues faced by the agents. Agents eligible for the mentorship program were spread across seven different sales divisions, corresponding to different companies' brands or products.

This setting has several attractive features for studying the efficacy of mentorship. Most importantly, the firm provided us with individual-level performance measures for each sales agent. Sales agents work independently on a call from start to finish, without subsequent hand-offs which could contaminate outcomes. Incoming calls are allocated to the next available agent within the appropriate division (each division receives calls from different phone numbers depending on the service being sold and the location of the callers, and the opportunities are then randomly allocated to agents). Agents generate revenue through each sale they make. The firm's focal productivity measure is revenue-per-call (RPC) because it allows managers to remove demand variation when comparing performance across workers. In addition, total revenue is important for workers, as the absolute amount of sales goes into workers' commission pay. At the end of each week, the total amount of revenue generated is multiplied by an agent's commission rate. The commission rate is a coarse function of the agent's selling efficiency (determined by RPC and revenue per hour worked), relative to other agents in the same division, and averages about 8% of total revenue.[4] Multiplying the worker's weekly revenue and commission rate determines their commission pay for that week. Sales agents also earn an hourly wage, which is above the federal minimum wage and increases with tenure.

3 Experimental Design

The experiment involves two high-level treatment conditions that were first assigned at the new-hire training class (cohort) level. Lower-level sub-treatments involving the assignment of mentors then occurred within each cohort. Training cohorts are specific to an office location and division. Cohorts joined the firm on a rolling basis during the experiment. We randomly assigned each cohort to either the Voluntary-Condition (probability 60%) or the Mandatory-Condition (probability 40%). Agents in the Voluntary-Condition were given the option to opt in or out of mentoring. Those who opted out did not receive a mentor. Agents in the Mandatory-Condition and those in the Voluntary-Condition who opted in were randomly assigned a mentor or

[4]There is mild relative performance evaluation in this setting, and commissions increase at each quintile of selling efficiency. Helping another agent is unlikely to change relative rankings across quintiles, as the probability is small that any two agents are pivotal at the commission rate kink.

not according to the following rule: if the supply of available mentors was greater than 50% of the cohort size, then approximately half of the agents would be assigned a mentor (the firm requested that we error on the side of randomly allocating mentors to more agents when possible, e.g., rounding up in instances when an odd number of agents was present); otherwise, the available mentors would be assigned at random to those eligible to receive a mentor.[5] The pairing of mentors and new hires always occurred at random.

Figure 1 displays the allocation of cohorts and agents to the different conditions and treatments in the experiment. There were 591 program-eligible sales agents spread across 52 new hire cohorts.[6] Thirty-one cohorts and their 327 sales agents were allocated to the Voluntary-Condition, whereas the other 21 cohorts and 264 sales agents were allocated to the Mandatory-Condition. In the Voluntary-Condition, 272 agents (83%) chose to opt in, of which 155 agents (57%) were randomized to receive a mentor, and the remaining 117 were not. The remaining 55 agents (17%) in the Voluntary-Condition chose to opt out of receiving a mentor.[7] Among the agents in the Mandatory-Condition, 127 agents (48%) were randomized to receive a mentor, and the remaining 137 were not.

3.1 Timeline for Administering the Program and Communicating Treatment Allocations

Prior to starting the two-week training protocol, each cohort was allocated to either the Voluntary- or the Mandatory-Condition, and the staff administering the program was made aware of the cohort's assignment. All new hires were asked to complete a

[5]When mentor supply fell below 50% of the number of eligible new hires, the most common reason was conflicting obligations to mentor other cohorts in the same division or office.

[6]Our prior working paper version reports 53 cohorts assigned to treatments. We erroneously coded one cohort that had no available mentors as eligible for the experiment.

[7]Across months, opt-out rates range from 6% to 43%, with no obvious time trend.

survey on the first day of training, which asked questions about their personality traits, work styles, and work experiences (specifically, whether they had call center and/or sales work experience). We use these survey responses to identify the characteristics of individuals who opted into versus opted out of mentoring.

For cohorts in the Voluntary-Condition, the staff described the mentoring program to the newly hired agents and told them that they could either opt in or opt out of participating. The agents were told that a randomly selected subset of those who opted in would receive a mentor at the end of the training period. The staff explained that the supply of mentors was limited and an outside research team would help with the randomization to ensure fairness in the assignment.[8] To avoid peer influence in program participation (Dahl et al., 2014), agents were asked to write on a piece of paper whether they wanted to opt in or out of the mentoring program, making their decision anonymous to their peers. Among those who opted in, agents were either randomly assigned a mentor or not, based on the assignment rule described above. For cohorts in the Mandatory-Condition, agents were either randomly assigned a mentor or not based on the assignment rule described above. Agents assigned a mentor were informed of this assignment by the within-firm mentoring staff during the last days of their training. To reduce the possibility of discouragement among agents in the Mandatory-Condition who were not assigned a mentor, the staff did not initially inform them about the mentorship program. If agents inquired about why they were or were not assigned a mentor, the staff told them that there was a limited supply of mentors, and that available mentors were already randomly allocated to new hires.[9]

[8]The staff members were asked to read the following statement to new hires in the Voluntary-Condition: "We have recently begun a mentorship program to help newly hired sales agents when they begin working on the sales floor. Agents who opt into the program and are chosen by [the research team] will be assigned a mentor. Your mentor will approach you during your first week on the sales floor to initiate the mentoring relationship. The program will run from your first week on the sales floor to your fourth week on the sales floor, and you and your mentor will meet once a week to discuss your progress."

[9]The staff reported to the authors on multiple check-in calls that they found no evidence of dis-

Across all treatment conditions, the two weeks of training remained exactly the same for all agents, regardless of their treatment assignment. After the two weeks of training, new hires graduated to work as regular agents, began taking customer calls, and had measurable sales productivity metrics. It was only then that meetings with mentors commenced. To facilitate meetings occurring, the firm built specific times to meet into mentors' and protégés' schedules. The mentoring relationships lasted for four weeks.

Mentors and protégés met once per week for approximately 30 minutes and completed a worksheet. They were free to discuss any topic, but the worksheet had to be completed for the mentor to receive credit for the meeting (as described below). Records of meeting occurrences and completed worksheets were kept by the staff and given to us. Shortly after their fourth and final week of mentorship meetings, protégés were asked to complete a post-mentorship survey about their experience. Although completion rates for the final survey were low, we use the data to provide insight into whether meetings continued after the formal program and whether agents viewed the experience as beneficial.

3.2 Identifying Mentors

The firm's staff sourced mentors by announcing to incumbent sales agents that a mentoring program for new hires would occur and that agents could volunteer to be a mentor. The staff directly asked some promising candidates to participate. If the staff and sales managers felt an agent was not suited to be a mentor, he or she was excluded from consideration. Mentors were given two main incentives to participate. First, in exchange for each pre-scheduled, confirmed meeting they held with their protégé, they received internal currency ("kudos" dollars) worth approximately $10. Second, incum-

couragement among the agents who did not receive a mentor.

bent sales agents were told that effective mentoring would help demonstrate leadership potential for future promotion considerations.

3.3 Hold-Out Cohorts to Test Stable Unit Treatment Value Assumption Violations

There were 217 agents hired throughout the experiment in cohorts that were ineligible for the mentorship program. Ineligibility largely arose because these cohorts entered at times when mentor supply was lacking. Insufficient mentor supply typically arose when the firm was hiring new cohorts back-to-back, but in some cases projected call volumes relative to available divisional staffing meant that potential mentors would have insufficient time to meet with new hires. Agents in these cohorts form hold-out groups that were not informed about the mentorship program. Variation in treatment eligibility at the cohort level allows us to test for discouragement effects in the control group and other possible violations of the Stable Unit Treatment Value Assumption (SUTVA). Although these hold-outs were not randomly assigned, they have similar characteristics as program-eligible cohorts in the same division and office. We leverage these hold-out cohorts to compare the productivity of hold-out new hires with non-treated agents in program-eligible cohorts, showing that SUTVA violations are not likely to be a concern in our setting (see the Online Appendix, Section O.A).

3.4 Pilot Data

We piloted our design in the firm from January to May of 2019 to ensure we could logistically implement the program. The pilot surfaced several virtues of the program while assuaging feasibility concerns: (i) there was sufficient interest amongst seasoned agents to mentor new hires, (ii) the firm could schedule meetings between mentors and protégés, (iii) mentors and protégés would engage with the protocol as designed, (iv) anecdotal evidence indicated that protégés felt they benefited from the mentorship, and (v) there were no indications of discouragement among non-mentored agents. As a result, we moved forward with the experimental design described thus far, which varied from the design of the pilot in only two ways. First, to accommodate scheduling, we changed the duration of the mentorship program from five meetings over five weeks (pilot) to four meetings over four weeks. Second, at the beginning of the pilot, the allocation of cohorts to the Voluntary-Condition and Mandatory-Condition was determined by the location of each cohort; i.e., all cohorts at one corporate campus were allocated to one condition, and those at the second campus were allocated to the other. This allocation was chosen to limit potential spillovers between the Voluntary- and Mandatory-Conditions (e.g., workers potentially talking about the choice to opt in). Within each condition, the firm's staff observed no discussion of program logistics among new hires or spillover effects within or across cohorts. There were also no complaints from agents in the Voluntary-Condition who requested but did not receive a mentor. Accordingly, we determined the risk of spillovers *across* conditions was small and the logistics were possible such that we could randomize Voluntary- and Mandatory-Condition assignment within offices as well. In all cases, the allocation of mentors to eligible agents was always random. No other changes were made between the pilot period and the later cohorts. The pre-registration text was finalized after the pilot and is documented in the Online Appendix (see Section O.B).

Based on power calculations and the hiring projections given to us by the firm, we expected the firm to hire 619 agents across 46 cohorts after the pilot period (May to December of 2019). The actual hiring at the firm was much less frequent and intense, with the firm only on-boarding 276 agents across 27 cohorts that were eligible for the mentorship program. We were not able to extend the mentorship program into 2020, as COVID-19 forced all employees to work remotely. Since the firm's actual hiring behavior was substantially less intense than expected, and given the similarity between the experimental design in the pilot period and the test period, our empirical analyses include the 315 agents and 25 cohorts from the pilot to improve statistical power. We detect no differences in treatment effects or imbalance in worker characteristics between the pilot cohorts and those from the post-pilot period (see Table A.1).

3.5 Balance Across Treatments

Agent characteristics are balanced across the different treatment groups and conditions of the experiment. Table 1 Panel A displays cohort-level balance tests across observable characteristics for the Voluntary-Condition compared to the Mandatory-Condition (the top level of randomization). There are no significant between-condition differences in average agent age, gender, marital status, hiring score (recruiters' evaluation of the worker's suitability for the position), and referral status. The average agent age in both groups is about 23 years old, women make up 40% of the agents in the Voluntary-Condition and 43% of agents in the Mandatory-Condition, and 13%–16% of agents are married in the two groups. The average hiring scores (which have a maximum value of 1) are 0.85 and 0.83, respectively. These scores are based on the recruiters' perceptions of applicants' sales experience, ability to adhere to the sales process, self-awareness, competitiveness, and personal motivation. We also report adjusted hiring scores, which take into account some recruiters' relative leniency compared to others—akin to curving grades. Throughout our analysis, we use the adjusted hiring score measure because it is a better predictor of opting out of the program relative to the raw hiring scores, but our results are not sensitive to the use of raw hiring scores, which we discuss in Section 4.2.[10]

[10] There are 15 recruiters in the data. Some recruiters systematically give higher scores than others conditional on observed performance of the workers they evaluate. We account for this using a procedure that adjusts for the stringency or leniency of each recruiter. We measure recruiter leniency using a sample of non-mentor eligible agents and regressing raw hiring scores on productivity (specifically,

Panel B of Table 1 considers the second level of randomization, the allocation of mentors to new hires *within* the Voluntary-Condition or Mandatory-Condition. Columns (1) and (2) show the agent-level average characteristics in the Voluntary-Condition for those who did and did not receive a mentor, respectively, conditional on opting into the program. These two groups are similar in age, gender, marital status, hiring scores, adjusted hiring scores, and referral status. Columns (3) and (4), and the associated p-values show that agents assigned mentors and those that were not in the Mandatory-Condition are similar across these observable characteristics as well.[11] We defer discussion of differences between agents who opt into and out of the program to Section 4.2.

4 Results

4.1 Treatment and Selection Effects of Mentoring on Productivity

We begin our results presentation by describing differences in productivity by high-level treatment condition (Voluntary or Mandatory) and low-level sub-treatment cell (assigned a mentor, not assigned a mentor, or opted out). We refer to agents assigned a mentor as "mentored," which we use as shorthand for treatment assignment in an intention-to-treat framework. Our main productivity outcomes of interest, $y_{i,t}$, are total

the inverse hyperbolic sine of revenue and the inverse hyperbolic sine of revenue-per-call), recruiter fixed effects, brand fixed effects, and time fixed effects. We then shrink the recruiter fixed effects (that are net of the productivity adjustment) using the procedure in Lazear et al. (2015). We subtract the adjusted recruiter fixed effects from the raw hiring scores to return the adjusted hiring scores.

[11] We also check for balance across mentor characteristics and across assignment to divisions based on estimated division-level productivity for non-mentor eligible workers. Although mentors were not designated exclusively to either the Voluntary- or Mandatory-Condition, their assignment across divisions could have been imbalanced. Table A.2 shows that the mentors of agents in the Voluntary- and Mandatory-Conditions were similar in age, gender, marital status, and tenure. Mentors were never informed about which condition their protégés were in. Table A.3 shows that assignment is balanced on the productivity metrics of non-mentor-eligible new hires (from hold-out cohorts) across divisions that each cohort in the Voluntary- and Mandatory-Conditions joined. Documentation of mentor instructions and the mentor-protégé worksheet can be found in the Online Appendix (see Section O.C).

daily sales revenue (Revenue) and daily revenue-per-call (RPC). Total daily revenue directly relates to the firm's profitability, while accounting for the opportunity cost of time spent meeting a mentor. RPC captures selling efficiency on a given opportunity. Different divisions have different base-levels of revenue and RPC. We pre-registered a natural way to capture gains from mentorship as percentage changes relative to base levels of productivity within a cohort. To do so, we use the inverse hyperbolic sine transformation (IHS) so that parameter estimates can be interpreted as approximate percentage changes.[12] We use a sample of agent-day productivity data for all program-eligible agents in their first two months on the job after completing training (estimates for months three through six are discussed in Section 4.1.4).

We estimate the following model using ordinary least squares separately for the Voluntary- and Mandatory-Conditions:

$$\mathrm{y}_{i,t} = \alpha + \beta_1 \mathrm{Mentored}_i + \gamma_j + \varepsilon_{i,t}, \tag{1}$$

where $Mentored_i$ equals one for agents who were randomly assigned to receive a mentor, and zero otherwise. The t subscript denotes the calendar date, and γ_j is a cohort fixed effect at the unit of randomization that absorbs product-/brand-level differences. The model also contains an idiosyncratic error term, $\varepsilon_{i,t}$, and we cluster standard errors by cohort.

4.1.1 Treatment Effects in the Voluntary-Condition

We estimate Equation (1) on the sample of Voluntary-Condition agents who signaled their interest in receiving a mentor. Due to random assignment of mentors to some opt-

[12]We use the inverse hyperbolic sine transformation because some agents have zero revenue days, but results are qualitatively similar if we use the natural logarithm of one plus revenue or one plus RPC. We display results using the logarithmic transformation in Table A.4.

in agents and not to others, the parameter β_1 is the average treatment effect of receiving a mentor in the Voluntary-Condition conditional on selection into program participation. We tabulate the results in Columns (1) and (2) of Table 2 for IHS(Revenue) and IHS(RPC), respectively. In both columns, we estimate effects that are statistically indistinguishable from zero. Random assignment to be mentored had a negligible effect on the productivity of workers in the Voluntary-Condition conditional on their opting into the program.

Given the ubiquity of mentorship program, the lack of positive treatment gains is surprising. This result could be due to the ineffectiveness of mentorship programs or to self-selection in who opts into the program, causing the treatment effect estimated in the Voluntary-Condition to deviate from the overall population-average treatment effect. We now pivot the analysis to focus on the Mandatory-Condition, identifying treatment effects *sans* self-selection.

4.1.2 Treatment Effects in the Mandatory-Condition

Due to random assignment of mentors to some Mandatory-Condition agents and not to others, when estimating Equation (1), the parameter β_1 is the overall average treatment effect of receiving a mentor across the *entire* population, not the treatment effect conditional on opting into the program. We tabulate the results in Columns (3) and (4) of Table 2 for IHS(Revenue) and IHS(RPC), respectively. In both columns, we estimate positive effects that are statistically significantly different from zero. The estimate in Column (3) implies that mentored agents in the Mandatory-Condition generated 18.6% ($= e^{0.171} - 1$, p-value $= 0.022$) more daily sales revenue than their non-mentored peers. The estimate in Column (4) implies that mentored agents generated 11.9% ($= e^{0.112} - 1$, p-value $= 0.015$) more in revenue-per-call. The large productivity gains here contrast with those in the Voluntary-Condition. Had the analysis been run only among those who selected into randomization, which is typical for many RCTs across disciplines ranging from medicine to economics, we would have falsely concluded that the program was not effective in the population. Instead, these results suggest that different procedures for administering the program can change inference. We turn now to assessing why estimates differ across conditions.

4.1.3 Self-Selection in the Voluntary-Condition

How much of the difference in treatment effects across the Voluntary- and Mandatory-Conditions arises from selection into who receives a mentor? To address this question, we estimate the bias in program participation as a function of workers' baseline productivity by comparing non-mentored agents in the Voluntary-Condition who opt into the program with those who opt out. We do this by estimating the following model using ordinary least squares:

$$\mathrm{y}_{i,t} = \alpha + \beta_1 \text{Voluntary Opt-Out}_i + \gamma_j + \varepsilon_{i,t}, \tag{2}$$

where *Voluntary Opt-Out*$_i$ equals one for agents who opted out of the mentorship program, and zero otherwise. The parameter β_1 captures the differences in productivity between agents who did, and those who did not, opt out of the program. The revenue and revenue-per-call results are reported in Columns (5) and (6) of Table 2, respectively. The estimate in Column (5) implies that opt-out agents generated 30.9% ($= e^{-0.369} - 1$, p-value $= 0.005$) *less* revenue per day than non-mentored, opt-in agents. The estimate in Column (6) implies that opt-out agents were 23.2% ($= e^{-0.264} - 1$, p-value < 0.001) less productive on a per-call basis. Combined, these results suggest that the agents who opted into program participation were significantly more productive, on average, than those who opted out.

4.1.4 Pooled Estimates, Additional Productivity Measures, Multiple Tests, and Long-Term Outcomes

In Columns (7)–(9) of Table 2, we estimate all three effects of interest simultaneously in a single model that includes all mentor-eligible agents across both the Mandatory-Condition and the Voluntary-Condition. We estimate the following model using ordinary least squares:

$$\begin{aligned} \mathrm{y}_{i,t} &= \alpha + \beta_1 \text{Mentored}_i + \beta_2 \text{Mentored}_i \times \text{Voluntary}_i \\ &\quad + \beta_3 \text{Voluntary Opt-Out}_i + \gamma_j + \varepsilon_{i,t}, \end{aligned} \tag{3}$$

where $Mentored_i$ equals one for agents who were randomly assigned to receive a mentor, and zero otherwise, $Voluntary_i$ equals one for agents in the Voluntary-Condition, and zero otherwise, and $Voluntary\ Opt\text{-}Out_i$ equals one for agents in the Voluntary-Condition who opted out of the mentorship program, and zero otherwise. In this case, the parameter β_1 captures the treatment effect of mentorship among agents in the Mandatory-Condition, β_2 captures the difference in treatment effect among opt-in agents in the Voluntary-Condition, relative to agents in the Mandatory-Condition, and β_3 captures the selection effect among non-mentored agents in the Voluntary-Condition. The baseline effects for the Voluntary- and Mandatory-Conditions are absorbed by the cohort fixed effects, which also control for differences in productivity that are specific to the time when agents entered the firm and the differing products sold.[13] Pooling the models allows us to test whether treatment effects differ between conditions.

The results of the pooled model are reported in Columns (7) and (8), with IHS(Revenue)

[13] All of our pre-registered specifications include cohort fixed effects, as we expected that between cohort variation would significantly increase minimum detectable effect sizes. With cohort fixed effects, calendar time and elapsed time since hire are co-linear. In a balanced panel with a short time window, cohort-fixed effects absorb time-fixed effects. We show in Figure A.1 that our results are robust to the inclusion of date fixed effects as well as to the use of several other alternative specifications.

and IHS(RPC) as the outcomes of interest, respectively. In the top row, the productivity treatment effects for the Mandatory-Condition are identical to the prior estimates. In the second row, the point estimates of productivity differences for those who opt out in the Voluntary-Condition are similar to the prior estimates, but they are not identical because the sample includes those who were randomized into receiving a mentor. The third row shows that the treatment effect of receiving a mentor in the Voluntary-Condition is statistically different than the treatment effect of receiving a mentor in the Mandatory-Condition. The bottom row of Table 2 reports tests of the null that treatment effects are zero for those mentored in the voluntary program, as their treatment effects are the sum of the coefficients on *Mentored* and *Mentored* $\times$ *Voluntary*.

We address multiple-hypothesis testing and potential alternative productivity measures in two related ways. First, in Column (9) we repeat the pooled analysis with an alternative dependent variable that factors in additional outcomes tracked by the firm: revenue-per-hour, which captures the revenue generated by an agent per working hour, and adherence, which is bounded by zero and one and captures how closely agents adhere to their pre-set schedules (e.g., having the requisite amount of up-time to take calls, while taking breaks at the correct time). In practice, these productivity measures are highly correlated with our primary outcomes, daily revenue and RPC. To account for this and to reduce the number of tests we perform, we follow Anderson (2008) and create the measure *Index*, which is the standardized, weighted summary index of all performance metrics—IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence—normalized to have mean zero and unit standard deviation for non-mentored agents in the Mandatory-Condition.[14] Using this omnibus measure as the dependent variable,

[14]The summary index approach has been used to evaluate education interventions when there are multiple potential outcomes (Deming et al., 2014). The procedure first demeans and standardizes each individual outcome by the control group standard deviation (in this case, non-mentored agents in the Mandatory-Condition). The index is then the weighted sum across inputs, where the weights come from the inverse of the covariance matrix of the standardized measures (GLS). Anderson (2008)

we continue to find that the program generally raised productivity when it was mandatory, that the program had no effect in the Voluntary-Condition, and that Voluntary-Condition participants who opted in were stronger overall. Note that the economic magnitudes of the point estimates in Column (9) must be interpreted differently, as they are in standard deviation units relative to the control mean.

We also correct for multiple hypothesis testing concerns using a second suggestion by Anderson (2008), wherein we report sharpened q-values. These values are analogous to a p-value after adjusting for the False Discovery Rate (FDR). The q-values are adjusted for tests on all regressors reported in Columns (7)–(9) of Table 2, as well as all regressors reported in Columns (7)–(9) of Table A.5, which capture the long-term treatment and selection effects of mentorship in months 3–6 of agents' tenure with the firm. The q-values indicate that inference regarding our main point estimates is robust to holding fixed the proportion of false positives as the number of tests increases.[15]

In Table A.5, we show that about 45% of the point estimates from months 1–2 persist through months 3–6 for mentored workers in the Mandatory-Condition, while the effect of having a mentor in the Voluntary-Condition remains close to zero. The longer-term point estimates have larger standard errors relative to the effects at 1–2 months of tenure for two primary reasons: a) there is an increase in residual variation as agents gain experience, causing productivity to fan out, and b) there are fewer agents who remain at the firm over longer time horizons. While we lose precision, the pattern of estimates suggests that the mentorship program helped treated workers in the Mandatory-Condition over the longer-term.

We study the mechanisms underlying the differing treatment effects between the

argues that this approach has three advantages: (i) it allows for a single test rather than multiple tests across different outcomes, (ii) it is a test of whether a program has a general effect, and (iii) the tests are potentially more powerful than multiple tests with marginal significance.

[15] The estimated sharpened q-values are conservative in our case because they do not account for the positive correlations across tests.

Voluntary-Condition and the Mandatory-Condition throughout the rest of the paper. While this initial evidence suggests that heterogeneous treatment effects and self-selection might be the cause of the differences, we push further using a few different approaches. In particular, because our experiment was done in the field, the greater effectiveness of mentorship in the Mandatory-Condition could have been caused by logistical differences, framing issues, or other unobserved factors that could have been encountered when implementing a mandatory versus a voluntary program within a real workplace.

4.2 Differences Between Agents who Opt Out and Those who Opt In

The heterogeneous treatment effects and self-selection explanation for the differences in observed treatment effects between conditions implies that the agents who opted out in the Voluntary-Condition had the most to gain from the program. Before we formally estimate these potential gains, we consider how agents who opt out differ from those who choose to opt into the program on observables. We restrict our sample to the 365 agents in the Voluntary-Condition who were given the choice to opt in or out, and we estimate logistic regressions of a *Voluntary Opt-Out* indicator on worker characteristics.[16]

Worker demographics from the firm's personnel records and personality traits obtained in subsequent on-boarding surveys do little to explain program participation. However, hiring scores and correlates of engagement—e.g. completing surveys—can partially explain participation decisions. In Column (1) of Table 3, we report no difference between agents that opt in and opt out based on age and marital status. The impact of gender on the opt-out decision is significant at the 10% level in Column (1),

[16] In this analysis, we include agents who did not complete training, in which case they do not have productivity data. This accounts for the difference in unique worker counts between this sample and that reported in Figure 1. Results are similar when we use only workers who completed training to examine the determinants of opting out.

but it loses explanatory power and decreases substantially in magnitude when controlling for additional covariates. Participation decisions do not depend on an agent's location (a fixed effect for one office compared to the other) or whether an existing employee referred the agent (following Friebel et al. (2023)). We find that agents with higher adjusted hiring scores—net of recruiter leniency (see Section 3.5)—are more likely to opt into the program. As an agent's hiring score is given to them by the recruiter who interviewed them for the job, this suggests that recruiters' assessments of agents' suitability for the job can predict their program engagement.[17] Computing marginal effects from the logit model, we find that an increase in the adjusted hiring score of 0.10 (approximately the interquartile range in the sample) is associated with a 6.9 percentage point decrease in the likelihood that the agent opts out of the program. Participation decisions also do not depend on whether the agent had prior call center experience or sales experience (which we collected from the new hire survey for 341 agents) or an agent's division (reported in Column (2)). Personality characteristics, also collected from the new hire survey, are weak predictors of the opt-out decision (reported in Column (3)). However, we find that agents that did not complete the new hire survey have a higher propensity to opt out, as reported in Column (4).

We also assess the extent to which the predictors of agent opt-out explain variation in agent productivity. We use a sample of agents from the Voluntary-Condition who were not mentored, and regress realized productivity on the factors that potentially explain program participation. Column (5) displays the baseline productivity regression results controlling only for agent demographics, hiring scores, referral status, and cohort fixed effects (which absorb the location dummy). Importantly, the coefficient on *Adjusted Hiring Score* is positive and statistically significant.

[17] We are missing hiring score data for 25 agents, so we set their hiring scores to zero and include an indicator variable that they had missing data. We find similar results in Table A.6 when using raw (non-adjusted) hiring scores.

A one standard deviation change in the adjusted hiring score (approximately 0.07 units), yields a 19% change in revenue. This suggests that both the opt-out decision and the observable agent characteristics (that go into opting out) help to explain on-the-job productivity. Column (6) adds data from the new hire survey and the *Missing Survey* dummy. The coefficient on *Adjusted Hiring Score* is larger in magnitude—even when other characteristics are included. The results are similar in Column (7) when the dependent variable is IHS(RPC), showing that hiring scores predict on-the-job performance. At this firm, and likely in others, workers with low pre-hire assessments are less productive than other workers with more favorable evaluations. As we show later, it is possible that workplace programs may help remediate this lower level of initial productivity.

4.3 Heterogeneous Treatment Effects For Agents Who Are Likely to Opt Out of the Program

Next, we conduct tests of heterogeneous treatment effects that are invariant to framing or logistical differences between the Voluntary and Mandatory-Conditions. Here we ask whether workers who were the most likely to opt out of the program have the largest *individual* treatment effects in the Mandatory-Condition. To do this, we use the estimates in Column (1) of Table 3 to impute opt-out propensity scores for workers in the Mandatory-Condition. We classify agents in the Mandatory-Condition as either $High_{Opt}$, if their opt-out propensity score is in the top tercile of the distribution, or Low_{Opt}, if their opt-out propensity score is in the bottom 66.6% of the distribution. We use a larger threshold than the opt-out rate in the Voluntary-Condition because (i) the individual propensity scores are less than 1, implying we need more workers to yield the total number of those who opt out in the Voluntary-Condition and (ii) we want a sample that is large enough for reliable inference. If treatment effects are monotone in the propensity to opt out of treatment, these choices are conservative. We then estimate Equation (1) on these subsets of the data along with pooled models that allow us to test for differences in treatment effects between workers with high and low propensities to opt out of participation.

We find that agents in the Mandatory-Condition with a high estimated likelihood of opting out had a significantly greater treatment effect of mentorship than did their peers who were less likely to opt out, reported in Table 4. The estimate in Column (1) shows that agents who were most likely to opt out of the program had revenue gains of over 38% ($= e^{0.324} - 1$), whereas the estimate in Column (2) shows that other agents had estimated gains of about 7%; zero is included in the confidence interval for the bottom two terciles. The pooled estimate on $Mentored \times High_{Opt}$ in Column (3) rejects equality of the treatment gains within the Mandatory-Condition between high and low opt-out agents, providing evidence in favor of heterogeneous treatment effects. Column (3) also shows that agents in the highest tercile of the opt-out propensity are about 21% ($= e^{-0.239} - 1$) less productive than other agents in the Mandatory-Condition. The results in Columns (4)–(6) report similar patterns when using IHS(RPC) as the dependent variable.[18] We note that the high opt-out propensity agents here are again less productive than those who are more likely to participate in the program.

We emphasize that the larger treatment gains for high opt-out propensity workers *within* the Mandatory-Condition points strongly to treatment effect heterogeneity and

[18]We also pre-registered a procedure for estimating heterogeneous treatment effects. We discuss this procedure in the Online Appendix (see Section O.D). The results of the pre-registered estimations provide additional evidence that the treatment effects of mentorship are greatest among agents who are most likely to have opted out of the program. Our pre-registered estimates of treatment gains for opt-out agents are larger than those given here because the pre-registered estimator imposes that the treatment effects for agents who opt in are constant across the voluntary and mandatory programs. The approach in Table 4 allows the treatment effects among those who are likely to opt in to differ across conditions, and we find that the treatment effects for agents who had a high probability of opting into the program are modestly positive (but insignificantly different from zero) in the Mandatory-Condition. The pre-registered approach is sensitive to this variation in the treatment effect estimates, and the results in Table 4 are conservative.

self-selection as a driver of the across-condition differences in estimates of program effectiveness. Our point estimates imply that had the Voluntary-Condition drawn in all workers (including those who opted out), overall revenue (revenue-per-call) gains for treated workers would have been 6% (5%) higher. Assuming the actual treatment effect in the Voluntary-Condition among opt-in agents is zero, adding opt-out workers to those eligible for treatment closes 33% and 38% of the gap in treatment effects between the Mandatory and Voluntary-Conditions for revenue and RPC, respectively.[19]

To summarize, we find evidence of both self-selection, where some workers are more likely to opt into the program than others, and heterogeneous treatment effects, where the mentorship program is most beneficial to those who tend to opt out. Taken together, these findings explain why we estimate negligible treatment effects of the voluntary mentorship program, but positive treatment effects of the mandatory program. As such, our findings show that the choice between making a workplace program voluntary versus mandatory is not trivial, as the returns to program implementation depend on the self-selection and heterogeneous treatment effects associated with the program.

4.4 The Mentorship Program Did Not Impact Worker Retention

Call centers have notoriously high levels of attrition (Hoffman et al., 2017), and retention is an important performance metric for the HR executives at the firm. To estimate retention effects from the mentorship program, we use data with a single observation per unique mentor-eligible agent among those who completed training,[20] and create an indicator variable $Tenure_{30}$ ($Tenure_{90}$) that equals one for agents who remain with the

[19] We note that this is a conservative estimate, as the propensity scores are measured with error, so we use estimates that average effects over the top third of the distribution of scores, covering more workers than the actual opt-out rate.

[20] The results are similar if we include individuals who did not complete training into the retention estimations.

firm for at least thirty (ninety) days after their hire date, and zero otherwise. We then re-estimate each of the models specified by Equations (1)–(3) with these two tenure achievement indicators as the dependent variables.

In Table 5, we find no evidence that mentorship impacted agents' retention, although agents who opt out of the program in the Voluntary-Condition are less likely to achieve ninety days of tenure than are non-mentored agents who opt in. There are no discernible retention effects among agents who were mentored, relative to those who were not mentored. Given that mentorship does not appear to have impacted agents' retention, it is unlikely that our estimated productivity treatment effects are driven by differences in retention between treatment conditions.

4.5 Addressing Alternative Explanations

Here we consider several alternative explanations could be supplied to explain the differences in treatment effects between the voluntary and mandatory programs. We discuss each of these explanations and our associated tests in detail in the Online Appendix (see. Section O.A). For brevity, we only report the conclusions of these tests here.

We begin in Section O.A.1, where we do not find any evidence that our results are driven by experimenter demand effects, Hawthorne effect, discouragement from treatment status, or information leakage. This suggests that violations of the Stable Unit Treatment Value Assumption (SUTVA) are not a major issue in our setting. In Section O.A.2, we show that the program does not appear to crowd out organic mentoring that may have occurred in its absence, as non-treated agents in experimental cohorts had similar productivity to agents entering the firm prior to the program's existence.

In Section 4.4, we showed that, across both conditions, the mentorship program had no impact on retention, which, therefore, cannot explain the observed sales revenue treatment effects. In Section O.A.3, we further show that productivity gains remain (i) when accounting for non-random attrition by filling in missing data after separations with the average productivity of replacements and (ii) when using Lee (2009) bounds estimators. These tests provide additional evidence that retention/attrition concerns are not likely driving the observed differences in treatment effects between the voluntary and mandatory programs.

Thus, the two most likely explanations for the remaining differences in treatment effect estimates are framing effects and sampling variation. We discuss these possibilities in Sections O.A.4–O.A.8. To summarize, we find small, positive treatment effects in the Mandatory-Condition for workers relatively far away from the opt-out margin, suggesting that these workers may have engaged differently with the program, consistent with framing effects. However, these results are imprecise, and we detect no measured engagement differences in worksheet contents across conditions. Ultimately, we find relatively little evidence in support of any of these alternative explanations.

5 Returns to the Program, Program Prevalence, and External Validity

In this section, we first conduct a net present value analysis and discuss the costs of misallocating mentors to agents with relatively small treatment gains. Then we discuss the results of a nationally representative survey that we conducted, which highlight the widespread prevalence and design variation of workplace programs. To finish, we comment on the external validity of our results.

5.1 Net Present Value of Mandatory Mentorship and the Costs of Misallocation

The net present value of the mandatory mentorship program to the firm is equal to approximately \$439,000. To arrive at this estimate, we calculate additional revenues of approximately \$536,000 over a 6-month period and costs of \$97,000, which include administrative and opportunity costs. These estimates use a mentorship slot as the unit of analysis, which accounts for the productivity of replacements when agents leave the firm. We provide details about the calculations in the Online Appendix (see Section O.F).

Had the firm allocated all workers to the Mandatory-Condition and had the treatment effects been the same across workers, the firm would have gained an additional \$602,000. If, instead, only about one-third of the treatment gains are due to selection and heterogeneous treatment effects (what our back-of-the-envelope calculation based on the opt-out propensity score recovers), the firm still would have gained approximately \$201,000 by reallocating treatment using the Mandatory-Condition protocol.

5.2 Prevalence of Workplace Programs

Beyond our study firm, we conducted a nationally representative worker survey to provide background context about workplace programs, with a focus on three questions: how prevalent are they, how is their participation determined (i.e., voluntary or mandatory), and which workers participate when programs are voluntary?

We administered the survey through the Lucid platform in June of 2022 and compensated respondents between \$1 and \$4. The survey took between 7–10 minutes to complete. Respondents had to be employed and pass attention checks to proceed through the survey. We asked respondents whether their current employer offers the following programs: (i) mentorship, (ii) training for new hires, and (iii) ongoing training or continuing education. We also asked whether the programs were required/mandatory or optional/voluntary and, if voluntary, whether they participated. We then probed for the reasons for their participation decisions. We display the results from this survey and details about the survey instrument in Table 6.[21]

The survey responses provided three main takeaways: (1) workplace programs are ubiquitous; (2) many are voluntary; and (3) many employees do not participate in voluntary programs. Specifically, 45% of the respondents said their employer offers a mentorship program, 87% said they offer new hire training, and 80% said they offer ongoing training or continuing education. About 59% of the mentorship programs and 43% of the continuing education programs offered are voluntary. New hire training is much more likely than the other programs to be mandatory. The last column in Table 6 shows substantial non-participation rates in voluntary programs. Roughly 27% of respondents did not participate in their employer's voluntary mentorship or ongoing training/continuing education program. Even for new hire training, rates of non-participation exceed 20% when training is optional. Time commitments and doubts about personal program benefits are the most common reasons for non-participation.[22] These survey results highlight the importance of studying the implications of the voluntary versus mandatory participation design choice that many mangers are faced with when they implement a new workplace program.

[21] We also included workplace wellness programs as a validation check. Sixty-five percent of our respondents indicated that their workplace has a wellness program. This is roughly comparable to numbers cited by Jones et al. (2019) from a 2016 Kaiser Family Foundation report, indicating that 53% of firms with more than 200 employees do biometric screening, 59% assess lifestyle health habits, and 83% have programs that encourage healthy lifestyles.

[22] Forty-seven percent of non-participants in mentorship, 36% in new hire training, and 42% in ongoing training cite time constraints or the inconvenience of program offerings as one of their reasons for not participating. "Didn't believe these programs would benefit me" (26% for mentorship, 28% for new hire training, and 31% for ongoing training) is the next most common reason. Other options such as, "Didn't plan to stay at the firm, so didn't invest," "Wanted to avoid interaction with coworkers or bosses," and "Felt the program would benefit my employer more than it would benefit me" were selected by 8%–13% of the respondents."

5.3 External validity

As part of the first wave of evidence on voluntary versus mandatory programs, we made multiple decisions to give us high internal validity (List, 2020). Several additional points, including performing our experiment in the field, suggest our results are likely to be externally valid for workers in other frontline or entry-level jobs. In particular, our representative worker survey found substantial rates of non-participation in workplace programs, implying that non-participation is a general phenomenon across firms. Furthermore, our participants are approximately representative of workers in similar occupations, e.g., average hourly earnings at the firm were about $21 per hour, while customer service representatives, telemarketers, and miscellaneous sales representatives earned about $23 per hour and $20 per hour at the national level and in the state in which the firm is located, respectively.[23] The task that agents performed in the mentorship program—reflecting on their work, sharing these thoughts with mentors, and acting on their mentors' advice—was a natural extension of their day-to-day activities. Finally, our intervention intentionally included features which would allow the treatment to be deployed at scale (permanently) both at the focal firm and in organizations more broadly.

6 Conclusion

Many firms dedicate substantial resources towards workplace training and development programs to augment workers' skills and human capital. But relative to the broad literature on investment (Stein, 2003), an important and understudied question is whether such human capital investments get allocated to the right workers. We consider a ubiquitous decision that managers face when deciding how to allocate human

[23]These figures come from the 2015–2019 5-year American Community Survey for SOC codes 43405, 41904, and 41309. To construct hourly earnings in the ACS data, we divide total individual income by the product of weeks worked last year and usual hours per week.

capital development resources: whether they should make human capital development programs mandatory or voluntary. We investigate the implications of this voluntary versus mandatory design choice by conducting a field experiment on mentorship in a U.S.-based inbound sales call center.

We find that the mandatory mentorship program significantly raised workers' productivity, with average sales gains on the order of about 19% over new hires' first two months on the job. By contrast, treatment gains were approximately zero when the program was voluntary. A substantial part of the difference between voluntary and mandatory programs arises because program treatment effects are negatively correlated with the propensity to participate in the program. Our design allows us to conclude that mentoring would most help workers who are the least likely to participate in the program. As such, our findings shed additional light on why wage inequality and performance differences may persist across workers and firms (Mueller, Ouimet, and Simintzi, 2017a,b). That these differences exist even in the presence of high-powered incentive pay suggests that managers may need to mandate worker participation in human capital development programs.

Given the growing interest among financial economists as to how human capital management impacts firm value (Agrawal, Hacamo, and Hu, 2021; Hacamo and Kleiner, 2022; Liu, Makridis, Ouimet, and Simintzi, 2023), our findings should motivate future researchers to determine other conditions under which human capital development resources are misallocated across workers. In our setting, training that leverages help from coworkers can lift lower-performing workers, but low-ability workers may be the *least* likely to seek out such help. Frictions around program participation deserve further investigation, as non-participation is likely a ubiquitous feature of training programs in both the public sector and in private firms. Beyond allocative efficiency, selection bias in program recruitment is likely to alter inferences about program efficacy, as demonstrated by negative sorting on gains in charter school enrollment (Walters, 2018) and the site selection bias identified by Allcott (2015). With remote work, these allocation questions may become more pronounced (Bojinov et al., 2021; Emanuel et al., 2023).

References

Acemoglu, Daron, Jörn-Steffen Pischke. 1999. The structure of wages and investment in general training. *Journal of Political Economy* **107**(3) 539–572.

Agrawal, Ashwini, Isaac Hacamo, Zhongchen Hu. 2021. Information dispersion across employees and stock returns. *The Review of Financial Studies* **34**(10) 4785–4831.

Allcott, Hunt. 2015. Site selection bias in program evaluation. *The Quarterly Journal of Economics* **130**(3) 1117–1165.

Allen, Tammy D, Lillian T Eby, Georgia T Chao, Talya N Bauer. 2017. Taking stock of two relational aspects of organizational life: Tracing the history and shaping the future of socialization and mentoring research. *Journal of Applied Psychology* **102**(3) 324.

Anderson, Michael L. 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association* **103**(484) 1481–1495.

Becker, Gary S. 1975. *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago press.

Benmelech, Efraim, Nittai Bergman, Hyunseob Kim. 2022. Strong employers and weak employees: How does employer concentration affect wages? *Journal of Human Resources* (S200-S250).

Benson, Alan, Danielle Li, Kelly Shue. 2019. Promotions and the peter principle. *The Quarterly Journal of Economics* **134**(4) 2085–2134.

Benson, Alan, Danielle Li, Kelly Shue. 2024. Potential and the gender promotions gap. *Available at SSRN* .

Bertrand, Marianne, Antoinette Schoar. 2003. Managing with style: The effect of managers on firm policies. *The Quarterly Journal of Economics* **118**(4) 1169–1208.

Bloom, Nicholas, Raffaella Sadun, John Van Reenen. 2016. Management as a technology? Tech. rep., National Bureau of Economic Research.

Bloom, Nicholas, John Van Reenen. 2007. Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics* **122**(4) 1351–1408.

Bojinov, Iavor, Prithwiraj Choudhury, Jacqueline N Lane. 2021. Virtual watercoolers: A field experiment on virtual synchronous interactions and performance of organizational newcomers. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (21-125).

Bourveau, Thomas, Maliha Chowdhury, Anthony Le, Ethan Rouen. 2022. Human capital disclosures. Tech. rep., Working Paper (Available at: SSRN link).

Campello, Murillo, Gaurav Kankanhalli, Hyunseob Kim. 2024. Delayed creative destruction: How uncertainty shapes corporate assets. *Journal of Financial Economics* **153** 103786.

Crouzet, Nicolas, Janice Eberly. 2023. Rents and intangible capital: A q+ framework. *The Journal of Finance* **78**(4) 1873–1916.

Crouzet, Nicolas, Janice C Eberly, Andrea L Eisfeldt, Dimitris Papanikolaou. 2022. The economics of intangible capital. *Journal of Economic Perspectives* **36**(3) 29–52.

Custódio, Cláudia, Miguel A Ferreira, Pedro Matos. 2019. Do general managerial skills spur innovation? *Management Science* **65**(2) 459–476.

Custodio, Claudia, Christopher Hansman, Diogo Mendes. 2022. Information frictions and firm take up of government support: A randomised controlled experiment. *Swedish House of Finance Research Paper* (21-15).

Custodio, Claudia, Diogo Mendes, Daniel Metzger. 2020. The impact of financial education of managers on medium and large enterprises-a randomized controlled trial in mozambique .

Dahl, Gordon B, Katrine V Løken, Magne Mogstad. 2014. Peer effects in program participation. *American Economic Review* **104**(7) 2049–2074.

Deming, David J, Justine S Hastings, Thomas J Kane, Douglas O Staiger. 2014. School choice, school quality, and postsecondary attainment. *American Economic Review* **104**(3) 991–1013.

Edmans, Alex. 2011. Does the stock market fully value intangibles? employee satisfaction and equity prices. *Journal of Financial Economics* **101**(3) 621–640.

Eisfeldt, Andrea L, Dimitris Papanikolaou. 2014. The value and ownership of intangible capital. *American Economic Review* **104**(5) 189–194.

Emanuel, Natalia, Emma Harrington, Amanda Pallais. 2023. The power of proximity .

Englmaier, Florian, Andreas Roider, Uwe Sunde. 2017. The role of communication of performance schemes: Evidence from a field experiment. *Management Science* **63**(12) 4061–4080.

Espinosa, Miguel, Christopher Stanton. 2021. Worker skills and organizational spillovers: Evidence from linked training and communications data. Tech. rep., Harvard Business School.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, Nick Zubanov. 2023. What do employee referral programs do? measuring the direct and overall effects of a management practice. *Journal of Political Economy* **131**(3) 633–686.

Fudenberg, Drew, Luis Rayo. 2019. Training and effort dynamics in apprenticeship. *American Economic Review* **109**(11) 3780–3812.

Ginther, Donna K, Janet M Currie, Francine D Blau, Rachel TA Croson. 2020. Can mentoring help female assistant professors in economics? an evaluation by randomized trial. *AEA Papers and Proceedings*, vol. 110. 205–09.

Gutner, Toddi. 2009. Finding anchors in the storm: Mentors. *The Wall Street Journal* .

Hacamo, Isaac, Kristoph Kleiner. 2022. Competing for talent: Firms, managers, and social networks. *The Review of Financial Studies* **35**(1) 207–253.

Harrison, Glenn W, John A List. 2004. Field experiments. *Journal of Economic Literature* **42**(4) 1009–1055.

Hoffman, Mitchell, Lisa B. Kahn, Danielle Li. 2017. Discretion in hiring. *The Quarterly Journal of Economics* **133**(2) 765–800.

Hong, Fuhai, Tanjim Hossain, John A List. 2015. Framing manipulations in contests: a natural field experiment. *Journal of Economic Behavior & Organization* **118** 372–382.

Hossain, Tanjim, John A List. 2012. The behavioralist visits the factory: Increasing productivity using simple framing manipulations. *Management Science* **58**(12) 2151–2167.

Hoxby, Caroline M, Jonah E Rockoff. 2005. The impact of charter schools on student achievement.

Investors, Institutional. 2018. *Dear Board of Directors*. https://corpgov.law.harvard.edu/2019/02/02/pay-ratio-disclosure-at-the-sp-500. Accessed: June 1st, 2020.

Iyer, Rajkamal, Antoinette Schoar, et al. 2008. The importance of holdup in contracting: Evidence from a field experiment. Tech. rep., Working Paper.

Johnson, Matthew S, David I Levine, Michael W Toffel. 2022. Improving regulatory effectiveness through better targeting: Evidence from osha. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (20-019).

Jones, Damon, David Molitor, Julian Reif. 2019. What do workplace wellness programs do? evidence from the illinois workplace wellness study. *The Quarterly Journal of Economics* **134**(4) 1747–1791.

Kim, Hyuncheol Bryant, Hyunseob Kim, John Y Zhu. 2022. The selection effects of part-time work: Experimental evidence from a large-scale recruitment drive .

Kogan, Leonid, Dimitris Papanikolaou, Amit Seru, Noah Stoffman. 2017. Technological innovation, resource allocation, and growth. *The Quarterly Journal of Economics* **132**(2) 665–712.

Larrain, Mauricio. 2015. Capital account opening and wage inequality. *The Review of Financial Studies* **28**(6) 1555–1587.

Lazear, Edward P. 2000. Performance pay and productivity. *American Economic Review* **90**(5) 1346–1361.

Lazear, Edward P, Kathryn L Shaw, Christopher Stanton. 2016. Making do with less: working harder during recessions. *Journal of Labor Economics* **34**(S1) S333–S360.

Lazear, Edward P, Kathryn L Shaw, Christopher T Stanton. 2015. The value of bosses. *Journal of Labor Economics* **33**(4) 823–861.

Lee, David S. 2009. Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies* **76**(3) 1071–1102.

Li, Danielle, Lindsey R Raymond, Peter Bergman. 2020. Hiring as exploration. Tech. rep., National Bureau of Economic Research.

List, John A. 2020. Non est disputandum de generalizability? a glimpse into the external validity trial. Tech. rep., National Bureau of Economic Research.

List, John A. 2022. *The voltage effect: How to make good ideas great and great ideas scale*. Currency.

Liu, Tim, Christos A Makridis, Paige Ouimet, Elena Simintzi. 2023. The distribution of nonwage benefits: maternity benefits and gender diversity. *The Review of Financial Studies* **36**(1) 194–234.

Lyle, David S, John Z Smith. 2014. The effect of high-performing mentors on junior officer promotion in the us army. *Journal of Labor Economics* **32**(2) 229–258.

Mueller, Holger M, Paige P Ouimet, Elena Simintzi. 2017a. Wage inequality and firm growth. *American Economic Review* **107**(5) 379–383.

Mueller, Holger M, Paige P Ouimet, Elena Simintzi. 2017b. Within-firm pay inequality. *The Review of Financial Studies* **30**(10) 3605–3635.

Nishesh, Naman, Paige Ouimet, Elena Simintzi. 2022. Labor and corporate finance. *Available at SSRN* .

Pan, Yihui, Elena S Pikulina, Stephan Siegel, Tracy Yue Wang. 2022. Do equity markets care about income inequality? evidence from pay ratio disclosure. *The Journal of Finance* **77**(2) 1371–1411.

Panousi, Vasia, Dimitris Papanikolaou. 2012. Investment, idiosyncratic risk, and ownership. *The Journal of Finance* **67**(3) 1113–1148.

Porter, Catherine, Danila Serra. 2020. Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics* **12**(3) 226–54.

Sandvik, Jason J, Richard E Saouma, Nathan T Seegert, Christopher T Stanton. 2020. Workplace knowledge flows. *The Quarterly Journal of Economics* **135**(3) 1635–1680.

SEC. 2020. Modernization of regulation s-k items 101, 103, and 105. *United States Securities and Exchange Commission* https://www.sec.gov/rules/final/2020/33-10825.pdf. Accessed: February 21st, 2024.

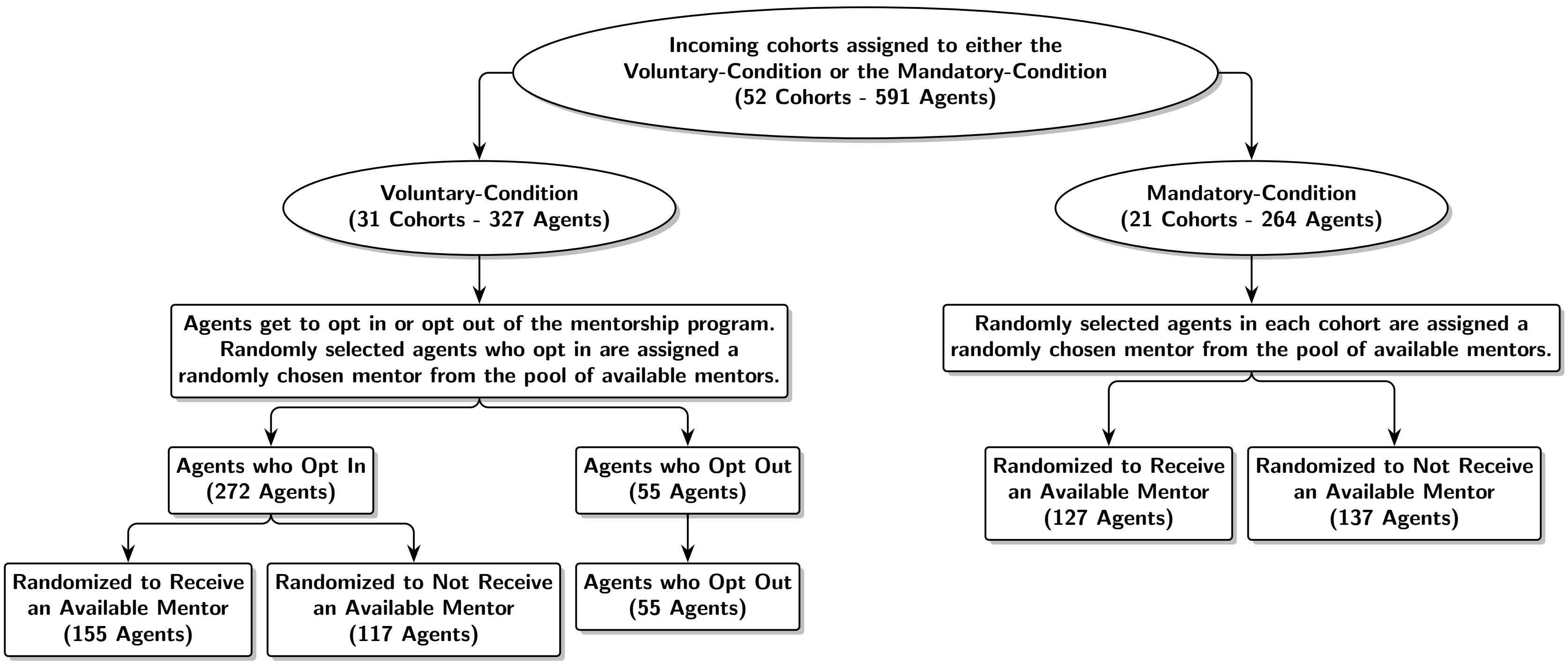
Statista. 2022. Total training expenditures in the united states from 2012 to 2022 (in billion u.s. dollars) [graph]. https://www-statista-com.ezp-prod1.hul.harvard.edu/statistics/788521/training-expenditures-united-states/. Retrieved on 2023-01-29.

Stein, Jeremy C. 2003. Agency, information and corporate investment. *Handbook of the Economics of Finance* **1** 111–165.

Walters, Christopher R. 2018. The demand for effective charter schools. *Journal of Political Economy* **126**(6) 2179–2223.

Zingales, Luigi. 2000. In search of new foundations. *The Journal of Finance* **55**(4) 1623–1653.

Figure 1: Allocation of Cohorts and Agents to Treatment Conditions



Notes. This figure displays the allocation of the 52 mentor-eligible cohorts to either the Voluntary-Condition or the Mandatory-Condition, our first level of variation. It then shows the allocation of the 591 mentor-eligible agents within these cohorts into different treatment conditions, our second level of variation. This is based on agents who complete training and are observed to have post-training productivity data.

Table 1: Balance Tests for Treatment Assignment

Panel A: Cohort-Level Balance in Agent Characteristics

	Voluntary-Condition	Mandatory-Condition	p-value
	(1)	(2)	(2)−(1)
Age (yrs.)			
Mean	22.80	22.70	0.887
Std Dev.	(2.34)	(2.40)	
Female			
Mean	0.40	0.43	0.624
Married			
Mean	0.16	0.13	0.522
Hiring Score			
Mean	0.85	0.83	0.207
Std Dev.	(0.04)	(0.04)	
Adjusted Hiring Score			
Mean	0.86	0.84	0.029
Std Dev.	(0.03)	(0.03)	
Referral			
Mean	0.58	0.57	0.746
N Cohorts	31	21	

Panel B: Balance in Agent Characteristics For Those Eligible for Mentor Assignment

	Voluntary-Condition			Mandatory-Condition		
	Mentored	Non-Mentored	p-value	Mentored	Non-Mentored	p-value
	(1)	(2)	(2)−(1)	(3)	(4)	(4)−(3)
Age (yrs.)						
Mean	22.47	22.51	0.945	22.40	23.51	0.193
Std Dev.	(5.54)	(6.18)		(4.46)	(8.60)	
Female						
Mean	0.45	0.38	0.318	0.46	0.40	0.303
Married						
Mean	0.15	0.17	0.722	0.09	0.15	0.150
Hiring Score						
Mean	0.85	0.86	0.508	0.83	0.84	0.432
Std Dev.	(0.08)	(0.08)		(0.09)	(0.08)	
Adj. Hiring Score						
Mean	0.86	0.86	0.522	0.83	0.84	0.322
Std Dev.	(0.07)	(0.07)		(0.08)	(0.07)	
Referral						
Mean	0.56	0.60	0.543	0.58	0.55	0.649
Number of Agents	155	117		127	137	

Notes. This table presents balance tests. Most characteristics are self-explanatory other than the Hiring Score, which is a recruiter-assigned measure of fit with the job, ranging from 0 to 1. The Adjusted (Adj.) Hiring Score accounts for individual recruiter leniency, estimated using productivity of non-mentor-eligible agents outside of the experiment, as described in Section 3.5. In Panel A, we report average agent characteristics at the cohort-level to test for assignment balance between the Voluntary- and Mandatory-Conditions. In Panel B, we test for balance in sub-treatment assignment to mentors. In Panel B, the Voluntary-Condition sample is not comparable to the Mandatory-Condition sample due to selection into the program (Table 3 compares the characteristics of those who opt in and opt out in the Voluntary-Condition). Standard deviations are in parentheses for continuous variables. The p-values come from difference-in-means tests across high-level treatment conditions in Panel A and for agents who do and do not receive mentors among those eligible for assignment in Panel B.

Table 2: Treatment and Selection Effects of Mentoring on Productivity

	Voluntary-Condition (Opt-In Agents)		Mandatory-Condition (All Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	-0.084	-0.084	0.171**	0.112**			0.171**	0.112***	0.145**
standard errors	(0.074)	(0.050)	(0.069)	(0.042)			(0.068)	(0.041)	(0.056)
sharpened q-value							[0.041]	[0.041]	[0.041]
Voluntary Opt-Out					-0.369***	-0.264***	-0.277**	-0.161**	-0.141*
standard errors					(0.120)	(0.070)	(0.110)	(0.071)	(0.070)
sharpened q-value							[0.041]	[0.041]	[0.066]
Mentored × Voluntary							-0.272***	-0.207***	-0.203***
standard errors							(0.099)	(0.065)	(0.070)
sharpened q-value							[0.041]	[0.041]	[0.041]
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.020	0.046	0.028	0.032	0.047	0.066	0.026	0.040	0.059
Observations	7,569	7,569	6,725	6,725	4,734	4,734	15,670	15,670	15,670
p-value: Mentored + Mentored × Voluntary							0.162	0.064	0.179

Notes. This table reports estimates of different treatment effects from the mentorship program. The sample is composed of agent-day productivity data across agents' first two months on the job after they complete training. IHS(.) indicates a variable that is transformed by the inverse hyperbolic sine. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (3) and (4) include all agents in the Mandatory-Condition. The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include all agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition (see the text for additional details). We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort and are reported in parentheses. Sharpened q-values that adjust for the false discovery rate are presented in brackets, following Anderson (2008). The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 3: Determinants of Program Opt-Out and the Relationship Between Opting Out, Productivity, and Worker Characteristics

Dep. Variable	= 1 if Opted Out				IHS(Revenue)		IHS(RPC)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	0.017	0.011	0.014	-0.011	-0.002	-0.004	-0.001
	(0.023)	(0.027)	(0.029)	(0.036)	(0.006)	(0.005)	(0.005)
Female	-0.453*	-0.098	-0.065	-0.169	-0.104	-0.093	-0.084
	(0.264)	(0.355)	(0.371)	(0.340)	(0.113)	(0.102)	(0.073)
Married	-0.084	-0.170	-0.155	0.040	0.067	0.119	0.105
	(0.343)	(0.459)	(0.439)	(0.400)	(0.169)	(0.191)	(0.131)
Adjusted Hiring Score	-4.903**	-6.059**	-6.172**	-5.959***	2.703***	2.931***	2.088***
	(2.104)	(2.488)	(2.406)	(2.233)	(0.946)	(0.898)	(0.564)
Location 1	-0.362	0.416	0.344	0.234			
	(0.351)	(0.407)	(0.417)	(0.428)			
Referral	-0.166	0.171	0.148	0.095	-0.188*	-0.183	-0.079
	(0.343)	(0.457)	(0.489)	(0.433)	(0.104)	(0.110)	(0.066)
Call Center Exp.		0.844	0.861*	0.886*		0.397**	0.313***
		(0.529)	(0.471)	(0.455)		(0.187)	(0.112)
Sales Experience		0.055	0.052	0.048		-0.017	-0.050
		(0.541)	(0.518)	(0.519)		(0.222)	(0.123)
High Extroversion			-0.227	-0.225		0.205	0.163
			(0.400)	(0.388)		(0.146)	(0.106)
High Agreeableness			-0.370	-0.363		-0.206**	-0.109
			(0.342)	(0.331)		(0.088)	(0.067)
High Conscientiousness			-0.540	-0.501		-0.035	-0.051
			(0.462)	(0.459)		(0.123)	(0.081)
High Emotional Stability			0.388	0.369		-0.146	-0.056
			(0.436)	(0.431)		(0.100)	(0.052)
High Openness			0.224	0.240		0.039	0.062
			(0.393)	(0.402)		(0.128)	(0.078)
Missing Survey				2.583***		-0.457***	-0.298**
				(0.571)		(0.161)	(0.118)
Division Fixed Effects		✓	✓	✓			
Cohort Fixed Effects					✓	✓	✓
(Pse.) R-Square	0.036	0.070	0.085	0.209	0.061	0.073	0.094
Observations	365	341	341	365	4,734	4,734	4,734

Notes. The sample in Columns (1)–(4) is restricted to the 365 agents in the Voluntary-Condition, including those who quit before they completed training. The dependent variable is an indicator that equals one if the agent opted out of the program. Coefficients come from logistic regressions of different potential predictors of the choice to opt out. Experience and personality factors were collected via survey. We split personality scores on the sample median. Column (4) includes agents who did not complete the new hire survey, which we account for with a *Missing Survey* indicator. In Columns (5)–(6), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(Revenue), the inverse hyperbolic sine of daily revenue, on agents' characteristics. In Column (7), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(RPC), the inverse hyperbolic sine of daily revenue-per-call, on agents' characteristics. Standard errors are clustered by cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 4: Treatment Effects of Mentoring in the Mandatory-Condition by Predicted Opt-Out Propensity

	IHS(Revenue)			IHS(RPC)		
	High_{Opt}	Low_{Opt}	All	High_{Opt}	Low_{Opt}	All
	(1)	(2)	(3)	(4)	(5)	(6)
Mentored	0.324**	0.069	0.063	0.262**	0.026	0.022
	(0.119)	(0.089)	(0.091)	(0.099)	(0.047)	(0.049)
Mentored × High_{Opt}			0.342**			0.285***
			(0.131)			(0.091)
High_{Opt}			-0.239**			-0.163**
			(0.107)			(0.063)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.038	0.036	0.030	0.044	0.035	0.035
Observations	2,244	4,481	6,725	2,244	4,481	6,725

Notes. This table reports heterogeneous treatment effect estimates for agents in the Mandatory-Condition. We estimate agents' opt-out propensity scores as described in Section 4.2. After estimating propensity scores, we place agents into $High_{Opt}$ if their propensity score of opting out is in the top 33.3% of the propensity score distribution, and we place agents with a propensity score in the bottom 66.7% into Low_{Opt}, indicating that they had a low likelihood to opt out. We use a larger threshold than the opt-out rate in the Voluntary-Condition because (i) the individual propensity scores are less than 1, implying we need more workers to approximate the total number of those who opt out in the Voluntary-Condition and (ii) we want a sample that is large enough for reliable inference. We then estimate Equation (1) within these subsets of the data with either IHS(Revenue) or IHS(RPC) as the dependent variable. To determine if the effect of mentorship is significantly different between the $High_{Opt}$ and Low_{Opt} agents, we pool the samples in Columns (3) and (6) and include a one-zero indicator for $High_{Opt}$ along with its interaction with *Mentored*. Standard errors are clustered by cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table 5: Treatment Effects on Retention

	Voluntary-Condition (Opt-In Agents)		Mandatory-Condition (All Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)	
	Tenure$_{30}$	Tenure$_{90}$	Tenure$_{30}$	Tenure$_{90}$	Tenure$_{30}$	Tenure$_{90}$	Tenure$_{30}$	Tenure$_{90}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mentored	-0.001	-0.057	-0.001	-0.061			-0.001	-0.061
	(0.045)	(0.064)	(0.041)	(0.075)			(0.040)	(0.074)
Voluntary Opt-Out					-0.117	-0.212*	-0.054	-0.242**
					(0.076)	(0.122)	(0.052)	(0.098)
Mentored × Voluntary							-0.018	0.003
							(0.058)	(0.098)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.003	0.004	0.041	-0.037	0.036	0.072	0.034	-0.003
Observations	272	272	264	264	172	172	591	591
p-value: Mentored + Mentored × Voluntary							0.650	0.368

Notes. The sample used is composed of a single observation per agent, among all mentor-eligible agents with post-training productivity data. $Tenure_{30}$ ($Tenure_{90}$) equals one for agents who remain with the firm for at least thirty (ninety) days after their hire date, and zero otherwise. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (3) and (4) include all agents in the Mandatory-Condition. The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7) and (8) include agents from both conditions. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort and are reported in parentheses. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

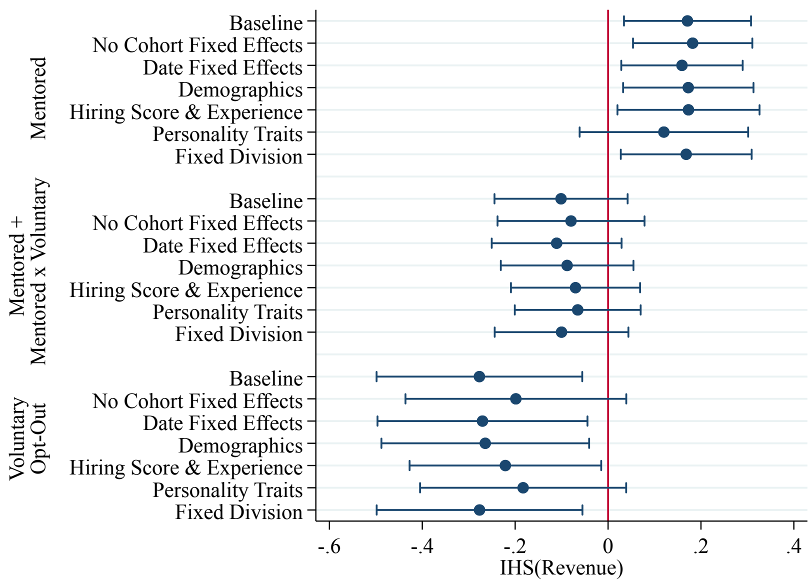
Table 6: Survey Data on Characteristics of Workplace Programs and Participation in Voluntary Programs

Program Type:	Is the Program Offered?	If It Is Offered, Is It Voluntary?	If It Is Voluntary, Do You Not Participate?
Formal Mentorship	0.45	0.59	0.27
	(0.01)	(0.01)	(0.02)
New Hire Training	0.87	0.22	0.21
	(0.01)	(0.01)	(0.02)
Ongoing Training or Cont. Ed.	0.80	0.43	0.28
	(0.01)	(0.01)	(0.01)
N = 3,191			

Notes. This table displays summary statistics on the prevalence and administrative choices for different workplace programs. Means and standard deviations (in parentheses) are reported. Data come from a nationally representative online survey conducted through the Lucid platform in June of 2022. The survey was restricted to respondents currently employed by others. Respondents were asked about whether their employer offers a particular program and whether it is voluntary or mandatory with the question: "Consider your current employer. Which of the following programs does your employer offer to you personally? If offered, are you required to participate (required/mandatory) or can you choose to participate or not (optional/voluntary)?" For each program, respondents chose between "Required or Mandatory," "Optional or Voluntary," or "Not offered." For the three core programs—mentorship, new hire training, and continuing education—if a respondent indicated that a program was voluntary, follow-up questions were asked about their participation and the reasons for their lack of participation, if applicable. As reported in the text, the survey also asked about workplace wellness programs to benchmark responses against other sources.

A Appendix Figures and Tables

Figure A.1: Robustness of the Treatment and Selection Effects of Mentoring on Productivity



Notes. This figure plots the regression coefficients (and 95% confidence intervals) on *Mentored*, the sum of *Mentored* and *Mentored* × *Voluntary*, and *Voluntary Opt-Out* from Equation (3). We use IHS(Revenue) as the dependent variable. The "Baseline" estimation replicates the result from Column (7) of Table 2. The second estimation excludes cohort fixed effects. The third includes date fixed effects. The fourth estimation includes controls for the agent's demographic characteristics: age, gender, and marital status. The fifth estimation includes additional controls for the agent's referral status, hiring score, previous call center experience, and previous sales experience. The sixth estimation layers on five more controls for the agent's personality traits: extroversion, agreeableness, conscientiousness, emotional stability, and openness. Note that these controls are frequently missing for non-mentored agents in the Mandatory-Condition, suggesting these are not good controls for estimating the average treatment effect because the data come from a highly selected sample of agents in the Mandatory-Condition who took the intake survey. The seventh estimation removes observations in which agents are no longer working in the division in which they were initially hired.

Table A.1: Treatment and Selection Effects of Mentoring (Pilot vs. Post-Pilot)

	Voluntary-Condition (Opt-In Agents)		Mandatory-Condition (All Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	-0.154	-0.152	0.180**	0.110**			0.180**	0.110**	0.174***
	(0.158)	(0.117)	(0.082)	(0.046)			(0.081)	(0.045)	(0.063)
Mentored × Post	0.099	0.096	-0.033	0.005			-0.033	0.005	-0.105
	(0.177)	(0.127)	(0.153)	(0.104)			(0.150)	(0.102)	(0.129)
Voluntary Opt-Out					-0.331**	-0.325***	-0.223	-0.177	-0.188
					(0.160)	(0.055)	(0.158)	(0.113)	(0.150)
Voluntary Opt-Out × Post					-0.063	0.102	-0.225	-0.043	0.028
					(0.236)	(0.116)	(0.260)	(0.169)	(0.164)
Mentored × Voluntary							-0.374**	-0.293**	-0.325**
							(0.170)	(0.122)	(0.138)
Mentored × Voluntary × Post							0.186	0.130	0.245
							(0.228)	(0.161)	(0.180)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.020	0.046	0.028	0.032	0.047	0.066	0.026	0.040	0.059
Observations	7,569	7,569	6,725	6,725	4,734	4,734	15,670	15,670	15,670

Notes. This table is structured similarly to Table 2, while allowing us to test whether estimates differ between the pilot and post-pilot data as described in Section 3.4. The sample is composed of agent-day productivity data for all mentor-eligible agents with post-training productivity data. The data covers agents' productivity across their first two months on the job after they complete training. IHS(.) indicates a variable that is transformed by the inverse hyperbolic sine. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. *Post* equals one for cohorts that entered the firm on or after May 27th (the post-pilot cohorts), and zero otherwise. The specifications in Columns (1) and (2) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (3) and (4) include all agents in the Mandatory-Condition. The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.2: Balance in Mentor Demographics

	Voluntary-Condition	Mandatory-Condition	p-value
	(1)	(2)	(2)−(1)
Mentor Age (yrs.)			
Mean	23.28	22.92	0.732
Std Dev.	(3.20)	(2.57)	
Mentor Female			
Mean	0.17	0.07	0.391
Mentor Married			
Mean	0.22	0.14	0.587
Mentor Tenure			
Mean	1.12	1.40	0.398
Std Dev.	(1.14)	(0.52)	
Number of Protégés	155	127	

Notes. In this table we report average characteristics of the agents who mentored protégés in the Voluntary-Condition in Column (1) and of the agents who mentored protégés in the Mandatory-Condition in Column (2). Mentors were not designated exclusively to either of the mentoring conditions. In other words, a mentor's first protégé could have been assigned to the Voluntary-Condition, whereas their second protégé could have been assigned to the Mandatory-Condition. Mentors were never informed as to whether their protégés were in the Voluntary-Condition or the Mandatory-Condition. We report standard deviations in parentheses, and we report p-values from difference in means tests to compare values across the different treatment conditions.

Table A.3: Balance in Division Performance

	Voluntary-Condition	Mandatory-Condition	p-value
	(1)	(2)	(2)−(1)
Revenue			
Mean	846.61	803.35	0.267
Std Dev.	(151.9)	(108.52)	
RPC			
Mean	51.61	48.50	0.305
Std Dev.	(11.39)	(9.31)	
RPH			
Mean	132.9	128.45	0.432
Std Dev.	(21.87)	(16.46)	
Calls			
Mean	17.41	17.44	0.928
Std Dev.	(1.42)	(1.06)	
Hours			
Mean	6.34	6.25	0.638
Std Dev.	(0.66)	(0.68)	
Adherence			
Mean	0.88	0.87	0.272
Std Dev.	(0.03)	(0.02)	
Conversion			
Mean	0.23	0.24	0.285
Std Dev.	(0.02)	(0.03)	
Number of Cohorts	31	21	

Notes. In this table, we take average productivity measures of agents who were not mentorship-eligible within each sales division. Cohorts are assigned to a particular sales division, so these tests estimate the balance in division-level productivity measures between cohorts in the Voluntary-Condition versus those in the Mandatory-Condition. We report standard deviations in parentheses, and we report p-values from difference in means tests to compare values across the different treatment conditions.

Table A.4: Treatment and Selection Effects of Mentoring on Productivity (Log-Transformed Dependent Variables)

	Voluntary-Condition (Opt-In Agents)		Mandatory-Condition (All Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	Log(Rev)	Log(RPC)	Log(Rev)	Log(RPC)	Log(Rev)	Log(RPC)	Log(Rev)	Log(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	-0.080	-0.079*	0.161**	0.101**			0.161**	0.101***	0.146**
	(0.071)	(0.046)	(0.066)	(0.038)			(0.065)	(0.037)	(0.056)
Voluntary Opt-Out					-0.346***	-0.241***	-0.261**	-0.145**	-0.143*
					(0.111)	(0.062)	(0.103)	(0.064)	(0.071)
Mentored × Voluntary							-0.257***	-0.189***	-0.205***
							(0.094)	(0.059)	(0.071)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.021	0.054	0.029	0.036	0.049	0.072	0.027	0.045	0.059
Observations	7,569	7,569	6,725	6,725	4,734	4,734	15,670	15,670	15,670
p-value: Mentored + Mentored × Voluntary							0.165	0.060	0.172

Notes. The sample used is composed of agent-day productivity data for all mentor-eligible agents with post-training productivity data. The data covers agents' productivity on their first two months on the job after they complete training. Log(.) indicates a variable that is transformed by the logarithm of one plus the value. Revenue ("Rev") is weekly total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to received an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (3) and (4) include all agents in the Mandatory-Condition. The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of Log(Revenue), Log(RPC), Log(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort and are reported in parentheses. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.5: Long-Term Treatment and Selection Effects of Mentoring on Productivity

	Voluntary-Condition (Opt-In Agents)		Mandatory-Condition (All Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	-0.016	-0.053	0.079	0.051			0.079	0.051	0.055
standard errors	(0.105)	(0.068)	(0.135)	(0.083)			(0.133)	(0.082)	(0.064)
sharpened q-value							[0.419]	[0.419]	[0.419]
Voluntary Opt-Out					0.127	0.075	0.037	0.051	-0.057
standard errors					(0.167)	(0.094)	(0.121)	(0.069)	(0.074)
sharpened q-value							[0.614]	[0.419]	[0.419]
Mentored × Voluntary							-0.099	-0.108	-0.069
standard errors							(0.167)	(0.105)	(0.084)
sharpened q-value							[0.419]	[0.389]	[0.419]
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.046	0.035	0.040	0.057	0.042	0.053	0.043	0.051	0.049
Observations	6,608	6,608	5,815	5,815	3,874	3,874	13,238	13,238	13,238
p-value: Mentored + Mentored × Voluntary							0.843	0.390	0.789

Notes. The sample used is composed of agent-day productivity data for all mentor-eligible agents with post-training productivity data. The data covers agents' productivity across their third to sixth months on the job after they complete training. IHS(.) indicates a variable that is transformed by the inverse hyperbolic sine. Revenue ("Rev") is daily total revenue and RPC is revenue per call. *Mentored* equals one for agents who were randomized to receive an available mentor, and zero otherwise, *Voluntary* equals one for agents in the Voluntary-Condition, and zero otherwise, and *Voluntary Opt-Out* equals one for agents who chose to opt out of possibly receiving a mentor, and zero otherwise. The specifications in Columns (1) and (2) include agents in the Voluntary-Condition who signaled their interest in receiving a mentor (i.e., those who opted in). The specifications in Columns (3) and (4) include all agents in the Mandatory-Condition. The specifications in Columns (5) and (6) include agents in the Voluntary-Condition who were not assigned a mentor, including those who opted out of the program. The specifications in Columns (7)–(9) include agents from both conditions. The dependent variable in Column (9), *Index*, is the standardized weighted index of IHS(Revenue), IHS(RPC), IHS(RPH), and Adherence normalized using data from the non-mentored agents in the Mandatory-Condition. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort and are reported in parentheses. Sharpened q-values that adjust for the false discovery rate are presented in brackets, following Anderson (2008). The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A.6: Determinants of Program Opt-Out and the Relationship Between Opting Out, Productivity, and Worker Characteristics (using Raw Hiring Scores)

Dep. Variable	= 1 if Opted Out				IHS(Revenue)		IHS(RPC)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Age	0.016	0.009	0.012	-0.012	-0.003	-0.005	-0.002
	(0.023)	(0.027)	(0.029)	(0.036)	(0.006)	(0.005)	(0.005)
Female	-0.431	-0.073	-0.038	-0.145	-0.113	-0.101	-0.092
	(0.265)	(0.350)	(0.366)	(0.333)	(0.114)	(0.102)	(0.072)
Married	-0.100	-0.195	-0.185	0.005	0.077	0.125	0.109
	(0.344)	(0.463)	(0.446)	(0.406)	(0.169)	(0.192)	(0.131)
Hiring Score	-4.415**	-4.825**	-4.848**	-4.784**	2.571**	2.766***	2.085***
	(1.854)	(2.187)	(2.149)	(2.048)	(1.023)	(0.862)	(0.506)
Location 1	-0.417	0.363	0.298	0.185			
	(0.363)	(0.410)	(0.416)	(0.430)			
Referral	-0.164	0.167	0.138	0.087	-0.199*	-0.189*	-0.084
	(0.341)	(0.451)	(0.481)	(0.429)	(0.106)	(0.111)	(0.067)
Call Center Exp.		0.838	0.845*	0.874*		0.400**	0.318***
		(0.540)	(0.479)	(0.464)		(0.187)	(0.113)
Sales Experience		0.002	0.003	0.005		0.005	-0.038
		(0.528)	(0.504)	(0.505)		(0.216)	(0.119)
High Extroversion			-0.279	-0.273		0.206	0.161
			(0.403)	(0.394)		(0.143)	(0.105)
High Agreeableness			-0.371	-0.364		-0.196**	-0.103
			(0.346)	(0.335)		(0.086)	(0.065)
High Conscientiousness			-0.514	-0.479		-0.045	-0.057
			(0.472)	(0.471)		(0.124)	(0.081)
High Emotional Stability			0.386	0.368		-0.151	-0.063
			(0.427)	(0.422)		(0.104)	(0.055)
High Openness			0.210	0.227		0.066	0.082
			(0.400)	(0.409)		(0.119)	(0.073)
Missing Survey				2.540***		-0.439**	-0.288**
				(0.558)		(0.163)	(0.119)
Division Fixed Effects		✓	✓	✓			
Cohort Fixed Effects					✓	✓	✓
(Pse.) R-Square	0.035	0.063	0.078	0.204	0.061	0.074	0.096
Observations	365	341	341	365	4,734	4,734	4,734

Notes. The sample in Columns (1)–(4) is restricted to the 365 agents in the Voluntary-Condition, including those who quit before they completed training. The dependent variable is an indicator that equals one if the agent opted out. We run logistic regressions on different potential predictors of the choice to opt out. Experience and personality factors were collected via survey. We split personality scores on the sample median. Column (4) includes agents who did not complete the new hire survey, which we account for with a *Missing Survey* indicator. In Columns (5)–(6), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(Revenue), the inverse hyperbolic sine of daily revenue, on agents' characteristics. In Column (7), we use the sample of agents in the Voluntary-Condition who were not mentored, and we regress IHS(RPC), the inverse hyperbolic sine of daily revenue-per-call, on agents' characteristics. Standard errors are clustered by cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Online Appendix to "Can Firms Use Self-Selection to Improve the Efficacy of Human Capital Investments? Evidence from a Field Experiment"

Contents

O.A Alternative Explanations for Treatment Effects

In this section, we evaluate threats to the validity of our treatment effect estimates, and we explore a range of possible alternative explanations for differences in treatment effects between the Voluntary- and Mandatory-Conditions. In discussing these possibilities, we distinguish between effects that come from *program design* decisions, like communication that a training or mentoring program is mandatory, and *experimental design* decisions, like randomization into treatment as a measurement tool.[1]

O.A.1 Experimenter Demand Effects, Hawthorne Effects, Discouragement, or Information Leakage

Our design is a natural field experiment (Harrison and List, 2004), where the participants never met the researchers, limiting experimenter demand effects (i.e., observer bias). To participants, the mentoring program appeared like a normal work activity. While participants were told that outside researchers were analyzing their survey and productivity data, the mentorship program was framed and experienced as a regular part of the firm's onboarding process. Furthermore, participants were not aware that differences between the Voluntary- and Mandatory-Conditions were the objects of researcher interest.[2] Similarly, Hawthorne effects were not likely in this setting, as sales managers monitor the same performance metrics that we study and provide workers with performance-related feedback. It is unlikely that subject behavior was impacted by the knowledge that outside researchers—with whom the agents never interacted—were tracking their performance. In our prior work with this firm, benign treatments with no productivity impact allowed us to test for Hawthorne and demand effects; we found no evidence of their importance in this setting (Sandvik et al., 2020).

Discouragement and the possible leakage of information due to the design of the program are always important mechanisms to consider in natural field experiments. In our setting, agents who did not receive a mentor may have become discouraged, reducing their performance. Discouragement effects could potentially result in a difference in productivity between mentored and non-mentored agents; we would expect agents who received a mentor to outperform those that did not. In the Voluntary-Condition, where we would expect discouragement effects to be most salient, we find no evidence that agents who opted in and received a mentor outperformed those who did not. In our implementation, we preemptively worked with the company to limit discouragement and information leakage. For example, the protocol called for the staff to privately notify treated workers of their mentor assignment—reducing the salience of unequal treatment and the potential for discouragement among non-mentored workers. We

[1]These dimensions are not always mutually exclusive, as resource or slot-constrained programs may have lotteries to determine allocations (Hoxby and Rockoff, 2005).

[2]Subjects were asked to provide informed consent when responding to the new hire survey. The survey was framed around understanding employees' preferences, work styles, and personality characteristics so that university researchers could help the firm better serve its workforce. The consent protocol did not specify that selection into or out of the mentoring program was the key metric being studied, as this decision was elicited by the firm's staff.

also asked the staff to monitor any complaints or concerns over not being matched to a mentor; no instances of discouragement were ever communicated to the research team. In addition, they never indicated that the content of the mentor-protégé meetings was shared with non-mentored agents.

We formally test the net effect of discouragement and information leakage by comparing the performance of new workers who joined the firm in experimental cohorts with new hires who were not part of the experiment. Under the null of no net discouragement (or encouragement), information leakage, or other SUTVA violations, we would expect the productivity of non-treated workers in experimental cohorts to be equal to the productivity of new hires outside of the experiment. To conduct this test, we compare new hires in experimental cohorts with 1) new hires who were in hold-out cohorts during the time of the experiment and 2) new hires who joined the firm prior to the experiment. Our tests examine the relative performance of new hires versus experienced veterans working in the same divisions at the same time, which enables us to make comparisons over time when sales conditions may differ. This approach removes common time series differences across cohorts, including those due to demand fluctuations, and improves the power of our tests.

We estimate the following model using ordinary least squares:

$$\begin{aligned} \mathrm{y}_{i,t} = {} & \alpha + \beta_1 \text{New Hire}_i + \beta_2 (\text{New Hire} \times \text{Mandatory})_i \\ & + \beta_3 (\text{New Hire} \times \text{Voluntary})_i \\ & + \beta_4 (\text{New Hire} \times \text{Mandatory} \times \text{Mentored})_i \\ & + \beta_5 (\text{New Hire} \times \text{Voluntary} \times \text{Mentored})_i + \zeta_{j,l,t} + \varepsilon_{i,t}, \end{aligned} \tag{O.A.1}$$

where *New Hire* is an indicator if the agent has tenure of two months or less, *Mandatory* and *Voluntary* are indicators for the Mandatory- and Voluntary-Conditions among program-eligible cohorts, respectively, and *Mentored* is an indicator for those assigned a mentor. $\zeta_{j,l,t}$ captures division-by-location-by-date fixed effects, absorbing fluctuations in call volumes across divisions on particular dates and location-specific shocks that may affect productivity; $\varepsilon_{i,t}$ is an idiosyncratic error term.

Our test of net discouragement/encouragement and information leakage is the joint test that $\beta_2 = \beta_3 = 0$, indicating that the productivity of new hires relative to veterans in the Mandatory- and Voluntary-Conditions is no different than the new hire-to-veteran productivity differences in hold-out cohorts. In the Voluntary-Condition, those who opt out of the program and those who opt in but are not mentored are a non-random group of agents who are pooled together in the (*New Hire* $\times$ *Voluntary*)$_i$ group indicator. We therefore conduct another test of the joint null of zero returns to mentorship and zero net discouragement and information leakage that is robust to non-random selection by testing whether $\beta_3 = \beta_5 = 0$. This test asks whether overall cohort-level productivity in the Voluntary-Condition differs from the productivity of hold-out cohorts.

The results in Panel A of Table O.A.1 report estimations of productivity relative to these hold-out cohorts. The negative and statistically significant coefficient on *New Hire* in Column (1) suggests that newly hired agents generate approximately 30% ($= e^{-0.354} - 1$) less daily revenue relative to veterans. The small and insignificant coefficients on *New Hire* × *Mandatory* and *New Hire* × *Voluntary* suggest that newly hired non-mentored agents in program-eligible cohorts perform like newly hired agents in hold-out cohorts. We thus fail to detect evidence of discouragement or leakage. In addition, we are unable to reject the null that $\beta_2 = \beta_3 = 0$, indicating that newly hired, non-mentored agents in the Mandatory-Condition performed like newly hired, non-mentored agents in the Voluntary-Condition. The significant coefficient on *New Hire* × *Mandatory* × *Mentored* and the insignificant coefficient on *New Hire* × *Voluntary* × *Mentored* align with the main treatment effect estimates discussed in Section 4. Column (2) shows that our results are robust when controlling for agent demographic characteristics—age, gender, and marital status—which is important, given that randomization of agents into treatments did not occur for veterans and hold-out cohort agents (we do not have data on referral status or hiring scores for many veteran agents). Columns (3)–(4) repeat this exercise while using IHS(RPC) as the dependent variable. The small, insignificant coefficients on *New Hire* × *Mandatory* and *New Hire* × *Voluntary* in both of the columns further support the notion that discouragement and leakage are unlikely drivers of our estimated mentorship treatment effects. In the last row of Panel A, we cannot reject the null that productivity in the voluntary cohorts equals the productivity of hold-out cohorts.

O.A.2 Possibility of Program Crowd-Out

Formal workplace programs have the potential to crowd out informal arrangements that fill similar functions. In our context, the program may have crowded out informal mentorship if the mentors were already providing informal guidance but stopped once the program began. To test whether our program crowded out mentoring that would have occurred in its absence, we assess whether non-mentored agents in program-eligible cohorts perform less well than new agents who joined the firm prior to the program. We again use the relative performance difference between new hires and veterans as the basis for comparison. If the mentorship program crowded out organic mentoring, then we would expect non-mentored new hires during the program to perform worse than new hires from prior years. We find no such evidence in Panel B of Table O.A.1. These estimations resemble those in Panel A, but the comparison group identifying the baseline *New Hire* indicator is now new hires who enter the firm prior to the experiment, rather than those from contemporaneous hold-out cohorts.[3] Non-mentored new hires during the time of the program had similar levels of productivity relative to new hires from before the program, suggesting that the program was unlikely to have crowded out organic mentoring.

[3] Contemporaneous cohorts are not a good comparison group because they would be subject to the same limited supply of informal mentors.

O.A.3 The Mentorship Program Did Not Impact Worker Retention

In Section 4.4, we showed that, across both conditions, the mentorship program had no impact on retention, which, therefore, cannot explain the observed sales revenue treatment effects. Still, retention is consequential for the firm and small differences in retention may change the unit economics of the program. To further investigate the role of retention in our setting, we estimate the treatment effects on a panel where a mentorship slot is the unit of analysis. Specifically, we fill in the productivity of agents who leave the firm with the expected productivity of a replacement for both mentored and non-mentored agents. The total productivity gain to the firm from a mentorship slot is the relative productivity gain of the treated agent while the agent is retained, followed by the productivity of a randomly drawn replacement post-separation. A similar approach is used for the productivity of non-mentored controls.[4] When using the mentorship slot as the unit of analysis and filling in replacement productivity, we find results that largely mirror our main results on the unbalanced panel (see Table O.A.2).[5] Using the mentorship slot as the unit of analysis, the per-worker benefits to the firm in the Mandatory-Condition remain positive and significant, whereas the analogous firm-benefit is negligible in the Voluntary-Condition.

O.A.4 Framing Effects and Encouragement Effects

Recall that Table 4 shows that agents in the Mandatory-Condition who are relatively likely to participate in the program have positive (but insignificant) treatment effect estimates. It is possible that framing effects or some other type of encouragement effect contributes to these positive point estimates. There are two separate potential mechanisms to consider, with different implications for implementing the program. The first is a direct framing effect where the mandatory nature of the program causes agents to infer something about its value and "buy in" or engage. This could be a valuable aspect of program design, and evidence in its favor would mean that effective communication—not just effective curriculum—can improve the implementation of training programs. Sections O.A.5 and O.A.6 discuss whether we can detect evidence of framing effects. While framing effects seem like the most plausible explanation for the positive point estimates among agents with low opt-out propensity scores in Table 4, we find relatively limited evidence for them based on proxies for workers'

[4] We do this by computing the average productivity of newly hired non-eligible agents in the same location-division-year-quarter as the departed agent. We then re-estimate our main regression models using the full sample plus the imputed data post-separation.

[5] We also estimate treatment effect bounds that account for non-random attrition, as proposed by Lee (2009). The key assumption when implementing this approach is that some mentored agents would have left the firm absent mentorship but no mentored agents left the firm *because* they were mentored (a traditional monotonicity assumption). Table O.A.3 reports the upper and lower bounds of the estimated treatment effects on productivity in months 1–2. The results suggest that our estimated treatment effects in the Mandatory-Condition are largely attributable to the intensive margin of agents becoming more productive, and not due to differential retention effects. Importantly, these estimations do not include cohort fixed effects due to limitations in the implementation of the *leebounds* command, reducing their comparability to our main results.

engagement with the program.

A different encouragement source may come from randomization, where treatment may cause agents to feel special or exceptional because of perceived inequality in access to benefits provided by the firm. These perceptions may be exacerbated if the randomized nature of treatment was less salient to agents in the Mandatory-Condition compared to the Voluntary-Condition. We find limited evidence that treatment salience differences drive divergence in productivity treatment effects across conditions. In Section O.A.7 we discuss tests for treatment effect heterogeneity for agents who were more or less likely to be aware of their differential treatment status. In this test, we compute the fraction of agent-days where an agent may have been aware of differential treatment by working alongside anyone in their hiring cohort who had a different treatment status. We find that agents in the Mandatory-Condition who are exposed only to coworkers with the same treatment status have slightly positive but insignificant treatment effects, reducing the concern that perceptions of feeling special from differential treatment drive our results.

O.A.5 Meeting Completion Rates and Framing Effects

To test for differences in compliance or buy-in between the Voluntary- and Mandatory-Conditions, we tabulate meeting completion rates between mentor-protégé pairs in Table O.A.4. Of the 155 agents assigned to mentorship in the Voluntary-Condition, 25 never completed a recorded meeting with a mentor, while 18 of the 127 treated agents in the Mandatory-Condition never met with their mentor. Mandatory-Condition protégés completed both more of their scheduled meetings (2.31 versus 2.11) and had a higher meeting completion ratio (74% versus 64%).[6] These differences could arise because the opt-in framing in the Voluntary-Condition may have portrayed the program as optional rather than a job requirement (Hossain and List, 2012; Hong et al., 2015; Englmaier et al., 2017). Yet compliance was relatively high in both the Voluntary- and Mandatory-Conditions, suggesting compliance differences are unlikely to be large enough to explain the gap in treatment effects across conditions.

To show this formally, we estimate Local Average Treatment Effects (LATE) of meetings for compliers and test for whether the effect of the meetings that did occur is heterogeneous across conditions. We use an instrumental variable strategy where we instrument for meeting completion with the assignment to a mentor indicator, *Mentored*. Panel A of Table O.A.5 reports separate estimates for the Voluntary-Condition and the Mandatory-Condition, using IHS(Revenue) as the dependent variable. The point estimates for those in the Voluntary-Condition, Columns (2) and (4), are not statistically different from zero, and the point estimates are negative. The results in the Mandatory-Condition, Columns (6) and (8), suggest completing all meetings

[6] While the pre-registered mentoring protocol called for one meeting per week for four weeks, there were instances in which either a mentor, protégé, or both were absent from work for an extended period of time (e.g., on vacation), reducing the number of possible scheduled meetings from four to three (or fewer, in some cases). As such, the denominator of the meeting completion ratio is occasionally less than four.

raised daily revenue for compliers by 24% ($= e^{0.216} - 1$) and that each additional meeting increased revenue by 6.5% ($= e^{0.063} - 1$). Panel B shows very similar results when using IHS(RPC) as the dependent variable.[7] The differences in IV estimates between the Voluntary- and Mandatory-Conditions indicate that it is not the lower number of meetings driving the gap in treatment effects across conditions. Instead, the effect of a meeting differs between the Voluntary- and Mandatory-Conditions among those who do comply; compliance differences in meeting completion appear too small to explain the gap in the intention-to-treat reduced form.

O.A.6 Meeting Contents Captured by Worksheets

The hypothesis that framing effects explain the remaining gap in treatment effects across conditions suggests that we might be able to capture differences in program engagement via the contents of worksheets recorded during sessions with mentors. We show that meeting contents were roughly similar for Voluntary-Condition and Mandatory-Condition agents, suggesting agents in both conditions supplied similar amounts of effort when meeting with their mentors. In particular, we consider the amount of content transcribed on agent's worksheets by counting the total number of words written. While this is an imperfect measure of the quality of the mentor-protégé meetings, it proxies for the agents' level of engagement. In our second approach, which is motivated by the worksheet analysis in Sandvik et al. (2020), we use a bag-of-words to determine how much of a response's content is focused on job-specific skills and knowledge, relative to how much is focused on receiving support or encouragement.[8] Specifically, we tabulate the number of "skill" words an agent uses in their responses, and we do the same thing for the number of "support" words. Words that are not classified as either support words or skill words are categorized as "other," including stop words.

In comparing the worksheet content of Voluntary-Condition agents and Mandatory-Condition agents (reported in Table O.A.6), we do not find meaningful differences in the number of total words, skill words, or other words recorded by agents in the two conditions. Mandatory-Condition agents use about 0.15 more support words than do Voluntary-Condition agents, but this effect is only marginally significant. The similarity of worksheet content suggests meeting engagement differences between the two conditions were likely small. This puts additional weight on the remaining explanation: that different agents benefited from similar program features, rather than differences in meeting contents or engagement by compliers across the program conditions.[9]

[7] The high degree of similarity between the Panel A and Panel B results is not a transcription error. The correctness of the point estimates has been verified.

[8] We list the words in each category in Appendix O.E, along with multiple example responses.

[9] Two weeks after mentors and protégés completed their final meeting, the staff asked protégés to complete a post-mentorship survey. The completion rates for this survey were quite low (less than 10%), as the firm did not monitor or provide incentives for completion. Figure O.A.1 shows that protégés, on average, felt like they benefited from the program. The average respondent reported that mentorship helped them to learn selling tactics and that the program increased their day-to-day

O.A.7 Salience of Differential Treatment

As an additional way to test for encouragement effects, we compare the treatment effects among mentored agents in teams where relatively more or fewer agents were also mentored. Specifically, if mentored agents felt they were chosen for mentorship for non-random reasons, then they might have changed their effort or buy-in in response to this perception of the firm's commitment to them. If such perceptions existed, we would expect to see larger treatment effects for agents with fewer mentored teammates, as exposure to more mentored teammates would likely moderate these feelings of being exceptional. We show in Table O.A.7 that the treatment effects of mentorship in the Mandatory-Condition do not vary based on whether or not mentored agents had more exposure to teammates from their same hiring cohort who were also mentored. While we cannot completely rule out the possibility that encouragement effects were at play in our setting, our tests do not detect any evidence of their influence.

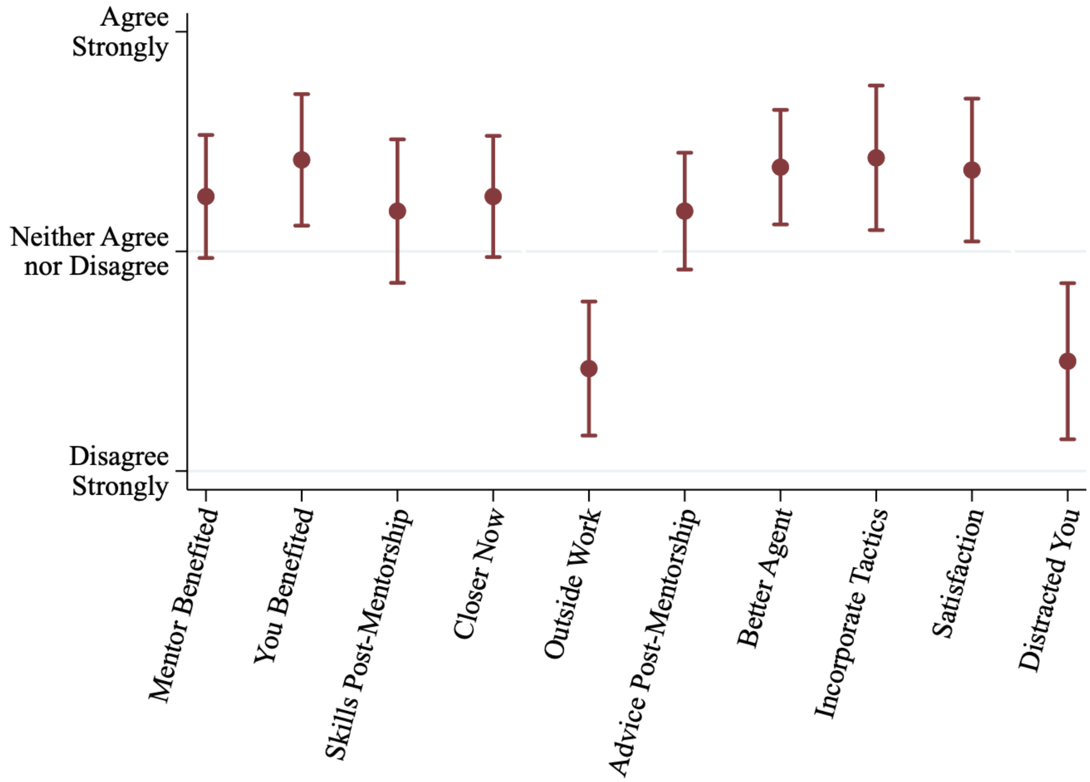
O.A.8 Belief- and Preference-Based Explanations for Opt-Out Behavior

It is possible that beliefs, preferences, or both explain why some agents opt out of the voluntary program. We cannot determine whether miscalibrated beliefs explain the opt-out decision, as we made a design decision to make the program appear as natural as possible. In particular, we did not ask agents in the Voluntary-Condition to report their prior beliefs about program efficacy before soliciting their decision to opt in or opt out due to the concern that reflecting on program benefits may have altered sorting patterns relative to most sign-up procedures that do not ask for such reports. However, our representative survey (see Section 5.2 in the main text) suggests that many workers, approximately 26%, believe similar programs would not benefit them. Accordingly, beliefs around program efficacy may play a role in limiting participation in voluntary training and mentoring programs.

We find limited evidence that workers' preferences explain the decision to opt out, albeit our tests are only indirect. First, to the extent that personality characteristics proxy for preferences across worker types, we show in Table 3 that variation in personality characteristics does little to explain opt-out decisions. Second, we show in Table O.A.4 that most agents in the Mandatory-Condition met with their mentors multiple times, suggesting that an aversion to meeting with more seasoned coworkers is also an unlikely explanation for the opt-out decisions.

satisfaction at work.

Figure O.A.1: Responses to Post-Mentorship Survey



Notes. This figure plots the average values (and 95% confidence intervals) for responses to the post-mentorship survey questions. All responses were made on a scale from -3 to 3, with -3 indicating "Disagree Strongly," 0 indicating "Neither Agree nor Disagree," and 3 indicating "Agree Strongly." The statements, from left to right, are as follows: "Your mentor benefited from the mentoring relationship"; "You benefited from the mentoring relationship"; "Since your formal meetings have ended, your mentor has continued to teach you skills to help you make more sales"; "You and your mentor are closer now than you were during the mentor program"; "Since your formal meetings have ended, you and your mentor have spent time together outside of the office"; "Since your formal meetings have ended, you have reached out to your mentor for help/advice"; "You have become a better sales agent as the result of being mentored"; "Being mentored helped you incorporate important selling tactics into your day-to-day work"; "Having a mentor increased your day-to-day satisfaction at work"; "Being mentored distracted you from reaching your potential each week." Seventeen protégés completed the post-mentorship survey.

Table O.A.1: Tests for SUTVA Violations Comparing New Hire Productivity Relative to Veteran Employees

Panel A: Discouragement/Leakage Tests Between Mentor-Eligible and Holdout Cohorts

	IHS(Revenue)		IHS(RPC)	
	(1)	(2)	(3)	(4)
New Hire	-0.354***	-0.275**	-0.431***	-0.393***
	(0.113)	(0.118)	(0.070)	(0.072)
New Hire × Mandatory	-0.025	-0.041	-0.033	-0.044
	(0.123)	(0.126)	(0.073)	(0.075)
New Hire × Voluntary	0.007	0.002	0.047	0.039
	(0.111)	(0.114)	(0.066)	(0.067)
New Hire × Mandatory × Mentored	0.191***	0.199***	0.111**	0.115**
	(0.068)	(0.069)	(0.046)	(0.047)
New Hire × Voluntary × Mentored	-0.067	-0.041	-0.067	-0.050
	(0.074)	(0.072)	(0.048)	(0.046)
Division-Location-Date FE	✓	✓	✓	✓
Demographic Controls		✓		✓
Adj. R-Square	0.073	0.076	0.096	0.099
Observations	47,803	47,766	47,803	47,766
New $\text{Hire}_{Man} = 0$, New $\text{Hire}_{Vol} = 0$	0.954	0.912	0.411	0.419
New $\text{Hire}_{Vol} = 0$, $\text{Mentored}_{Vol} = 0$	0.647	0.839	0.351	0.536

Panel B: Crowdout Tests Between Mentor-Eligible and Pre-Experimental Cohorts

	IHS(Revenue)		IHS(RPC)	
	(1)	(2)	(3)	(4)
New Hire	-0.400***	-0.370***	-0.399***	-0.394***
	(0.088)	(0.091)	(0.057)	(0.059)
New Hire × Mandatory	0.009	0.008	-0.074	-0.070
	(0.119)	(0.121)	(0.073)	(0.074)
New Hire × Voluntary	0.027	0.026	-0.015	-0.012
	(0.114)	(0.114)	(0.071)	(0.071)
New Hire × Mandatory × Mentored	0.166**	0.170**	0.093**	0.094*
	(0.071)	(0.073)	(0.047)	(0.049)
New Hire × Voluntary × Mentored	-0.070	-0.057	-0.069	-0.060
	(0.072)	(0.072)	(0.047)	(0.047)
Division-Location-Date FE	✓	✓	✓	✓
Demographic Controls		✓		✓
Adj. R-Square	0.102	0.103	0.131	0.132
Observations	75,094	75,094	75,094	75,094
New $\text{Hire}_{Man} = 0$, New $\text{Hire}_{Vol} = 0$	0.969	0.970	0.504	0.548
New $\text{Hire}_{Vol} = 0$, $\text{Mentored}_{Vol} = 0$	0.614	0.732	0.276	0.378

Notes. This table reports tests of the net effect of discouragement, leakage, and crowd-out by comparing the performance of three groups of agents to seasoned veterans who began working at the firm prior to the onset of the experiment: (1) new hires who were in mentor-eligible hiring cohorts; (2) new hires who were in hold-out cohorts that were not eligible for mentorship during the time of the experiment (Panel A); and (3) new hires who entered the firm before the experiment (Panel B). We test whether the performance of treatment-eligible new hires differs from those of non-treatment eligible new hires. Agents who opt out in the Voluntary-Condition are included (see the text for parameter interpretation given this sample). The dependent variable is IHS(Revenue) in Columns (1)–(2) and IHS(RPC) in Column (3)–(4). In the bottom two rows, $NewHire_{Man}$ stands for new hire in the Mandatory-Condition, $NewHire_{Vol}$ stands for new hire in the Voluntary-Condition, and $Mentored_{Vol}$ stands for new hire who was mentored in the Voluntary-Condition. All specifications include division-by-location-by-date fixed effects. Columns (2) and (4) control for agent age, gender, and marital status. Standard errors are clustered by hiring cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table O.A.2: Productivity Treatment Effects When a Mentorship Slot is the Unit of Analysis

	Voluntary-Condition (Opt-In Agents)		Mandatory-Condition (All Agents)		Voluntary-Condition (Non-Mentored Agents)		Both Conditions (All Agents)		
	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	IHS(Rev)	IHS(RPC)	Index
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mentored	-0.070	-0.050	0.140*	0.107**			0.140*	0.107**	0.136**
	(0.069)	(0.043)	(0.076)	(0.043)			(0.075)	(0.043)	(0.056)
Voluntary Opt-Out					-0.281***	-0.166***	-0.201**	-0.098	-0.094
					(0.083)	(0.058)	(0.091)	(0.059)	(0.063)
Mentored × Voluntary							-0.234**	-0.171***	-0.177**
							(0.101)	(0.061)	(0.068)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓	✓
Adj. R-Square	0.025	0.043	0.023	0.034	0.036	0.046	0.024	0.036	0.078
Observations	9,986	9,986	10,216	10,216	6,588	6,588	22,451	22,451	22,451
p-value: Mentored + Mentored × Voluntary							0.173	0.142	0.293

Notes. The results in this table show estimates of the treatment effects of mentorship when a slot (i.e., an occupied position as a sales agent within the firm) is the unit of analysis regardless of the agent's tenure. For this analysis, we form a balanced panel of agents made up of the observed productivity of those who remain at the firm and imputed productivity of a replacement for agents who separate before the indicated time horizon. In other words, for mentor-eligible agents who leave the firm before the two-month mark, we extend the time series of their productivity provision to two months and replace their post-termination productivity values with the average productivity of a newly hired replacement agent. We do this by computing the average productivity of newly hired non-mentor-eligible agents in the same location-division-year-quarter as the departed agent. We then re-estimate our main intention-to-treat regression models. We estimate ordinary least squares regressions with cohort fixed effects in all columns. Standard errors are clustered by cohort and are reported in parentheses. The bottom row reports the p-values from post-estimation tests that the sum of the coefficients on *Mentored* and *Mentored* × *Voluntary* equals zero. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table O.A.3: Lee Bounds Estimates of the Effect of Mentoring on Productivity

	Voluntary-Condition		Mandatory-Condition	
	IHS(Revenue)	IHS(RPC)	IHS(Revenue)	IHS(RPC)
	(1)	(2)	(3)	(4)
Mentored$_{lower}$	-0.222***	-0.131***	0.169***	0.094***
	(0.076)	(0.048)	(0.046)	(0.035)
Mentored$_{upper}$	-0.058	-0.018	0.208**	0.122**
	(0.042)	(0.031)	(0.083)	(0.051)
Observations	9,231	9,231	8,408	8,408

Notes. This table uses agent-day productivity data for agents in the Voluntary-Condition, excluding those agents who opt out of the program, in Columns (1) and (2) and for agents in the Mandatory-Condition in Columns (3) and (4). The sample is a balanced panel of agents' productivity through their first two months of tenure, augmented with missing data for those who exit the firm during the panel. We estimate treatment effect bounds that account for non-random attrition as proposed by Lee (2009). These estimations do not include cohort fixed effects, as the *leebounds* command cannot accommodate fixed effects, which causes the estimates to differ from those in our main results. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table O.A.4: Meeting Completion Rates Across Conditions in the Experiment

	Voluntary-Condition	Mandatory-Condition	p-value
	(1)	(2)	
Number of Agents	155	127	
At Least One Recorded Meeting	130	109	
No Recorded Meeting	25	18	
Number Recorded Meetings (avg.)	2.11	2.31	0.260
	(1.36)	(1.58)	
Meeting Completion Ratio (avg.)	0.64	0.74	0.031
	(0.38)	(0.38)	

Notes. In this table we report the mentor meeting completion details of protégés in the Voluntary-Condition and the Mandatory-Condition. *No Recorded Meeting* indicates that there is no record that the mentor-protégé pair ever met with one another. The *Meeting Completion Ratio* measure is based on the number of possible meetings the mentor-protégé pair could have had. While the pre-registered mentoring protocol called for one meeting per week for four weeks, there were instances in which either a mentor or protégé or both were absent from work for an extended period of time (e.g., on vacation), reducing the number of possible scheduled meetings from four to three (or fewer, in some cases). As such, the denominator of the meeting completion ratio is occasionally less than four.

Table O.A.5: Instrumental Variables Estimates of Meetings with Mentors Across Different Conditions

Panel A: Daily Revenue

	Voluntary-Condition				Mandatory-Condition			
	First Stage (1)	IV (2)	First Stage (3)	IV (4)	First Stage (5)	IV (6)	First Stage (7)	IV (8)
Mentored	0.747***		2.314***		0.791***		2.724***	
	(0.062)		(0.159)		(0.045)		(0.218)	
Meeting Completion Ratio		-0.113				0.216**		
		(0.100)				(0.086)		
Number Recorded Meetings				-0.036				0.063**
				(0.032)				(0.025)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Cragg-Donald Wald F	20,788		10,489		26,622		14,556	
Centered R-Square	0.618	0.000	0.569	0.000	0.769	0.002	0.657	0.004
Observations	7,569	7,569	7,569	7,569	6,725	6,725	6,725	6,725

Panel B: Daily Revenue-per-Call

	Voluntary-Condition				Mandatory-Condition			
	First Stage (1)	IV (2)	First Stage (3)	IV (4)	First Stage (5)	IV (6)	First Stage (7)	IV (8)
Mentored	0.747***		2.314***		0.791***		2.724***	
	(0.062)		(0.159)		(0.045)		(0.218)	
Meeting Completion Ratio		-0.113				0.141**		
		(0.067)				(0.054)		
Number Recorded Meetings				-0.036*				0.041**
				(0.021)				(0.016)
Cohort Fixed Effects	✓	✓	✓	✓	✓	✓	✓	✓
Cragg-Donald Wald F	20,788		10,489		26,622		14,556	
Centered R-Square	0.618	0.001	0.569	0.001	0.769	0.001	0.657	0.003
Observations	7,569	7,569	7,569	7,569	6,725	6,725	6,725	6,725

Notes. In Panel A of this table we present IV regressions of daily IHS(Revenue) on measures of mentor-protégé meeting completion using *Mentor* assignment as an instrumental variable. IHS(RPC) is the dependent variable in Panel B. Columns (1)–(4) consider only agents in the Voluntary-Condition who did not choose to opt out of the mentorship program, and Columns (5)–(8) consider only agents in the Mandatory-Condition. The dependent variable in the first stage regressions in Columns (1) and (5) is *Meeting Completion Ratio*, the fraction of possible (i.e., scheduled) mentor-protégé meetings that the protégé completed. The dependent variable in the first stage regressions in Columns (3) and (7) is *Number Recorded Meetings*, the number of mentor-protégé meetings that the protégé completed. In Columns (2), (4), (6), and (8) of Panel A (Panel B), we present IV regressions of IHS(Revenue) (IHS(RPC)) as a function of the completion ratio or the number of completed meetings, using *Mentored* as an instrument. Note, the high degree of similarity between the Panel A and Panel B results is not a transcription error. The correctness of the point estimates has been verified. Agents who opt out in the Voluntary-Condition are not included in this sample. All specifications include cohort fixed effects. Standard errors are clustered by cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table O.A.6: Differences in Worksheet Content Across Conditions in the Experiment

	Total Words per Worksheet	Skill Words per Worksheet	Support Words per Worksheet	Other Words per Worksheet
	(1)	(2)	(3)	(4)
Mandatory-Condition	-0.023	0.253	0.145*	-0.422
	(2.686)	(0.311)	(0.079)	(2.584)
Adj. R-Square	-0.006	-0.002	0.013	-0.006
Observations	159	159	159	159
Mean DV	47.36	4.27	0.448	42.65

Notes. This table considers differences in worksheet content between protégés in the Mandatory-Condition and those in the Voluntary-Condition. We have completed worksheet data for 159 mentored agents, as some worksheets that were turned in to the internal mentoring staff were never returned to us. For each agent, we compute the number of words written on all of their completed worksheets, and we divide this by the number of worksheets received from that agent. For each worksheet, we identify the fraction of words in the responses that relate to job-specific skills or knowledge (*Skill*), those that relate to receiving support, encouragement, and friendship (*Support*), and those that are neither related to skill nor support (*Other*), which include stop words. These become the dependent variables in our regression specifications of worksheet content on mentorship type. Robust standard errors are reported in parentheses. The mean of the dependent variable is listed below the observation count line. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table O.A.7: Tests for Variation in Mentoring Treatment Effects Based on Exposure to Teammates with the Same or Different Treatment Status

	IHS(Revenue)	IHS(RPC)
	(1)	(2)
Mentored	0.154*	0.101**
	(0.075)	(0.046)
Mentored × Same Treatment	0.030	0.022
	(0.229)	(0.129)
Cohort FE	✓	✓
Adj. R-Square	0.033	0.036
Observations	6,725	6,725

Notes. The sample consists of agents in the Mandatory-Condition. After training, agents joined different teams, after which they would begin the mentoring program. For each agent, mentored or not, we generate an intermediate variable (not the regressor), *Same Teammate*, which equals one if, on that day, all the agent's teammates from their hiring cohort had the same mentored/non-mentored treatment designation as their own. Otherwise, *Same Teammate* equals 0 for that agent-day. We then compute the rolling average of the *Same Teammate* variable for each agent, from their first day on the job after training to the present day, and we label this *Same Treatment*. The variable *Same Treatment* captures the fraction of days on the job historically that a new hire worked alongside teammates from their same hiring cohort who had the exact same mentored/non-mentored treatment designation as themselves. So, a mentored (non-mentored) agent with *Same Treatment* $= 1$ has never worked alongside a non-mentored (mentored) teammate from their same hiring cohort and would have no perceptions of differential treatment. The mean of *Same Treatment* in the sample is 0.22. All specifications include cohort fixed effects. Standard errors are clustered by cohort and are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

O.B AEA Pre-Registration Text

Here we replicate the AEA pre-registration text. Differences between the AEA pre-registration and our actual implementation are denoted in footnotes.[10]

O.B.1 Abstract

Mentoring is increasingly encouraged in workplaces, and a number of firms have implemented formal programs. While a growing body of research suggests that mentoring relationships benefit those being mentored (protégés), there is scant evidence to delineate whether these favorable outcomes are driven by the mentoring experience on average, by the self-selection of protégés into mentoring who anticipate having the largest gains (selection based on gains), or by the self-selection of protégés who would have performed well in the absence of mentoring (selection based on levels). We use a field experiment to evaluate a workplace mentoring program inside a large sales organization.

Experienced employees opt-in as mentors, and new hires are slated as potential protégés. The project objective is to study the mentoring consequences across protégés who actively elect to be formally mentored relative to those who are randomly allocated a mentor. We estimate treatment effects on sales productivity and turnover for those who select into mentoring and for those who opt out.

O.B.2 Intervention(s)

We analyze the effectiveness of a workplace mentoring program where employees opt-into mentoring or are randomly assigned a mentor. More details are provided in the design field.

O.B.3 Intervention Start Date

2019-05-27

O.B.4 Intervention End Date

2019-12-20

O.B.5 Primary Outcomes (end points)

Log revenue-per-call (RPC), an indicator for worker turnover, log completed tenure, the firm's internal adherence to schedule measure (e.g. time spent working whilst at work), and the firm's internal engagement metrics (online surveys asking for willingness to recommend employment at firm, comfort with leadership, etc.).[11]

[10] We received IRB approval at all of our respective institutions prior to the onset of the mentoring program. See IRB19-0769.

[11] Following our prior paper, we use an approximate logarithmic transformation of RPC (and revenue). In this case, we use IHS() rather than logs because we have daily data that includes cases where these newly hired agents make zero sales. Our results, as shown in Table A.4, are similar if we use the original pre-registered log() measure.

O.B.6 Primary Outcomes (explanation)

Agent's weekly RPC is a measure of sales productivity that removes demand variation outside of the worker's control. RPC is the primary productivity measure used by the firm, combining both agent's firm-specific knowledge and their individual effort.[12] Worker turnover measures whether the interventions changed the agents' propensity to leave the firm. Log of completed tenure is a different measure of retention that has been used in the prior literature and the attendance measure provides an adjacent measure of agent effort.[13] Finally, engagement measures are hypothesized to be forward looking measures of productivity.[14]

O.B.7 Experimental Design Details

Seasoned sales agents are invited to apply as internal mentors to incoming recruits (the firm "qualifies" mentors as having sufficient sales experience). New mentorship opportunities are periodically announced, and prior mentors are permitted to re-enter the mentor pool. The firm communicates that serving as a mentor is a useful first step to being considered for a managerial position. New mentors complete a survey asking them about their personality, interests, work preferences, and values. Mentors are randomly assigned with probability 50% to receive a set of instructions emphasizing that mentoring is about teaching protégés how to do the job. The remaining mentors receive instructions emphasizing that mentoring is about providing protégés support. Sales agents are hired in batches (cohorts). Newly hired sales agents complete two weeks of training, primarily in a classroom or listening in on other agents' sales calls. New agents then complete the same personality and preference survey that mentors take. At the end of their two-week training, each cohort of agents is eligible for randomization into a mentoring treatment arm. Any mentoring relationship commences as soon as the agent completes their training.

The randomization procedure is as follows:

O.B.8 Cohort Level Randomization

The initial level of randomization is cohorts of new hires (potential protégés). Each cohort (a group of new hires who are joining the firm at the same time, are in the same training group, and will be working in the same sales division and office location) will be randomized into one of two conditions: Mandatory-Condition or Voluntary-Condition. 40% of the cohorts will be in the Mandatory-Condition group and 60% of the cohorts will be in the Voluntary-Condition group.

[12] RPC was the primary endpoint based on our experience analyzing the productivity of veteran agents within the firm (Sandvik et al., 2020), but seminar participants asked us about other margins of adjustment (like time use) that is captured by total revenue and is more relevant for new agents compared to more experienced workers, which is why we report both metrics.

[13] In practice, we find no differences in completed tenure and simplify the analysis by using indicators for completed tenure rather than log completed tenure.

[14] As noted below, the firm changed the cadence of collecting engagement metrics (which were planned for 5 weeks after training completion), so we do not have these measures for many cohorts.

O.B.9 Within Cohort Randomization

For cohorts in Mandatory-Condition, new hires will receive a mentor with probability 50%. This will be communicated privately between sales floor staff and the individual workers. Agents in the Mandatory-Condition who do not receive a mentor will not receive communication regarding the program. For cohorts in Voluntary-Condition, sales floor staff verbally explain the firm's mentorship program, answer questions, and provide each agent a confidential ballot where they can decide whether or not to enter a lottery which randomly determines whether the agent is allocated a randomly assigned mentor, or no mentor at all. Of the agents who enter the lottery, approximately 50% will be assigned a mentor. Agents who choose not to be mentored will never be assigned a mentor.

O.B.10 Compliance Tracking

The firm's training staff will track whether mentors and protégés meet. This tracking will be aided by worksheets. Upon completion of the worksheets, the firm will reward "kudos" points that can be accumulated to purchase items from the company store. As mentioned earlier, mentors may participate more than once, however they will never have more than one protégé at a time.[15] Eligible protégés and mentors will each take an electronic survey at the end of the formal program. The survey for protégés will ask about the protégé's initial excitement when told about the mentoring program, their perceived engagement with their mentor, and an estimate of the effectiveness of mentoring. This question will be phrased as: "What was your average RPC last week? What do you think your average RPC would have been had you not been working with a mentor?" The survey for mentors will ask about the protégé's enthusiasm for the mentorship program and an estimate of the mentor's perceived treatment effect on the protégé. This question will be phrased as: "If your protégé had not received mentoring, his/her RPC would have been [40% lower — slider — 40% higher].[16] Note that numbers greater than zero mean that mentoring was not effective for improving protégé performance. Please be candid, as your responses will not be shared with management."

O.B.11 Edit June 4, 2019

To assess the potential for spillovers, we have revisited the design in consultation with the company such that there will be "hold out" cohorts for one division-office who

[15] As the program progressed, the internal mentoring staff felt that many of the mentors could effectively mentor multiple protégés as once. As a result, we adjusted the protocol such that it was possible for a single individual to mentor multiple new hires concurrently, but mentor-protégé pairs always met individually, meaning the protocol was the exact same from the point of view of the protégé.

[16] The post-mentorship survey completion rates of mentors and protégés were very poor, so we do not have meaningful data for this question. Anecdotally, the average responses of both sets of individuals suggests that protégés' RPC would have been lower in the absence of mentorship, but the inference is not precise.

never receive mentoring. Any cohorts/individuals who are switching brands also will be held-out. Work-from-Home cohorts will also present a possible "hold out" group for comparison and all cohorts in a smaller third office (which no longer exists, but for whom historical data is available) were "hold out" cohorts who knew nothing about mentoring.[17] A "sentiment survey" will be administered to all agents in their 5th week on the sales floor.[18] This will be one week after mentored agents finish hiring. We will gather information on their feelings towards the onbaording process and ask questions, common in the literature, to solicit their sentiment towards the firm, their perceptions of their ability, their enthusiasm about the job, etc. We will use this survey to test for spillovers based on survey responses.

O.B.12 Randomization Method

Randomization done by computer. Participants will be informed if randomized in.

O.B.13 Randomization Unit

Clustered randomization of cohorts in a first level, with individual randomization within the cohort. See design details.

O.B.14 Was the treatment clustered?

Yes

O.B.15 Sample size: planned number of clusters

The exact sample size is stochastic and depends on the firm's actual hiring. We have 46 planned clusters.

O.B.16 Sample size: planned number of observations

In one office, the firm has projected 269 new hires in 22 cohorts. There are 350 new hires in 24 cohorts projected in the second office.

O.B.17 Sample size (or number of clusters) by treatment arms

Please see design field.

O.B.18 Minimum detectable effect size for main outcomes

Using pre-intervention data to estimate the intra-class correlation coefficient and residual variation, the minimum detectable effect size for log RPC between those randomized into and out of mentoring is 0.07 (accounting for sample design and clustering).

[17] Further consultation with the firm's staff meant that the hold-out procedure was not isolated to a single division or office but instead involved rotation of some cohorts out of the program. More detail about hold-out cohorts is provided in the main text.

[18] We were not able to administer this survey. The firm had several of its own survey initiatives occurring simultaneously, so additional surveys connected to the mentorship program were not conducted due to the concern of "survey fatigue" among the sales agents.

O.B.19 Analysis Plan

The Treatment Effect of Mentoring on those who opt in is:

$$\beta_{OptInMentor} = mean(Y_{OptInMentor}) - mean(Y_{OptInNoMentor}).$$

We will estimate this mean difference using a regression of Y on an indicator for receiving a mentor along with cohort fixed effects and indicators for the type of instructions mentors receive.[19] The sample will be the workers in the voluntary treatment cohorts who opt into mentoring.

The Treatment Effect of Mentoring on those who opt out can then be derived by writing the average gain from mentoring in the population as:

$$mean(Y_{RandomMentor}) - mean(Y_{NoMentor}) = \beta_{OptInMentor}\pi_{OptIn} + \beta_{OptOutMentor}\pi_{OptOut}.$$

The β parameters are the heterogeneous treatment effects and the π are the population fraction who opt in and opt out. This yields:

$$\beta_{OptOutMentor} = [mean(Y_{RandomMentor}) - mean(Y_{NoMentor}) - \beta_{OptInMentor}\pi_{OptIn}]/\pi_{OptOut},$$

where the difference in means is net of cohort fixed effects and indicators for mentoring instruction type. The population average treatment effect (ATE) of mentoring can be estimated from a regression of Y on a dummy for receiving a mentor and cohort fixed effects in cohorts that have (entirely) randomly assigned mentoring. This yields:

$$\beta_{OptOutMentor} = [ATE - \beta_{OptInMentor}\pi_{OptIn}]/\pi_{OptOut}.$$

Inference for $\beta_{OptOutMentor}$ will come from block bootstrapping the statistic. Selection bias will be measured among voluntary treatment cohorts as the regression analogue of:

$$mean(Y_{OptInNoMentor} - mean(Y_{OptOut}),$$

where the means are net of cohort fixed effects. This procedure allows us to estimate sales productivity differences among protégés who opt into mentoring and those who do not. We use the sample of agents in the voluntary cohorts who did not receive a mentor. We regress Y on an indicator that the agent opted into mentoring along with cohort fixed effects and their mentor instruction-type fixed effects. Other regressions will look at opt-in as a function of early sales and demographic characteristics (gender, age, office location) and past experience (prior sales or call center experience).

[19] Mentors received instructions that either put more emphasis on the supportive nature of the program or the skills-building nature of the program. We detect no differences between instruction type. Because of this and for brevity, we omit this indicator from the models in our heterogeneous treatment effects tests.

We plan to validate these estimates using the electronic survey responses collected after the protégé graduates from the formal mentoring program, approximately 4 weeks following the initial onboarding instruction (e.g. how to use the systems, enroll for benefits, etc.).[20] We will compare average perceived gains from mentors and protégés to the actual estimated treatment effects across different assignment conditions. We will then assess whether the effectiveness of the mentoring pair differs based on characteristics of the mentor and protégé. We will regress protégé sales on fully saturated interactions of demographic characteristics for the mentor-protégé pair (old/young based on coarse buckets; gender) as well as similarity in survey responses on the intake survey.[21]

Finally, to assess whether mentoring detracts from–or improves sales–for the mentor, we will regress mentor log RPC and other sales measures on indicators demarking whether the mentor is eligible to mentor but has not yet done so, whether they have previously mentored in the program, or whether they are actively mentoring a protégé. This regression will include mentor fixed effects and mentor tenure.[22]

[20] As mentioned earlier, we were not able to administer this survey.

[21] This is a very high-dimensional exercise, and the most interesting potential differences (like females mentoring females) ended up having very small cell sizes, reducing statistical power for these tests.

[22] Tests that compare the characteristics of mentors and protégés, and those that look at the impact of mentorship on mentor productivity, are likely to be discussed in a separate article.

O.C Documentation of Instructions to Mentors and Example Worksheet for Structuring Conversations

Mentor Instructions

What is a Mentor?

In *The Odyssey*, Odysseus prepared to fight in the Trojan War. Before leaving home to fight in the war, he asked his trustworthy friend, named Mentor, to train and educate his son, Telemachus. Similarly, mentors today are meant to train and educate their protégés. Management at has chosen you to be a mentor---a source of further skill development---for newly hired sales agents. You have been selected specifically because you've demonstrated a willingness to teach other sales agents and help them become a successful and productive sales agent.

The responsibility to mentor a newly hired sales agent should not be taken lightly. Management strongly believes new agents will benefit from the additional training and the insider knowledge received as a result of being mentored by a talented, more seasoned agent. Because of this, has devoted significant resources to give mentors and protégés the best opportunity to spend productive time together, so please take your mentorship responsibilities seriously.

What will You Do as a Mentor?

As a mentor, you will do the following:

1. You will meet with your protégé at **least once a week**.
    a. Before meeting, your protégé will complete the Protégé Worksheet.
        i. If he/she has not completed it, you will kindly help him/her do so.
    b. During your meeting, you and your protégé will discuss his/her responses. You should also take this time to do the following:
        i. Impart knowledge and skill by explaining, giving useful examples, and demonstrating processes, and asking thought-provoking questions.
        ii. Discuss actions you've taken to become a successful sales agent.
        iii. Provide him/her with any tips and sales tactics that help you overcome customer concerns and that help you up-sell to better services.
        iv. Practice the designated sales protocol with them and help them gain a strong understand of the products, services, and bundles available.

2. After meeting with your protégé, you will deliver the finished worksheet to .
    a. will initial and timestamp the worksheet and make a record that you completed your weekly meeting responsibility.

3. Every two weeks, you will be asked to complete an on-line survey.
    a. These questions are meant to gauge the progress of your protégé and the overall benefit of the mentoring relationship.
    b. Please answer these questions honestly, as they are not meant to punish but, instead, to help assess the effectiveness of the mentorship program.

Protégé Worksheet (Week 1)

Protégé: ____
Mentor: ___________________________________ Number of times mentor has reached out: _____
Date: ____

Weekly Self-Reflection:

What are **your expectations** regarding your sales ability? Does your mentor know this?

What may **prevent you from** having a successful first week? Does your mentor know this?

Think of the **MOST** successful call you had recently. What made it **successful**?

Think of the **LEAST** successful call you had recently. What made it **unsuccessful**?

Weekly Goal:
What **ONE goal** are you setting for yourself for this coming week?

What will you do to **reach this goal**? Have you told your mentor about this goal? ___________
____

For Mentors to Respond:
How have you, as a mentor, been a source of **skill development** for your protégé? What have you done so far to help him/her succeed on the sales floor here at ?

Protégé's Initials	Mentor's Initials	Intern's Initials & Timestamp

O.D Pre-Registered Estimation of Heterogeneous Treatment Effects

We can use the estimated Mandatory-Condition and Voluntary-Condition productivity treatment effects, along with the data on the fraction of Voluntary-Condition agents who opt out of receiving a mentor, to estimate the treatment effect of mentorship among opt-out agents. We pre-registered the following procedure for this purpose. Using productivity measure Y, we define the conditional average treatment effect of mentoring given selection into participation as the difference in expected production between mentored and non-mentored agents conditional on opting in:

$$ATE|OptIn = \beta_{OptInMentored} = E(Y_{OptInMentored}) - E(Y_{OptIn\sim Mentored}).$$

We can then express the unconditional average treatment effect of mentorship as the weighted average of heterogeneous effects with shares π:

$$\begin{aligned} ATE = E(Y_{MandatoryMentored}) - E(Y_{Mandatory\sim Mentored}) = \\ \beta_{OptInMentored} \times \pi_{OptIn} + \beta_{OptOutMentored} \times \pi_{OptOut}. \end{aligned}$$

Rearranging terms, we get,

$$\beta_{OptOutMentored} = \{ATE - \beta_{OptInMentored} \times \pi_{OptIn}\}/\pi_{OptOut}.$$

We use the estimated treatment effect in the Mandatory-Condition as the estimated ATE, and we use the estimated treatment effect in the Voluntary-Condition as the estimated $ATE|OptIn$.[23] The values of π come from the proportion of agents who opted out in the Voluntary-Condition. We show the estimated treatment effect for opt-out agents in Table O.D.1, where standard errors come from 500 block-bootstrap iterations by cohort. The point estimate of 1.207 in Column (1) of Panel A implies that opt-out agents would have more than doubled their overall revenue generation, on average, had they received mentorship. Based on this analysis, opt-out agents were those who would have benefited the most from receiving mentorship. That is, program participation is negatively correlated with treatment gains.

[23] We include cohort fixed effects in estimating these treatment effects.

Table O.D.1: Estimated Treatment Effect of Mentoring Among Opt-Out Agents

Pre-Registered Estimates of Opt-Out Treatment Effects		
	IHS(Revenue)	IHS(RPC)
	(1)	(2)
Opt-Out Mentored Effect	1.422**	1.159***
	(0.648)	(0.417)

Notes. This table reports estimates of the treatment effect of mentorship among agents who would have opted out of the program using the pre-registered estimator described above. This estimator imposes that the full difference between the Voluntary- and Mandatory-Condition comes from heterogeneous treatment effects and selection. To estimate standard errors, we block-bootstrap by cohort (N = 52) over the whole procedure, with 500 bootstrap replications for each column. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

O.E Worksheet Response Examples

Panel A: Think of the most successful call you had recently. What made it successful?	
Skill	· I pitched TV really well · Having different examples of pitches from my coach to fall back on
Support	· I was confident and tried to connect · The person I spoke with was very nice

Panel B: Think of the least successful call you had recently. What made it unsuccessful?	
Skill	· Customer didn't want to pay the deposit, [I] didn't rebuttal · Not doing call flow, not caring, not enough discover
Support	· Not being confident in my ability to rebuttal · The person was rude and wanted me fired

Panel C: What will you do to reach this goal? Have you told your mentor about this goal?	
Skill	· I will create better pitches · Be better with the triple play, use what [the] mentor told [me] · My mentor is going to help me pitch DTV by giving me her tips on what helped her · Practice on every unserviceable call · Try upsell technique
Support	· [Goal to achieve] 1500 a day, build confidence in it · Be more positive · Stay positive · Stay in communication with [my coach] · Check in with my coach and be confident

Panel D: Words Associated with Sales Skills and Knowledge
Adherence, Conversion, Customer, Direct, Dish, Double, DPI, DTV, Internet, Knowledge, Phone, Pitch, Price, Pricing, Process, Revenue, RPC, RPH, Sale, Security, Sell, Skill, Sold, System, Television, Triple, TV

Panel E: Words Associated with Receiving Support
Annoy, Breath, Confidence, Confident, Cool, Encourage, Encouraging, Friend, Introduce, Kind, Laugh, Mean, Motivate, Motivation, Nice, Patience, Patient, Positive, Rude, Social, Support, Welcome, Welcoming

O.F Calculation Details for the Value to the Firm

This section details how we calculate the NPV to the firm from the mentorship program. First, we consider the additional revenues. Using a slot as the unit of analysis, taking the estimated treatment effect, and multiplying the average revenue of agents in non-mentored slots yields a daily revenue increase of about \$41 over the first 6 months of tenure. There were 127 mentored agents in the Mandatory-Condition, resulting in \$860 more in revenue per agent-month based on agents' total days of work. The firm earns this additional revenue net of an 8% commission rate paid to sales agents. Hourly pay is invariant to productivity and did not change with the program. We multiply these monthly net-revenue amounts by six, the number of months, and by 127, the number of mentorship slots. We conservatively assume this additional revenue is realized at the end of the year and discount the future cash flow using a 12.5% discount rate, which gives us a present value of the additional revenue earned by mentored agents equal to approximately \$536,000.

Second, we consider the costs. These costs include the estimated time costs of taking the mentors off the phone (protégés' opportunity costs are included in the revenue treatment effects) and administrative costs. Mentors and protégés spent 30 minutes in the mentorship meetings each week. Revenue per hour for mentors averaged \$148 and they were paid an additional \$10 of "kudos" points for completing each meeting. Together this implies a cost of \$84 per meeting or \$30,000 total, assuming that each treated agent met with their mentor 2.5 times. We include the administrative costs of the two internal mentorship staff members who oversaw the program in the two locations, estimated to total approximately \$67,000, (generously) assuming that mentoring administration accounted for 50% of their workload. This leads to total costs of about \$97,000. Together, this leads to a net present value of the program equal to approximately \$439,000, which may be a lower bound if more productive agents allow sales managers to have larger spans of control (Espinosa and Stanton, 2021).

The estimate of potential gains from reallocating agents in the Voluntary-Condition to the Mandatory-Condition is done as follows: There are 155 agents to reallocate. Assuming the full treatment effects translate, we can use the previous estimates, which allocate overhead and opportunity costs, and scale by the relative number of agents to get roughly $\frac{155}{127} \times \$439,000 = \$535,000$. We then account for the fact that overhead was already allocated to the Mandatory-Condition, yielding a gain of \$602,000. This number can be scaled by the treatment effect difference due to self-selection and treatment effect heterogeneity.

Finally, the switch from a voluntary to a mandatory program would have led to an increase in gross commissions to agents of around \$64,000 ($= \$41 \times 21 \times 6 \times 0.08 \times 155$).