Can Financial Incentives to Firms Improve Apprentice Training? Experimental Evidence from Ghana

Gabriel Brown	Morgan Hardy
University of British Columbia	New York University-Abu Dhabi

Isaac Mbiti	Jamie McCasland	Isabelle Salcher
University of Virginia	University of British Columbia	New York University

June 2023

Abstract

We use a field experiment to test whether financial incentives can improve the quality of apprenticeship training. Trainers (firm owners) in the treatment group participated in a tournament incentive scheme where they received a payment based on their apprentices' rank-order performance on a skills assessment. Trainers in the control group received a fixed payment based on their apprentices' participation in the assessment. Performance on the assessment was higher in the treatment group. Two years later, treated apprentices scored 0.15σ higher on a low-stakes oral skills test and earned 24% more in total earnings, driven by higher self-employment profits.

JEL Classifications: D22, M53, O12
Word Count: ~ 6150

1 Introduction

Apprenticeships are viewed as a promising pathway to augment the skills, productivity, and labor market outcomes of youth (Wolter and Ryan, 2011). By combining on-the-job training with practical work experience, apprenticeships can deliver market relevant skills to apprentices by harnessing the knowledge and experience of firms (or firm-owners) in a scalable and potentially cost-effective manner (ILO, 2019). While apprenticeships in developed nations generally occur in the formal sector, in developing countries they are typically conducted within informal sector small-scale enterprises with the firm owner serving as the trainer (Teal, 2016).[1] Although comparable survey data on participation in apprenticeship training across low- and middle-income countries is limited, available evidence suggests (informal) apprenticeships are the most important source of skills training for many countries in Africa and South-Asia (ILO, 2011). For example, focusing on six African countries with harmonized survey data, Filmer and Fox (2014) find that 20% of youth (aged 24-35) had participated in apprenticeships, whereas only 4% had ever attended a vocational training program. In addition to providing training, informal apprenticeships are an important and growing form of employment, especially in more urbanized settings in Sub-Saharan Africa (Teal, 2016).

Despite the importance of informal apprenticeships, there are concerns that they may not be able to deliver high-quality training due to the limited nature of monitoring, formal contracting, standardization (e.g. a syllabus), and skills certification (Filmer and Fox, 2014; Darvas and Palmer, 2014; Teal, 2016; ILO, 2019). Trainers (firm-owners) may prolong the apprenticeship duration to extract additional rents from apprentices who are typically paid much less than regular workers. Further, because apprentices tend to start their own business after completing their training by replicating their trainer's business practices (Frazer, 2006), trainers may be reluctant to exert effort to train a future potential competitor (Ackah et al., 2015).

In principle, well-designed incentive contracts could encourage trainers to exert higher effort levels toward training their apprentices, potentially increasing the skills, productivity, and livelihood of many youth in developing countries. Experimental evidence from primary school settings in developing countries shows that financial incentives for teachers can be a cost-effective policy

---

[1]See Frazer (2006) and Hardy and McCasland (2023) for detailed descriptions of informal apprenticeships in Ghana.

option to improve the quality of education (Breeding et al., 2021; Pham et al., 2021; Ganimian and Murnane, 2016; Muralidharan and Sundararaman, 2011). However, it is less clear if (modest) incentives for trainers can improve apprentices' outcomes given the potential (opportunity) costs of training and the challenges of recovering training (effort) costs.[2] In the context of skills training programs, a number of government sponsored programs in developing countries incorporate trainer incentives using outcome-based payment schemes. In these contexts, trainers payments are conditional on trainee outcomes such as training completion or verified internship or job placement.[3] However, despite the growing prevalence of outcome-based schemes in training programs, there is little empirical evidence of their effectiveness (Filmer and Fox, 2014; Adoho et al., 2014).

In this paper, we report the results of a field experiment in Ghana that provided (tournament-style) financial incentives to firms training apprentices in the context of a government-sponsored program. In the absence of the program, aspiring apprentices would be required to find a training firm and pay an upfront entrance fee. The government program waived the entrance fee and matched apprentices with trainers. Trainers participating in the program were then randomly assigned to either a treatment group or a control group. Control trainers received a fixed payment for every apprentice that participated in a trade-specific skills assessment, while treatment trainers received payments that varied based on their apprentices' rank-order performance on the same test. The average payout from both payment schemes was equalized to ensure the results were not driven by differential income effects. The design was intended to encourage both treatment and control trainers to avail their apprentices for testing approximately eight months after the start of the intervention.

While only about two-thirds of apprentices in our study ultimately took the incentivized assessment, the assessment sample is balanced along baseline observables, and neither treatment status nor other covariates associated with ability predict participation in the assessment. We augment our assessment results with two additional low-stakes skills measures from the endline survey – an oral craft skills tests and a measure of sales skills. These were collected two years

---

[2]See Acemoglu and Pischke (1999) for a discussion on the importance of recouping training costs in determining a firm's training decisions.

[3]For example, see Chakravorty et al. (2021) in India, Attanasio et al. (2011) in Colombia, Chakravarty et al. (2019) in Nepal, and Adoho et al. (2014) in Liberia.

after the assessment for over 90% of participants.

All three measures of skills reveal that apprentices in the treatment group attained more skills. Practical test scores in the assessment were 0.13σ (p-value < 0.05) higher in the treatment group. Two years after the intervention, apprentices from treated firms performed 0.15σ better on the low-stakes craft skills test and 0.12σ better on an index of sales skills. Although we observe no changes in firm owner time use, treatment firm owners are 10 percentage points more likely to report using a formal syllabus in their training.

At endline, apprentices who trained with treatment firms earn 24% more each month on average from all sources, driven by higher self-employment profits. We find no effect of treatment on the probability of any self-employment or on overall labor supply. However, apprentices in the treatment group were 20% more likely to have left the firm by the time of the endline survey.

We make three main contributions. First, we use an experiment to show that trainer incentives can improve the quality of training in informal sector apprenticeships. While there is an extensive literature on the potential for incentives to improve education and health services in developing countries, to the best of our knowledge, we are not aware of any other paper that experimentally tests the impact of financial incentives on firm training behavior.

Second, our study contributes to the literature on training within firms (Becker, 1964; Acemoglu and Pischke, 1998, 1999; Caicedo et al., 2022), where many of the papers highlight the proclivity of firms to under-invest in (general) training.[4] Our results show that relatively small incentives, averaging 3.3% of average annual firm profits in the control group, improve training quality offered by firms, which in turn increases the probability that the apprentice exits the firm (a proxy for "graduation"). Due to data limitations, we are unable to directly estimate impacts on firms during the period in which treatment firms experienced accelerated apprentice exit. However, Hardy and McCasland (2023) show that firms in this context are labor constrained and are able to capture rents from apprentice productivity. This suggests that apprentice turnover is costly not just in terms of forgone rents, but also in terms of the costs of replacing departed apprentices. The increase in apprentice separation we document here could therefore suggest that absent intervention, firms strategically slow-walk training to retain apprentices.[5]

---

[4]See Leuven (2005) for a discussion on how factors such as credit constraints among workers and the probability of worker turnover (or poaching by other firms) can lead to inefficient investment in general training.

[5]See Fudenberg and Rayo (2019) and Fudenberg et al. (2021) for theoretical insights on the length of apprenticeships.

Finally, our results contribute to the broader job (and vocational) training literature in developing countries. While some high-quality programs have large medium-term impacts on earnings (e.g. Alfonsi et al. (2020)), most rigorous evaluations of job (or vocational) training programs find limited impacts on employment and earnings, coupled with high program costs (Mckenzie, 2017; Blattman and Ralston, 2015). The small but growing literature on apprenticeships in developing countries is also mixed, but the evidence suggests that higher quality programs can improve trainee labor market outcomes (Alfonsi et al., 2020; Crepon and Premand, 2019; Hardy et al., 2019; Cho et al., 2013; Monk et al., 2008).

Our results highlight that trainer incentives can be a cost-effective approach to improve the quality of training programs in developing countries. Indeed, our incentive program pays for itself in returns to apprentices in less than two years. The results suggest that there are significant returns to improving the technical and sales (or soft) skills gained through apprenticeships. However, consistent with Frazer (2006), these returns accrue in self-employment rather than in wage employment.

2 Study Design

2.1 Study Background

This study was conducted in collaboration with the Council for Technical and Vocational Education and Training (COTVET), a Ghanaian government agency that develops technical and vocational training policy and programs, and works to improve the informal apprenticeship sector. COTVET recognized that informal apprenticeships suffer from misaligned incentives between apprentices and trainers. They were concerned that trainers slow-walk skill development and retain workers at low wages for excessive periods of time, with most apprenticeships lasting about three years. COTVET was keen on implementing reforms to accelerate the pace of training.

Our partnership with COTVET yielded a nested set of three experiments, embedded in the implementation of the National Apprenticeship Program (NAP)—a nationwide program that aimed to place unemployed youths into apprenticeships with firms (trainers) in the informal sector. From August 2012 to January 2013, a pool of NAP applicants were randomly selected to

5

participate in the program, on a district by district basis.[6] These selected apprentices were then (conditionally) randomly assigned to firms and commenced training between October 2013 and January 2014.[7] The sample for the present study is composed of apprentices that were training in NAP firms at the time of the intervention in January of 2015, about a year after training began.[8]

Hardy and McCasland (2023) and Hardy et al. (2019) provide a detailed discussion of the external validity of the sample. Hardy et al. (2019) show the apprentices recruited under the NAP were observably similar to traditional (non-NAP) apprentices along all dimensions (including cognitive ability and schooling) except that NAP apprentices were poorer. They also find no differences between the self-reported training experiences of NAP apprentices and traditional apprentices. Hardy and McCasland (2023) show that firms (trainers) in the NAP program were broadly representative of firms that typically train traditional apprentices in Ghana.

2.2 Sample Summary Statistics

The NAP program included five trades, three in skilled construction, cosmetology, and garment-making. Due to the unavailability of updated training curricula for the construction trades at the time of the intervention, the incentive program implementation was restricted to garment-making and cosmetology. These were also the two most popular trades, accounting for almost 80% of NAP applicants at baseline (Hardy et al., 2019). This restriction yields a sample of 797 apprentices across 467 training firms at the start of the intervention. Due to occupational gender segregation, 94% of the apprentice sample is female. Consequently, we report results for both the full sample and also for the sub-sample of female apprentices consisting of 751 apprentices training at 446 firms. Due to the small sample size, we cannot report results for the sub-sample of males.

The sample comes from 32 districts around Ghana. Appendix Table A1 reports baseline (in 2012) characteristics of apprentices. Apprentices were 23 years old, with 7.5 years of schooling. Roughly one-third were married and 48% had at least one child. 22% of apprentices in the sample were self-employed and just 4% were wage-employed at baseline.

---

[6]See Hardy et al. (2019) for a short-run analysis of returns to apprenticeship training.

[7]Hardy and McCasland (2023) examines the firm-level impacts of access to the apprentice placement program, focusing on the first year of training (2014) which predates the incentives scheme and any payments to firms.

[8]In December 2014 just before the roll-out of this study, in an effort to foster ongoing goodwill and trust, the research team provided all trainers (including those not in this study) with 100 (GhC) per apprentice. Payment was not conditioned on participation in the incentives scheme presented in this study and is not directly related to the analysis presented in this paper.

2.3 Experimental Design

Trainers (firms) in the experiment were randomized into treatment and control, stratified by district, trade, and a measure of the number of NAP apprentices at the firm.[9] These three dimensions generate 128 strata, although 21 cells were empty so our study has 107 strata in total. The intervention roll-out was conducted one-on-one and in person. A member of the research team explained the program (either the treatment or control incentive structure), provided the firm owner with a copy of COTVET's occupational standards handbook (the syllabus used to design the incentivized tests), and officially enrolled apprentices in the program.

Control firm owners would receive a flat payment of 100 GhC for each apprentice who participated in the skills test. Treatment firm owners would receive payments based on the decile ranking of their apprentices' performance within their trade and region.[10] Apprentices in the lowest decile earned their firm owners 25 GhC, which increased to 200 GhC in the top decile. The payment schedule is presented in Appendix Table A2. In addition, across every region and trade, the trainer of the highest-performing treatment apprentice received an additional 250 GhC "bonus" (on top of their regular payout), in an effort to encourage effort even among the highest performers. Both in expectation and ex-post, treatment firm payouts averaged 100 GhC.[11] Because the average payouts are equalized across groups, we mitigate concerns that outcomes could be driven by income effects. The skills test was implemented a little more than eight months after the intervention roll-out, in August and September of 2015. Incentive payments were made in October of 2015.

2.4 Data

Baseline data for apprentices and firms was collected in 2012 and 2013, respectively, prior to randomization. Instruction hours, syllabus use, wages, and profits during the intervention period

---

[9]Randomizations were conducted separately for the South (24 sample districts) and the North of Ghana (8 sample districts). In the South, where the average apprentice in our sample trained with 1.3 other NAP apprentices, firms were split into those training a single NAP apprentice and those training more than one NAP apprentice. In the North, where the average apprentice in our sample trained with 1.6 other NAP apprentices, firms were split into those training one or two NAP apprentices and those training more than two NAP apprentices. This stratification ensured that we equalized the numbers of apprentices in each experimental group.

[10]Our 32 districts, by design, have representation across all 10 regions as of 2015, generating 20 performance scheme groups. Group size ranges from 10 to 100; mean and median group size is 40 apprentices.

[11]Median payouts were also close to 100 GhC (averaging the fifth and sixth deciles).

were collected via a firm midline survey in June 2015, before the assessment.[12] All monetary outcomes for both apprentices and firms are inflated or deflated to October 2015, when firm owners received their payments from the intervention.

The assessment consisted of a theoretical component, administered verbally by an assessor, and a practical component, which required apprentices to physically demonstrate a series of skills, including a role play with a model client. Both sections were conducted at the firm and scored by the assessor, an expert in the craft. The assessment was developed in collaboration with a team of experts, and conformed to the national occupational standards developed by COTVET. The assessment and scoring protocols were tested and refined in a pilot district. We therefore drop the pilot district from the analysis of the assessment below. Because COTVET mandated that all the exams had to be anchored to the national standards, the theoretical component of the exam contained a large share of questions that are unlikely to measure remunerable craft skills.[13]

The practical component of the assessment, on the other hand, tested core competencies in each craft. Garment-makers were required to demonstrate several different types of stitches and seams, and asked to measure, cut, and sew a garment for a human model. Cosmetologists could choose to either shampoo and braid a human model or to relax, roller-set, and blowout the model's hair. Assessors scored each step, taking the quality of work into account. Apprentices who passed both the practical and theoretical tests earned an official skills certification.

We use the scoring schemes that determined whether or not an apprentice passed the exam as our measure of performance on the assessment.[14] From the theoretical section, these scoring schemes dropped questions for which more than 25% of the variation was explained by the assessor. The theoretical score is then the sum of the remaining correct answers. For the practical section, the certificate scoring scheme up-weighted the most technical components of the exam (like the quality of the final garment) and down-weighted less technical components (like a score for greeting the client in a professional manner).[15]

The endline survey, conducted from July 2017 to May 2018, provides data on long-term skills

---

[12]See Figure S1 for a timeline.

[13]For example, (1) *State three reasons why it is important to create an organized and tidy workspace.*, (2) *State three duties involved in tidying the workplace. For each duty, explain how regularly you should complete it and why?*, (3) *What are the broader societal benefits of the hairdressing and beauty industry in Ghana?*, and (4) *Name three items that should be found in a first aid box.*

[14]The scoring schemes were developed in conjunction with COTVET.

[15]Results using Item Response Theory (IRT) models are available upon request.

outcomes, earnings, labor supply, apprenticeship completion, and exit from the firm (turnover). Over 90% of the sample was interviewed during this survey. Due to data limitations, we do not have firm-level outcomes past 2015, and therefore use apprentice turnover to generate an estimate of the indirect costs of the program to firm owners. The craft skills quiz was developed by the research team, in consultation with local experts, and focused on core competencies in each craft. It was an oral quiz and did not require practical demonstration of skills, but we believe it to be a high-quality measure of craft competencies as it significantly predicts higher earnings cross-sectionally in the endline survey and is positively correlated with practical test performance (see Appendix Table A3).

We also present results on a measure of sales skills, composed of two indices. First, we asked the respondent to discuss skills training in Ghana, and the surveyor rated the respondent on the audibility, animation, engagement, and coherency of the response. Second, we asked respondents to demonstrate how they would sell a pen to a surveyor. Respondents were scored on standard sales strategies such as describing the attributes of the pen, describing how others like the pen, and using humor to engage the enumerator. We take the first principal component of all ratings from both sections. We believe this index can be interpreted as a measure of soft skills.

The key labor market outcomes come from a retrospective earnings module that asked respondents to relay their wage-employment, self-employment, farming, and apprenticeship earnings (and labor supply) 1 month, 4 months, 7 months, and 10 months prior to the endline interview. Our measure of total earnings is the sum across all four earned income sources.[16] Profits in self-employment follow self-reported question structures from De Mel et al. (2009). We stack the four retrospective observations in all earnings and labor supply specifications. Data on apprenticeship completion and exit from the firm are self-reported.

2.5 Attrition and Balance Tests

Appendix Tables A4 and A5 present apprentice-level balance tests for the two samples that generate variation in our outcome analysis: those apprentices with a follow-up survey and those who participated in the assessment. In both tables, 19 of 20 covariates suggest a balanced sample

---

[16]Wage- and self-employment earnings are about 87% of total control group earnings.

and F-tests of the joint significance of all 20 covariates fail to reject joint orthogonality.[17] Only our measure of household assets, the first principal component of a set of housing quality and household asset measures, is imbalanced. Because the baseline household asset index predicts some follow-up outcomes of interest (namely earnings and apprentice turnover), we control for it (and a dummy for cases in which it is missing) in all apprentice-level specifications. Baseline controls also include an indicator variable for apprentice gender in the full sample specifications, as this covariate is imbalanced in the full sample (see Appendix Table A1). As a robustness check, we report results excluding these controls in Appendix Tables A21- A25.

We test for firm-level balance along baseline firm characteristics in Tables A9 and A10. We find that treatment and control groups are balanced along (baseline) firm-level characteristics such as profits, firm size, firm owner cognitive ability, firm owner experience training apprentices, and several self-reported measures of firm owner instruction.

Appendix Table A11 presents information on attrition for each of the five data sources. We find that attrition rates are balanced across treatment and control groups in each piece of data we exploit in the paper. Columns (1) and (2) of Table 1 show that about 65% of apprentices in both the treatment and the control group participated in the skills assessment, in both the full sample and the sample excluding the pilot district. Appendix Table A12 shows that there are no statistically significant apprentice-level correlates of participation in the assessment, either in the full sample or differentially by treatment. Appendix Table A13 conducts the same analysis using firm-level characteristics and finds some suggestive evidence that firm owners with more experience training apprentices were more likely to avail their apprentices for testing. This effect, however, does not vary by treatment status.

3 Results

3.1 Skill Development

We present results on skills outcomes in Table 1. Intent-To-Treat (ITT) incentivized test score outcomes are estimated using the following least squares regression:

---

[17]Appendix Tables A6-A8 present analogous balance tests for the female apprentice sample, with similar conclusions.

$$Y_i = \beta_0 + \beta_1 * T_i + \gamma_s + \delta_a + \alpha_b * X_i + u_i \tag{1}$$

where γ_s are strata dummies, δ_a are assessor dummies (to control for more and less generous grading), X_i are controls for apprentice gender, the baseline household asset index, and a dummy for cases in which that variable is missing, T_i is a treatment indicator, and u_i is an error term. Skills outcomes measured in the endline survey are estimated using the same specification, with the exception that the assessor fixed effects are removed. Our sales skills measure was graded by surveyors, so specifications on sales skills are estimated including surveyor fixed effects. Errors are clustered at the firm level. We report results for the full sample (Panel A) and also the sample of females (Panel B). The sample size of men is too small to examine a male sub-sample.

The estimated treatment effects on the practical component of the assessment are 0.13σ in the whole sample and 0.18σ for females (Column 3). We find no statistically significant effect of the performance-based incentives on test performance in the theoretical component of the assessment (Column 4), likely due to its content. COTVET requested relatively lenient pass thresholds, which were collaboratively set at 50% for the garments practical, 45% for the cosmetology practical, and 50% for the theoretical exam, yielding pass rates of 80% in the control group.[18] The observed improvement in performance on the practical exam is associated with a noisy 5 percentage point increase in the probability of passing that exam and earning a certificate (Panel A, Column (5)). The probability of passing the practical exam and earning a certificate is more precisely estimated among the female sample (Panel B, Column (5)).[19]

Columns (6) through (9) present results on our measures of craft skills and sales skills collected at endline, in the full sample and the sub-sample of apprentices that both took the incentivized exam and participated in the endline survey. The treatment increased craft skills by 0.15σ (Column 6) in this low-stakes exam. These results help validate the results from the assessment, allowing us to rule out potential gaming and concerns about attrition in the assessment sample.[20] The treatment also increased sales skills by 0.12σ. The improvements in verbal and soft skills

[18]Extremely strict marking of exams paired with relatively lenient pass thresholds is culturally common in Ghana, as in many low- and middle-income countries.

[19]Although there were a few test-takers who failed the theoretical portion of the exam, the theory section was never binding; test-takers who failed the theory section also failed the practical section.

[20]We observe craft skills at endline for 93% of the analysis sample.

associated with this measure could be the result of direct sales skill training or indirectly through improvements in confidence or experience with self-employment. Point estimates for both craft skills and sales skills in the incentivized test sub-sample are noisier but of a similar magnitude as the full sample estimates.

3.2 Labor Market Outcomes

In Table 2, we present our second set of main outcomes. Earnings and labor supply outcomes at endline are a stacked panel of four retrospective rounds, estimated with the following specification:

$$Y_{it} = \beta_0 + \beta_1 * T_i + \gamma_s + \alpha_b * X_i + e_{it} \tag{2}$$

where γ_s are strata dummies, X_i are controls for apprentice gender, the baseline household asset index, and a dummy for cases in which that variable is missing, T_i is a treatment indicator, and e_{it} is an error term. We focus on ITT estimates because LATE (2SLS) estimates would require very strong assumptions about the exclusion restriction. Errors are clustered at the firm level. Only a small fraction of the apprentice sample had any positive earnings at baseline and earnings data were measured differently. We therefore exclude baseline values of the dependent variable from our outcomes analysis. Since the retrospective panel elicits labor market outcomes 1, 4, 7, and 10 months prior to the endline survey that was collected between July 2017 and May 2018, β_1 is an average of the treatment effects five months prior to the survey date, and about two years from the time of the assessment.[21]

ITT treatment effects on earnings are estimated at 10.43 GhC per month (Panel A, Column 1). This is a meaningful increase as it translates to a 24% increase in earnings relative to the control group mean. Earnings effects are driven by higher self-employment profits—the treatment increased unconditional self-employment profits by 8.41 GhC, a 35% increase compared to the control group mean (Column (2)), and conditional profits by a noisy 15.15 GhC (Column (4)). We find no treatment effects on wage-employment earnings (Columns (3) and (5)).

In terms of labor supply, we do not find any significant changes in the probability of working in any sector, working in self-employment, or working for a wage (Appendix Table A14). However,

---

[21]Recall, all nominal earnings in our data are converted to October 2015 GhC real values.

we do see a small significant decrease in unskilled self-employment with a corresponding small significant increase in skilled self-employment. Following Attanasio et al. (2011), we can generate bounds that suggest that our results are driven by the productivity effects from increased skill levels among the treated group rather than by a composition effect reflecting changes in the sector of work.[22]

3.3 Heterogeneity

An important concern is that "weaker" participants may be discouraged from exerting effort if they believe they are competing with "stronger" participants (Dechenaux et al., 2015; Charness and Kuhn, 2011). We test for this possibility by examining the heterogeneity in treatment effects by (baseline) trainer and apprentice characteristics such as cognitive ability. Overall, we do not find evidence consistent with discouragement. We also do not find (strong) evidence of differential treatment effects by baseline trainer or apprentice ability (see Appendix Tables A15-A20).

An additional concern is that the effects might be driven by the gains of the best performing apprentices. Using quantile regressions, we find that the gains from the intervention were broad-based and not driven by the best performers (see Appendix Figures S2 and S3).[23]

Finally, in Appendix Figures S4 and S5 we show treatment effects on earnings in each of the four retrospective rounds in the endline survey. We find that point estimates are quite similar across the retrospective rounds and never statistically different from each other. This finding suggests that earnings gains are relatively stable in the range of 1.5 to 2.5 years after the assessment.

4 Mechanisms

4.1 Training Inputs and Intermediate Financial Outcomes

Using data from the firm midline survey that was conducted with firm owners after the intervention was launched, but prior to the assessments, we explore shifts in the training environment in Table 3. At the firm level, we find no evidence of a shift in firm owner hours worked last week or

---

[22]See Appendix B for details.
[23]The exception is sales skills, where we see larger effects at higher quantiles.

firm owner time spent yesterday on instruction (Columns 1 and 2). At the apprentice level, we find no evidence of a shift in firm owner instructional time, but we do find that apprentices in treated firms worked more hours last week, about an 8% increase in hours worked, potentially allowing for more learning by doing or learning by observing (Columns 3 and 4).[24] Despite the estimated increase in apprentice hours and the measured improvement in apprentice skills, point estimates on contemporaneous profits and wages paid to apprentices during the intervention period (Columns 8 and 9) are noisy and not statistically different from zero.

Informal apprenticeships are often criticized for lacking standards or formal syllabi. While the program provided trainers with a physical copy of the COTVET occupational standards handbook, it was unclear if they would be used. In Columns (5) through (7) we show that treatment increased the self-reported use of any formal syllabus, driven about evenly by the use of the COTVET occupational standards syllabus and syllabi previously supplied by trade associations (some firm owners reported using both). The increased adherence to a formal syllabus could potentially improve the quality of training.

4.2 Completion

Although training completion and promotion are sometimes fluid concepts in this setting (39% of certificate-holders self-report non-completion) and retrospective data on completion and exit timing in the endline survey naturally generate some data quality issues, treatment increased completion rates. In Table 4, we show that apprentices from treated firms were 22% (7 percentage points) more likely to self-report completion after the start of the program in 2015 (Column (1)) and 67% (10 percentage points) more likely to self-report paying a non-zero ceremony or exit fee to formally celebrate the completion of the apprenticeship (Column (2)).[25] Treatment increased the probability of completion and exit, while having no impact on the probability of completion with retention (Column (3) and (4)). Apprentices at treated firms are 20% (8 percentage points) more likely to exit the firm in the two years after the assessment, as reported by apprentices in the endline survey and presented in Column (5). Earnings gains may accrue in part due to a higher probability of completion.

---

[24]Appendix Table A28 shows that we also find no evidence of changes in other apprentice time use categories collected, including instruction time from other workers.

[25]These ceremonies sometimes are just parties financed by the trainee. Other times they have traditional elements as well.

4.3 Is Quality Training Costly to Firms?

Although we do not find any reductions in firm profits during the intervention period (Table 3), we do find that treatment apprentices are 20% more likely to exit the firm in the medium term (Table 4), generating a potentially salient cost to firms of providing high-quality training. Due to data limitations (we do not have firm-level outcomes past 2015), we cannot directly test whether the program impacted medium-term profit outcomes. However, using insights from a sister experiment, we argue that turnover is costly.

Hardy and McCasland (2023) show that firms in this context experience an increase in both firm size and profits when experimentally exposed to a worker placement program that gave firms access to additional apprentices. In other words, (i) firms in this context are labor constrained and may not be able to readily replace apprentices who exit and (ii) firms in this context profit from having access to additional apprentices. Together, these pieces of evidence suggest that accelerated apprentice exit is costly to firms. In Column (6) of Table 4, we estimate the number of months that firm owners missed out on due to accelerated apprentice exit, based on the apprentice's self-reported month and year of exit from the firm (if any) and the timing of the endline survey, assuming zero missed months for apprentices still at the firm. On average, we estimate apprentices exit 1.5 months earlier in the treatment group than the control group. Hardy and McCasland (2023) estimate that firm profits increase by 11% per apprentice. Applying 11% to the profit measures in the firm midline survey in Table 3 (251 GhC per month in the control group), we estimate a loss due to exit of 27.6 GhC per month. Multiplied by 1.5 months, our back-of-the-envelope calculation suggests firm owners lost a little over 40 GhC in the medium-term due to higher apprentice turnover rates. It is unclear if trainers fully anticipated these costs. However, it is possible they would learn to better account for these costs with repeated exposure to incentive programs.[26]

5 Analysis of Program Costs and Benefits

The largest costs associated with the implementation of this demonstration program were (1) the cost of the financial incentives, and (2) the cost of developing and administering the assessment.

---

[26]It is also possible that they anticipated these costs but overestimated their incentive payouts but would learn to better predict the payouts with more experience.

Auxiliary costs such as program enrollment and certificate distribution were subsumed into the larger administrative and survey costs associated with the research study and ongoing administrative follow-up through COTVET.

We make some simplifying assumptions in our analysis of costs. First, although financial incentives were only actually paid out for the 65% of the sample who participated in the exam, we assume an incentive cost of 100 GhC per trainee for 100% of the trainees in the program. Second, we exclude the cost of developing the exam, since it is a fixed cost and instead focus on variable costs per apprentice. Third, we treat the auxiliary administrative costs mentioned above as fixed costs and also exclude them from our analysis.[27]

After these restrictions and simplifications, we are left with 100 GhC per apprentice in financial incentive costs and the variable costs associated with testing. We spent about 40,000 GhC on the testing program, with assessor salaries, per diems, and travel as the largest costs. Allocated on a per apprentice basis across all 797 apprentices at the start of this study, this amounts to about 50 GhC per apprentice, yielding a total cost per apprentice of roughly 150 GhC. Taking monthly returns from Table 2, an average across all 748 apprentices for whom we have endline data, the 12-month return to the program is estimated at about 125 GhC, easily paying for itself in returns to apprentice in less than two years, using any reasonable discount rate. Although these calculations are quite rough, they suggest a highly cost-effective program in terms of earnings outcomes.

6 Conclusion

Skills-upgrading through apprenticeship training offers a potentially low-cost, high-access alternative to institutional vocational training. However, the quality of training provided by these programs could be low due to lack of formalization and monitoring. In this paper, we show that trainer incentives can encourage firms to invest more in training, which can yield large earnings gains likely associated with substantial welfare improvements in people's lives.

We explore potential determinants of trainer choice to provide quality training and find little empirical evidence supporting immediate training cost considerations, with no detectable impacts

---

[27]Note that we also exclude the flat cash payment that was paid to firm owners immediately prior to the program in December 2014, since it was largely the result of political issues unrelated to the performance-based incentive and testing scheme.

on (firm- owner) training time, apprentice wages, nor firm productivity during training. Instead, we find that apprentices increased their labor supply in response to the incentives and trainers were more likely to consult a formal syllabus during training. We also find a significant increase in apprentice turnover, echoing existing literature from high-income country contexts that suggests firms strategically withhold training of general skills and are more willing to provide training in monopsonistic labor markets (Acemoglu and Pischke, 1998). Because our study is not repeated over multiple time periods, we cannot examine changes in apprentice and trainer responses (and outcomes) that might occur as trainers gain experience with incentives schemes. Further, because of the relatively small scale of our experiment, it is unclear if our main results would change if the program was scaled up. These are two promising areas for future research.

References

Acemoglu, Daron, and Jörn-Steffen Pischke. 1998. "Why Do Firms Train? Theory and Evidence." *Quarterly Journal of Economics*.

Acemoglu, Daron, and Jörn-Steffen Pischke. 1999. "Beyond Becker: Training in Imperfect Labour Markets." *The Economic Journal*.

Ackah, Charles, Abena Oduro, and Richard Boateng. 2015. "Gender and Enterprise Development in Africa: The Case of Ghana."

Adoho, Franck, Dala T. Chakravarty, Shubha, Mattias Lundberg, and Afia Tasneem. 2014. "The impact of an adolescent girls employment program : the EPAG project in Liberia."

Alfonsi, Livia, Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali. 2020. "Tackling Youth Unemployment: Evidence from a Labor Market Experiment in Uganda." *Econometrica*.

Attanasio, Orazio, Adriana Kugler, and Costas Meghir. 2011. "Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics*.

Becker, Gary. 1964. "Human Capital."

Blattman, Christopher, and Laura Ralston. 2015. "Generating employment in poor and fragile states: Evidence from labor market and entrepreneurship programs." *Available at SSRN 2622220*.

Breeding, Mary E., Tara Beteille, and David K. Evans. 2021. "Teacher Pay-for-Performance: What Works? Where? and How?." `http://documents.worldbank.org/curated/en/183331619587678000/Teacher-Pay-for-Performance-What-Works-Where-and-How`.

Caicedo, Santiago, Miguel Espinosa, and Arthur Seibold. 2022. "Unwilling to Train?—Firm Responses to the Colombian Apprenticeship Regulation." *Econometrica* 90 (2): 507–550.

Chakravarty, Shubha, Mattias Lundberg, Plamen Nikolov, and Juliane Zenker. 2019. "Vocational training programs and youth labor market outcomes: Evidence from Nepal." *Journal of Development Economics* 136 71–110.

Chakravorty, Bhaskar, Wiji Arulampalam, Clement Imbert, and Roland Rathelot. 2021. "Can information about jobs improve the effectiveness of vocational training? Experimental evidence from India." *Working Paper*.

Charness, Gary, and Peter Kuhn. 2011. "Lab labor: What can labor economists learn from the lab?" In *Handbook of labor economics*, Volume 4. 229–330, Elsevier.

Cho, Yoonyoung, Davie Kalomba, Ahmed Mushfiq Mobarak, and Victor Orozco-Olvera. 2013. "Gender differences in the effects of vocational training: Constraints on women and drop-out behavior." *World Bank Policy Research Working Paper* (6545): .

Crepon, Bruno, and Patrick Premand. 2019. "Direct and Indirect Effects of Subsidized Dual Apprenticeships." *Working Paper*.

Darvas, Peter, and Robert Palmer. 2014. "Demand and Supply of Skills in Ghana: How Can Training Programs Improve Employment and Productivity?".

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2009. "Measuring Microenterprise Profits: Must We Ask How The Sausage Is Made?" *Journal of Development Economics* 88, pages 19-31.

Dechenaux, Emmanuel, Dan Kovenock, and Roman M Sheremeta. 2015. "A survey of experimental research on contests, all-pay auctions and tournaments." *Experimental Economics* 18 (4): 609–669.

Filmer, Deon, and Louise Fox. 2014. "Youth Employment in Sub-Saharan Africa."

Frazer, Garth. 2006. "Learning the Master's Trade: Apprenticeship and Human Capital in Ghana." *Journal of Development Economics* 81 259–298.

Fudenberg, Drew, George Georgiadis, and Luis Rayo. 2021. "Working to learn." *Journal of Economic Theory* 197 105347.

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

Ganimian, Alejandro J, and Richard J Murnane. 2016. "Improving education in developing countries: Lessons from rigorous impact evaluations." *Review of Educational Research* 86 (3): 719–755.

Hardy, Morgan, Isaac Mbiti, Jamie McCasland, and Isabelle Salcher. 2019. "The Apprenticeship-to-Work Transition." *World Bank Policy Research Working Paper 8851*.

Hardy, Morgan, and Jamie McCasland. 2023. "Are Small Firms Labor Constrained? Experimental Evidence from Ghana." *American Economic Journal: Applied Economics*.

ILO. 2011. "Upgrading informal apprenticeship systems."

ILO. 2019. "Quality Apprenticeships: Addressing skills mismatch and youth unemployment."

Leuven, Edwin. 2005. "The economics of private sector training: a survey of the literature." *Journal of economic surveys* 19 (1): 91–111.

Mckenzie, David. 2017. "How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence." *World Bank Research Observer*.

Monk, Courtney, Justin Sandefur, and Francis Teal. 2008. "Does Doing an Apprenticeship Pay Off? Evidence from Ghana." *CSAE Working Paper Series 2008-08* Centre for the Study of African Economies, University of Oxford.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): pp. 39–77, `http://www.jstor.org/stable/10.1086/659655`.

Pham, Lam D, Tuan D Nguyen, and Matthew G Springer. 2021. "Teacher merit pay: A meta-analysis." *American Educational Research Journal* 58 (3): 527–566.

Teal, Francis. 2016. "Are apprenticeships beneficial in sub-saharan africa?" *IZA World of Labor*.

Wolter, Stefan C, and Paul Ryan. 2011. "Apprenticeship." In *Handbook of the Economics of Education*, Volume 3. 521–576, Elsevier.

Table 1: Skills

	Assessment					Endline survey			
	Took incentivized test (0/1)	Took incentivized test (0/1)	Practical component (z-score)	Theoretical component (z-score)	Earned certificate (0/1)	Craft skills quiz (z-score)	Craft skills quiz (z-score)	Sales skills (z-score)	Sales skills (z-score)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Full sample									
Treatment	-0.01	-0.01	0.13*	-0.08	0.05	0.15**	0.14	0.12**	0.10
	(0.04)	(0.04)	(0.07)	(0.07)	(0.04)	(0.07)	(0.09)	(0.06)	(0.07)
Observations	797	763	488	488	488	743	466	748	468
Mean of Dep Variable T=0	0.65	0.65	0.00	0.00	0.80	0.00	0.05	0.00	-0.11
Incentivized Test Sample						Yes	Yes		Yes
Panel B: Women only									
Treatment	-0.00	-0.01	0.18**	-0.05	0.07*	0.15**	0.13	0.13**	0.12
	(0.04)	(0.04)	(0.08)	(0.07)	(0.04)	(0.07)	(0.09)	(0.06)	(0.08)
Observations	751	717	457	457	457	703	439	706	439
Mean of Dep Variable T=0	0.64	0.64	0.00	0.00	0.81	0.00	0.06	0.00	-0.09
Incentivized Test Sample						Yes	Yes		Yes

Notes: This table shows the results from estimating equation 1 for different skill measures. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Full sample specifications include a control for apprentice gender. Specifications from the assessment include assessor dummies. Sales skills specifications include surveyor fixed effects. Columns 2 through 5, 7, and 9 exclude test-takers from the first district, because the assessment (being implemented in collaboration with our government partners) differed from the test given in all other districts. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Earnings

	Unconditional			Conditional	
	Total earnings (GhC)	Self-employment profits (GhC)	Wage employment earnings (GhC)	Self-employment profits (GhC)	Wage employment earnings (GhC)
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample					
Treatment	10.43*	8.41*	3.41	15.15	-4.10
	(5.56)	(4.70)	(3.40)	(12.23)	(19.55)
Observations	2,992	2,992	2,992	869	301
Mean of Dep Variable T=0	44.34	24.18	12.98	83.58	136.89
Panel B: Women only					
Treatment	12.89**	8.93*	3.02	22.22*	0.21
	(5.46)	(4.75)	(3.57)	(12.67)	(21.40)
Observations	2,824	2,824	2,824	806	287
Mean of Dep Variable T=0	41.52	22.99	13.10	80.62	135.34

Notes: This table shows the results from estimating equation 2 for different labor market outcomes. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Full sample specifications include a control for apprentice gender. Outcomes are stacked across a four-round retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Total earnings includes earnings from wage-employment, self-employment, farming, and apprenticeship. Profits in self-employment follow self-reported question structures from De Mel et al. (2009). All earnings are in October 2015 Ghana Cedis. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Impacts During the Intervention

	Firm level labor inputs		Apprentice level labor inputs		Firm level pedagogy			Firm Midline Survey Financial outcomes	
	Firm owner hours last week	Firm owner instruction hours per day	Apprentice hours last week	Firm owner instruction hours per day	Any formal syllabus	Occupational standards syllabus	Trade association syllabus	Firm profits last month	Apprentice wages last month
	(1)	(2)	(3)	(4)	(0/1) (5)	(0/1) (6)	(0/1) (7)	(GhC) (8)	(GhC) (9)
Panel A: Full sample									
Treatment	-1.24	-0.09	4.13**	0.10	0.10*	0.05	0.07	12.60	1.55
	(1.85)	(0.21)	(1.91)	(0.09)	(0.06)	(0.05)	(0.05)	(31.11)	(3.08)
Observations	339	341	565	595	341	341	341	335	566
Mean of Dep Variable T=0	52.33	3.42	44.18	1.29	0.41	0.23	0.23	251.39	22.17
Panel B: Women only									
Treatment	-0.74	-0.04	4.24**	0.08	0.10*	0.06	0.07	9.64	1.63
	(1.76)	(0.22)	(2.01)	(0.09)	(0.06)	(0.05)	(0.05)	(32.72)	(3.28)
Observations	323	325	527	556	325	325	325	319	528
Mean of Dep Variable T=0	52.46	3.33	43.73	1.32	0.41	0.24	0.22	237.20	21.85

Notes: All specifications include strata fixed effects. Specifications in Columns (3), (4) and (9) include imbalanced apprentice-level controls for the baseline household asset index of the apprentice (including a dummy for where the variable is missing), with full sample specifications in these columns including a control for apprentice gender. The results on instructional time in Column (2) make the conservative assumption that, in firms with multiple NAP trainees, no instruction time was simultaneous (summing the instruction time reported for all NAP trainees in a firm). Firm profits follow self-reported question structures from De Mel et al. (2009). Wages for each apprentice are reported by the firm owner. All financial outcomes are in October 2015 Ghana Cedis. Standard errors in Columns (4) and (5) are clustered at the firm level. $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$

Table 4: Apprentice Completion and Exit

	Apprentice completion				Apprentice turnover	
	Completed	Paid exit/ceremony	Completed and exited	Completed and retained	Exited	Missed months
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
Treatment	0.07**	0.10***	0.08**	-0.01	0.08**	1.50*
	(0.04)	(0.03)	(0.04)	(0.01)	(0.04)	(0.84)
Observations	748	748	748	748	748	748
Mean of Dep Variable T=0	0.31	0.15	0.28	0.03	0.39	7.73
Panel B: Women only						
Treatment	0.07*	0.11***	0.08**	-0.00	0.08**	1.63*
	(0.04)	(0.03)	(0.04)	(0.01)	(0.04)	(0.86)
Observations	706	706	706	706	706	706
Mean of Dep Variable T=0	0.32	0.16	0.30	0.02	0.40	7.99

Notes: All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Full sample specifications include a control for apprentice gender. Exited in Column (5) includes those who exit and complete and those who self-report exiting without completing. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

24

A Online Appendix

Table A1: Balance: Apprentice Characteristics (Full Sample—Men and Women)

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	797	0.56	0.54		
North	797	0.34	0.35		
Program apprentices (incl. respondent)	797	2.51	2.45	-0.18	(0.17)
Female	797	0.92	0.96	0.04	(0.10)*
Age	774	22.87	22.96	0.10	(0.81)
Married (0/1)	772	0.37	0.34	-0.03	(0.38)
Number of children	781	0.74	0.79	-0.01	(0.85)
Lives with parent(s) (0/1)	778	0.49	0.45	-0.01	(0.81)
Household asset score (z-score)	766	0.00	0.16	0.13	(0.04)**
Years of schooling	760	7.50	7.38	-0.17	(0.44)
Ability index (z-score)	780	0.00	0.01	0.01	(0.94)
Soft skills index (z-score)	778	0.00	0.08	0.09	(0.19)
Food security (z-score)	776	0.00	-0.03	-0.06	(0.40)
Self-reported health (z-score)	779	0.00	0.01	-0.01	(0.94)
Mother years of schooling	647	3.28	3.53	0.32	(0.35)
Father years of schooling	567	5.55	5.66	0.08	(0.86)
Prior apprenticeship experience (0/1)	780	0.26	0.27	0.01	(0.74)
Wage-employed (0/1)	781	0.04	0.04	0.00	(0.92)
Wage-employment earnings (GhC)	781	1.44	1.39	0.15	(0.82)
Self-employed (0/1)	781	0.22	0.18	-0.02	(0.59)
Self-employment profits (GhC)	781	11.74	10.59	0.36	(0.90)
Total weekly hours worked	772	18.54	16.61	-0.70	(0.72)
Desires self-employment (0/1)	781	0.55	0.52	-0.04	(0.22)

Notes: *North* indicates that the apprentice lives in the the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas of 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and program apprentices (including respondent) to be balanced by construction. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Errors are clustered at the firm level. 781 of 797 apprentices participated in the baseline survey between August and December of 2012. This table includes all available data from the baseline survey. *Household Asset* is the first principal component of a set of housing quality (floor material, roof material, wall material, sanitation access, water source, lighting source, primary cooking implement, and number of people who sleep in the same room) and asset measures (mattress, tv, radio, refrigerator, car, motorbike, bicycle, working mobile phones, and shoes). *Ability Index* is the normalized sum of the normalized scores on each of four cognitive tests: Digit Span Recall, four math questions we developed ourselves, Ravens Matrices Group B, and a fifteen word oral English vocabulary recognition test. *Soft Skills Index* is the normalized sum of the normalized scores on each of two (adapted to context) non-cognitive tests: the Rosenberg Self-Esteem scale and the Rotter Locus of Control scale. *Food Security* is the first principal component of two food consumption measures: meals eaten per day and meals per day that included meat (excluding eggs). *Self-Reported Health* is the normalized response to *All in all, how would you describe your state of health these days?* on a four-point Likert scale. The top 0.5% of wage job earnings and self-employment profits have been winsorized; these variables are measured in October 2015 Ghana Cedis. *Desires Self-Employment* is an indicator variable for the respondent reporting that the reason they are interested in apprenticeship training is that it will lead to self-employment. An F test of the joint significance of all 20 covariates (from female to desires self-employment) in predicting treatment yields of p-value of 0.39. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Pay for Performance Scheme

Group (Decile)	Payment to MCP (Ghana Cedis)
1	200
2	140
3	130
4	120
5	105
6	90
7	80
8	70
9	35
10	25

Notes: The above lists the payouts by apprentice performance decile for each apprentice in the treatment group. In addition, firm owners in the treatment group whose apprentice performed best in the region and craft earned their trainer 250 GhC. Firm owners in the control group received 100 GhC for every apprentice who took the test regardless of performance. Payouts for both control and treatment firms averaged 100 GhC per apprentice, both in expectation and ex-post.

Table A3: Correlation between practical test and endline measures

	Full sample	Women only
	(1)	(2)
Craft skills quiz (z-score)	0.10**	0.10**
	(0.04)	(0.04)
Observations	466	439

Notes: This Table shows the results from estimating the correlation between the practical test and one of the endline measures: Craft skills. The estimation includes strata fixed effects. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: Balance: Apprentice Characteristics (Endline Sample—Men and Women)

	Observations (1)	Mean Control (2)	Mean Treatment (3)	Treatment Coefficient (4)	Coefficient p-value (5)
Garments	748	0.56	0.53		
North	748	0.34	0.35		
Program apprentices (incl. respondent)	748	2.52	2.46	-0.16	(0.23)
Female	748	0.92	0.96	0.03	(0.13)
Age	727	22.91	22.98	0.04	(0.92)
Married (0/1)	725	0.36	0.35	-0.02	(0.57)
Number of children	734	0.74	0.80	-0.02	(0.82)
Lives with parent(s) (0/1)	732	0.49	0.44	-0.01	(0.77)
Household asset score (z-score)	720	-0.01	0.14	0.15	(0.03)**
Years of schooling	713	7.49	7.40	-0.11	(0.62)
Ability index (z-score)	733	-0.02	0.00	0.02	(0.82)
Soft skills index (z-score)	731	0.00	0.08	0.07	(0.30)
Food security (z-score)	730	-0.01	-0.02	-0.04	(0.58)
Self-reported health (z-score)	732	0.00	-0.01	-0.02	(0.85)
Mother years of schooling	611	3.34	3.44	0.10	(0.76)
Father years of schooling	536	5.60	5.74	0.00	(1.00)
Prior apprenticeship experience (0/1)	733	0.26	0.28	0.02	(0.50)
Wage-employed (0/1)	734	0.04	0.04	0.00	(0.86)
Wage-employment earnings (GhC)	734	1.46	1.36	0.12	(0.86)
Self-employed (0/1)	734	0.22	0.19	-0.02	(0.61)
Self-employment profits (GhC)	734	12.23	11.03	0.10	(0.98)
Total weekly hours worked	726	19.06	17.32	-0.41	(0.84)
Desires self-employment (0/1)	734	0.56	0.52	-0.05	(0.17)

Notes: *North* indicates that the apprentice lives in the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghana as of 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and program apprentices (including respondent) to be balanced by construction. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Errors are clustered at the firm level. 781 of 797 apprentices participated in the baseline survey between August and December of 2012. 748 apprentices participated in the endline survey. 734 participated in both and thus have baseline covariates against which to test for balance and endline survey data, from which most of our treatment effects are measured. This table restricts the sample to those who participated in the endline survey. *Household Asset* is the first principal component of a set of housing quality (floor material, roof material, wall material, sanitation access, water source, lighting source, primary cooking implement, and number of people who sleep in the same room) and asset measures (mattress, tv, radio, refrigerator, car, motorbike, bicycle, working mobile phones, and shoes). *Ability Index* is the normalized sum of the normalized scores on each of four cognitive tests: Digit Span Recall, four math questions we developed ourselves, Ravens Matrices Group B, and a fifteen word oral English vocabulary recognition test. *Soft Skills Index* is the normalized sum of the normalized scores on each of two (adapted to context) non-cognitive tests: the Rosenberg Self-Esteem scale and the Rotter Locus of Control scale. *Food Security* is the first principal component of two food consumption measures: meals eaten per day and meals per day that included meat (excluding eggs). *Self-Reported Health* is the normalized response to *All in all, how would you describe your state of health these days?* on a four-point Likert scale. The top 0.5% of wage job earnings and self-employment profits have been winsorized; these variables are measured in October 2015 Ghana Cedis. *Desires Self-Employment* is an indicator variable for the respondent reporting that the reason they are interested in apprenticeship training is that it will lead to self-employment. An F test of the joint significance of all 20 covariates (from female to desires self-employment) in predicting treatment yields of p-value of 0.56. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Balance: Apprentice Characteristics (Assessment Sample—Men and Women)

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	488	0.59	0.57		
North	488	0.47	0.48		
Program apprentices (incl. respondent)	488	2.65	2.51	-0.24	(0.16)
Female	488	0.91	0.96	0.03	(0.23)
Age	474	22.69	22.71	-0.07	(0.89)
Married (0/1)	473	0.41	0.42	-0.01	(0.80)
Number of children	480	0.75	0.88	0.01	(0.90)
Lives with parent(s) (0/1)	479	0.50	0.48	0.02	(0.60)
Household asset score (z-score)	472	-0.12	0.06	0.21	(0.01)**
Years of schooling	463	7.16	7.42	0.29	(0.33)
Ability index (z-score)	480	-0.05	0.02	0.03	(0.79)
Soft skills index (z-score)	478	-0.07	-0.05	0.07	(0.50)
Food security (z-score)	477	-0.04	-0.09	-0.03	(0.76)
Self-reported health (z-score)	478	0.01	-0.04	0.00	(0.96)
Mother years of schooling	416	2.75	2.86	0.15	(0.71)
Father years of schooling	367	4.59	5.11	0.40	(0.47)
Prior apprenticeship experience (0/1)	479	0.30	0.26	-0.02	(0.65)
Wage-employed (0/1)	480	0.04	0.03	0.00	(0.89)
Wage-employment earnings (GhC)	480	1.28	1.28	0.18	(0.84)
Self-employed (0/1)	480	0.24	0.20	-0.03	(0.52)
Self-employment profits (GhC)	480	12.12	11.79	1.98	(0.62)
Total weekly hours worked	476	21.54	17.24	-2.69	(0.33)
Desires self-employment (0/1)	480	0.53	0.51	-0.03	(0.55)

Notes: *North* indicates that the apprentice lives in the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas of 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and program apprentices (including respondent) to be balanced by construction. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Errors are clustered at the firm level. 781 of 797 apprentices participated in the baseline survey between August and December of 2012. 488 apprentices participated in the assessment (excluding the pilot district). 480 participated in both and thus have baseline covariates against which to test for balance and assessment data, from which some of our skills treatment effects are measured. This table restricts the sample to those who participated in the assessment. *Household Asset* is the first principal component of a set of housing quality (floor material, roof material, wall material, sanitation access, water source, lighting source, primary cooking implement, and number of people who sleep in the same room) and asset measures (mattress, tv, radio, refrigerator, car, motorbike, bicycle, working mobile phones, and shoes). *Ability Index* is the normalized sum of the normalized scores on each of four cognitive tests: Digit Span Recall, four math questions we developed ourselves, Ravens Matrices Group B, and a fifteen word oral English vocabulary recognition test. *Soft Skills Index* is the normalized sum of the normalized scores on each of two (adapted to context) non-cognitive tests: the Rosenberg Self-Esteem scale and the Rotter Locus of Control scale. *Food Security* is the first principal component of two food consumption measures: meals eaten per day and meals per day that included meat (excluding eggs). *Self-Reported Health* is the normalized response to *All in all, how would you describe your state of health these days?* on a four-point Likert scale. The top 0.5% of wage job earnings and self-employment profits have been winsorized; these variables are measured in October 2015 Ghana Cedis. *Desires Self-Employment* is an indicator variable for the respondent reporting that the reason they are interested in apprenticeship training is that it will lead to self-employment. An F test of the joint significance of all 20 covariates (from female to desires self-employment) in predicting treatment yields of p-value of 0.62. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Balance: Apprentice Characteristics (Full Sample—Women Only)

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	751	0.53	0.52		
North	751	0.34	0.33		
Female	751	1.00	1.00		
Program apprentices (incl. respondent)	751	2.44	2.48	-0.15	(0.24)
Age	729	22.62	22.86	0.29	(0.43)
Married (0/1)	727	0.38	0.34	-0.05	(0.17)
Number of children	736	0.75	0.79	-0.03	(0.72)
Lives with parent(s) (0/1)	733	0.48	0.45	-0.01	(0.69)
Household asset score (z-score)	722	0.00	0.17	0.15	(0.03)**
Years of schooling	715	7.40	7.30	-0.17	(0.45)
Ability index (z-score)	735	0.00	0.03	0.02	(0.80)
Soft skills index (z-score)	733	0.00	0.09	0.09	(0.18)
Food security (z-score)	731	0.00	-0.02	-0.05	(0.49)
Self-reported health (z-score)	734	0.00	0.03	0.01	(0.85)
Mother years of schooling	611	3.37	3.59	0.26	(0.47)
Father years of schooling	531	5.60	5.67	0.10	(0.84)
Prior apprenticeship experience (0/1)	735	0.25	0.25	0.01	(0.86)
Wage-employed (0/1)	736	0.02	0.03	0.01	(0.48)
Wage-employment earnings (GhC)	736	0.57	1.11	0.56	(0.19)
Self-employed (0/1)	736	0.21	0.18	-0.01	(0.68)
Self-employment profits (GhC)	736	11.43	9.03	-1.38	(0.60)
Total weekly hours worked	728	18.27	15.85	-0.95	(0.62)
Desires self-employment (0/1)	736	0.55	0.52	-0.05	(0.21)

Notes: *North* indicates that the apprentice lives in the the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas of 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and program apprentices (including respondent) to be balanced by construction. All apprentices in this table are female. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Errors are clustered at the firm level. 736 of the 751 women in the analysis sample participated in the baseline survey between August and December of 2012. *Household Asset* is the first principal component of a set of housing quality (floor material, roof material, wall material, sanitation access, water source, lighting source, primary cooking implement, and number of people who sleep in the same room) and asset measures (mattress, tv, radio, refrigerator, car, motorbike, bicycle, working mobile phones, and shoes). *Ability Index* is the normalized sum of the normalized scores on each of four cognitive tests: Digit Span Recall, four math questions we developed ourselves, Ravens Matrices Group B, and a fifteen word oral English vocabulary recognition test. *Soft Skills Index* is the normalized sum of the normalized scores on each of two (adapted to context) non-cognitive tests: the Rosenberg Self-Esteem scale and the Rotter Locus of Control scale. *Food Security* is the first principal component of two food consumption measures: meals eaten per day and meals per day that included meat (excluding eggs). *Self-Reported Health* is the normalized response to *All in all, how would you describe your state of health these days?* on a four-point Likert scale. The top 0.5% of wage job earnings and self-employment profits have been winsorized; these variables are measured in October 2015 Ghana Cedis. *Desires Self-Employment* is an indicator variable for the respondent reporting that the reason they are interested in apprenticeship training is that it will lead to self-employment. An F test of the joint significance of all 19 covariates (from age to desires self-employment) in predicting treatment yields of p-value of 0.24. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Balance: Apprentice Characteristics (Endline Sample—Women Only)

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	706	0.53	0.52		
North	706	0.33	0.34		
Female	706	1.00	1.00		
Program apprentices (incl. respondent)	706	2.45	2.50	-0.14	(0.29)
Age	686	22.66	22.89	0.21	(0.58)
Married (0/1)	684	0.37	0.35	-0.04	(0.35)
Number of children	693	0.75	0.80	-0.03	(0.73)
Lives with parent(s) (0/1)	691	0.48	0.44	-0.01	(0.73)
Household asset score (z-score)	680	-0.01	0.16	0.15	(0.02)**
Years of schooling	672	7.40	7.33	-0.10	(0.67)
Ability index (z-score)	692	-0.02	0.03	0.04	(0.63)
Soft skills index (z-score)	690	0.01	0.08	0.07	(0.36)
Food security (z-score)	689	-0.01	0.00	-0.04	(0.63)
Self-reported health (z-score)	691	0.00	0.01	0.01	(0.95)
Mother years of schooling	578	3.43	3.52	0.07	(0.83)
Father years of schooling	502	5.71	5.74	-0.03	(0.95)
Prior apprenticeship experience (0/1)	692	0.25	0.26	0.01	(0.70)
Wage-employed (0/1)	693	0.02	0.03	0.01	(0.38)
Wage-employment earnings (GhC)	693	0.60	1.18	0.67	(0.16)
Self-employed (0/1)	693	0.22	0.18	-0.01	(0.65)
Self-employment profits (GhC)	693	11.83	9.35	-2.09	(0.44)
Total weekly hours worked	686	18.84	16.62	-0.72	(0.72)
Desires self-employment (0/1)	693	0.56	0.52	-0.06	(0.14)

Notes: *North* indicates that the apprentice lives in the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas of 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and program apprentices (including respondent) to be balanced by construction. All workers in this table are female. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Errors are clustered at the firm level. 736 of the 751 women in the analysis sample participated in the baseline survey between August and December 2012. 700 participated in the endline survey. 706 participated in both and thus have baseline covariates against which to test for balance and endline survey data, from which most of our treatment effects are measured. This table restricts the sample to those who participated in the endline survey. *Household Asset* is the first principal component of a set of housing quality (floor material, roof material, wall material, sanitation access, water source, lighting source, primary cooking implement, and number of people who sleep in the same room) and asset measures (mattress, tv, radio, refrigerator, car, motorbike, bicycle, working mobile phones, and shoes). *Ability Index* is the normalized sum of the normalized scores on each of four cognitive tests: Digit Span Recall, four math questions we developed ourselves, Ravens Matrices Group B, and a fifteen word oral English vocabulary recognition test. *Soft Skills Index* is the normalized sum of the normalized scores on each of two (adapted to context) non-cognitive tests: the Rosenberg Self-Esteem scale and the Rotter Locus of Control scale. *Food Security* is the first principal component of two food consumption measures: meals eaten per day and meals per day that included meat (excluding eggs). *Self-Reported Health* is the normalized response to *All in all, how would you describe your state of health these days?* on a four-point Likert scale. *Desires Self-Employment* is an indicator variable for the respondent reporting that the reason they are interested in apprenticeship training is that it will lead to self-employment. An F test of the joint significance of all 19 covariates (from age to desires self-employment) in predicting treatment yields of p-value of 0.23. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Balance: Apprentice Characteristics (Assessment Sample—Women Only)

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	457	0.57	0.55		
North	457	0.46	0.46		
Female	457	1.00	1.00		
Program apprentices (incl. respondent)	457	2.61	2.55	-0.22	(0.22)
Age	443	22.40	22.61	0.16	(0.75)
Married (0/1)	442	0.43	0.41	-0.04	(0.45)
Number of children	449	0.76	0.87	0.00	(0.99)
Lives with parent(s) (0/1)	448	0.50	0.47	0.01	(0.88)
Household asset score (z-score)	441	-0.12	0.07	0.23	(0.01)***
Years of schooling	432	6.96	7.32	0.36	(0.24)
Ability index (z-score)	449	-0.07	0.06	0.07	(0.47)
Soft skills index (z-score)	447	-0.07	-0.05	0.05	(0.63)
Food security (z-score)	446	-0.04	-0.07	-0.02	(0.81)
Self-reported health (z-score)	447	0.01	-0.04	0.00	(0.98)
Mother years of schooling	391	2.82	2.93	0.15	(0.73)
Father years of schooling	342	4.65	5.12	0.43	(0.47)
Prior apprenticeship experience (0/1)	448	0.28	0.24	-0.03	(0.53)
Wage-employed (0/1)	449	0.02	0.02	0.00	(1.00)
Wage-employment earnings (GhC)	449	0.40	0.88	0.40	(0.46)
Self-employed (0/1)	449	0.23	0.19	-0.03	(0.51)
Self-employment profits (GhC)	449	11.17	9.02	-1.12	(0.73)
Total weekly hours worked	446	21.03	16.25	-3.42	(0.20)
Desires self-employment (0/1)	449	0.53	0.52	-0.02	(0.69)

Notes: *North* indicates that the apprentice lives in the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas of 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and program apprentices (including respondent) to be balanced by construction. All workers in this table are female. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Errors are clustered at the firm level. 736 of the 751 women in the analysis sample participated in the baseline survey between August and December of 2012. 457 participated in the assesment (excluding the women from the pilot district where the assessment differed from other districts). 449 participated in both and thus have baseline covariates against which to test for balance and assessment data, from which some of our key skills treatment effects are measured. This table restricts the sample to those who participated in the assessment. *Household Asset* is the first principal component of a set of housing quality (floor material, roof material, wall material, sanitation access, water source, lighting source, primary cooking implement, and number of people who sleep in the same room) and asset measures (mattress, tv, radio, refrigerator, car, motorbike, bicycle, working mobile phones, and shoes). *Ability Index* is the normalized sum of the normalized scores on each of four cognitive tests: Digit Span Recall, four math questions we developed ourselves, Ravens Matrices Group B, and a fifteen word oral English vocabulary recognition test. *Soft Skills Index* is the normalized sum of the normalized scores on each of two (adapted to context) non-cognitive tests: the Rosenberg Self-Esteem scale and the Rotter Locus of Control scale. *Food Security* is the first principal component of two food consumption measures: meals eaten per day and meals per day that included meat (excluding eggs). *Self-Reported Health* is the normalized response to *All in all, how would you describe your state of health these days?* on a four-point Likert scale. The top 0.5% of wage job earnings and self-employment profits have been winsorized; these variables are measured in October 2015 Ghana Cedis. *Desires Self-Employment* is an indicator variable for the respondent reporting that the reason they are interested in apprenticeship training is that it will lead to self-employment. An F test of the joint significance of all 19 covariates (from age to desires self-employment) in predicting treatment yields of p-value of 0.26. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

33

Table A9: Balance: Firm Characteristics—Firms Training Men and Women

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	467	0.55	0.56		
North	467	0.31	0.32		
Program apprentices	467	1.70	1.71	-0.10	(0.16)
Female	467	0.86	0.90	0.03	(0.34)
Age	457	35.47	35.05	-0.29	(0.69)
Years of schooling	462	8.52	8.48	-0.09	(0.78)
Completed apprenticeship (0/1)	461	0.97	0.97	0.01	(0.62)
Ability index (z-score)	421	0.00	0.02	0.01	(0.95)
Firm age (years)	463	11.22	10.72	-0.37	(0.59)
Revenues last month (GhC)	419	372.00	373.40	11.36	(0.78)
Profits last month (GhC)	419	187.70	199.10	21.48	(0.33)
Firm size (incl. owner)	421	4.43	4.52	-0.05	(0.86)
Wage bill (GhC)	421	55.11	56.16	-3.84	(0.72)
Firm assets (GhC)	421	2,870.00	3,029.00	249.05	(0.32)
Trade association member (0/1)	421	0.85	0.86	0.00	(0.94)
Past completed apprentices	420	12.39	10.29	-1.94	(0.28)
Customer service instruction (0/1)	406	0.98	0.97	0.00	(0.98)
Inventory management instruction (0/1)	406	0.59	0.65	0.05	(0.38)
Financial management instruction (0/1)	406	0.54	0.56	0.01	(0.90)
Business confidence instruction (0/1)	406	0.60	0.65	0.00	(0.97)
Provides more instruction to slow apprentices (0/1)	402	0.30	0.30	-0.01	(0.84)
Months until apprentices work with clients	406	9.96	9.62	-0.14	(0.87)

Notes: *North* indicates that the firm is located in the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and number of program apprentices to be balanced by construction. This table includes all firms in our sample. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. *Ability Index* is the normalized sum of the normalized scores on two cognitive tests: Digit Span Recall and four math questions we developed ourselves. All financial variables are in October 2015 GhC. An F test of the joint significance of all 19 covariates (from female to months until apprentices work with clients) in predicting treatment yields a p-value of 0.98. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A10: Balance: Firm Characteristics—Firms Training Women Only

	Observations	Mean Control	Mean Treatment	Treatment Coefficient	Coefficient p-value
	(1)	(2)	(3)	(4)	(5)
Garments	446	0.53	0.54		
North	446	0.31	0.31		
Program apprentices	446	1.71	1.73	-0.08	(0.22)
Female	446	0.91	0.92	0.01	(0.61)
Age	436	35.37	35.20	-0.11	(0.88)
Years of schooling	441	8.46	8.44	-0.07	(0.83)
Completed apprenticeship (0/1)	440	0.97	0.97	0.01	(0.67)
Ability index (z-score)	403	0.00	0.03	0.02	(0.86)
Firm age (years)	442	11.20	10.80	-0.24	(0.73)
Revenues last month (GhC)	401	378.10	377.80	14.23	(0.73)
Profits last month (GhC)	401	191.80	201.00	20.67	(0.37)
Firm size (incl. owner)	403	4.44	4.52	-0.03	(0.92)
Wage bill (GhC)	403	55.18	54.70	-5.80	(0.59)
Firm assets (GhC)	403	2,849.00	3,054.00	240.90	(0.35)
Trade association member (0/1)	403	0.85	0.87	0.01	(0.84)
Past completed apprentices	402	12.81	10.68	-1.98	(0.30)
Customer service instruction (0/1)	389	0.98	0.97	0.00	(0.97)
Inventory management instruction (0/1)	389	0.59	0.67	0.07	(0.20)
Financial management instruction (0/1)	389	0.53	0.56	0.02	(0.77)
Business confidence instruction (0/1)	389	0.62	0.65	-0.02	(0.66)
Provides more instruction to slow apprentices (0/1)	385	0.32	0.29	-0.03	(0.58)
Months until apprentices work with clients	389	9.68	9.38	-0.11	(0.90)

Notes: *North* indicates that the firm is located in the the northern regions of the country as of 2015, which are socio-culturally distinct from the 7 southern regions of Ghanas 2015. The randomization was stratified by district, trade, and a measure of the number of NAP apprentices training at the firm, so we expect garments, north, and number of program apprentices to be balanced by construction. This table includes all firms training female workers in our sample. Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. *Ability Index* is the normalized sum of the normalized scores on two cognitive tests: Digit Span Recall and four math questions we developed ourselves. All financial variables are in October 2015 GhC. An F test of the joint significance of all 19 covariates (from female to months until apprentices work with clients) in predicting treatment yields of p-value of 0.97. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A11: Attrition

| | Targeted Sample | = 1 if Surveyed | | Treatment Coefficient | Coefficient p-value |
| | | Control | Treatment | | |
	(1)	(2)	(3)	(4)	(5)
Panel A: Full sample					
Apprentice Baseline Survey	751	0.98	0.98	0.00	(0.88)
Apprentice Incentivized Test	797	0.65	0.63	-0.01	(0.87)
Apprentice Endline Survey	797	0.94	0.94	-0.01	(0.71)
Firm Baseline Survey	467	0.90	0.91	0.01	(0.72)
Firm Midline Survey	467	0.72	0.74	0.01	(0.81)
Panel B: Women only					
Apprentice Baseline Survey	751	0.98	0.98	0.00	(0.88)
Apprentice Incentivized Test	751	0.64	0.63	0.00	(0.97)
Apprentice Endline Survey	751	0.94	0.94	-0.01	(0.66)
Firm Baseline Survey	446	0.90	0.90	0.00	(0.96)
Firm Midline Survey	446	0.72	0.73	0.01	(0.85)

Notes: Each treatment coefficient in column (4) comes from a separate regression that includes strata fixed effects. Standard errors in apprentice-level regressions are clustered at the firm level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A12: Correlates of Attrition—Apprentice Characteristics

	Full Sample		Women Only	
	Took incentivized test (0/1)	Took incentivized test (0/1)	Took incentivized test (0/1)	Took incentivized test (0/1)
	(1)	(2)	(3)	(4)
Treatment	-0.00	0.12	0.01	0.12
	(0.04)	(0.12)	(0.04)	(0.13)
Age	-0.01	-0.00	-0.01	-0.00
	(0.00)	(0.00)	(0.00)	(0.01)
Age*Treatment		-0.01*		-0.01
		(0.01)		(0.01)
Married (0/1)	0.01	-0.02	0.01	-0.03
	(0.04)	(0.05)	(0.04)	(0.06)
Married*Treatment		0.07		0.08
		(0.07)		(0.08)
Number of Children	-0.01	-0.02	-0.01	-0.01
	(0.02)	(0.04)	(0.02)	(0.04)
Children*Treatment		0.01		-0.00
		(0.04)		(0.04)
Household Asset score (z-score)	0.01	0.01	0.02	0.02
	(0.02)	(0.03)	(0.02)	(0.03)
Asset Score*Treatment		-0.01		-0.01
		(0.04)		(0.04)
Years Schooling	0.01	0.00	0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Years Schooling*Treatment		0.01		0.01
		(0.01)		(0.01)
Ability Index (z-score)	-0.00	-0.01	-0.00	-0.02
	(0.02)	(0.03)	(0.02)	(0.03)
Ability*Treatment		0.02		0.04
		(0.04)		(0.04)
Soft Skills Index (z-score)	-0.02	0.01	-0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Soft Skills*Treatment		-0.05		-0.05
		(0.03)		(0.04)
Food Security (z-score)	-0.02	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Food Security*Treatment		-0.00		-0.00
		(0.03)		(0.04)
Self-Reported Health (z-score)	-0.01	0.00	-0.01	0.01
	(0.02)	(0.03)	(0.02)	(0.03)
Health*Treatment		-0.03		-0.05
		(0.03)		(0.04)
Desires Self-Employment (0/1)	0.00	-0.00	0.01	0.01
	(0.03)	(0.04)	(0.03)	(0.05)
Self Employment*Treatment		0.02		0.02
		(0.06)		(0.07)
Observations	797	797	751	751
Joint F-statistic p-value	0.47	0.41	0.56	0.28

Notes: Regressions include strata fixed effects. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A13: Correlates of Attrition—Firm Characteristics

	Full Sample		Women Only	
	Took incentivized test (0/1) (1)	Took incentivized test (0/1) (2)	Took incentivized test (0/1) (3)	Took incentivized test (0/1) (4)
Treatment	-0.00	-0.03	0.01	-0.04
	(0.04)	(0.19)	(0.04)	(0.19)
Years Schooling	-0.01	-0.01	-0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
Years Schooling*Treatment		-0.01		-0.01
		(0.01)		(0.01)
Ability Index (z-score)	0.01	-0.00	0.00	-0.01
	(0.02)	(0.04)	(0.02)	(0.04)
Ability*Treatment		0.02		0.03
		(0.05)		(0.05)
Firm Size (Incl Owner)	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.01)	(0.01)
Firm Size*Treatment		0.00		-0.00
		(0.02)		(0.02)
Above Median Apprentices Trained (0/1)	0.08*	0.09	0.07	0.08
	(0.05)	(0.07)	(0.05)	(0.07)
Above Med Apps Trained*Treatment		-0.01		0.00
		(0.09)		(0.10)
Trade Association Member (0/1)	0.08	0.11	0.10	0.15
	(0.07)	(0.09)	(0.08)	(0.10)
Association Member*Treatment		-0.08		-0.11
		(0.12)		(0.13)
Profits Last Month (log)	0.01	-0.03	0.01	-0.00
	(0.03)	(0.04)	(0.04)	(0.05)
Profits*Treatment		0.07		0.04
		(0.07)		(0.08)
Revenues Last Month (log)	-0.01	0.02	-0.03	-0.02
	(0.04)	(0.06)	(0.04)	(0.06)
Revenues*Treatment		-0.06		-0.01
		(0.07)		(0.07)
Wage Bill (log)	-0.00	0.00	-0.00	-0.00
	(0.01)	(0.02)	(0.01)	(0.02)
Wage Bill*Treatment		-0.01		-0.01
		(0.03)		(0.03)
Firm Assets (log)	0.00	-0.01	0.01	0.00
	(0.02)	(0.03)	(0.02)	(0.03)
Firm Assets*Treatment		0.02		0.02
		(0.04)		(0.04)
Observations	797	797	751	751
Joint F-statistic p-value	0.638	0.852	0.681	0.886

Notes: Regressions include strata fixed effects. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A14: Labor supply

	Working	Self-employment	Skilled self-employment	Unskilled self-employment	Wage employment	Skilled wage employment	Unskilled wage employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Full sample							
Treatment	0.04	0.01	0.05*	-0.03*	0.00	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	2992	2992	2992	2992	2992	2992	2992
Mean of Dep Variable T=0	0.62	0.29	0.16	0.13	0.10	0.05	0.05
Panel B: Women only							
Treatment	0.03	0.00	0.04	-0.03*	-0.00	-0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	2824	2824	2824	2824	2824	2824	2824
Mean of Dep Variable T=0	0.62	0.29	0.16	0.13	0.10	0.05	0.05

Notes: This table shows the results from estimating equation 2 for different labor market outcomes. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Full sample specifications include a control for apprentice gender. Outcomes are stacked across a four-round retrospective panel that asked respondents to report on labor supply 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Working includes wage-employment, self-employment, farming, and apprenticeships. Wage employment and apprenticeships are treated as separate categories of work. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A15: Heterogeneity in treatment effects (Practical test (z-score))

	Full Sample					Women Only				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Panel A: Apprentice characteristics

	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment*Covariate	-0.09	0.31**	-0.02	-0.12**	0.05	-0.04	0.31**	0.00	-0.12**	0.09
	(0.07)	(0.14)	(0.07)	(0.06)	(0.15)	(0.07)	(0.14)	(0.07)	(0.06)	(0.14)
Observations	488	463	480	478	480	457	432	449	447	449

Panel B: Firm owner characteristics

	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment*Covariate	-0.06	0.10	-0.03	-0.08	-0.00	0.10	0.08	-0.00	-0.07	-0.01
	(0.18)	(0.08)	(0.03)	(0.17)	(0.01)	(0.17)	(0.08)	(0.03)	(0.17)	(0.01)
Observations	486	464	464	462	464	455	434	434	432	434

Notes: The outcome variable is the practical component z-score. Each column shows the heterogeneous treatment effect by different worker (Panel A) and firm (Panel B) characteristics. For both panels, Columns 1-5 use the full apprentices sample and Columns 6-10 only the female apprentices sample. For apprentices characteristics: *Asset score* is the z-score of the apprentice's household asset score; *Less than grade 8* is a dummy variable that equals 1 if the apprentice has less than eight years of formal schooling; *Ability* is the z-score of the apprentice's ability index; *Soft skills* is the z-score of the apprentice's soft skills index; *Desires self-employment* is a dummy variable if the apprentice answered yes to the prompt. For firm owner characteristics: *Less than grade 8* is a dummy variable that equals 1 if the firm owner has less than eight years of formal schooling; *Ability* is the z-score of the firm owner's ability index; *Total workers* is the baseline firm size (including owner); *Above median sales* is a dummy variable that equals 1 if the firm earned more than median revenues last month; *Apprentices trained* is the total number of past completed apprentices. All specifications include strata fixed effects, assessor dummies and controls for the baseline household asset index (including a dummy for where the variable is missing). All specifications exclude pilot district. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A16: Heterogeneity in treatment effects (Craft skills (z-score))

	Full Sample					Women Only				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Panel A: Apprentice characteristics

	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment
Treatment*Covariate	0.09	0.27*	-0.12	-0.04	-0.17	0.07	0.29*	-0.13	-0.02	-0.18
	(0.08)	(0.16)	(0.08)	(0.07)	(0.16)	(0.08)	(0.16)	(0.08)	(0.08)	(0.17)
Observations	743	710	730	728	731	703	671	691	689	692

Panel B: Firm owner characteristics

	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained
Treatment*Covariate	0.23	-0.02	-0.03	-0.02	0.01	0.22	-0.00	-0.03	-0.07	0.01
	(0.18)	(0.09)	(0.03)	(0.17)	(0.00)	(0.18)	(0.09)	(0.03)	(0.17)	(0.00)
Observations	738	693	693	691	692	698	654	654	652	653

Notes: The outcome variable is the craft skills quiz z-score. Each column shows the heterogeneous treatment effect by different worker (Panel A) and firm (Panel B) characteristics. For both panels, Columns 1-5 use the full apprentices sample and Columns 6-10 only the female apprentices sample. For apprentices characteristics: *Asset score* is the z-score of the apprentice's household asset score; *Less than grade 8* is a dummy variable that equals 1 if the apprentice has less than eight years of formal schooling; *Ability* is the z-score of the apprentice's ability index; *Soft skills* is the z-score of the apprentice's soft skills index; *Desires self-employment* is a dummy variable if the apprentice answered yes to the prompt. For firm owner characteristics: *Less than grade 8* is a dummy variable that equals 1 if the firm owner has less than eight years of formal schooling; *Ability* is the z-score of the firm owner's ability index; *Total workers* is the baseline firm size (including owner); *Above median sales* is a dummy variable that equals 1 if the firm earned more than median revenues last month; *Apprentices trained* is the total number of past completed apprentices. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A17: Heterogeneity in treatment effects (Sales skills (z-score))

	Full Sample					Women Only				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Panel A: Apprentice characteristics

	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment
Treatment*Covariate	0.10	0.05	0.08	0.08	-0.20	0.13^{*}	0.06	0.07	0.05	-0.22
	(0.07)	(0.14)	(0.07)	(0.06)	(0.13)	(0.06)	(0.14)	(0.07)	(0.06)	(0.14)
Observations	748	713	733	731	734	706	672	692	690	693

Panel B: Firm owner characteristics

	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained
Treatment*Covariate	-0.04	0.16^{**}	0.02	0.02	0.01^{***}	-0.03	0.17^{**}	0.01	0.00	0.01^{***}
	(0.14)	(0.07)	(0.03)	(0.15)	(0.00)	(0.14)	(0.08)	(0.03)	(0.15)	(0.00)
Observations	743	696	696	694	695	701	656	656	654	655

Notes: The outcome variable is the sales skills z-score. Each column shows the heterogeneous treatment effect by different worker (Panel A) and firm (Panel B) characteristics. For both panels, Columns 1–5 use the full apprentices sample and Columns 6–10 only the female apprentices sample. For apprentices characteristics: *Asset score* is the z-score of the apprentice's household asset score; *Less than grade 8* is a dummy variable that equals 1 if the apprentice has less than eight years of formal schooling; *Ability* is the z-score of the apprentice's ability index; *Soft skills* is the z-score of the apprentice's soft skills index; *Desires self employment* is a dummy variable if the apprentice answered yes to the prompt. For firm owner characteristics: *Less than grade 8* is a dummy variable that equals 1 if the firm owner has less than eight years of formal schooling; *Ability* is the z-score of the firm owner's ability index; *Total workers* is the baseline firm size (including owner); *Above median sales* is a dummy variable that equals 1 if the firm earned more than median revenues last month; *Apprentices trained* is the total number of past completed apprentices. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A18: Heterogeneity in treatment effects (Total earnings (GhC))

	Full Sample					Women Only				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Panel A: Apprentice characteristics

	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment*Covariate	16.32*	-31.07	3.89	14.57**	18.62	17.98*	-26.58	-0.66	14.47**	16.94
	(9.38)	(22.79)	(5.31)	(6.37)	(16.05)	(9.69)	(23.69)	(5.25)	(6.57)	(16.13)
Observations	2,992	2,852	2,784	2,924	2,936	2,824	2,688	2,624	2,760	2,772

Panel B: Firm owner characteristics

	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Treatment*Covariate	-34.45***	3.89	-0.41	38.77***	0.08	-28.79**	-0.66	0.87	25.70**	-0.10
	(13.13)	(5.31)	(2.26)	(12.41)	(0.40)	(12.55)	(5.25)	(2.17)	(11.23)	(0.40)
Observations	2,972	2,784	2,784	2,776	2,780	2,804	2,624	2,624	2,616	2,620

Notes: The outcome variable is the total earnings. Each column shows the heterogeneous treatment effect by different worker (Panel A) and firm (Panel B) characteristics. For both panels, Columns 1-5 use the full apprentices sample and Columns 6-10 only the female apprentices sample. For apprentices characteristics: *Asset score* is the z-score of the apprentice's household asset score; *Less than grade 8* is a dummy variable that equals 1 if the apprentice has less than eight years of formal schooling; *Ability* is the z-score of the apprentice's ability index; *Soft skills* is the z-score of the apprentice's soft skills index; *Desires self-employment* is a dummy variable if the apprentice answered yes to the prompt. For firm owner characteristics: *Less than grade 8* is a dummy variable that equals 1 if the firm owner has less than eight years of formal schooling; *Ability* is the z-score of the firm owner's ability index; *Total workers* is the baseline firm size (including owner); *Above median sales* is a dummy variable that equals 1 if the firm earned more than median revenues last month; *Apprentices trained* is the total number of past completed apprentices. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Outcomes are stacked across a four-round retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Total earnings includes earnings from wage-employment, self-employment, farming, and apprenticeship. Earnings are in October 2015 Ghana Cedis. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A19: Heterogeneity in treatment effects (Self-employment business profits (Unconditional; GhC))

Panel A: Apprentice characteristics

	Full Sample					Women Only				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment
Treatment*Covariate	10.60	-30.69	-4.89	14.06**	17.64	11.81	-30.35	-5.22	13.86**	21.20
	(8.22)	(22.84)	(4.38)	(5.90)	(13.55)	(8.54)	(23.98)	(4.36)	(6.18)	(14.06)
Observations	2,992	2,852	2,784	2,924	2,936	2,824	2,688	2,624	2,760	2,772

Panel B: Firm owner characteristics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained
Treatment*Covariate	-0.96	-4.89	0.29	15.35	-0.02	-0.63	-5.22	0.84	13.44	-0.13
	(10.22)	(4.38)	(1.66)	(10.05)	(0.38)	(10.63)	(4.36)	(1.63)	(10.11)	(0.39)
Observations	2,972	2,784	2,784	2,776	2,780	2,804	2,624	2,624	2,616	2,620

Notes: The outcome variable is the unconditional self-employment business profits. Each column shows the heterogeneous treatment effect by different worker (Panel A) and firm (Panel B) characteristics. For both panels, Columns 1-5 use the full apprentices sample and Columns 6-10 only the female apprentices sample. For apprentices characteristics: *Asset score* is the z-score of the apprentice's household asset score; *Less than grade 8* is a dummy variable that equals 1 if the apprentice has less than eight years of formal schooling; *Ability* is the z-score of the apprentice's ability index; *Soft skills* is the z-score of the apprentice's soft skills index; *Desires self-employment* is a dummy variable if the apprentice answered yes to the prompt. For firm owner characteristics: *Less than grade 8* is a dummy variable that equals 1 if the firm owner has less than eight years of formal schooling; *Ability* is the z-score of the firm owner's ability index; *Total workers* is the baseline firm size (including owner); *Above median sales* is a dummy variable that equals 1 if the firm earned more than median revenues last month; *Apprentices trained* is the total number of past completed apprentices. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Outcomes are stacked across a four-round retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Total earnings includes earnings from wage-employment, self-employment, farming, and apprenticeship. Earnings are in October 2015 Ghana Cedis. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

44

Table A20: Heterogeneity in treatment effects (Self-employment business profits (Conditional; GhC))

	Full Sample					Women Only				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)

Panel A: Apprentice characteristics

	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment	Asset score	Less than grade 8	Ability	Soft skills	Desires self-employment
Treatment*Covariate	7.86	-17.25	6.97	25.97**	51.90**	12.32	-28.82	3.09	24.04*	64.24**
	(11.54)	(19.84)	(11.01)	(11.48)	(25.10)	(12.07)	(20.74)	(10.62)	(12.28)	(26.31)
Observations	869	831	787	856	860	806	771	727	796	800

Panel B: Firm owner characteristics

	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained	Less than grade 8	Ability	Total workers	Above median sales	Apprentices trained
Treatment*Covariate	-6.66	6.97	-5.46	30.26	0.00	-18.70	3.09	-6.12	23.20	-0.35
	(24.12)	(11.01)	(4.44)	(31.47)	(1.22)	(25.70)	(10.62)	(4.90)	(31.08)	(1.24)
Observations	864	787	787	785	787	801	727	727	725	727

Notes: The outcome variable is the conditional self-employment business profits. Each column shows the heterogeneous treatment effect by different worker (Panel A) and firm (Panel B) characteristics. For both panels, Columns 1–5 use the full apprentices sample and Columns 6–10 only the female apprentices sample. For apprentices characteristics: *Asset score* is the z-score of the apprentice's household asset score; *Less than grade 8* is a dummy variable that equals 1 if the apprentice has less than eight years of formal schooling; *Ability* is the z-score of the apprentice's ability index; *Soft skills* is the z-score of the apprentice's soft skills index; *Desires self-employment* is a dummy variable if the apprentice answered yes to the prompt. For firm owner characteristics: *Less than grade 8* is a dummy variable that equals 1 if the firm owner has less than eight years of formal schooling; *Ability* is the z-score of the firm owner's ability index; *Total workers* is the baseline firm size (including owner); *Above median sales* is a dummy variable that equals 1 if the firm earned more than median revenues last month; *Apprentices trained* is the total number of past completed apprentices. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Outcomes are stacked across a four-round retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Total earnings includes earnings from wage-employment, self-employment, farming, and apprenticeship. Earnings are in October 2015 Ghana Cedis. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

45

Table A21: Skills—Robustness

	Assessment					Endline survey			
	Took incentivized test (0/1)	Took incentivized test (0/1)	Practical component (z-score)	Theoretical component (z-score)	Earned certificate (0/1)	Craft skills quiz (z-score)	Craft skills quiz (z-score)	Sales skills (z-score)	Sales skills (z-score)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Full sample (Baseline controls)									
Treatment	-0.01	-0.01	0.13*	-0.08	0.05	0.15**	0.14	0.12**	0.10
	(0.04)	(0.04)	(0.07)	(0.07)	(0.04)	(0.07)	(0.09)	(0.06)	(0.07)
Observations	797	763	488	488	488	743	466	748	468
Mean of Dep Variable T=0	0.65	0.65	0.00	0.00	0.80	0.00	0.05	0.00	-0.11
Incentivized Test Sample							Yes		Yes
Panel B: Full sample (No controls)									
Treatment	-0.01	-0.01	0.13*	-0.06	0.05	0.16**	0.16*	0.12*	0.10
	(0.04)	(0.04)	(0.07)	(0.07)	(0.03)	(0.07)	(0.09)	(0.06)	(0.07)
Observations	797	763	488	488	488	743	466	748	468
Mean of Dep Variable T=0	0.65	0.65	0.00	0.00	0.80	0.00	0.05	0.00	-0.11
Incentivized Test Sample							Yes		Yes
Panel C: Women only (No controls)									
Treatment	0.00	0.00	0.18**	-0.04	0.07*	0.16**	0.15*	0.14**	0.13*
	(0.04)	(0.04)	(0.07)	(0.07)	(0.04)	(0.07)	(0.09)	(0.06)	(0.08)
Observations	751	717	457	457	457	703	439	706	439
Mean of Dep Variable T=0	0.64	0.64	0.00	0.00	0.80	0.00	0.05	0.00	-0.09
Incentivized Test Sample							Yes		Yes

Notes: This table shows the results from estimating equation 1 for different skill measures. All specifications include strata fixed effects. Specifications from the assessment include assessor dummies. Sales skills specifications include surveyor fixed effects. Columns 2 through 5 exclude test-takers from the first district, because the assessment (being implemented in collaboration with our government partners) differed from the test given in all other districts. The first panel mirrors specifications in Table 1, including controls for the baseline value of the household asset index (and a dummy for where that variable in missing) in the full sample of 797 apprentices (including both men and women). Full sample specifications also include a control for apprentice gender. The second and third panels test for robustness to the exclusion of the baseline controls in the full sample (men and women) and the analysis sample (women only). Columns 2–5 exclude the pilot district. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A22: Earnings—Robustness

	Unconditional			Conditional	
	Total earnings (GhC) (1)	Self-employment profits (GhC) (2)	Wage employment earnings (GhC) (3)	Self-employment profits (GhC) (4)	Wage employment earnings (GhC) (5)
Panel A: Full sample (Baseline controls)					
Treatment	10.43*	8.41*	3.41	15.15	-4.10
	(5.56)	(4.70)	(3.40)	(12.23)	(19.55)
Observations	2,992	2,992	2,992	869	301
Mean of Dep Variable T=0	44.34	24.18	12.98	83.58	136.89
Panel B: Full sample (No controls)					
Treatment	11.82**	9.36**	3.92	14.87	-12.65
	(5.43)	(4.58)	(3.44)	(12.09)	(24.53)
Observations	2992	2992	2992	869	301
Mean of Dep Variable T=0	44.34	24.18	12.98	83.58	136.89
Panel C: Women only (No controls)					
Treatment	14.84***	10.54**	3.42	22.79*	-10.60
	(5.34)	(4.65)	(3.62)	(12.58)	(26.86)
Observations	2824	2824	2824	806	287
Mean of Dep Variable T=0	41.52	22.99	13.10	80.62	135.34

Notes: This table shows the results from estimating equation 2 for different labor market outcomes. All specifications include strata fixed effects. Outcomes are stacked across a four-round retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Total earnings includes earnings from wage-employment, self-employment, farming, and apprenticeship. Profits in self-employment follow self-reported question structures from De Mel et al. (2009). All earnings are in October 2015 Ghana Cedis. The first panel mirrors specifications in Table 2, including controls for the baseline value of the household asset index (and a dummy for where that variable in missing) in the full sample of 797 apprentices (including both men and women). Full sample specifications also include a control for apprentice gender. The second and third panels test for robustness to the exclusion of the baseline controls in the full sample (men and women) and the analysis sample (women only). Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A23: Impacts During the Intervention—Robustness

	Apprentice level labor inputs		Firm Midline Survey Financial Outcomes
	Apprentice hours last week (1)	Firm owner instruction hours per day (2)	Apprentice wages last month (GhC) (3)
Panel A: Full sample (Baseline controls)			
Treatment	4.13**	0.10	1.55
	(1.91)	(0.09)	(3.08)
Observations	565	595	566
Mean of Dep Variable T=0	44.18	1.29	22.17
Panel B: Full sample (No controls)			
Treatment	3.66*	0.11	1.53
	(1.91)	(0.09)	(3.04)
Observations	565	595	566
Mean of Dep Variable T=0	44.18	1.29	22.17
Panel C: Women only (No controls)			
Treatment	4.03**	0.08	1.94
	(2.03)	(0.09)	(3.20)
Observations	527	556	528
Mean of Dep Variable T=0	43.73	1.32	21.85

Notes: All specifications include strata fixed effects. Wages for each apprentice are reported by the firm owner. Wages are in October 2015 Ghana Cedis. The first panel mirrors specifications in Table 3, including controls for the baseline value of the household asset index. The second and third panels test for robustness to the exclusion of the imbalanced apprentice-level controls in the full sample (men and women) and the analysis sample (women only). Full sample specifications also include a control for apprentice gender. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A24: Apprentice completion and turnover—Robustness

	Apprentice completion				Apprentice turnover	
	Completed	Paid exit/ceremony	Completed and exited	Completed and retained	Exited	Missed months
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample (Baseline Controls)						
Treatment	0.07**	0.10***	0.08**	-0.01	0.08**	1.50*
	(0.04)	(0.03)	(0.04)	(0.01)	(0.04)	(0.84)
Observations	748	748	748	748	748	748
Mean of Dep Variable T=0	0.31	0.15	0.28	0.03	0.39	7.73
Panel B: Full sample (No controls)						
Treatment	0.07*	0.10***	0.08**	-0.01	0.08**	1.47*
	(0.04)	(0.03)	(0.04)	(0.01)	(0.04)	(0.83)
Observations	748	748	748	748	748	748
Mean of Dep Variable T=0	0.31	0.15	0.28	0.03	0.39	7.73
Panel C: Women only (No controls)						
Treatment	0.07*	0.11***	0.07**	-0.00	0.08**	1.60*
	(0.04)	(0.03)	(0.04)	(0.01)	(0.04)	(0.86)
Observations	706	706	706	706	706	706
Mean of Dep Variable T=0	0.32	0.16	0.30	0.02	0.40	7.99

Notes: All specifications include strata fixed effects. The first panel mirrors specifications in Table 4, including controls for the baseline value of the household asset index (and a dummy for where that variable in missing) in the full sample of 797 apprentices (including both men and women). The second and third panels test for robustness to the exclusion of the baseline controls in the full sample (men and women) and the analysis sample (women only). Full sample specifications also include a control for apprentice gender. Standard errors, clustered at the firm level, are in parentheses. $*$ $p < 0.10$, $**$ $p < 0.05$, $***$ $p < 0.01$

Table A25: Labor supply—Robustness

	Working	Self-employment	Skilled self-employment	Unskilled self-employment	Wage employment	Skilled wage employment	Unskilled wage employment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Full sample (Baseline controls)							
Treatment	0.04	0.01	0.05*	-0.03*	0.00	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	2992	2992	2992	2992	2992	2992	2992
Mean of Dep Variable T=0	0.62	0.29	0.16	0.13	0.10	0.05	0.05
Panel B: Full sample (No controls)							
Treatment	0.04	0.01	0.04	-0.03	0.00	-0.00	0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	2992	2992	2992	2992	2992	2992	2992
Mean of Dep Variable T=0	0.62	0.29	0.16	0.13	0.10	0.05	0.05
Panel B: Women only (No controls)							
Treatment	0.03	0.01	0.04	-0.03	-0.00	-0.00	0.00
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Observations	2824	2824	2824	2824	2824	2824	2824
Mean of Dep Variable T=0	0.62	0.29	0.16	0.13	0.10	0.05	0.05

Notes: This table shows the results from estimating equation 2 for different labor market outcomes. All specifications include strata fixed effects and controls for the baseline household asset index (including a dummy for where the variable is missing). Full sample specifications also include a control for apprentice gender. Outcomes are stacked across a four-round retrospective panel that asked respondents to report on labor supply 1 month ago, 4 months ago, 7 months ago, and 10 months ago. Working includes wage-employment, self-employment, farming, and apprenticeships. Wage employment and apprenticeships are treated as separate categories of work. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A26: Summary statistics of time use (Firm owners)

	Mean	SD	Min	Max	N
Panel A: Percentage of total work hours					
Instruction	0.06	0.09	0.00	1.00	1,060
Supervising	0.04	0.09	0.00	0.69	1,060
Trade association	0.01	0.05	0.00	0.44	1,060
Learning new skills	0.00	0.03	0.00	0.38	1,060
Work in shop	0.62	0.23	0.00	1.00	1,060
Other	0.01	0.04	0.00	0.44	1,060
Idle time	0.26	0.19	0.00	1.00	1,060
Panel B: Percentage of idle hours					
Errands outside shop	0.28	0.29	0.00	1.00	1,006
Break from work	0.57	0.34	0.00	1.00	1,006
Waiting for customers	0.15	0.29	0.00	1.00	1,006

Table A27: Summary statistics of time use (Apprentices)

	Mean	SD	Min	Max	N
Panel A: Percentage of total work hours					
Instruction MCP	0.10	0.13	0.00	1.00	882
Instruction workers	0.01	0.04	0.00	0.88	882
Observation	0.06	0.12	0.00	0.75	882
Practice	0.04	0.10	0.00	1.00	882
Work in shop	0.57	0.23	0.00	1.00	882
Other	0.01	0.05	0.00	1.00	882
Idle time	0.21	0.12	0.00	1.00	882
Panel B: Percentage of idle hours					
Errands outside shop	0.45	0.30	0.00	1.00	843
Break from work	0.45	0.33	0.00	1.00	843
Non shop tasks	0.10	0.26	0.00	1.00	843

Table A28: Apprentice time use

	Firm owner instruction hours per day	Other worker instruction hours per day	Apprentice level Observation hours per day	Practice hours per day	Work on customer orders hours per day	Errand hours per day
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Full sample						
Treatment	0.10	0.05	0.03	0.01	-0.09	0.04
	(0.09)	(0.06)	(0.12)	(0.10)	(0.23)	(0.12)
Observations	595	595	595	595	595	595
Mean of Dep Variable T=0	1.29	0.34	1.39	0.96	3.93	0.59
Panel B: Women only						
Treatment	0.08	0.05	0.05	0.02	-0.08	0.08
	(0.09)	(0.07)	(0.12)	(0.11)	(0.24)	(0.12)
Observations	556	556	556	556	556	556
Mean of Dep Variable T=0	1.32	0.36	1.31	0.99	3.90	0.57

Notes: All specifications include strata fixed effects and controls for the baseline household asset index of the apprentice (including a dummy for where the variable is missing). Full sample specifications also include a control for apprentice gender. Standard errors, clustered at the firm level, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure S1: Timeline

Apprentice
Incentivized
Test and Incentive
Payments to Firms
(Aug-Oct)

Firm
Baseline
Survey
(May-Sept)

Rollout
Incentives
Program
(Jan)

Apprentice
Baseline
Survey
(Aug-Dec)

Training
Begins
(Oct-Jan)

Firm
Midline
Survey
(June)

Apprentice
Endline
Survey
(July-May)

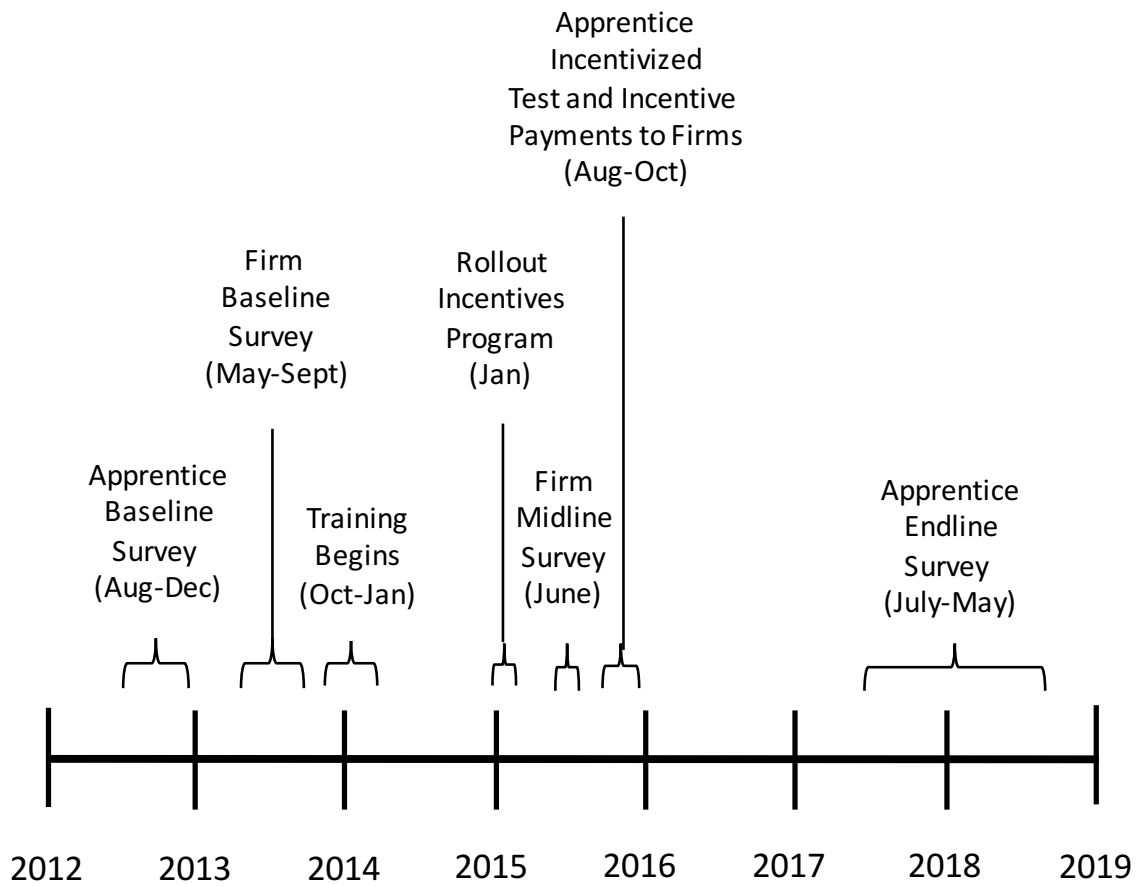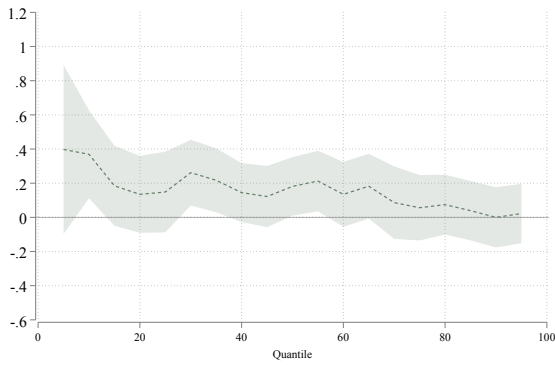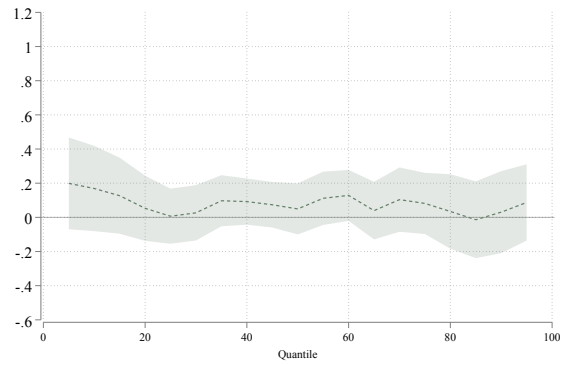2012   2013   2014   2015   2016   2017   2018   2019
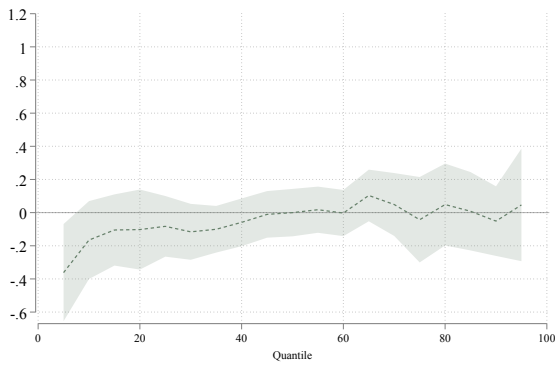
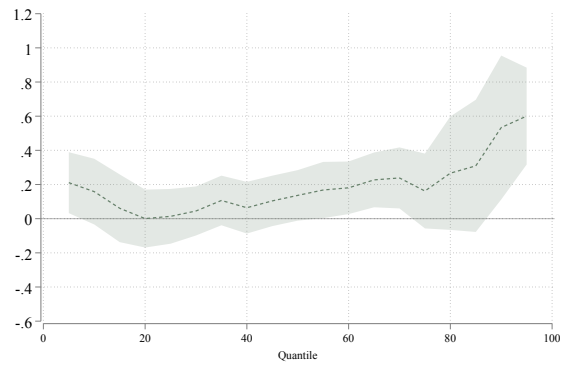Figure S2: Quantile treatment effects (Full sample)

(a) Craft skills (z-score)

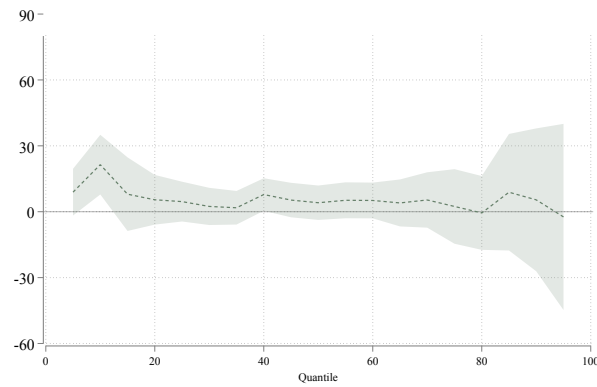(b) Practical component (z-score)

(c) Theoretical component (z-score)
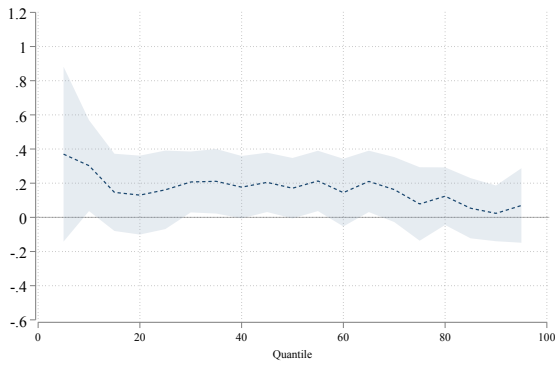
(d) Sale skills (z-score)
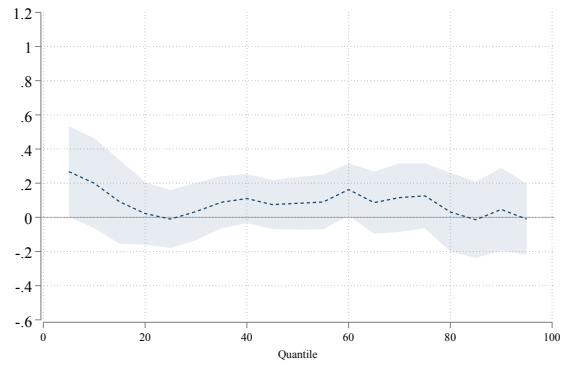

(e) Total earnings (GhC)

Note: This figures plots the coefficients from a quantile regression on several outcomes and their 95% confidence intervals. Standard errors are clustered at the firm level.

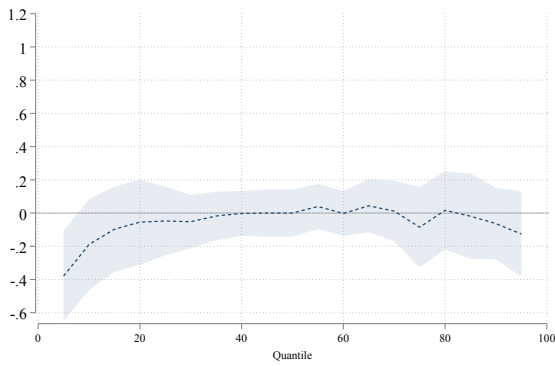Figure S3: Quantile treatment effects (Women only)
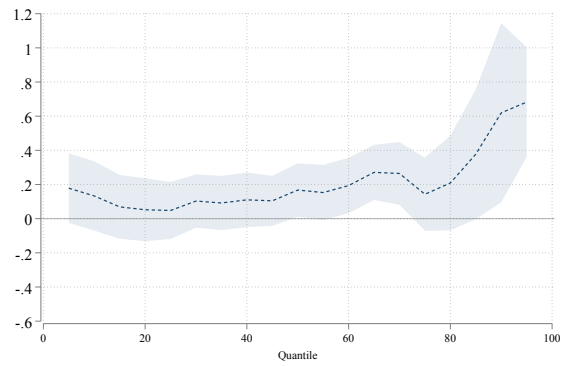
(a) Craft skills (z-score)
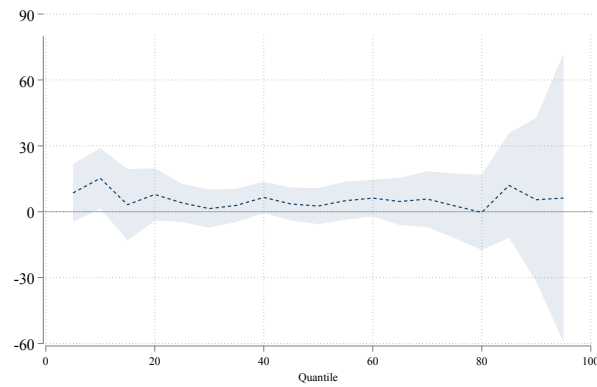
(b) Practical component (z-score)

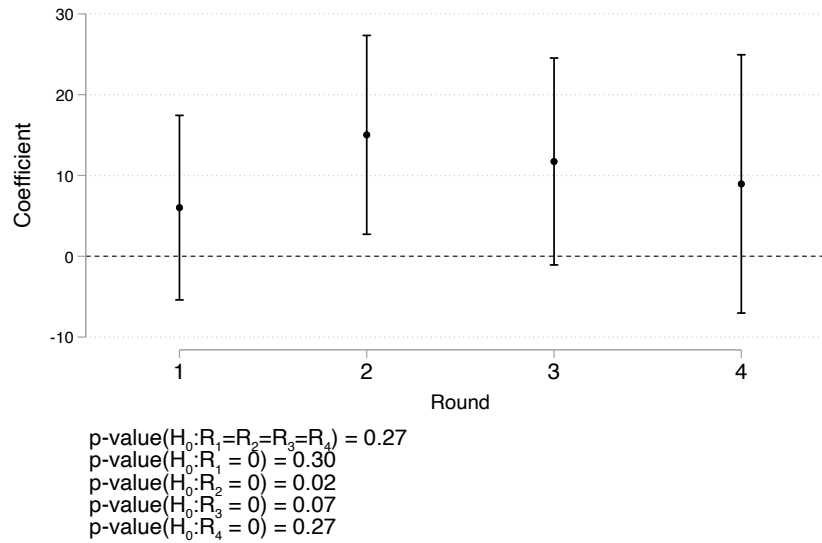(c) Theoretical component (z-score)

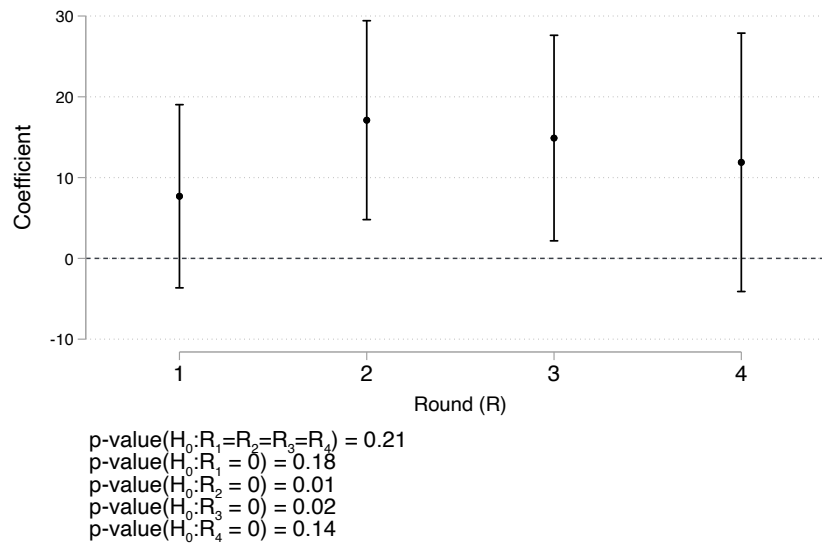(d) Sale skills (z-score)

(e) Total earnings (GhC)

Note: This figures plots the coefficients from a quantile regression on several outcomes and their 95% confidence intervals. Standard errors are clustered at the firm level.

Figure S4: Treatment effects on earnings (Full sample)

p-value(H_0:R_1=R_2=R_3=R_4) = 0.27
p-value(H_0:R_1 = 0) = 0.30
p-value(H_0:R_2 = 0) = 0.02
p-value(H_0:R_3 = 0) = 0.07
p-value(H_0:R_4 = 0) = 0.27

Note: This figure shows the dynamic treatment effects (and their 95% confidence interval) on earnings from the four rounds of the retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago from the endline survey.

Figure S5: Treatment effects on earnings (Women only)

p-value(H_0:R_1=R_2=R_3=R_4) = 0.21
p-value(H_0:R_1 = 0) = 0.18
p-value(H_0:R_2 = 0) = 0.01
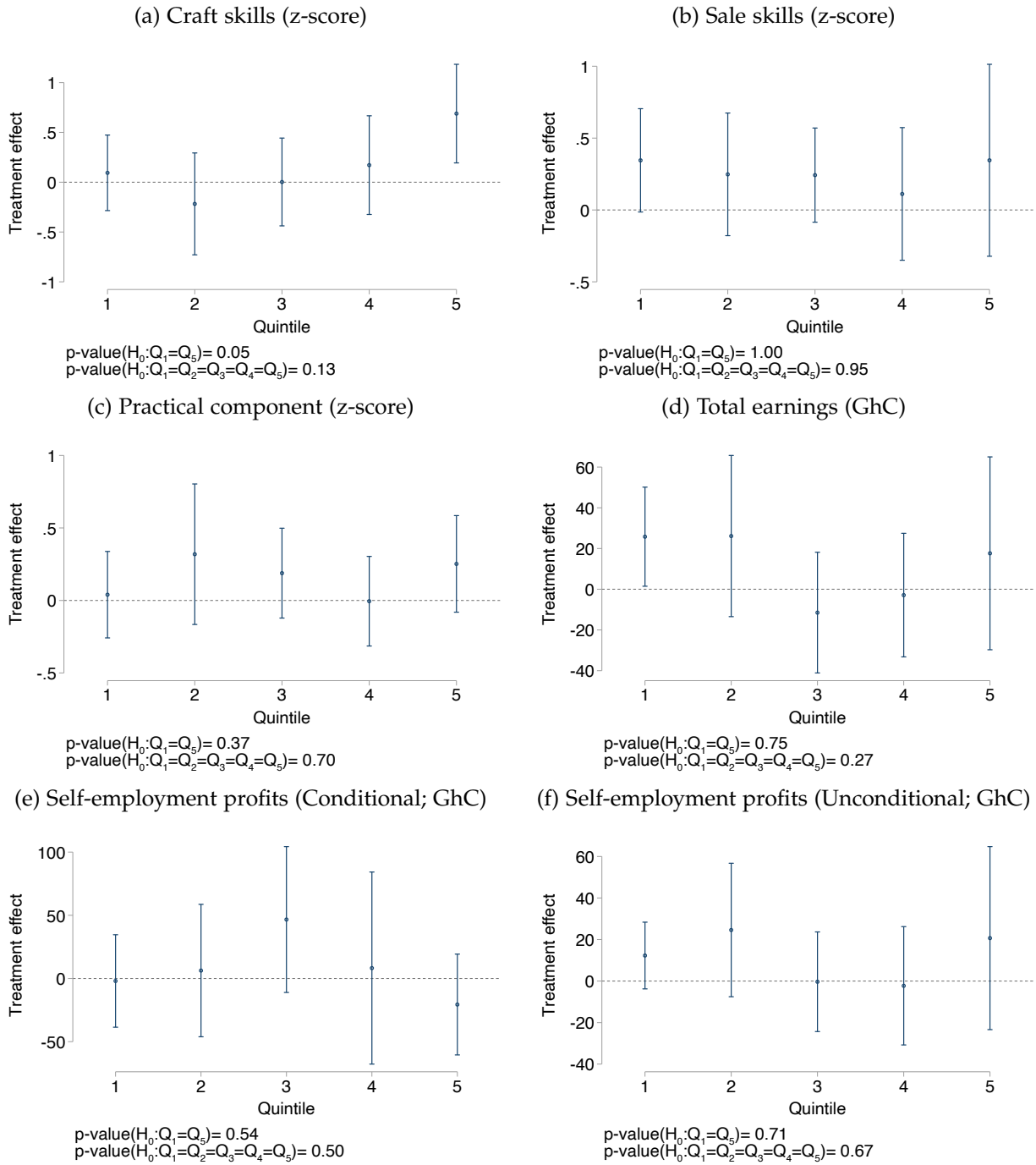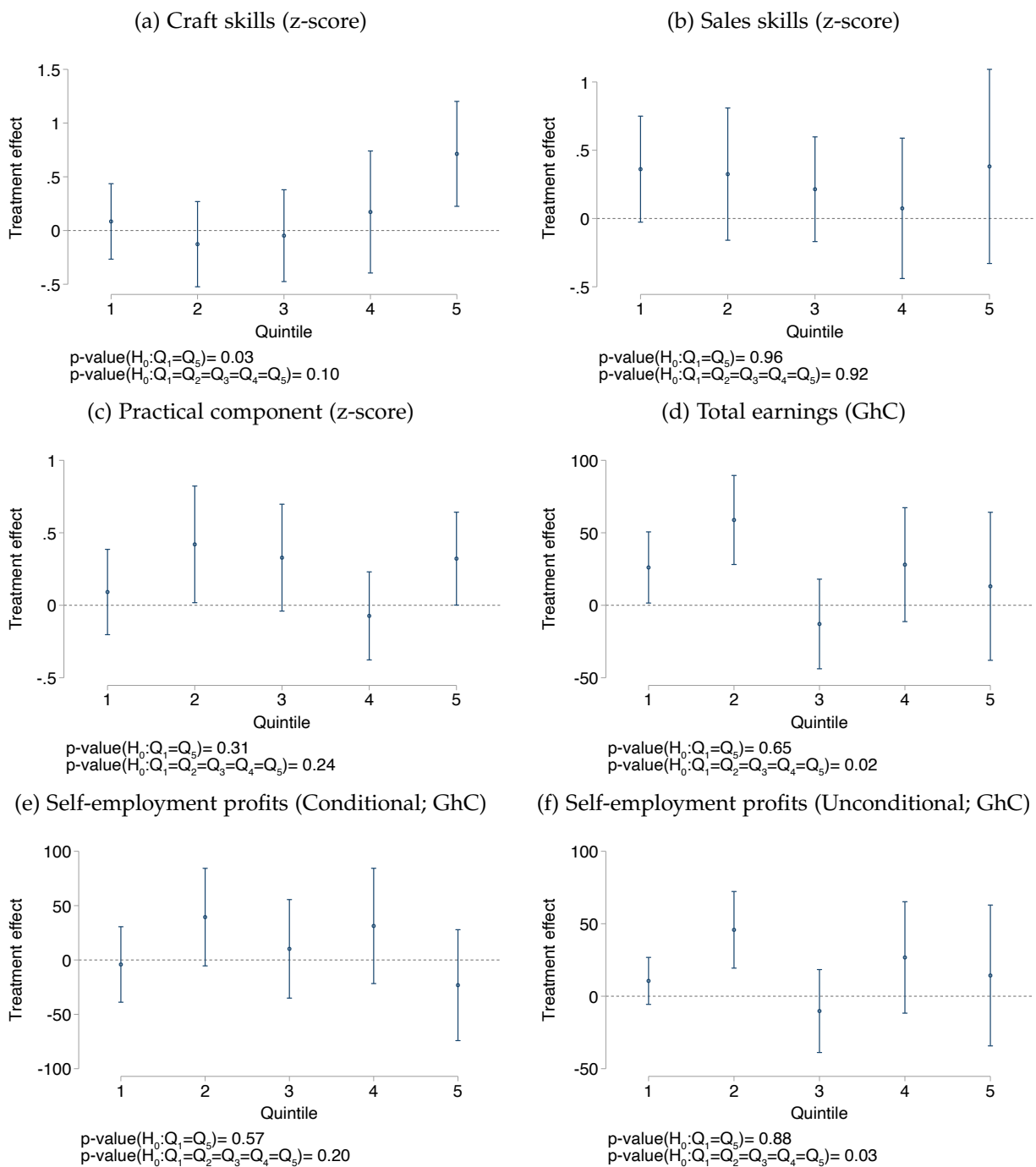p-value(H_0:R_3 = 0) = 0.02
p-value(H_0:R_4 = 0) = 0.14

Note: This figure shows the dynamic treatment effects (and their 95% confidence interval) on earnings from the four rounds of the retrospective panel that asked respondents to report on earnings 1 month ago, 4 months ago, 7 months ago, and 10 months ago from the endline survey.

Figure S6: Outcome heterogeneity by idle time (Full sample)

(a) Craft skills (z-score)

p-value(H_0:Q_1=Q_5)= 0.05
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.13

(b) Sale skills (z-score)

p-value(H_0:Q_1=Q_5)= 1.00
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.95

(c) Practical component (z-score)

p-value(H_0:Q_1=Q_5)= 0.37
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.70

(d) Total earnings (GhC)

p-value(H_0:Q_1=Q_5)= 0.75
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.27

(e) Self-employment profits (Conditional; GhC)

p-value(H_0:Q_1=Q_5)= 0.54
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.50

(f) Self-employment profits (Unconditional; GhC)

p-value(H_0:Q_1=Q_5)= 0.71
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.67

Note: This figure shows the treatment effect (and their 95% confidence intervals) in a series of outcomes (y-axis) by firm owners' quintile in idle time (x-axis). We also show p-values for testing whether the treatment effect for firm owners in the first quintile (least idle time) are the same as the effects for those in the last quintile (most idle time), and for testing whether the treatment effect is the same across all five quintiles.

Figure S7: Outcome heterogeneity by idle time (Women)

(a) Craft skills (z-score)

p-value(H_0:Q_1=Q_5)= 0.03
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.10

(b) Sales skills (z-score)

p-value(H_0:Q_1=Q_5)= 0.96
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.92

(c) Practical component (z-score)

p-value(H_0:Q_1=Q_5)= 0.31
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.24

(d) Total earnings (GhC)

p-value(H_0:Q_1=Q_5)= 0.65
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.02

(e) Self-employment profits (Conditional; GhC)

p-value(H_0:Q_1=Q_5)= 0.57
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.20

(f) Self-employment profits (Unconditional; GhC)

p-value(H_0:Q_1=Q_5)= 0.88
p-value(H_0:Q_1=Q_2=Q_3=Q_4=Q_5)= 0.03

Note: This figure shows the treatment effect (and their 95% confidence intervals) in a series of outcomes (y-axis) by firm owners' quintile in idle time (x-axis). We also show p-values for testing whether the treatment effect for firm owners in the first quintile (least idle time) are the same as the effects for those in the last quintile (most idle time), and for testing whether the treatment effect is the same across all five quintiles.

B   Bounds on Productivity Effects

Attanasio et al. (2011) evaluate a randomized vocational training program in Colombia and find that access to the program increased significantly both employment and earnings. They develop a method to disentangle productivity effects among workers from composition effects driven by increased employment. They arrive at the following bounds for the causal effects of productivity:

$$E(S|L = 1, R = 1) - E(S|L = 1, R = 0) \pm$$
$$[E(S(p(0.90))) - E(S(p(0.10)))]$$
$$\frac{Pr(L = 1|R = 1) - Pr(L = 1|R = 0)}{Pr(L = 1|R = 1)}$$

Where S is earnings (salary), L is labour supply and R is treatment status (randomized). $E(S(p(0.90)))$ and $E(S(p(0.10)))$ are the earnings of the control group among those working at the 90th and 10th percentiles respectively. Intuitively, it's the earnings difference between treatment and control groups bounded by how much the introduction of compliers (only working because of the program) may affect the composition of workers. Refer to pages 202-204 and Appendix B (pages 214-219) of their paper for derivation.

In our study, as per table A14, we find no significant effects on labour supply or even self-employment but we do see treated apprentices moving from unskilled self-employment into skilled self-employment. Thus, the earnings gains could be driven either by this shift into skilled employment or increased profits among those working in skilled self-employment. Note that both of these factors can be interpreted as productivity gains since all participants in this sub-sample of self-employed people at endline are working. Still, we can show that there are productivity gains even among those working in skilled self-employment. Adapting the equation above by setting L to be skilled self-employment and S to be profits in skilled self-employment, we obtain the following bounds. For women, 9.23 GhC to 65.80 GhC and in the full sample, 0.72 GhC to 67.62 GhC. Both bounds exclude 0 and are rather large suggesting strong productivity effects.