Apprentice Labor and Firm Output:
Evidence from a random match experiment

Morgan Hardy[*] and Jamie McCasland[†]

March 7, 2014

This document describes the analysis plan for the paper with the working title "Apprentice Labor and Firm Output: Evidence from a random match experiment". This draft was completed and registered prior to the analysis of any midline or follow-up data and therefore provides a useful reference in evaluating the final results of the study.

The plan is outlined as follows: Section 1 reviews the motivation, context, sample selection, and randomization procedures; Section 2 describes the data; Section 3 enumerates the hypotheses to be tested; and Section 4 discusses the specifications to be used in analyzing the data.

1 Study Design

Microenterprises with few employees dominate both the firm size distribution and labor market participation throughout much of the developing world, including in our study country, Ghana (Sandefur (2010)). Employment growth among microenterprises is minimal and could be hampered by social norms which make it difficult to pay low or negative wages during on-the-job training, and search frictions which make it costly to find and hire suitable workers. Recent research finds that wage subsidies effectively drove short-term but not long-term wage employment among microenterprises in Sri Lanka (De Mel, McKenzie and Woodruff (2013)). The West African apprenticeship context differs in that it hinges more critically on the relative costs of on-the-job training, and the persistence of firm monopsony power (Frazer (2006)). Further, though general skill in the manufacturing and services trades common in the Ghanaian informal sector is quickly demonstrable, aptitude for work and ability to learn are difficult to measure and potentially quite important in the determination of employment growth in microenterprises (Acemoglu and Pischke (1998)).

First, this study seeks to experimentally measure the effect of apprentice labor, defined both as additional apprentices in a firm and additional apprentice labor hours, on firm output, adding to the literature on labor inputs in microenterprises with a novel context. Second, the study seeks to unpack the mechanisms and dynamics driving any effect, using random variation in the baseline characteristics of training firms, apprentices, and the match between the two, and detailed repeated data on training inputs, apprentice progress, and firm output during the training period.

1.1 Context

This study randomizes the match between apprentices and firms recruited to participate in the National Apprenticeship Program (NAP), a collaborative project initiated by the parliamentary Coun-

---

[*]Department of Economics, Brown University. Email: mlhardy@gmail.com

[†]Department of Economics, UC Berkeley. Email: jamie.mccasland@gmail.com

cil for Technical and Vocational Education and Training (COTVET), implemented by district-level Technical and Vocational Education and Training (TVET) coordinators of the Ghana Education Service (GES), and evaluated by Innovations for Poverty Action-Ghana (IPA), in close partnership with the Ghana Tailors and Dressmakers Association (GHABA), the Ghana Hairdressers and Beauticians Association (GHABA), and other craft-specific trade associations. The authors have developed partnerships with each of these institutions and gratefully acknowledge their support. The randomization procedure also creates random variation in the number of apprentices assigned to each participating firm (with some being assigned zero apprentices).

The NAP mirrors the widespread traditional apprenticeship institution in Ghana, which includes an up front fee at the commencement of the apprenticeship and often requires apprentices to bring tools for training to the training firm. The slated time period for the NAP is one year, but we anticipate many apprentices extending their tenure at training firms in the capacity, at least in title, of apprentice (something that itself is somewhat common). Because the government program has myriad implementation challenges, distribution of tools and fee payment is indefinitely delayed (again, something that itself is not uncommon among apprentices unaffiliated with the government). At present, apprentices are training in participating firms despite this delay.

Recruitment of apprentices and training firms took place independently and required a structure for placing apprentices with firms in an arguably fair and geographically feasible way. To facilitate this process, the project partners undertook 149 district and craft level "matching meetings" where interested firms were asked to provide location, craft specialty, and languages spoken to program apprentices. The matching meeting activities were heavily monitored, though unmonitored communication between participants was common. After all formal information sharing on the part of firms, apprentices were asked to list and rank firms with which they would like to train. The "feasible sets" used in the randomization therefore include both geographic feasibility (walkability, generally) and idiosyncratic preference. No minimum or maximum was placed on the number of ranked firms per apprentice, though anecdotally feasible sets appear to be an honest revelation of preferences. The mean feasible set is quite small, at 2.2 firms. Firms are assigned between zero and eight apprentices as an artifact of the randomization to be described below, with a mean and median of one apprentice per firm.

1.2   Sample

The NAP evaluation takes place in 32 districts nationwide, where 1,197 apprentices and 1,087 firms participated in matching meetings. Logistical challenges on the part of the implementing partners led to significant delays in all districts, and the rolling out of the start of apprenticeships in three phases. 21 districts, 657 apprentices, and 684 firms made up Phase 1; 7 districts, 388 apprentices, and 280 firms made up Phase 2; and 4 districts, 152 apprentices, and 123 firms made up Phase 3. Because Phase 3 apprentices started their training in early 2014, they were excluded from the first January 2014 midline survey. At present, their inclusion in future surveys is undecided, again for purely logistical reasons.

1.3   Randomization

The randomization procedure was quite simple. Given the feasible set of each apprentice, a random training firm was chosen using a computer generated random number. No re-randomization or stratification beyond individual apprentice was done. An example district and trade is replicated in Figure 1.

Expressing random variation in the treatment assignment as the actual treatment assignment

less the expected treatment assignment implicit in the feasible sets-based randomization, Figures 2 through 5 represent the variation in the treatment variables. Cognitive test scores are calculated as the mean of assigned (or potential for the expected values) apprentices.

2   Data

Under current planning, data will come from four sources: (1) Apprentice baseline surveys, (2) Firm baseline surveys, (3) Feasible sets and other data collected at the matching meetings, and (4) Four midline surveys, to be carried out in January, March, May, and July 2014, though funding for the fourth of these midline surveys is pending.

Of the 1,197 participating apprentices, 1,165 of them participated in a baseline survey which included educational, training, and work background, a series of cognitive tests, risk preferences, non-cognitive measures of fatalism and self-esteem, information on credit access and network strength, expectations over future earnings, household assets, and intergenerational exposure to the trade of choice. Of the 1,087 firms, 1,074 of them participated in a baseline survey which included personal background, digit span recall, firm capital, firm labor inputs, firm sales and profits, managerial aptitude questions, and information on apprenticeship training experiences. Each participant in the matching meeting gave a set of "walkable" firms, and a set of "feasible" firms with whom they would like to train, as well as answered a series of questions about each of these firms.

Training commencement was staggered across districts and began between October 2013 and March 2014. Midline surveys capture training inputs, apprentice progress, owner, apprentice and other worker labor inputs, firm sales and profits, and stated evaluation of apprentice performance by the firm owner to date.

3   Hypotheses and Outcomes of Interest

We group hypotheses as follows:

1. *Quantity Hypotheses*: This group uses random variation in the number of apprentices assigned to each firm to test for the effect of apprentice labor on firm output and other outcomes of interest.

2. *Quality Hypotheses*: This group uses random variation in the baseline characteristics of apprentices and firms to experimentally test for heterogeneity in outcomes.

3. *Training Inputs and Dynamics Hypotheses*: This group of hypotheses pertains to mechanisms and training quality more than firm outcomes.

4. *Compliance Hypotheses*: This group of hypotheses explores the structure of non-compliance and its relationship to feasible sets data.

5. *Auxiliary Hypotheses*: This group of hypotheses looks at relationships outside basic firm output and firm-apprentice match.

The midline survey which was administered in January 2014, less the tracking and personally identifiable information sections in included in Appendix 1, and referred to throughout this section with the pre-notation M. Future midline/follow-up surveys are expected to be quite similar. In addition, the firm baseline and apprentice baseline are likewise attached as Appendices 2 and 3, and will be referred to with the pre-notations F and A, respectively.

Note that apprentice labor will be defined as number of NAP apprentices or total NAP apprentice labor hours throughout.

3.1 Quantity Hypotheses

H1: Apprentice labor increases sales, profits, and number of customers on average

Profits, Sales, and Number of Customers are questions M:C2/M:C6/FD:45, M:C1/M:C15/F:D44 and M:C3/M:C7/F:D46 respectively and follow De Mel, McKenzie and Woodruff (2008) in wording and administration. Our experience in the field suggests that profits and sales are easier to collect for the time period LAST MONTH, while hours are easier to collect for the time period LAST WEEK. However, weekly profits, sales and customers was added to the midline survey to reflect the potential timing mismatch associated with surveys conducted past the first week of any given month. Each of these measures will be trimmed at 99%.

H2: Apprentice labor increases sales, profits, and number of customers more in firms closer to production capacity

M:C4 and M:C8 are meant to directly measure how close a firm is to production capacity at the time of the midline survey. Of course, this is also potentially affected by treatment. Accordingly, near production capacity will also be proxied with an indicator variable for above median customers per worker in the baseline survey using the sum of F:D28a-F:D28f and F:D46.

H3: Apprentice labor decreases firm owner labor hours (and more so for women and those caring for children)

Firm owner labor hours are captured with F:D25 and F:D26 at baseline and M:A5 and M:A6 at midline. We expect women, particularly those in cosmetology, to substitute away from work in the firm and toward household work which is captured by M:A7. Additionally, the firm baseline captured a series of questions on motivation for self-employment which follow De Mel, McKenzie and Woodruff (2008), F:D65-F:D68. Particularly relevant to this hypothesis is F:D65, which asks whether the following is an important motivation for self-employment: "The ability to look after children while I am working."

H4: Apprentice labor increases total firm size over the course of the study

This hypothesis is essentially the first stage of an instrumental variables specification using random variation induced by treatment assignment as an instrument for actual number of workers or total labor hours in the firm. Two threats to the strength of the first stage are important. The first is standard non-compliance by which apprentices do not report to their assigned firms (or to any firm), which will be discussed below. The second is where firms substitute away from accepting traditional apprentices or hiring workers during the NAP apprentice training period. We expect very limited substitution early in the training period and perhaps more over the course of the training period. However, the first stage we expect to be strong throughout the study period. Number of "traditional" apprentices is captured at baseline by F:D28d and at midline by M:A12d. Total number of workers is the sum F:D28a-F:D28f at baseline and the sum M:A12a-M:A12f at midline.

3.2 Quality Hypotheses

H5: Apprentices with experience in the trade have a larger effect on firm output

Direct apprentice experience with the trade is captured by A:B7 and A:B10. Indirect apprentice experience with the trade, via siblings or parents, is captured by A:C5, A:C11, A:C15, A:C18, A:C21, and A:C24, which will be used to create a dummy variable for exposure to the trade.

H6: Apprentices with more formal education have a larger effect on firm output

Apprentice formal education is captured by A:C3. Literacy and numeracy are directly measured in the apprentice baseline survey and are anecdotally quite important to ability to learn in trades where measurements and accuracy are important. Formal education is highly correlated with these skills and captures them to some degree, but also captures perhaps other important non-cognitive skills related to experience with formal education.

H7: Apprentices with better cognitive and non-cognitive test scores have a larger effect on firm output

Cognitive and non-cognitive measures are captured in sections D, F, K, L, and M of the apprentice baseline. We expect cognitive performance to be associated with higher productivity and that productivity to benefit firm sales and profit. Non-cognitive measures of fatalism and self-esteem (measured by Rotter Locus of Control and Rosenberg Self-Esteem scales) have been shown to be associated with labor market outcomes in the United States and we expect them to be relevant in this setting as well (Heckman and Urzua (2005)).

H8: Older firms and firms run by people with more apprenticeship training experience benefit more from apprentice labor

Baseline firm age is measured as F:D1 and F:D2. The years of operation and year started business are both asked to increase accuracy for a variable which is prone to measurement error. About 1.5% of respondents reported values for F:D1 and F:D2 that do not accord. For these cases, we averaged the implied firm age between the two noisy measures. Apprenticeship training experience is measured by F:D32 and F:D33 at baseline.

H9: Firms run by people with more formal education benefit more from apprentice labor

Firm owner formal education is measured by F:B7. We expect firm owners with better formal education to be better able to develop and execute structured syllabi. We also expect well educated firm owners to be able to communicate and teach especially well with well educated apprentices. In addition, there is unfortunately no direct measure of literacy in the firm baseline, so formal education will proxy for literacy.

H10: Firms run by people with higher cognitive test scores benefit more from apprentice labor

Cognitive performance by firm owners is measured by F:C1a-F:C7b and F:C10-F:C13. Correct digits forward values will be summed and standardized. F:C8 and F:C9 are single question measured of the Rotter Locus of Control and Rosenberg Self Esteem measures captured with more questions in the apprentice baseline.

H11: Firms with a higher capital to labor ratio benefit more from apprentice labor

Firm capital at baseline is measured by F:D14, F:D16, F:D18, F:D19, F:D20, F:D22, and F:D23. Two sums will be used: the full total, and the full total less land and buildings, measured by F:D14 and F:D16. The sums will be trimmed at 99%. In addition, each of these questions comes with an innovative measure of the certainty of the respondent with respect to these difficult to estimate values. We will test how these certainty measures affect our estimates.

H12: Firms run by people with better measured managerial aptitude benefit more from apprentice labor

Managerial aptitude is measured by questions F:D52-F:D60, from which we will create an index. In addition, apprentice training perspectives of the firm owner may be directly relevant to the apprenticeship training experience and the ability of the firm owner to improve apprentice productivity early in the training period. Several of these measures are captured in the firm baseline survey (e.g. F:D41 "How long does an apprentice train with you before having access to all tools and machinery in your business?")

H13: Firm benefit varies by craft

The nature of the businesses in the five crafts in this study (Garmentmaking, Cosmetology/Hairdressing, Welding, Blocklaying/Construction, and Carpentry/Construction) varies quite a bit. Some have much more variable work assignments, such as the construction trades. Others have a more important retail component. There is also variation in the degree to which work can be accomplished in the absence of clients (Cosmetology is essentially fully retail, in the sense that no projects continue when the shop is empty of clients). These issues will likely alter the findings by trade in predictable ways. In main specifications, craft will simply be controlled for, but sub-group analysis will also be done by craft.

H14: Firm benefit varies by gender

A growing literature finds that female-owned firms are less profitable and less productive than male-owned firms in the developing country microenterprise context. Though trade subsumes the majority of gender variation, garmentmaking is a dual gender trade, and we expect to find that female owned enterprises benefit less from being assigned additional apprentices and from being assigned particularly cognitively or experientially high quality apprentices.

H15: Matches that pair a high-ability apprentice with a high-ability firm owner will lead to the largest effect on firm output

Ability here will be measured by experience, formal education, cognitive performance, and non-cognitive attitudes. This is one of the central hypotheses of the study and the most novel potential contribution, in the sense that it unpacks the relatively limited literature on labor in microenterprises by skill/ability type. It may be the case that we find these high-ability/high-ability matches to be the only ones that truly benefit firms, which would carry with it important policy implications.

3.3 Training Inputs and Dynamics Hypotheses

H16: Apprentices who expect a higher return to training invest more training hours

Questions A:J6-A:J9 measure expectations over returns to apprenticeship training. Training hours depend on both apprentice time investment and firm time investment (from the owner or other

workers or senior apprentices at the firm). This is an important intermediate hypothesis in that it could explain apprentice labor investment. It follows that firms which are assigned apprentices who expect high returns to apprenticeship training benefit more.

H17: Firm owners invest more training hours in higher ability apprentices

We expect firm owner and other worker investment in training hours to vary with apprentice ability, though the sign of the effect is difficult to predict. It could be the case that firm owners invest more heavily in lower ability or struggling apprentices out of a social or altruistic norm to help them succeed. However, it would be rational to invest more heavily in higher ability or excelling apprentices who we expect to benefit more from training investment in terms of adding to productivity and general progress in the trade. Training investment is measured by M:B1.9/M:B2.9/etc.

H18: High ability firm owners are able to appropriately score apprentice performance

Asymmetric information and search costs are an important theoretical motivation for this study. We seek to understand the information revelation processes over the course of the training via M:B1.16/M:B2.16/etc. and expect higher ability firm owners to be better predictors of apprentice quality. Here apprentice quality can be measured by apprentice cognitive scores at baseline, but a more accurate measured proxy of apprentice quality will depend on findings in other parts of the study.

H19: High ability apprentices receive more tips from firm owners

Tips and wages are measured by M:B1.12/M:B2.12/etc. and M:B1.13/M:B2.13/etc. In previous work on apprenticeships in West Africa, wages are essentially assumed to be zero during the training period. In our observation, this is not entirely accurate and could mask some important co-trending of increases in productivity and increases in tips. Though the firm holds some monopsony power over the apprentice during the training period, exiting the apprenticeship and finding work without the completed certification is definitely a decision open to apprentices.

H20: More training hours lead to faster skill progression

This will include both training hours measured in M:B1.9 and total work hours measured in M:B1.6-M:B1.8. Skill progression will be measured in M:B1.14 and MB1.15.

H21: Higher ability apprentices will progress in the craft more quickly

Ability here will be measured by experience, formal education, cognitive performance, and non-cognitive attitudes. Skill progression will be measured in M:B1.14 and MB1.15.

H22: Firm output will decrease during the early heavy training periods and increase in the later periods

A simple 2 period model of on-the-job training during the apprenticeship would suggest that heavy training investment early in the apprenticeship could draw firm owner attention away from firm profitability/customers. Where high-ability apprentices have a higher productivity return to training investment, firms would benefit in the second period from investments in high-ability apprentices made in period 1.

H23: Training hours will be largest early in the training period

This hypothesis also relates to the model in H22.

H24: The firm output gap between firms with low and high ability apprentices will increase over time

This hypothesis also relates to the model in H22.

3.4 Compliance Hypotheses

H25: Firms owners with higher cognitive test scores were ranked more highly on average

This set of hypotheses relates to a secondary analysis of compliance and the preferences captured by the feasible sets. Based on other findings in the study, we will be able to characterize firms that are able to provide good training or pay high tips. We can then back out whether those are more preferred firms, and if so by whom. One hypothesized predictor of firm training quality is firm owner cognitive ability.

H26: Apprentices who were placed with a firm they ranked more highly are more likely to comply with firm assignment

We predict that compliance will depend on apprentice preference. This analysis will allow us to understand matching in the absence of the randomization, and potentially simulate counterfactuals.

H27: Apprentices who were placed with a firm they ranked more highly invest more training hours

Here training hours can be interpreted as another measure of compliance.

3.5 Auxiliary Hypotheses

H28: Apprentices placed with peers of high ability progress faster and invest more training hours

Peer effects could be potentially very important and are also essentially randomized by this design. We will use the measures delineated above for apprentice compliance, progression and ability to measure whether being placed with a higher ability peer (or any peer) improves apprentice outcomes.

H29: Apprentices placed in firms with more existing senior apprentices progress more quickly

NAP apprentices' peers will not only be other NAP apprentices but also wage workers and other apprentices in the firm. Qualitative interviews suggest that in firms with experienced senior apprentices, a large fraction of the training is done by these senior apprentices rather than the firm owner. The sign of this effect is difficult to predict, as more experienced workers in the firm could lead to specialization on the part of the NAP apprentice in rudimentary tasks, but overall we expect the access to more coaching to benefit NAP apprentices.

H30: Apprentices placed in firms with a better educated existing workforce progress faster and invest more training hours

Skill complementaries suggest that groups of high quality workers are associated with multiplier effects, and we will look for variants of that pattern in this dataset as well.

H31: Apprentices placed in firms with a better educated existing workforce benefit firm output more

This hypothesis relates to hypothesis H30.

4   Specifications

At present, the precise specifications are still under review. We seek to ensure that we properly control for feasible sets and thereby generate unbiased estimates. Additionally, we seek to maximize power, measure average effects, and measure differential effects over time. The current specification under review for the most basic quantity hypothesis, which assumes that effects are constant across rounds, follows McKenzie (2012) pooling data across midline/follow-up rounds 1-3 (or 1-4) as follows:

$$Y_{it} = \beta_0 + \beta_1 T_i + \beta_2 E[T_i] + \beta_3 Var[T_i] + \beta_4 Y_{i0} + \beta_5 NT_i + \eta_t + \epsilon_{it}$$

where Y_{it} is the outcome variable in round t, T_i is the treatment assignment (number of apprentices), $E[T_i]$ is the expected value of the treatment assignment, $Var[T_i]$ is the variance of the treatment assignment, Y_{i0} is the baseline measure of the outcome variable, NT_i is the number of apprentices who ranked firm i, and η_t are round dummies. Additionally, we will run tests to determine whether effects are constant across rounds, specifications by round, and specifications which include a vector of other controls. Adjustments to the standard errors based on the multiple hypotheses in each hypothesis group will also be used.

We hope the reader will interpret the present specification indecision as indicative of a steep graduate student learning curve rather than intentional opaqueness.

Non-compliance will be defined as failure to report to the apprenticeship at the assigned firm, whether the apprentice reports to a different firm or does not report at all. We will account for non-compliance with an instrumental variables specification.

Balance across treatment groups is not straightforward given this randomization, but will be estimated using individual and joint tests of independence between firm characteristics and treatment assignment in firm output hypotheses, and between apprentice characteristics and treatment assignment in apprentice outcome hypotheses.

Survey attrition is not yet available as the midline data from the first round has just become available and future rounds have not yet been collected.

References

Acemoglu, Daron, and Jorn-Steffen Pischke. 1998. "Why Do Firms Train? Theory and Evidence." *Quarterly Journal of Economics*, 113: 79 – 119.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *Quarterly Journal of Economics*, 123(4): 1329–1372.

De Mel, Suresh, David McKenzie, and Christopher Woodruff. 2013. "Making the Leap from Self-Employed to Employer? What matters - capital, labor, or training?" *Working Paper.*

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

Heckman, James, Jora Stixrud, and Sergio Urzua. 2005. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *IZA Working Paper.*

McKenzie, David. 2012. "Beyond Baseline and Follow-up: The Case for more T in Experiments." *Journal of Development Economics.*

Sandefur, Justin. 2010. "On the Evolution of the Firm Size Distribution in an African Economy." *CSAE Working Paper.*

Figure 1: Randomization Example - Awutu Senya District, Garments

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
Firm 1	Ranked 1	Ranked 3										
Firm 2	Ranked 2	Ranked 2	Ranked 3									
Firm 3	Ranked 3											
Firm 4		Ranked 4	Ranked 2	Ranked 2								
Firm 5			Ranked 4									
Firm 6					Ranked 1							
Firm 7		Ranked 1		Ranked 1		Ranked 1						
Firm 8			Ranked 1				Ranked 6					
Firm 9	Ranked 4		Ranked 6									
Firm 10			Ranked 5					Ranked 4				
Firm 11							Ranked 2		Ranked 4			
Firm 12						Ranked 3						
Firm 13						Ranked 2						
Firm 14						Ranked 4						
Firm 15							Ranked 4			Ranked 1		
Firm 16							Ranked 3			Ranked 2		
Firm 17							Ranked 5	Ranked 2	Ranked 2	Ranked 3		
Firm 18							Ranked 1	Ranked 3	Ranked 1	Ranked 4	Ranked 2	
Firm 19								Ranked 1	Ranked 3	Ranked 5	Ranked 1	Ranked 1

11

Figure 2

Variation in Labour Quantity Assignment

Figure 3

Variation in Average Ravens Score

Figure 4

Variation in Average Literacy

Figure 5

Variation in Average Digits Forward Score