The Selection and Causal Effects of Work Incentives on Labor Productivity: Evidence from a Two-stage Randomized Controlled Trial in Malawi*

Bryant Hyuncheol Kim†, Seonghoon Kim‡, and Thomas T. Kim§

30 January 2016

Abstract

Incentives are essential to promote labor productivity. We implemented a two-stage experiment to measure effects of career and wage incentives on productivity through self-selection and causal-effect channels. First, workers were hired with either career or wage incentives. After employment, a random half of workers with career incentives received wage incentives and a random half of workers with wage incentives received career incentives. We find career incentives attract higher-performing workers than wage incentives but do not promote productivity for existing workers. Instead, wage incentives promote productivity for existing workers. Observable characteristics and training performance are limited in explaining the selection effect.

Keywords: Career Incentive, Wage Incentive, Internship, Self-selection, Labor Productivity

JEL Classification: J33, O15, M52

* We are grateful to the following staff members of Africa Future Foundation for their excellent field assistance: Narshil Choi, Jungeun Kim, Seungchul Lee, Hanyoun So, and Gi Sun Yang. In addition, we thank Jim Berry, Syngjoo Choi, Andrew Foster, Dan Hamermesh, Guojun He, Kohei Kawaguchi, Asim Khwaja, Etienne Lalé, Kevin Lang, Suejin Lee, Pauline Leung, Zhuan Pei, Cristian Pop-Eleches, Victoria Prowse, Imran Rasul, Nick Sanders, Slesh Shrestha, and Armand Sim as well as seminar participants at Cornell University, Hitotsubashi University, Korea Development Institute, National University of Singapore, Singapore Management University, Seoul National University, NEUDC 2016, SJE International Conference on Human Capital and Economic Development, and UNU-WIDER Conference on Human Capital and Growth for their valuable comments. This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant. All errors are our own.

† hk788@cornell.edu; Department of Policy Analysis and Management, Cornell University

‡ seonghoonkim@smu.edu.sg; School of Economics, Singapore Management University

§ krutaeks@gmail.com; Department of Economics, Yonsei University

1. Introduction

Work incentives are an essential means of human resource management to recruit productive workers and motivate them to exert more effort. Work incentives can affect labor productivity through selection and causal effect channels. For example, Lazear (2000) shows that performance pay, as opposed to fixed hourly wages, increases labor productivity not only because more productive workers self-select into jobs but also because it motivates workers to increase their effort levels. A better understanding of these channels through which incentives affect labor productivity would allow firms to design optimal hiring and compensation strategies to maximize labor productivity and reduce the need for costly screening processes. However, empirical evidence on these channels is limited because it is difficult to separately isolate the productivity-enhancing effect of work incentives (*causal effect*) and the change in productivity through endogenous worker sorting (*selection effect*).[5]

In this study, we provide experimental evidence on how career and wage incentives affect labor productivity through worker selection and causal effect channels. We conduct a two-stage field experiment to separately isolate the selection and causal effects of these incentives in collaboration with Africa Future Foundation (AFF), an international non-governmental organization (NGO).[6] The career incentives we study are a future job prospect and a recommendation letter, which are typical benefits of an internship position.[7] Wage incentives in our setting are a lump-sum salary and performance-related bonus payment.[8]

Our experimental design was implemented in the context of a recruitment drive of census enumerators. To hire enumerators for a population census of Chimutu, a rural district in Malawi, AFF approached 536 recent high school graduates. As shown in Figure 1, in the first stage, study subjects were randomly assigned to one of three groups: (i) those who received a job opportunity with career incentives of an internship (hereafter the *Internship* group), (ii) those who received a job opportunity with wage incentives (hereafter the *Wage* group), and (iii) those who did not receive any job opportunity

[5] The *causal* effect refers to the difference in labor productivity when incentives affect performance holding employee composition constant. The *selection* effect refers to the difference in labor productivity driven by workers' self-selection into the job.

[6] Our two-stage experimental design is similar to recent experimental studies in development economics (Karlan and Zinman, 2009; Ashraf et al., 2010; Cohen and Dupas, 2010; and Beaman et al., 2015).

[7] An internship is a temporary position that can be paid or unpaid, and is distinguished from a short-term job in that it emphasizes on-the-job training for students or entry-level workers. According to a 2011 survey of the US-based National Association of Colleges and Employers, more than 50% of graduating college students had internship experiences (Nunley et al., 2016). Internship programs are also widely available in Malawi in the public, private, and NGO sectors. For example, about 20% of regular workers in AFF are hired through the internship program.

[8] While promotion and a pay raise often occur jointly in the workplace, we study career and financial incentives as distinct components of work incentives.

(hereafter the control group).[9] Those assigned to the *Internship* group received an internship opportunity that comes with (a) a potential long-term employment opportunity at AFF as a regular employee and (b) a recommendation letter specifying their relative job performance.[10] A one-time temporary work opportunity was offered to those assigned to the *Wage* group with a lump-sum wage and bonus payment based on job performance. Individuals who accepted the job opportunity in the first stage proceeded to enumerator training and the second-stage randomization. In the second stage, a randomly selected half of the trainees in the *Internship* group additionally received the same wage incentives of the *Wage* group by surprise after completing the training. In the same manner, a randomly selected half of the trainees in the *Wage* group additionally received the same internship incentives as the *Internship* group by surprise.

This research design allows us to obtain two sub-groups, Group 2 (hereafter G2) and Group 3 (hereafter G3), which have identical incentives (both career and wage incentives) during the work, but the channels through which they were attracted to the job are different. As a result, we isolate the selection effect on labor productivity by comparing G2 and G3.[11] In addition, we estimate the causal effects of career and wage incentives on job performance by comparing job performance between Group 1 (hereafter G1) who only have career incentives and G2, and G3 and Group 4 (hereafter G4) who only have wage incentives. G1 and G2 workers became enumerators through career incentives but only G2 received additional wage incentives. Any difference in performance between G1 and G2 can be interpreted as a causal effect of wage incentives in the *Internship* group. Similarly, any difference in performance between G3 and G4 can be interpreted as a causal effect of career incentives in the *Wage* group.[12]

The nature of an enumerator job is multidimensional. We measure job performances by number of surveys conducted per day (survey quantity), survey error rate (survey quality), and subjective performance evaluations (SPEs) by census respondents as well as supervisors in AFF.

We used three data sources: AFF's administrative data of candidates' training and enumerators' daily job performance, Chimutu population census data, and AFF's surveys of the study participants which collect rich information on observable individual characteristics. Our rich and high frequency data enables us to estimate selection and causal effects of work incentives precisely and to explain its potential mechanism.

---

[9] Job opportunity was conditional on the successful completion of training.

[10] An entry-level regular position (enumerator or data entry clerk) at AFF has career advancement prospects that lead to more advanced positions, such as head enumerator, junior project assistant, senior project assistant, and project manager. AFF did not explicitly state the actual probability of hiring for the *Internship* group. We acknowledge that changing probabilities of hiring after the internship might affect effort levels, but we do not compare different levels of the same incentive, but rather two different types of incentives.

[11] Due to the nature of our experimental design, the selection effect of either career incentives or wage incentives can be evaluated against the other type of incentive.

[12] We discuss potential threats to the identification in Section 4.

Of 536 recent high school graduates AFF approached for the baseline survey of this study without a prior notice of a job opportunity, 443 (83%) participated in the baseline survey. Of 176 participants assigned to the *Wage* group, 74 (39.8%) accepted a job opportunity by joining the training session. 74 (42.0%) of 186 participants assigned to the *Internship* group took up the job opportunity. The fact that the take-up rates were similar suggests that we are comparing two distinct types of incentives whose perceived market values are similar. 11 out of 148 trainees dropped out from the training. Thus, 137 enumerators worked in the field on average 18 days interviewing 21,561 households.[13]

We reach three main conclusions using high frequency and quality data on multidimensional labor productivity. First, we find that career incentives of an internship, compared to wage incentives, attract workers with higher labor productivity through self-selection in terms of survey quality and quantity as well as SPEs by census respondents. However, we find that career incentives are limited in improving labor productivity for existing workers. Second, wage incentives causally increase labor productivity which corresponds to the gift exchange model (Akerlof, 1984). As a result, job performances measured by survey quality and quantity as well as SPE by respondents are highest among G2 enumerators who were hired through the internship and additionally received wage incentives. Third, observable individual characteristics are limited in explaining the selection effect suggesting a limitation of screening based on observable characteristics and a need for a self-selection mechanism that can attract productive workers with desirable unobserved characteristics.

Our study is related to several strands of the literature. First, it is closely related to the literature that estimates the selection and causal effects of incentives on job performances (Lazear 2000, Gagliarducci and Nannicini, 2013, Guiteras and Jack, 2015). Lazear (2000) separately isolates worker selection and causal effects of a financial incentive using non-experimental panel data on job performance from a large manufacturing factory in the U.S. which changed from a fixed salary to a piece rate. Guiteras and Jack (2015) separately isolate worker selection and causal effects of a financial incentive by experimentally varying the wage level of daily workers and identifying their reservation wages in rural Malawi.[14] Gagliarducci and Nannicini (2013) also separately identify the

[13] Throughout this paper, *study participants* refer to 443 individuals who participated in the baseline survey; *trainees* refer to 148 individuals who joined the training; and *enumerators* refer to 137 individuals who worked for the census.

[14] They use the Becker–DeGroot–Marschak (BDM) mechanism (Becker et al., 1964) to identify reservation wage. By comparing the work performance of workers with an identical reservation wage but different actual wages for bean-sorting work, they isolate the causal effect of a higher wage on labor productivity from worker selection. However, Berry et al. (2015) argue that the BDM mechanism could measure reservation prices (willingness to pay) successfully while Bohm et al. (1997) and Horowitz (2006) discuss that BDM may not be incentive compatible in practice, thus could be biased in measuring reservation prices. Moreover, since revealing a reservation wage is not part of the ordinary employment process, job applicants might not be comfortable revealing their true reservation wage. Our research design separately isolates the selection and causal effects without relying on the indirect inference of an unobserved worker characteristic.

selection and causal effects of a financial incentive on the performance of politicians by exploiting policies that discontinuously change salaries by population size and a term limit on re-election. All of these studies shed light on the relative importance of the selection and causal effects of incentives on labor productivity.

Some studies focus on the effects of incentives at the recruitment stage on worker selection and job performance (i.e., selection effect). Dohmen and Falk (2011) show that sorting of workers largely explains higher labor productivity under a variable-payment scheme (a piece rate and a tournament) compared to a fixed-payment scheme in a laboratory experiment setting with payment based on a simple math test. In a natural field experiment, Dal Bó et al. (2013) show that a higher wage attracts more qualified applicants without the cost of losing workers with strong public service motivation, but lacks data on job performance. Ashraf et al. (2016) similarly show that salient career incentives attract more productive workers without discouraging those with pro-social preferences from applying for a job in the context of a recruitment drive of community health workers. On the other hand, Deserranno (2016) finds that the expectation of a higher salary discourages job applications from candidates with pro-social preferences.

In addition, other studies focus on causal effects of work incentives.[15] Gneezy and List (2006) empirically test the gift exchange theory originally developed by Akerlof (1984) and show that workers exert more efforts when they receive a financial incentive ("gift") from their employers. Shearer (2004) presents experimental evidence from Canadian tree planters that piece rates induce more effort than fixed wages. In addition, several studies estimate the causal effects of performance pay on productivity of agents in the public service or NGO (Glewwe et al., 2010; Duflo et al., 2012; Fryer, 2013; Ashraf et al., 2014). The literature on causal effects of work incentives focuses mainly on financial incentives.

Lastly, our study is related to the literature on internships. Most existing studies on internships are mainly descriptive (Brooks et al., 1995; D'Abate et al., 2009; Liu et al., 2014). A rare exception is Nunley et al. (2016) which sends out fake résumés with randomly changed characteristics of applicants. They find that a résumé with internship experience receives 14% more callbacks from potential employers but a major limitation of the résumé audit study is that it cannot analyze job performance.

We contribute to the existing literature by providing real-world evidence on selection and causal effects of both career and wage incentives on labor productivity. To the best of our knowledge, our study is the first to compare career and financial incentives directly in the same setting. In addition, we carefully separate selection and causal effect channel of work incentives through two-stage randomization. Last but not least, this study provides the first empirical evidence that credibly examines the selection and causal effects of an internship on job performance.

[15] Oyer and Schaefer (2005) and Bandiera et al. (2011) provide an excellent survey of the literature.

The remainder of the paper is structured as follows. Section 2 outlines the research design and project chronology. Section 3 describes the data and reports sample statistics. Section 4 presents the main results and discusses the findings. Section 5 presents the results of additional analysis including the impact of supervisor visits and short-term effects of enumerator job experience. Section 6 concludes.

2. Research Design and Project Chronology

2.1. Research Context

Malawi is one of the least developed countries in the world with GDP per capita in 2014 of US$255 (World Bank, 2015). Among 20-29 years old males, 19.6% completed secondary school education according to the 2010 Malawi Demographic and Health Survey. Employment in the official sector is 11% and the median monthly income is about US$28.8 (13,400 MWK) (National Statistical Office of Malawi, 2014).[16] About 10% of our study participants worked for pay in official sectors and about 60% were actively searching for jobs at the time of the baseline survey.

AFF conducted a district-wide population census of Chimutu, a rural district located outside of the capital city of Malawi, to collect demographic and socio-economic information of households in January 2015. Chimutu district consists of 52 smaller catchment areas and there are about 90,000 people in 23,000 households. To conduct a census within about a month, AFF needed to hire over 130 enumerators.

The enumerator position confers career-advancing incentives because it is an entry-point job that leads to a regular worker position in the NGO. For example, AFF's many regular staff members were initially recruited as enumerators. A primary role of the census enumerators was to interview household heads to collect basic demographic, socioeconomic, and health information. During the census period, enumerators stayed at the house in the assigned catchment area that AFF rented. Enumerators were asked to survey at least eight households per day. They work alone in the field but supervisors visit them periodically. Since enumerators interviewed many residents in remote villages to collect a variety of personal and complex information, the job required both cognitive and interpersonal skills as well as physical endurance.

Study participants were recruited from the sample who participated in the 2011 secondary school student survey of 7,971 secondary school students in four districts in Malawi, including Chimutu. This 2011 survey was conducted to determine eligibilities for

[16] MWK denotes Malawi Kwacha. As of January 1, 2015, US$1 was equivalent to 466 MWK. Throughout the paper, we use this as the currency exchange rate.

the AFF's random provision of HIV/AIDS education, male circumcision, and financial support for female education programs.[17] Of those participated in the 2011 secondary school survey, AFF approached 536 males who graduated from secondary schools in July 2014 who live near Chimutu.[18] AFF invited them to come for a survey (i.e., baseline survey of this study) without the notice of a potential job offer.

Our sample recruitment strategy has the following advantages. First, we observe the population of a young cohort whose members are potentially interested in a job opportunity in the local labor market, contrary to most existing studies which observe only job applicants. This sampling feature allows us to have a better sense of the external validity of our findings. Second, approaching those who just graduated from secondary school is relevant to an internship, which mainly targets young and entry-level workers.

2.2. Experimental Design

As discussed in the introduction, we implement a two-stage randomized trial to separately identify selection and causal effects of career and wage incentives. As shown in Figure 1, in the first stage, study participants were randomly assigned to the *Wage* group, the *Internship* group, or the control group.[19] In the second stage, of those who accepted AFF's conditional job opportunity, a randomly selected half of the *Internship* group and the *Wage* group additionally received career and wage incentives, respectively.[20]

This process creates five study groups: G1 receives a job opportunity with career incentives only while G2 has an opportunity with both career and wage incentives. The compensation scheme of G1 and G2 mimics unpaid and paid internship arrangements, respectively. Similarly, G4 receives a job opportunity with wage incentives only while G3 has a job opportunity with career and wage incentives. While the work incentives across the study groups were different, the enumerators performed the same tasks during the census. The fifth group, which participated only in the surveys, serves as the control group.

Comparing G2 and G3 isolates the selection effect of the career incentives evaluated against the wage incentives because they have identical incentives at work, but the

[17] The four districts are Chimutu, Chitukula, Tsbango, and Kalumb. For details of AFF programs, see Data Appendix.

[18] This hiring approach allowed AFF to hire workers familiar with the catchment area. The NGO considered only males due to security concerns in the field. In addition, the NGO required secondary school graduation as proof of minimum cognitive skill requirements.

[19] Note that participants were randomly pre-assigned to either one of these three groups before they participated in the baseline survey. At the end of the survey, they were given a job opportunity according to the pre-assigned group. Job opportunities were given regardless of participants' employment and schooling status at the time of the survey.

[20] A job opportunity was valid conditional on the successful completion of the training. For the sake of simplicity, we refer to a conditional job opportunity simply as a job opportunity henceforth.

channels they were recruited were different.[21] In addition, comparing G1 and G2 isolates the causal effects of the wage incentives and comparing G3 and G4 isolates the causal effects of the career incentives. The differences in productivity between G1 and G4 can be interpreted as the combination of the selection and causal effects. The control group allows us to study impacts of short-term work experience 1 year after the completion of the census enumeration work.

2.2.1. Recruitment and baseline survey

We describe the research stages in chronological order as shown in Table 1. As stated above, AFF invited 536 males who met the eligibility criteria for the baseline survey without revealing a job opportunity (Row A). 443 (83%) out of 536 participated in the baseline survey (Row B).[22] In addition, AFF invited study subjects again soon after the census was completed (between April and June in 2015) to measure time and risk preferences and rationality.[23] We discuss the data collected from these surveys further in Section 3.

To minimize unexpected peer effects among workers with different incentives, the baseline survey was conducted separately for the *Internship* group (G1 and G2), the *Wage* group (G3 and G4), and the *Control* group. In addition, the training was separately provided for the *Internship* group and the *Wage* group.

2.2.2. First-stage randomization

AFF supervisors explained the details of an enumerator job to participants at the end of the baseline survey. As described in the Introduction, those randomly assigned for an enumerator position received a job offer with different work incentives.[24] 176 participants (80.0%) out of 220 assigned to the *Wage* group showed up for the baseline survey and were given a short-term job offer, each with a fixed salary of 10,000 MWK (US$21.5) for 20 working days and performance pay of 500 MWK (US$1.1) for every extra 8 households after the first 160 households. 186 participants (84.6%) out of 220 assigned to the *Internship* group showed up for the baseline survey and were given a job opportunity with career incentives which consists of a recommendation letter and the prospect of a job opportunity at AFF as a regular staff member.[25] One-time transportation support, on average about 1500 MWK (US$3.2), was given to both *Wage*

[21] In addition, the comparison of G2 and G3 can be interpreted as the selection effect of the wage incentives evaluated against the career incentives, but for the sake of convenience we mainly focus on the career incentives.

[22] Those who did not participate in the survey (93 individuals) were unreachable (45%), or refused to participate because they were unwilling (13%), enrolled in school (32%), or working (10%).

[23] This survey to measure time and risk preference was conducted after the census was completed under the assumption that these measures are not affected by our interventions. 403 (75%) out of 536 individuals joined the survey.

[24] Note that the offers given in the first-stage randomization were conditional because those who take-up the opportunities must pass the minimum training requirement.

[25] We randomized the first-stage incentives in advance and then invited different first-stage groups on different dates. Thus, study participants were not aware of other types of incentives when they received an opportunity.

and *Internship* groups depending on the distance from the worker's home and the dispatched village. Out of 96 people in the control group, 81 (84.4%) participated in the baseline survey. The base wage of 10,000 MWK (US$21.5) was competitive for young workers who have just graduated from secondary schools. The median monthly salary of secondary school graduates in 2013 was 12,000 MWK (US$25.8), according to the Malawi Labor Force Survey (NSO, 2014). In addition, the prospect of a regular entry-level staff position at AFF whose entry-level monthly salary is 26,000 MWK (US$55.8) could be attractive.[26] AFF notified the *Internship* group that there would be a chance of a long-term contract, without specifying the precise probability, depending on job performance during the contract period and AFF's job vacancies.[27] The recommendation letter was signed jointly by the director of AFF and the head of the Chimutu district.

2.2.3. Training

The job opportunity takers were required to participate in the 1-week training program. It was designed to equip trainees with the necessary skills and knowledge for the census work.[28] The training outcome was measured by a quiz score and the proportion of erroneous entries in a practice survey. We provide the details of the training in Subsection 3.2.

Out of 186 participants in the *Internship* group, 74 (39.8%) completed the 1-week training session, as did 74 out of 176 (42%) participants in the *Wage* group. The training completion rates between the *Internship* group and the *Wage* group were not statistically different, which could imply that the perceived value of career and wage incentives evaluated by the job take-up rate were similar. However, 11 trainees from the *Internship* group was not hired because they did not meet the minimum requirement, while no one failed from the *Wage* group. As a result, in total, 137 enumerators were finally hired, 63 of which were from the *Internship* group and 74 from the *Wage group*.[29]

2.2.4. Second-stage randomization

Second-stage randomization was conducted during the training and the results were announced after the training completion and before the dispatch to the catchment area. Trainees were not aware of additional incentives in the second stage during the training. The wage incentives were given to a randomly selected half of the *Internship* group. Similarly, career incentives were given to a randomly selected half of the *Wage* group. No one refused to accept additional incentives, which implies that the composition of worker characteristics between G1 and G2 and between G3 and G4 remain the same.

---

[26] Those promoted to a project manager position at AFF were paid between US$100 and US$160 (MK 46,600 and MK 74,560) per month during the study time.

[27] Working as an intern without knowing the exact probability of hiring is close to the general internship setting.

[28] During the training, meals and reimbursed transportation costs were provided.

[29] Therefore, we do not observe the job performance of 11 trainees from the *Internship* group who failed the training requirement. We discuss this further in Section 4.

Different incentives for different study groups were described clearly in the contract as shown in Figures A.1, A.2, and A.3. For example, the contract of G1 explicitly states that enumerators will not be given any financial compensation and will be provided with a recommendation letter and a potential job opportunity based on their performance. It is also noteworthy that the contract explains three main job performance measures: speed and accuracy of the survey as well as SPE by survey respondents.

2.2.5 Census and post-enumeration survey

Enumerators were dispatched to 52 catchment areas in January 2015. Enumerators were randomly assigned to catchment areas stratified by population and land size and worked independently. As stated in Subsection 2.2, enumerators in the same catchment area have the same incentives to prevent unexpected peer effects. In addition, enumerators were not assigned to areas from which they originally came, as locality could affect their performance. Finally, we assigned enumerators to have the same incentives in the same dispatched village during the census period to avoid unnecessary peer interaction across different study groups.

The specifics of the census questionnaire include a variety of individual- and household-specific characteristics such as demographics, wealth, employment and income, health, etc. It took about 25 minutes on average to interview a household.

Enumerators as a whole surveyed 21,561 households during the contract. After the contract period, AFF surveyed additional 2,561 households newly found or failed to reach during the original census period. In this paper, we only use 21,561 surveys collected during the original census contract period to analyze job performance of enumerators.

Supervisors of AFF visited enumerators to monitor and guide enumeration work on randomly selected dates without prior notice.[30] Supervisors met enumerators at least once during the census period. 37% of the enumerators met supervisors twice and the remaining 63% met supervisors once.

Shortly after the completion of the census, AFF conducted a post enumeration survey (PES) to correct errors in the census and to find omitted households by revisiting all households in Chimutu. Another purpose of PES was to prevent enumerators from outright cheating or fabricating census interview sheets. AFF informed census enumerators that there would be a PES before they were dispatched to the field. Lastly, PES was conducted to collect information on the survey proficiency of census enumerators by PES enumerators.[31]

---

[30] Supervisors are AFF's regular staff members who have at least 3 years of experience conducting field surveys.

[31] Hiring enumerators as regular staff members required the careful calculation of job performance after the completion of the census which requires time and manpower. Meanwhile, as AFF was computing the job performance, it hired 43 PES enumerators among 98 census enumerators with career incentives (G1, G2, and G3) on a temporary basis (2–3 months) through a simple performance evaluation based on subjective evaluation of work attitude by supervisors and error rates measured from five randomly selected census surveys.

As stated in the contract, AFF provided recommendation letters to the enumerators with career incentives (G1, G2, and G3) in May, 2015. The letter specified the job description of an enumerator and his relative job performance measured by survey quality, survey quantity, and SPEs.[32]

2.2.6. Follow-up Survey

AFF conducted follow-up surveys with the study participants after 1 year (May 2016) over the telephone. The telephone survey was administered to investigate whether the short-term job experience affected participants' future labor market outcomes. Specifically, it asked whether they were involved in any kinds of activities after the census such as additional education, vocational training, and employment.

3. Data

We use data from various sources, including a) baseline and follow-up surveys, b) AFF's administrative data on training and job performance outcomes, and c) the Chimutu population census.[33]

Data from the surveys include the followings. First, we use data from the 2011 secondary school student survey. It contains rich information on a variety of areas covering demographics, socioeconomic status, health, and cognitive ability. Second, we use data from the 2014 baseline survey, which collects information on demographics, education, employment history, cognitive abilities, non-cognitive traits, and HIV/AIDS related outcomes. We measure cognitive ability by a cognitive ability index, defined as the average z-score of the Raven's matrices test score, the math and English scores of the 2014 Malawi School Certificate of Education (MSCE) test, and the verbal and clerical ability test scores of the O*NET ability test, following the approach of Kling et al. (2007).[34] Non-cognitive traits include self-esteem, intrinsic motivation, extrinsic motivation, and the Big Five personality test (extraversion, openness, conscientiousness, agreeableness, and

[32] For example, if an enumerator has higher job performance than the average, the letter specifies very strong recommendation. If an enumerator has performance below the average, the letter specifies somewhat lukewarm recommendation.

[33] We calculate average characteristics of the catchment area based on the census data. These average catchment area-specific characteristics were used as the control variables in the main regression analysis.

[34] AFF had access to the administrative MSCE score data via the cooperation of the Ministry of Education of the Republic of Malawi. Raven's progressive matrices test is a non-verbal test of thinking and observation skills. The MSCE is a test that all Malawian students must take to graduate from secondary school. The MSCE score we use is a standardized test score of mathematics and English which are mandatory subjects of the test. The O*NET® test is a tool for career exploration developed through the U.S. Department of Labor. We use verbal and clerical perception ability test scores of O*NET®, which are directly related to enumerator job characteristics. Data Appendix A.1 provides the definitions of cognitive ability measures.

neuroticism).[35] Third, the additional baseline survey conducted in April–June 2015 collected data on risk and time preferences and rationality using the tests recently developed by Choi et al (2014).[36] Finally, we use data from the follow-up survey to measure labor market activities after the census.

Columns (3) to (5) of Table 2 present the baseline characteristics of each study group. The results of first- and second-stage randomization balance are presented in Columns (6) and (7) and Columns (8) and (9), respectively. Panel A represents individual baseline characteristics and Panel B represents the catchment area characteristics where enumerators were dispatched. Study participants are about 20 years old and only 9% work in the official sector reflecting weak labor demand in Malawi.[37] Column (6) compares the *Internship* and *Wage* groups, and Column (7) compares the treatment groups (both *Internship* and *Wage* groups) and the control group. Columns (8) and (9) compare G1 and G2, and G3 and G4, respectively. The results confirm that the study groups are well balanced: the proportion of statistically significant mean difference at the 10% significance level is 2 out of 27 (7.4%) in Column (6), 5 out of 20 (25%) in Column (7), 3 out 27 (11.1%) in Column (8), and 4 out of 27 (14.8%) in Column (9). This suggests that our randomization was generally successful in creating balanced study groups.[38] In addition, we examine whether the baseline survey participants and nonparticipants are systematically different. Table A.1 shows that they are not statistically different from each other in most dimensions except for the household asset index.

Training outcomes are measured by a quiz score and the proportion of erroneous entries in a practice survey.[39] The quiz tested specific knowledge on the census details.[40] Job performance during the census is measured based on the census survey data in three dimensions: 1) quantity, 2) quality, and 3) subjective performance evaluations (SPE). Quantity is measured by the number of households surveyed by each enumerator per day.

---

[35] Data Appendix A.2 provides the definitions of non-cognitive traits.

[36] As explained in Section 2.2.1, risk preference, time preferences, and rationality were measured after the census was completed. We included these measures in the randomization balance test under the assumption that these traits were not affected by our experiment. Data Appendix A.1 provides the details how we measure them.

[37] The employment rate of survey non-participants is also similar. We reached non-participants via phone calls and 9.7% of them did not attend because they were working.

[38] The number of siblings, the only unbalanced individual variable in Column (6), eligibility for AFF's past interventions and catchment area characteristics controls are included in all specifications of the main analysis.

[39] The purpose of the practice survey was to practice interview skills before enumerators were dispatched to the field. The practice survey performance was evaluated as follows. First, we randomly matched two trainees. Each trainee in a randomly assigned pair received a pre-filled census questionnaire sheet and a blank survey questionnaire sheet. Then, one trainee interviewed the other matched trainee in the same pair and the latter answered based on the assigned survey sheet. There were two different types of pre-filled questionnaire sheets with different hypothetical household information. Thus, trainees in the same pair acted as if they were two different households. Each trainee in every pair conducted this practice survey by changing roles. After conducting practice survey sessions, supervisors collected the survey sheets and calculated the error rate.

[40] The quiz consists of 12 questions, a mixture of open-ended and true/false type questions. The full text of the quiz is presented in Figure A.4.

Quality is measured by the proportion of systematically inconsistent or incorrect entries in the census questionnaire specific to each household.[41] SPEs are measured by census respondents as well as supervisors, respectively. A census respondent was asked to evaluate how carefully the enumerator explained the questions during the PES.[42] In addition, SPE is also measured by AFF supervisors. After the completion of the census, 12 supervisors jointly evaluated the overall work attitude of each enumerator. Enumerators were aware of job performance evaluation because the employment contract states that job performance will be measured by survey quantity and quality as well as SPE by the respondents (but not SPE by the supervisors).

4. Main Results

4.1. Job Opportunity Take-up

Column (1) of Table 3 shows that the job opportunity take-up rates between the *Internship* and the *Wage* groups are not different. This suggests that the average perceived market values of an opportunity with career incentives and an opportunity with wage incentives are similar. Even though the take-up rates are similar across the two groups, it is possible that the composition of job takers between the two groups could be different if different incentives attract workers with different observable and unobservable characteristics. We test multidimensional sorting discussed in Dohmen and Falk (2011) by exploring whether career and wage incentives attract those with different observable characteristics. Columns (2) to (18) of Table 3 show the regression results of the following equation:

$$Accept_i = \alpha + \delta \cdot Internship_i + \lambda \cdot Trait_i + \varphi \cdot Internship_i \cdot Trait_i + \epsilon_i \qquad (1)$$

$Accept_i$ is a binary indicator that equals 1 if individual i accepted a job opportunity, and 0 otherwise. $Internship_i$ is a binary indicator if individual i belongs to the *Internship* group and the omitted category is the *Wage* group.[43] Trait is an individual characteristic variable which we evaluate one by one. ϵ_i is an error term. We test whether career incentives differently attracts workers with different characteristics as opposed to wage incentives over a variety of individual characteristics including demographic and socioeconomic characteristics, cognitive ability index, and non-cognitive traits.

[41] For example, if a respondent has a child, the information about her child should be filled in. If not, it was counted as an error. The Data Appendix provides the details about how we calculate the survey error rate.

[42] The question asked was "Whenever you were confused or could not understand the meaning of any question, did the enumerator carefully explain the meaning of the questions to you?"

[43] We restrict the sample to *Internship* and *Wage* groups.

Table 3 shows the regression results of equation (1). Our coefficient of interest is φ, which captures whether there is differential take-up of a job opportunity between the *Internship* group and the *Wage* group by individual traits. We find that none of the estimates of φ across individual traits is statistically significant at the 5% level. This finding implies that observable characteristics are not likely to predict self-selection. Table A.2 provides additional evidence on self-selection by the observable characteristics. The results in Table A.2 confirm the results in Table 3, that job-takers of the two groups are not systematically different in terms of both statistical and economic significances (the p-value of joint F-test is 0.53). We find two statistically significant results (extroversion and risk preference) out of 21 traits.[44] In addition, only four out of 21 variables have the mean differences larger than the 0.2 standard deviations.

We acknowledge that study participants could have responded the self-reported non-cognitive tests in a way they believe desirable from a perspective of a potential employer. This possibility might explain little differences in terms of non-cognitive traits between the two groups.[45] However, participants did not know about the possibility of receiving a job offer and AFF never used these non-cognitive trait tests to hire an employee.

No systematic differences in observable characteristics do not necessarily mean that unobservable characteristics, training outcome, and job performance would not be different if some of unobservable characteristics affect training outcomes and job performance.

4.2. Training Outcomes

Even though we do not find any differences in observable individual characteristics between job takers of the two groups, we might find a difference in training outcomes if career and wage incentives attract people with different unobservable characteristics. Figure 2 displays the kernel density estimates of the training outcomes measured by the quiz score (Panel A) and the practice survey error rate (Panel B). Table 4 shows the corresponding results from the following specification:

$$Training_i = \alpha + \beta \cdot Internship_i + \omega_i \tag{2}$$

[44] We acknowledge that this finding should be interpreted with caution due to the concern about multiple hypothesis testing.

[45] It is consistent with real world where job seekers are not able to manipulate test scores (cognitive ability) in a pre-employment test but may try to response a personality test in a way that they have a desirable non-cognitive skill.

where $Training_i$ is the training outcome for individual i.[46] Panel A of Figure 2 shows that the *Wage* group performs better than the *Internship* group in terms of quiz score and survey error. Panel A of Table 4 provides corresponding results from the regression. It confirms that quiz score of the *Internship* group trainees are 2.0 points (24%) lower than the *Wage* group trainees (Column (1)). Similarly, the survey error rate is 10.4 percentage points (38%) higher among the *Internship* group trainees than the *Wage* group trainees (Column (3)). As a result, 11 trainees of *Internship* group and none of *Wage* group fail to complete training.

Panel B presents the training outcomes of enumerators dispatched to the field by excluding 11 training failures. The regression results between the two panels are qualitatively similar, but the magnitude of the coefficient estimates is larger in Panel A than in Panel B because those who failed training are all from the *Internship* group.

The specification used in Columns (2) and (4) is to test whether individual observable characteristics can explain the differences in the training outcomes between the two groups. In addition to number of siblings and past eligibility status of AFF's HIV/AIDS program included in Columns (1) and (3), they additionally include age, household asset score, cognitive ability index, and non-cognitive traits such as self-esteem, intrinsic and extrinsic motivation, and Big 5 personality scales. We find similar coefficient estimates between Columns (1) and (2), and between Columns (3) and (4), implying that observable characteristics are limited in explaining the difference in the training outcomes.

In summary, we find that those attracted by a job opportunity with wage incentives outperformed those attracted by an internship opportunity in the training. This difference could be because workers with different characteristics might be attracted by different work incentives, thereby creating the difference in the training outcomes (selection effect). On the other hand, the difference could be because those in the *Internship* group have an incentive to exert more effort than the *Wage* group due to the future job prospect nature of the career incentives (causal effect). However, we are not able to disentangle which channel is more likely to explain the observed difference in the training outcomes.

4.3. Selection effect of career incentives on labor productivity (G2 vs. G3)

In this section, we examine the selection effect of career incentives evaluated against wage incentives on job performance. As previously discussed, G2 and G3 have the

[46] We control for number of siblings, which is not balanced in the baseline, and binary indicator variables of the past eligibility status of AFF's program.

same incentives at work but the channels by which they were recruited are different. Therefore, we argue that differences in performance are driven by the *selection effect*.

An implicit identifying assumption in this analysis is that perceived work incentives of enumerators in G2 and G3 are identical even though the sequences of the provision of career and wage incentives were different. The different sequence could form different perceived expectations of the level of compensation package. If this is the case, it could affect the feeling about what actually occurs, leading to different levels of work efforts, and our estimates of the selection effect would be biased as Abeler et al. (2011) discussed. However, we argue that this is unlikely: if there were such a difference in feeling, we expect that differences in job performance become smaller over time because the difference in feeling might diminish over time. However, Figure A.5. shows that the difference in job performances are fairly constant over time.

Figure 3 suggests that G2 performs better than G3 in terms of survey quality and quantity, and SPE by census respondents. It is surprising because the *Wage* group had better training outcomes than the *Internship* group.[47] We test this graphical evidence formally by estimating the following equation:

$$Y_{ijklt} = \alpha + \beta \cdot G2_j + \gamma \cdot H_i + \emptyset \cdot Z_k + V_{lt} + \sigma_t + \psi_{ijklt} \quad (3)$$

where Y_{ijktl} is job performance measured in the survey collected from household i by enumerator j whose supervisor is l, in catchment area k, surveyed on the t-*th* work day. $G2_j$ is 1 if enumerator j belongs to Group 2 and 0 if he belongs to Group 3. H_i is a vector of respondents' household characteristics and Z_k is a vector of catchment area characteristics.[48] σ_t is the survey date fixed effect. V_{lt} is the supervisor-specific post-visit effect.[49] Standard errors are clustered at the enumerator level. For the dependent variable, we consider survey quality measured by the survey error rate ($Error_{ijktl}$), survey quantity measured by the number of surveys per day ($Survey_{jktl}$), SPE by census respondents ($SPErespondent_{ijktl}$), and SPE by supervisors ($SPEsupervisor_{jk}$). [50]

Panel A of Table 5, which corresponds to Figure 3, presents the regression results from equation (3). It shows that G2 outperforms G3 in most job performances even though G3 outperforms G2 during the training.[51] For example, the error rate is 2.0 percentage points (26%) lower in G2 than G3, as shown in Column (1). Survey quantity of

[47] We further explain the difference in job performances and training outcomes in Section 4.5.

[48] Respondent's household characteristics include the fixed effect for the number of household members. Catchment area characteristics include the total number of households, asset score, birth rate, malaria incidence, rate of birth with the assistance of a health professional, and death rate.

[49] $V_{lt} = \eta_1 I(t>First) + \eta_2 I(t>Second)$ where First and Second are the dates of supervisor l's first and second visits, respectively, to enumerator j.

[50] We do not control for σ_t and V_{jlt} when we analyze SPE(supervisor).

[51] See Figure A.6 for the training outcomes of G2 and G3.

G2 is higher by 1.48 households per day (13.8%) than G3 as shown in Column (4). In addition, G2 has a 37% higher SPE score by the respondents (Column (7)) while the SPE score by supervisors is higher in G3 than in G2. We find similar results in the specification with a linear time trend instead of a flexible control in the survey date (survey date fixed effect), as shown in Columns (2), (5), and (8).

To assess how much observable individual characteristics can explain the selection effect estimated in Columns (1), (4), (7), and (10), we additionally control for demographic and socioeconomic status, cognitive ability, and non-cognitive traits. As shown in Columns (3), (6), (9), and (11), observable individual characteristics explain little in selection effect on survey quality and quantity. For example, the inclusion of observed individual characteristics does not explain the estimated selection effect of an internship on survey quality at all and it explains survey quantity only by 5% (=(1.48-1.41)/1.41).[52] This result is consistent with the fact that observable characteristics of job takers between the *Internship* and the *Wage* group are not different.

In Table A.3, we report additional performance outcomes to decompose the objective performance outcomes (survey quality and quantity). In terms of survey quality, we decompose the types of errors into incorrectly entered entries (e.g., 179 for a person's age) and incorrectly blank entries (e.g., a child is present in the household but his/her age is missing). To better understand how survey quantity changes, we run a regression on several time–use variables such as total work hours per day, average survey time per household, and intermission time between surveys.[53] Work hours per day are the difference between the beginning time of the first survey and the end time of the last survey of the day. Survey time per household is the length of each survey. Intermission time is defined as the difference between the beginning time of a survey and the ending time of the previous survey.[54]

Panel A of Table A.3 indicates that the selection effect of an internship on survey quality in Table 5 is mostly driven by the decrease in errors of incorrectly blank entries (Column (3)). In addition, the selection effect of an internship on survey quantity mainly comes from a decrease in intermission time between surveys: the average intermission time for G2 enumerators is shorter by 4 minutes (a 20.8% decrease) compared to G3 enumerators (Column (9)). In other words, those recruited by career incentives increases survey quantity by exerting more efforts to reduce time to move to the next households, not increasing total work hours or decreasing survey time. Again, observable enumerator characteristics do not explain differences between G2 and G3.

[52] However, the individual characteristics can explain SPEs about 13% and 27% of the selection effect on SPEs by census respondents and by supervisors, respectively.

[53] Since there were sizable number of missing values in the survey beginning and end times in each survey, we impute the missing values. See Data Appendix A.3 for the imputation process.

[54] The survey beginning and end times were recorded as a part of the census questionnaire.

In summary, we find that career incentives of an internship improve labor productivity through the self-selection of workers. In addition, the observable characteristics do not play an important role in explaining the selection effect, which implies the importance of unobservable characteristics in self-selection. However, we find that productivity difference between two groups is persistent over time from the first workday. Our findings suggest that screening via the observables by an employer could be limited unless they can measure worker's productivity directly. Pre-employment screening for highly productive employees could be particularly difficult when hiring entry-level workers who have no record of accomplishment of past job history or credentials to verify their unobserved productivity.

4.4. Causal effects of work incentives on labor productivity

To measure causal impacts of career incentives on labor productivity, we compare enumerators who receive both wage and career incentives (G3) and enumerators with wage incentives only (G4). Similarly, we measure causal impacts of wage between enumerators with only career incentives (G1) and enumerators with both career and wage incentives (G2).[55]

Panels B and C of Table 5 report the causal effects of career and wage incentives on job performance, respectively. Figures 4 and 5 present the corresponding graphical evidence. In Panel B of Table 5, we find no evidence that career incentives improve job performance (survey quality, survey quantity, and SPE by respondents) although career incentives could motivate enumerators to exert more effort. Contrary to expectation, we find a decrease of survey quantity (a decrease of 0.594 households per day, or 5.2%), mainly driven by an increase of intermission time (by 6.72 minutes) as shown in Panel B of Table A.3.[56] It leads to a total increase of intermission time of 70.5 (=6.72*(11.5-1)) minutes per day, which dominates a 39.7-minute increase of total work hours. Lastly, we find that SPE measured by supervisors significantly increases by 52.3%. In sum, career incentives given to existing workers hired through the wage incentive channel do not improve labor productivity, but they induce enumerators to have better evaluation from supervisors. A possible explanation is that it is difficult (costly) for those hired through wage incentives, whose work performance is lower compared to those hired through career

[55] We estimate the causal effects of different work incentives from different types of workers after self-selection. Specifically, any difference in job performance between G1 and G2, and G3 and G4 is the causal effect of career incentives (among enumerators recruited through the wage incentive channel) and wage incentives (among enumerators recruited through the career incentive channel), respectively.

[56] We are slightly under-powered in the regression analysis of Panel B in that the size of the standard errors is not small enough to capture the small effect (if any) of the work incentives. For example, we are able to capture causal impacts of career incentive on survey quality and quantity only if the change is greater than 26.3% (=0.011×1.96/0.082) and 10.1% (=0.594×1.96/11.5), respectively.

incentives, to improve work performance at least in the short-run. They rather exerted efforts on the relationship with supervisors.

Panel C of Table 5 shows that wage incentives improve job performance. We find that survey errors decrease by 2.8 percentage points (37.3%) and survey quantity increases by 1.2 households per day (12.2%) without finding statistically significant changes in SPEs. Panel C of Table A.3 shows that the decrease in the survey error rate is mainly explained by a decrease in incorrectly blank entries (Column (3)). In addition, we find that an increase in survey quantity is driven by a decrease in actual survey time (i.e., each survey was completed more quickly) as shown in Column (7). This finding is consistent with the gift exchange model of the efficiency wage theory formulated by Akerlof (1984), by which a worker exerts more efforts upon receiving a gift from an employer that exceeds the minimum level of compensation for the minimum level of effort.

4.5 Discussion of Selection and Causal Effects

As stated in the introduction, a major contribution of our study is to separately identify the worker's self-selection (G2 vs G3) and causal effects of work incentives (G1 vs G2 and G3 vs G4). Panel D of Tables 5 and 6, which compare G1 versus G4, resembles the combined effects of selection and causal effects on productivity because participants were attracted to accept a job opportunity via different incentives and the incentives at work also remained different. We find no significant difference in the combined effects between G1 and G4 except for SPE by supervisors, which shows the importance of separating selection and causal effects. It is noteworthy that the combined effects of career incentives (Panel D) are not necessarily to be a simple sum of the selection effect (Panel A) and causal effect (Panel B). In addition, the study sample used in Panel D is different from that in Panels A and B.

Another point to consider is unbalanced dropouts of the trainees. All 11 trainees who were dropped were from the *Internship* group.[57] One may argue that if the labor productivity of the dropouts were lower than that of the hired enumerators, the performance-improving selection effects of an internship would be overestimated. However, we do not consider that any particular adjustment is necessary in the main analysis because screening out trainees who did not meet the minimum requirement was a regular business practice. Nevertheless, we re-estimate equation (2) after dropping 11 trainees with the lowest training scores from the *Wage* group.[58] Panel A of Table A.4 shows that the results on selection effect remain mostly robust; the size of the coefficients for the selection effect on survey quality becomes smaller while that on survey quantity becomes

[57] Trainees, including 11 trainees who were dropped, were not aware of the second-stage randomization.

[58] We dropped six from G3 and five from G4 from the *Wage* group enumerators.

larger. A possible explanation is that training performance is not a precise predictor of job performance. We find similar results on causal effects (Panels B and C) and combined effects (Panel D).

A possible reason that training and job performance are different is different skill sets required in each setting. The test taken during the training was in a classroom setting while job performances resulted from actual interactions with respondents in the field. Thus, it is plausible that those selected through the internship could have comparative advantages in on-the-job performances but may be not in tests in the classroom setting. A critical characteristic of an enumerator is the skill to ask strangers sensitive questions about their households. This kind of skill might not be captured easily in a test taken in a laboratory setting. Another explanation is that screening out 11 trainees could serve as a reminder or a credible threat to those with career incentives that only some of them will be hired as regular workers in AFF.[59]

5. Additional Analysis

In this section, we present empirical evidence on the effect of supervisor visits on job performance and the 1-year effect of short-term job experience on employment.

5.1. Impacts of Supervisor Visits on Job Performance

As stated in Subsection 2.3.5, supervisors visited enumerators on randomly selected dates to monitor and advise enumerators during the survey. Enumerators were aware of supervisor visits, but did not know the exact date. Supervisors joined each enumerator for interviews of about three households, addressed common errors, and provided overall comments at the end of the supervision visit.

Figure 6 shows how a supervisor visit affected job performance of enumerators before and after the visit. Panels A and B illustrate changes in job performance over time around the first and second supervisor visits, respectively. A vertical line at day zero represents the day of a supervisor visit. It seems that the survey error rate decreases over time, especially right after the supervisor visit, but survey quantity and SPE remain similar. To formally quantify the changes in job performance after supervisor visits, we estimate the following equation:

[59] It is also possible that the *Internship* group initially had a lower performance in the training but they caught up the *Wage* group later in the field due to a steeper learning curve. However, this is less likely, as we find no evidence for the catch-up. Job performance between the *Internship* group and the *Wage* group over the survey period remained constant over time (see Figure A.5. for the performance trend).

$$Y_{ijklt} = \alpha + \eta_1 \cdot First_{lt} + \eta_2 \cdot Second_{lt} + \gamma \cdot H_i + \emptyset \cdot Z_k + \delta \cdot X_j + \psi \cdot t + \mu_{ijktl} \quad (3)$$

where Y_{ijklt} is a job performance outcome for the survey collected from household i by enumerator j whose supervisor is l, in catchment area k, observed on the t-th work day. *First* and *Second* are binary indicators of surveys collected on the dates after the supervisor's first and second visit, respectively. X is individual characteristics including the standard set of control variables used above, and study group dummy.

Column (1) of Table 6 reports the regression results of equation (3). Column (2) includes survey date fixed effect instead of time trend. Columns (1) and (3) show that the error rate decreases by 1.5 percentage points (a 20% improvement) and survey quantity increases by 1.36 household per day (a 12.5% increase) after the first supervisor visit. This finding is not surprising because the supervisor visit could help enumerators accumulate job-specific human capital. However, the effects on both survey quality and quantity become smaller and insignificant when the survey date fixed effect is included. In addition, we find that SPE measured by respondents is negatively affected by the supervisor visit and the effect become statistically significant and large when we control for the survey date fixed effect.

Lastly, the second visit do not lead to any statistically significant change in productivity. A possible explanation is that the additional visit does not lead to much skill acquisition. Another possible explanation is that 22% of enumerators had the second visit so the lack of statistical power could make it difficult to precisely estimate the impact of the second supervisor visit.

5.2. One-year Impact on Employment

The presence of the control group who participated in the baseline survey but did not receive a job offer allows us to estimate the impacts of short-term job opportunity on employment.[60] Specifically, Panel A of Table 7 presents the 1-year intent-to-treat effects of a short-term job offer on labor market activities by comparing G1 through G4 with the

[60] As stated in footnote 31, AFF rehired 43 enumerators in a temporary basis among the 98 individuals with career incentives. The extended contract offered a daily wage of 500 MWK to conduct the PES or other surveys implemented by AFF. After the completion of the PES, some enumerators who remained for other surveys went on a labor strike asking for a steep wage increase, while AFF was still digitizing job performance. AFF decided not to rehire any enumerator from the study sample as a regular worker. As a result, at the time of the 1- year follow-up phone survey, there was no one working for AFF.

control group. We find no evidence that those offered enumeration jobs are more likely to work for pay. Panel B shows that coefficients for the *Wage* group are even negative, implying that a change of short-term job experience did not help to find a job in the future.[61]

Next, we compare the *Internship* and *Wage* group directly. Panel C reports that the employment rate of the *Internship* is higher than that of the *Wage* group by 5.4 percentage points (a 132% increase) (Column 1 in Panel B). This result is robust after controlling for individual characteristics (Column 2). If the increase in employment is driven by career incentives of an internship, we conjecture that the impact is larger among individuals who accepted an internship opportunity. As expected, those who accepted an internship opportunity have a 9.9 percentage point higher employment rate than those who received a job offer with wage incentives. However, there is no significant impact for those declined the internship offer.

A possible explanation of why the *Internship* group outperforms the *Wage* group in the labor market after 1 year of the census enumeration work is that a recommendation letter provided to the *Internship* group enumerators could have helped them to find a job. Another possible explanation is that the *Internship* group has accumulated more skills useful in the labor market during the census because the nature of the career incentives made them work harder. However, to understand the specific mechanisms behind how internship experience affect future labor market outcomes is the avenue for future research.

6. Concluding Remarks

There are two major channels through which an organization can increase job performance of employees. The first is to offer attractive incentives to hire effective workers (selection effect) and the second is to motivate existing employees to exercise more efforts via providing more incentives at work (causal effect). This study analyzes how career incentives of an internship and wage incentives affect labor productivity through worker selection and causal effects. We separately estimate the selection and causal effects of work incentives on job performance through a two-stage randomized controlled trial in the context of a recruitment drive of census enumerators in Malawi.

We find that even though the training performances of the *Wage* group is better than that of the *Internship* group, the *Internship* group outperformed the *Wage* group at

[61] We acknowledge that the characteristics between the treatment group (G1–G4) and the control group are not perfectly balanced (Column (6) in Table 2), therefore results on Panels A and B should be interpreted with caveat.

work. In addition, observable individual characteristics are limited in explaining the difference in labor productivity due to self-selection. The fact that neither observable characteristics nor training outcomes predict actual job performances implies that screening via observable characteristics is imperfect, particularly when hiring entry-level workers who have no track record of past job history or credentials to verify their unobserved productivity. This finding also implies that it is difficult for an employer to foresee desirable attributes of an effective enumerator. These implications in turn suggests the importance of a recruitment strategy to attract workers with strong unobservable skills via self-selection (e.g., an internship). Regarding the causal effect of career incentives, we do not find positive evidence on the causal effects of career incentives on job performances except for the subjective evaluation of work attitude by supervisors. Our findings suggest that career incentives of internship are effective in improving labor productivity mainly through the selection effect channel. Lastly, we find that financial incentives in the form of a cash bonus can be an effective means to improve job performance of existing workers as the gift exchange model predicts (Akerlof, 1984).

There are limitations to our study. First, we acknowledge that the approach by which we estimate the causal effects of incentives might not perfectly characterize the real world. In the real world, workers usually do not receive additional incentives by surprise. Even if there were performance pay, workers are usually aware of it when they join the workplace. Second, the length of a job we study is relatively short-term (less than 1 month). As such, we cannot study whether the estimated selection and causal effects of career and wage incentives remain constant over longer periods. In addition, we cannot study the effects of work incentives on retention. Third, we do not directly observe the individual's perception of the value of work incentives. Nevertheless, explicitly asking a question on such subjective perception could be problematic if it affects job performance. To measure the willingness to pay for each type of job opportunity could be a solution, but this subjective evaluation could also be imprecise. Fourth, non-cognitive traits used in this study are self-reported psychometric scales measured based on a paper test. It would be interesting to know whether such paper-based and self-reported non-cognitive traits are highly correlated with non-cognitive traits measured in other settings.

A better understanding of selection and causal effects of work incentives would allow employers to design optimal employment strategies. One of the major challenges of economic development is low labor productivity. In addition to the problem of absenteeism, difficulty of efficient job applicant screening and lack of motivation among existing workers are key drivers of low labor productivity in developing countries. Based on our findings, we argue that active adoption of career incentives in the workplace as a hiring strategy could be an effective means to increase labor productivity of an organization. In addition, an additional unexpected cash bonus could be also an effective means to further motivate existing employees. Even if the findings of this study

specifically reflect the selection and causal effects of an enumerator job in Malawi, our analysis has several implications for jobs in different settings in which employers have difficulties screening out productive workers and motivating existing workers.

References

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. Reference points and effort provision. *The American Economic Review* 101, no. 2 (2011): 470-492.

Akerlof, George A. 1984. Gift Exchange and Efficiency–Wage Theory: Four Views. *American Economic Review: Papers and Proceedings*, 74(2): 79–83.

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 2016. Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery. Harvard Business School Working Paper.

Ashraf, Nava, Oriana Bandiera, and Kelsey Jack. 2014. No Margin, No Mission? A Field Experiment on Incentives for Public Services Delivery. *Journal of Public Economics*, 120 (Dec 2014): 1-17.

Ashraf, Nava, James Berry, and Jesse M. Shapiro. 2010. Can Higher Prices Stimulate Product Use? Evidence from a Field Experiment in Zambia. *American Economic Review*, 100(5): 2383-2413.

Bandiera, Oriana, Iwan Barankay and Imran Rasul. 2011. Field Experiments with Firms. *Journal of Economic Perspectives*, 25(3): 63-82.

Beaman, Lori, Dean Karlan, Bram Thuysbaert, and Christopher Udry. 2015. Self-Selection into Credit Markets: Evidence from Agriculture in Mali. NBER Working Paper No. 20387. National Bureau of Economic Research.

Becker, Gary, Morris H. DeGroot, and Jacob Marschak. 1964. Measuring Utility by a Single-response Sequential Method. *Behavioral Science*, 9: 226–236.

Berry, James, Greg Fischer, and Raymond Guiteras. 2015. Eliciting and Utilizing Willingness to Pay: Evidence from Field Trials in Northern Ghana. BREAD Working Paper No. 459. Bureau for Research and Economic Analysis of Development.

Brooks, L., Cornelius, A., Greenfield, E. and Joseph, R., 1995. The relation of career-related work or internship experiences to the career development of college seniors. *Journal of Vocational Behavior*, 46(3), pp.332-349.

Bohm, P., J. Lindén, and J. Sonnegård. 1997. Eliciting Reservation Prices: Becker–Degroot–Marschak Mechanisms vs. Markets. *The Economic Journal*, 107(443): 1079–1089.

Callen, Michael, Saad Gulzar, Ali Hasanain, Yasir Khan, and Arman Rezaee. 2015. Personalities and Public Sector Performance: Evidence from A Health Experiment In Pakistan. NBER Working Paper No. 21180. National Bureau of Economic Research.

Choi, Syngjoo, Shachar Kariv, Wieland Müller, and Dan Silverman. 2014. Who is (More) Rational? *American Economic Review* 104, No. 6: 1518-1550.

Cohen, J. and Dupas, P., 2010. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. *The Quarterly Journal of Economics*, pp.1-45.

D'abate, C.P., Youndt, M.A. and Wenzel, K.E., 2009. Making the most of an internship: An empirical study of internship satisfaction. *Academy of Management Learning & Education*, 8(4), pp.527-539.

Dal Bó, Ernesto, Frederico Finan, and Martin A. Rossi. 2013. Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service. *Quarterly Journal of Economics*, 128(3): 1169–1218.

Deserranno, Erika. 2016. Financial Incentives as Signals: Experimental Evidence from the Recruitment of Health Workers. Working Paper.

Dohmen, T. and Falk, A., 2011. Performance pay and multidimensional sorting: Productivity, preferences, and gender. *American Economic Review*, 101(2), pp.556-590.

Duflo, E., Hanna, R. and Ryan, S.P., 2012. Incentives work: Getting teachers to come to school. The *American Economic Review*, 102(4), pp.1241-1278.

Fryer Roland. 2013. Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*. 31 (2): 373-427.

Gagliarducci, Stefano, and Tommaso Nannicini. 2013. Do Better Paid Politicians Perform Better? Disentangling Incentives From Selection. *Journal of the European Economic Association*.

Glewwe, P., Ilias, N., and Kremer, M., 2010. Teacher incentives. *American Economic Journal: Applied Economics*, 2(3), pp. 205-227.

Gneezy, Uri, and John A. List. 2006. Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments, *Econometrica*, 74(5): 1365-1384.

Guiteras, Raymond P., and B. Kelsey Jack. 2015. Incentives, selection and productivity in labor markets: Evidence from rural Malawi. NBER Working Paper No. 19825. National Bureau of Economic Research.

Horowitz, J.K. 2006. The Becker–DeGroot–Marschak Mechanism is Not Necessarily Incentive Compatible, Even for Non-random Goods. *Economics Letters*, 93(1): 6–11.

Karlan, D. and Zinman, J., 2009. Expanding credit access: Using randomized supply decisions to estimate the impacts. *Review of Financial studies*, 23 (1): 433-464.

Kling, Jeffrey, Jeffrey Liebman and Lawrence Katz. 2007. Experimental Analysis of Neighborhood Effects. *Econometrica*, 75(1): 83–119.

Lazear, Edward P. 2000. Performance Pay and Productivity. *American Economic Review*, 90(5): 1346–1361.

Liu, Y., Ferris, G.R., Xu, J., Weitz, B.A. and Perrewé, P.L., 2014. When ingratiation backfires: The role of political skill in the ingratiation–internship performance relationship. *Academy of Management Learning & Education*, 13(4), pp.569-586.

National Statistical Office (NSO). 2014. Malawi Labour Force Survey 2013. Zomba, Malawi.

National Statistical Office (NSO). 2010. Malawi Demographic and Health Survey.

Nunley, J.M., Pugh, A., Romero, N. and Seals, R.A., 2016. College major, internship experience, and employment opportunities: Estimates from a résumé audit. *Labour Economics*, 38, pp.37-46.

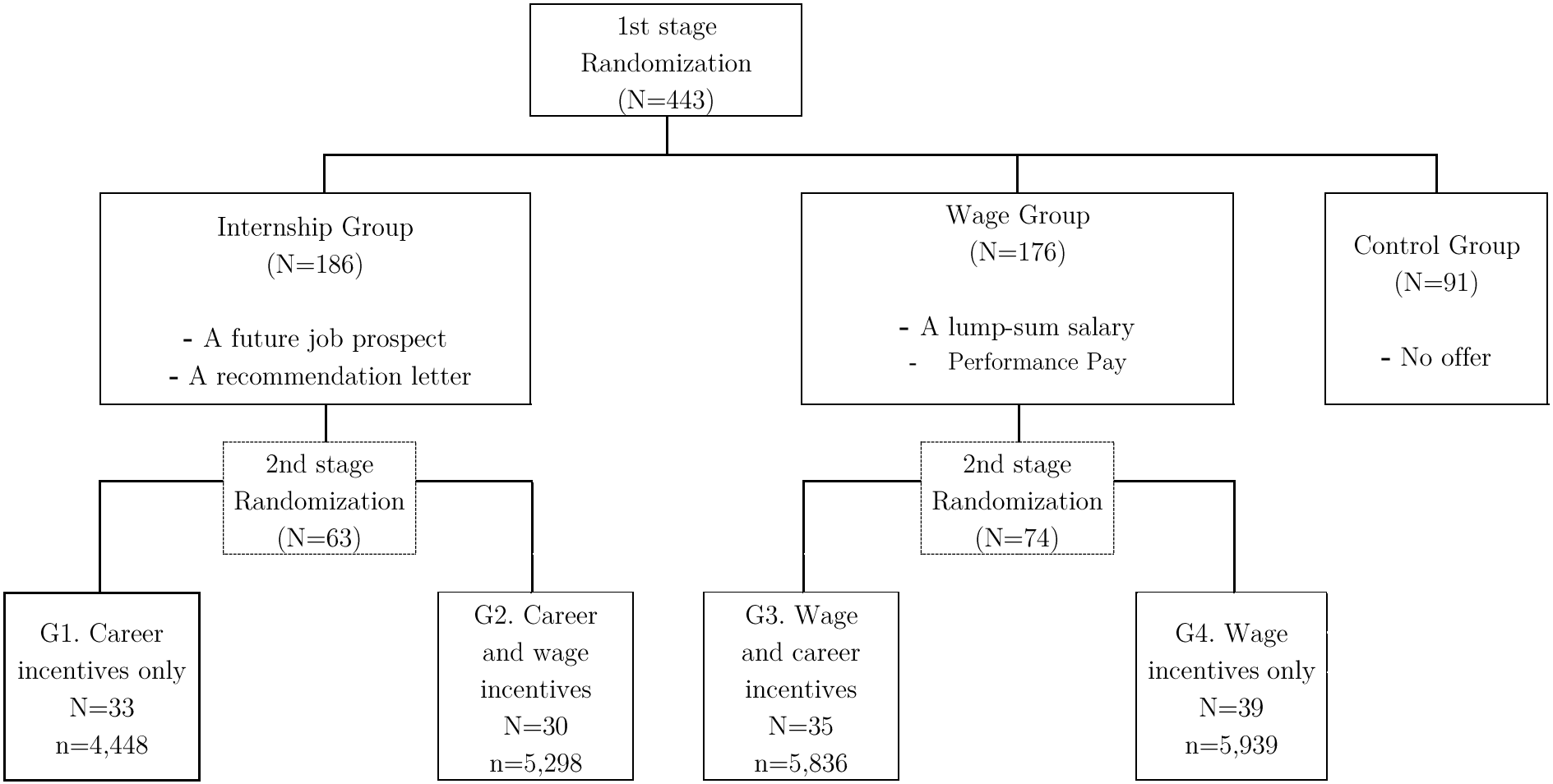
Malawi National Examination Board (MANEB). 2014. Malawi School Certificate of Education Examination- Grades and Awards for Candidates.

Oyer, Paul and Schaefer, S., 2011. Personnel Economics: Hiring and Incentives, In: Orley Ashenfelter and David Card, editors: *Handbook of Labor Economics*, Vol 4b, Great Britain, North Holland, pp. 1769-1823.

Shearer, B., 2004. Piece rates, Fixed Wages and Incentives: Evidence from a Field Experiment. *Review of Economic Studies*, 71(2), pp.513-534.

World Bank. 2015. World Development Indicators.
http://data.worldbank.org/country/malawi

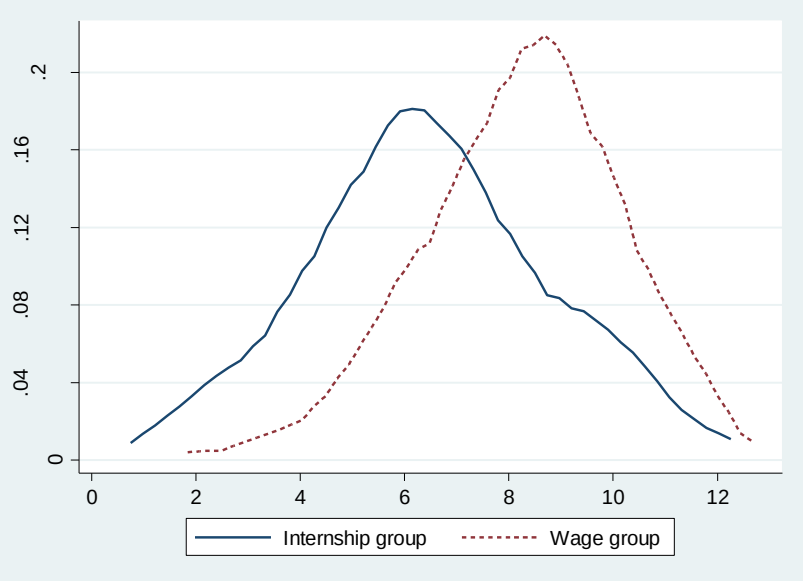
Figure 1: Experimental Design



Notes: Upper case N indicates the number of participants in the each phase. Lower case n indicates the number of surveys conducted by census enumerators in the each group.

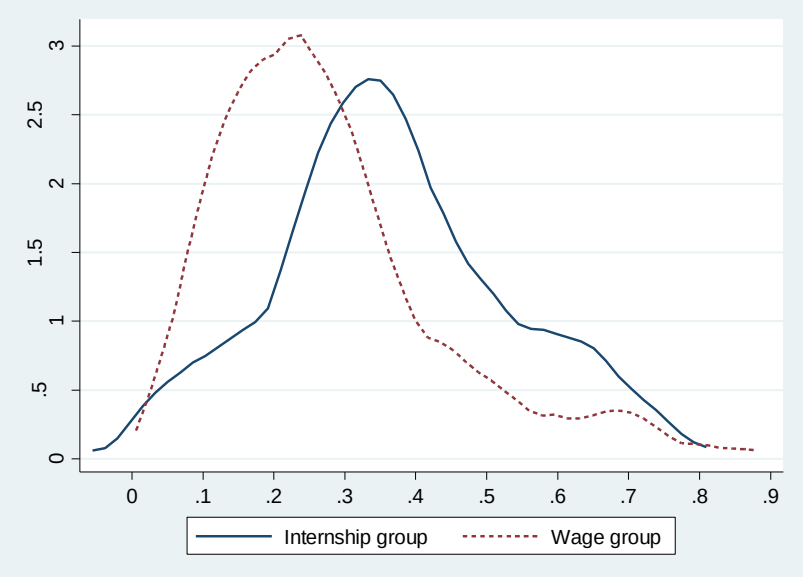
Figure 2: Training Performance

Panel A. Quiz score



Mean (Standard deviation)	Internship group (G1 & G2)	6.46 (2.27)
	Wage group (G3 & G4)	8.24 (1.86)

Panel B. Practice survey error rate

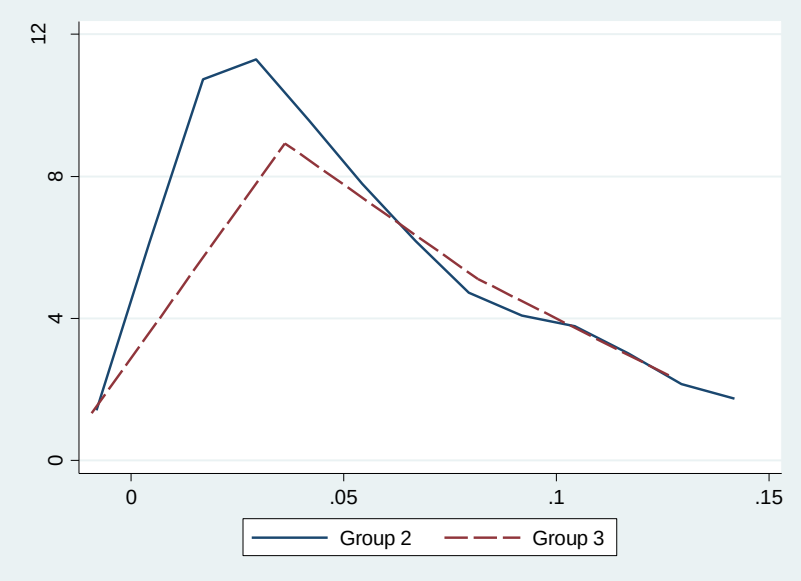


Mean (Standard deviation)	Internship group (G1 & G2)	.365 (.165)
	Wage group (G3 & G4)	.277 (.159)

Notes: The figures present Kernel density estimate (KDE) of quiz score and practice survey error rate. Maximum quiz score is 12. Internship group received an internship offer in the first stage and Wage group received a wage offer in the first stage.

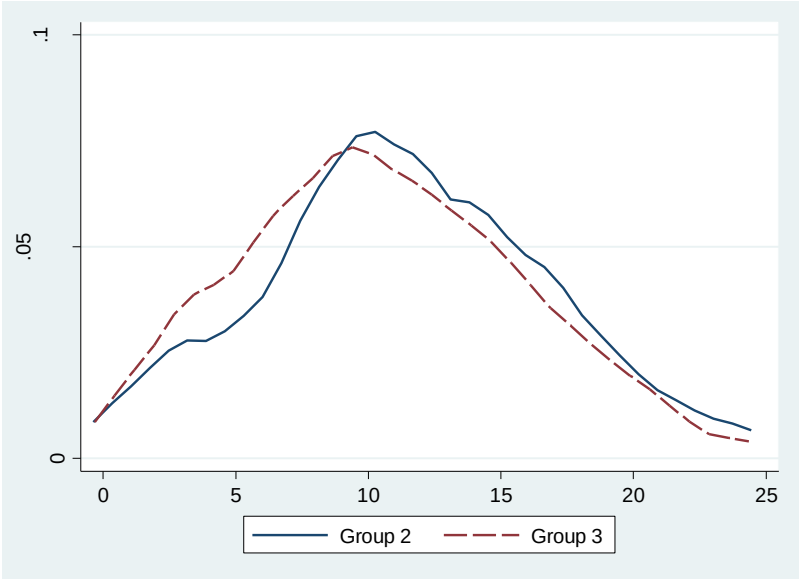
Figure 3: Selection Effect (G2 vs. G3)

Panel A. Survey quality (error rate)



Mean (Standard Deviation)	Group 2	.067 (.061)
	Group 3	.080 (.077)

Panel B. Survey quantity (number of surveys per day)

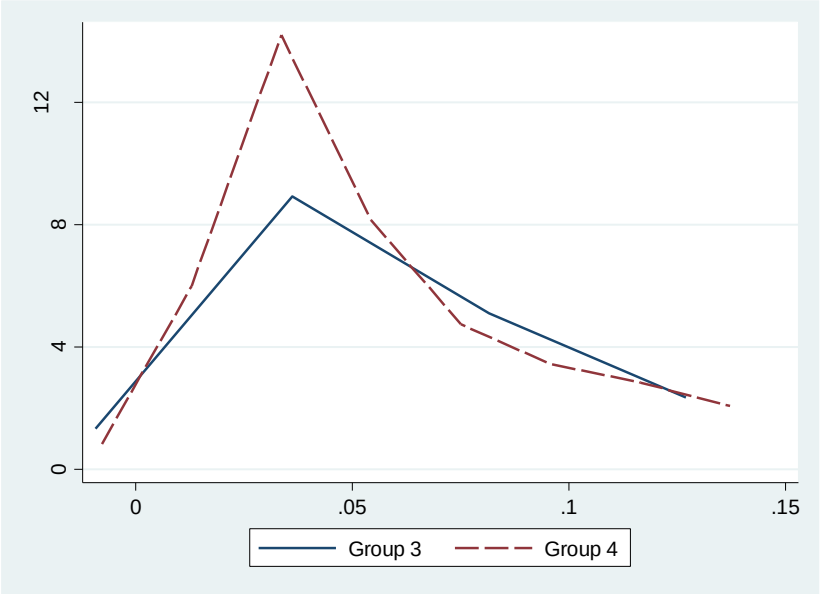


Mean (Standard Deviation)	Group 2	11.6 (5.51)
	Group 3	10.7 (5.45)

Notes: The figures present Kernel density estimate (KDE) of survey quality and survey quantity for Group 2 and Group 3. Group 2 received an internship offer in the first stage and additional wage incentives in the second stage. Group 3 received a wage offer in the first stage and an additional career incentive (internship) in the second stage.

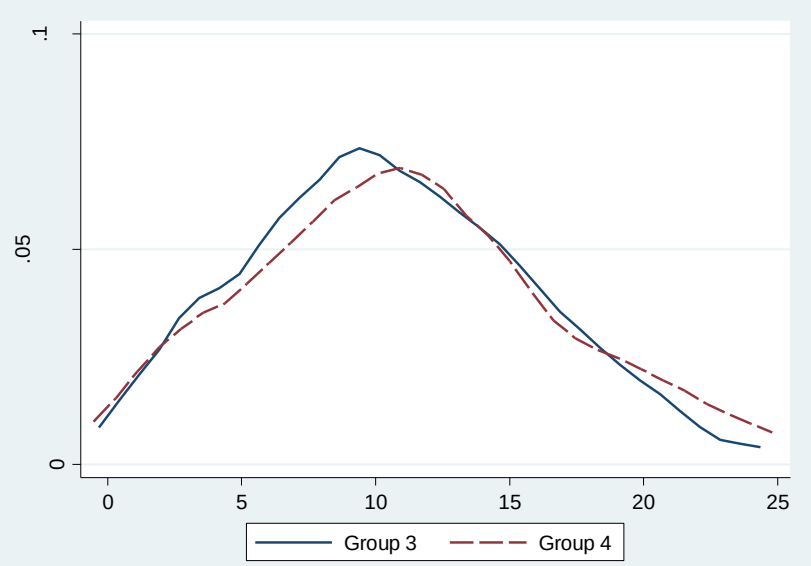
Figure 4: Causal Effect of Career Incentives (G3 vs. G4)

Panel A. Survey quality (error rate)



Mean (Standard Deviation)	Group 3	.080 (.077)
	Group 4	.075 (.070)

Panel B. Survey quantity (number of surveys per day)

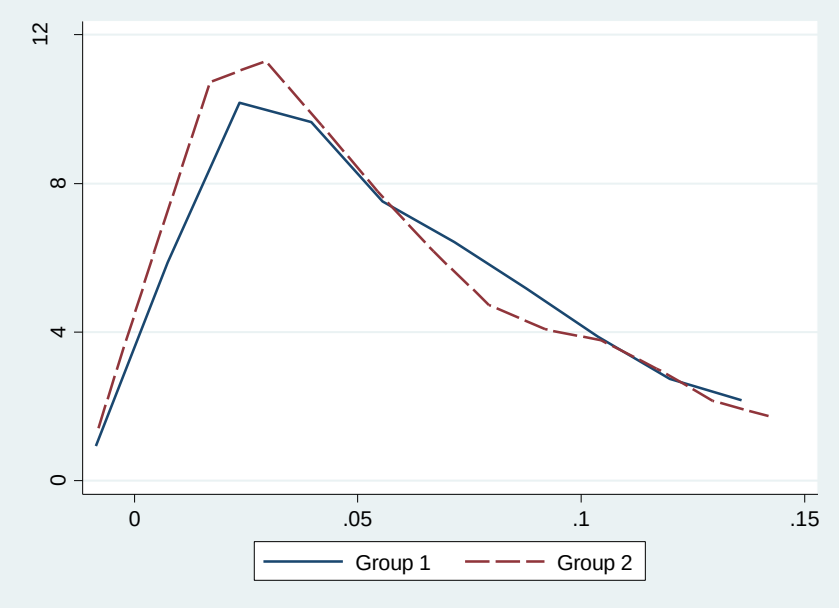


Mean (Standard Deviation)	Group 3	10.7 (5.45)
	Group 4	11.5 (6.35)

Note: The figures present Kernel density estimate (KDE) of survey quality and survey quantity for Group 3 and Group 4. Both Groups 3 and 4 received a wage offer in the first stage but only Group 3 received additional career incentives in the second stage.

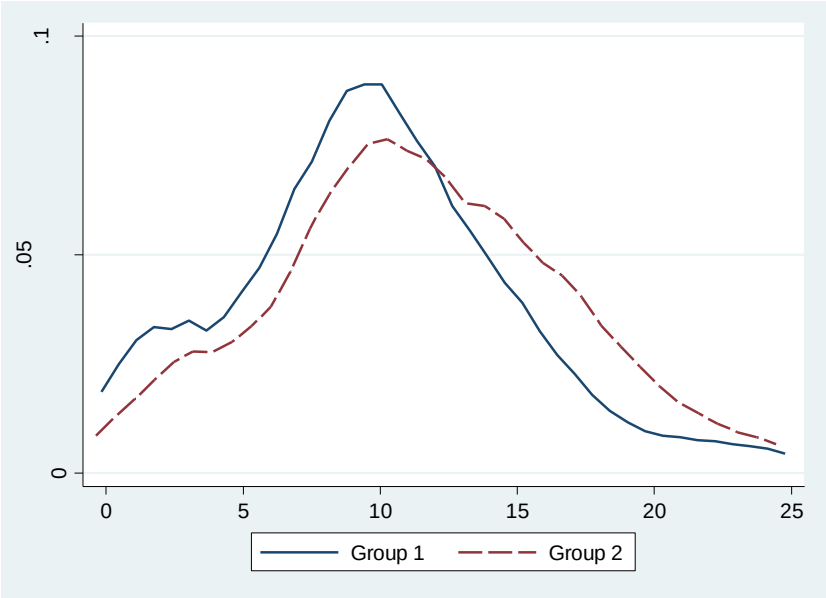
Figure 5: Causal Effect of Wage Incentives (G1 vs. G2)

Panel A. Survey quality (error rate)



Mean (Standard Deviation)	Group 1	.075 (.068)
	Group 2	.067 (.061)

Panel B. Survey quantity (number of surveys per day)

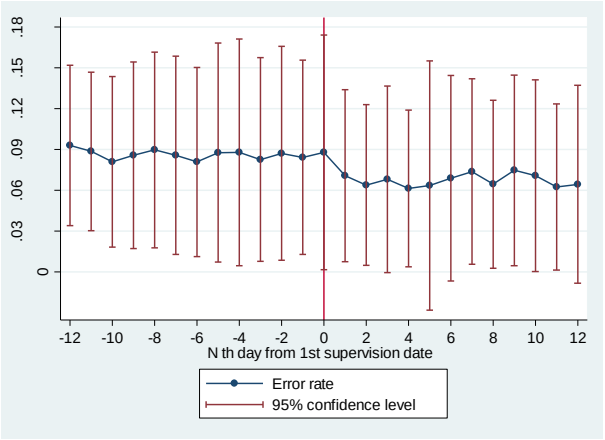


Mean (Standard Deviation)	Group 1	9.85 (5.19)
	Group 2	11.6 (5.51)

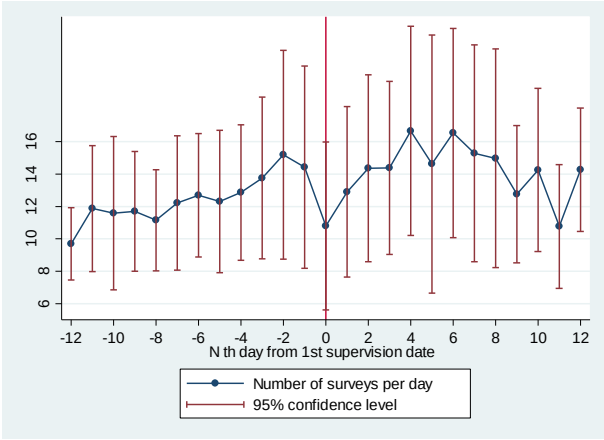
Note: The figures present Kernel density estimate (KDE) of survey quality and survey quantity for Group 1 and Group 2. Both Groups 1 and 2 received an internship offer in the first stage, but only Group 2 received additional wage incentives in the second stage.

Figure 6: Impact of Supervisor Visit on Job Performance

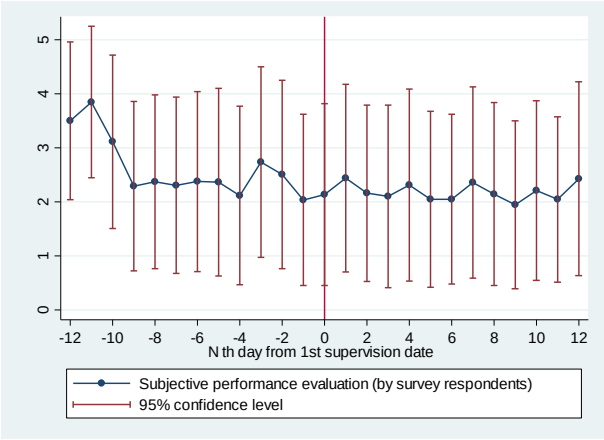
Panel A: Impacts of the first supervisor visit



Error rate

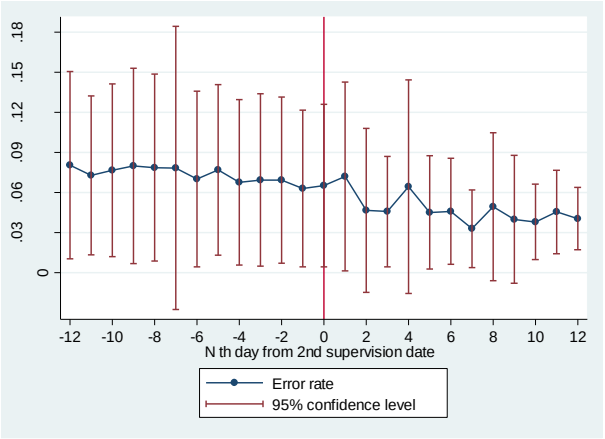


The number of surveys per day

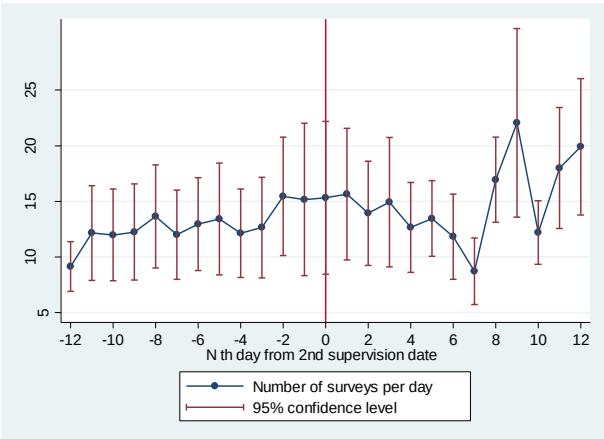


Subjective performance evaluation by survey respondents

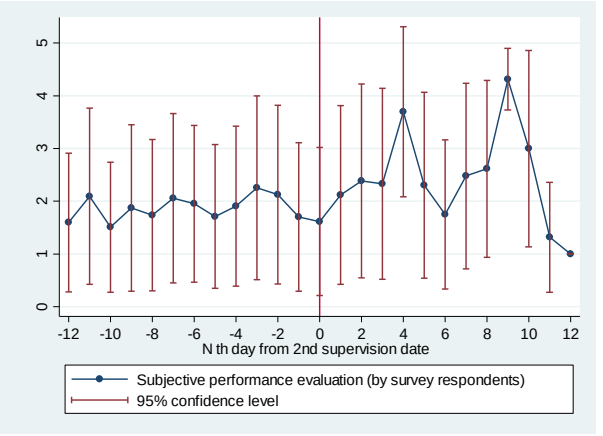
Panel B. Impacts of the second supervisor visit



Error rate



The number of surveys per day



Subjective performance evaluation by survey respondents

Notes: The blue horizontal lines in each panel indicates daily mean of error rate, daily mean of number of surveys, and daily mean of subjective performance evaluation by survey respondents. The red vertical lines indicates 95% confidence interval for each daily values.

Table 1 Experiment Stages

Experiment Stage			Number of individuals						
			G1 (career incentive only)	G2 (career incentives and additional wage incentives)	G3 (wage incentives and additional career incentives)	G4 (wage incentives only)	Control	p-value	Total
A	Target study subjects	2011 Dec	220		220		96		536
B	Baseline survey participants	2014 Dec	186		176		81	.402 (F-stat)	443
C	Conditional job offer takers (Joined the training)	2015 Jan	74 (39.8%)		74 (42.0%)		-	.663 (t-stat)	148
D	Training failures		11 (5.9%)		0 (0.0%)		-	-	11
E	Enumerators	2015 Jan-Feb	63 (33.9%)		74 (42.0%)		-	-	137
			33	30	35	39			

Note: The proportions of individuals remaining at experiment stage C, D, and E are based on baseline survey participants.

Table 2 Randomization Balance Check

Variable	Obs	Full sample	Internship group	Wage group	Control group	Mean difference (p-value) Internship vs Wage	Mean difference (p-value) Internship + Wage vs Control	Mean difference (p-value) G2 vs G1	Mean difference (p-value) G3 vs G4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: 2014 baseline survey									
Age	443	20.3	20.5	20.4	20.0	.065	.395**	-.200	-.207
		(1.61)	(.120)	(.126)	(.159)	(.707)	(.031)	(.629)	(.520)
Number of siblings	443	4,38	4.60	4.17	4.49	.432**	.071	5.00	-.158
		(1.84)	(.132)	(.134)	(.243)	(.022)	(.771)	(.315)	(.650)
Asset score	443	1.15	1.09	1.19	1.22	-.102	-.084	.133	.048
		(.901)	(.066)	(.067)	(.102)	(.282)	(.457)	(.489)	(.799)
Currently working	442	.088	.097	.074	.099	.023	-.014	.036	-.006
		(.284)	(.022)	(.020)	(.033)	(.455)	(.697)	(.514)	(.913)
Self-esteem	443	19.5	19.4	19.3	20.1	-.158	-.706	.441	-.768
		(3.72)	(3.86)	(3.51)	(.421)	(.684)	(.134)	(.662)	(.341)
Intrinsic motivation	443	3.10	3.10	3.09	3.10	.010	-.005	.033	-.075
		(.346)	(.330)	(.351)	(.038)	(.644)	(.912)	(.642)	(.372)
Extrinsic motivation	442	2.84	2.84	2.84	2.81	.000	-.026	.031	.004
		(.286)	(.281)	(.285)	(.031)	(.896)	(.480)	(.646)	(.956)
Extroversion	433	3.53	3.61	3.47	3.44	.140	.103	.055	-.246
		(1.18)	(1.12)	(1.20)	(.136)	(.237)	(.523)	(.851)	(.393)
Agreeableness	443	5.17	5.10	5.10	5.46	.008	-.356**	.035	-.268
		(1.40)	(.106)	(.103)	(.149)	(.955)	(.034)	(.927)	(.408)
Conscientiousness	442	5.77	5.69	5.68	6.17	.010	-.487***	.094	-.054
		(1.34)	(1.34)	(1.36)	(.147)	(.908)	(.002)	(.778)	(.850)
Emotional stability	439	5.12	5.08	5.06	5.31	.020	-.237	.064	-.190
		(1.46)	(1.49)	(1.42)	(.164)	(.905)	(.207)	(.866)	(.591)
Openness to experiences	443	5.18	5.14	5.10	5.45	.043	-.332	-.094	-.268
		(1.52)	(.114)	(.103)	(.194)	(.778)	(.115)	(.779)	(.408)
Time preference	402	.391	.394	.398	.366	-.004	.030	.072*	.013
		(.143)	(.011)	(.011)	(.016)	(.783)	(.101)	(.050)	(.697)
Risk preference	403	.639	.629	.642	.656	-.012	-.020*	.008	-.033*
		(.084)	(.007)	(.006)	(.011)	(.181)	(.089)	(.714)	(.077)
Rationality	402	.819	.817	.836	.786	-.019	.040*	.037	-.007
		(.153)	(.012)	(.011)	(.020)	(.234)	(.068)	(.353)	(.820)
Cognitive ability index	443	-.001	-.019	.049	-.068	-.068	.084	.092	.001
		(.645)	(.047)	(.049)	(.049)	(.314)	(.297)	(.556)	(.995)

Table 2 Randomization Balance Check (continued)

Variable	Obs	Full sample	Internship group	Wage group	Control group	Mean difference (p-value) Internship vs Wage	Mean difference (p-value) Internship + Wage vs Control	Mean difference (p-value) G2 vs G1	Mean difference (p-value) G3 vs G4
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Male circumcision treatment	443	.454	.425	.460	.506	-.035	-.064	-.006	-.226*
		(.498)	(.036)	(.038)	(.056)	(.498)	(.300)	(.962)	(.042)
HIV/AIDS education treatment	443	.483	.511	.443	.506	.068	-.028	-.009	.030
		(.500)	(.037)	(.038)	(.056)	(.199)	(.648)	(.943)	(.800)
Scholarship treatment	443	.458	.414	.500	.469	-.086	-.013	.021	-.024
		(.499)	(.036)	(.038)	(.056)	(.101)	(.829)	(.868)	(.838)
Transportation reimburse	443	1531	1525	1547.7	1511.1	-22.7	24.9	-103.9	-57.2
		(584.4)	(43.8)	(41.8)	(69.2)	(.708)	(.742)	(.516)	(.707)
Panel B: Characteristics of dispatched catchment areas									
Number of households	137	227.1	213.0	239.1	-	-26.2	-	-161.6***	63.8
		(189.4)	(20.2)	(24.6)		(.412)		(.000)	(.213)
Number of family members	137	3.86	3.94	3.79	-	.148	-	.017	.114
		(.633)	(.068)	(.081)		(.165)		(.170)	(.486)
Household asset score	137	.248	.241	.253	-	-.012	-	-.017	.028*
		(.058)	(.006)	(.007)		(.201)		(.170)	(.058)
Birth rate	137	.068	.071	.065	-	.006**	-	.005	.010**
		(.017)	(.002)	(.002)		(.019)		(.119)	(.026)
Malaria incidence	137	.518	.525	.513	-	.012	-	-.063**	-.018
		(.144)	(.014)	(.019)		(.615)		(.025)	(.651)
Death rate	137	.006	.006	.006	-	.000	-	.001	-.001
		(.007)	(.001)	(.001)		(.981)		(.590)	(.717)
Catchment area size	137	3.29	3.11	3.45	-	-.335	-	-.361	.238
		(1.77)	(.133)	(.255)		(.248)		(.178)	(.657)
Number of Observations		433	186	176				63	74

Notes: ***, **, * denote the significance level at 1%, 5%, and 10% respectively. Asset score consists of dummy variables whether to have improved toilet, refrigerator, and bicycle. For non-cognitive traits from self-esteem through openness to experiences, a higher score means a stronger the personality trait of a person is. Time preference is 1 if a person is most impatient and 0 if he is patient. Risk preference is an increasing function of risk-lovingness. Rationality is also an increasing function of economically rational decision making ability. A cognitive ability index is defined as the average z-score of the Raven's matrices test score, the math and English scores of the 2014 Malawi School Certificate of Education (MSCE) test, and the verbal and clerical ability test scores of the O*NET ability test. See the appendix B for the definitions of cognitive and non-cognitive trait variables in details. Male circumcision treatment, HIV/AIDS education treatment, and scholarship treatment are a binary indicator variable of whether to receive male circumcision offer, HIV/AIDS education, and to be belonged to the same class of female scholarship beneficiaries, respectively. Number of households is the average number of households per enumerator. The number of family members is the average number of family members per enumerator. Household asset score is the number of items owned out of the followings: improved toilet, bicycle, lamp, radio, cell phone, bed, and table and chair. Birth rate is the average number of births in the last 3 years per household. Malaria incidence is a proportion of under-3 which had infected with malaria per household. Death rate is a proportion of deaths in the last 12 months per household. Catchment area size is subjectively reported by HSAs and supervisors with a scale 1 to 10.

Table 3 Job Offer Acceptance by Individual Trait

Dependent Variable (Job offer acceptance)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Age	BMI	Number of siblings	Asset score	Currently working	Self-esteem	Intrinsic motivation	Extrinsic motivation
Trait		.042	-.028	.038*	-.068*	-.107	-.024**	-.012	-.019
		(.030)	(.018)	(.019)	(.040)	(.136)	(.010)	(.108)	(.136)
Internship group	-.024	-.323	-.901*	-.029	-.023	-.025	-.321	.521	.733
	(.052)	(.747)	(.489)	(.131)	(.085)	(.055)	(.278)	(.491)	(.520)
Trait * Internship group		.015	.044*	-.002	-.009	.028	.015	-.176	-.266
		(.037)	(.025)	(.028)	(.054)	(.180)	(.014)	(.157)	(.182)
Constant	.481***	-.372	1.03***	.326***	.558***	.491***	.931***	.517	.537
	(.055)	(.613)	(.357)	(.094)	(.073)	(.057)	(.205)	(.336)	(.387)
Observations	362	362	360	362	362	362	362	362	361
R-squared	.018	.046	.028	.036	.036	.021	.034	.027	.031
Mean (Standard Deviation)		20.4(1.65)	19.8(2.13)	4.39(1.80)	1.14(.896)	.086(.280)	19.3(3.69)	3.09(.340)	2.84(.282)

Dependent Variable (Job offer acceptance)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	Extroversion	Agreeableness	Conscientiousness	Emotional stability	Openness to experiences	Time preference	Risk preference	Rationality	Cognitive ability index
Trait	-.058*	-.001	.046*	.011	-.001	.196	.288	.019	-.126**
	(.032)	(.027)	(.026)	(.027)	(.027)	(.284)	(.498)	(.274)	(.053)
Internship group	-.297*	.025	.251	.145	.041	-.096	.388	-.228	-.034
	(.173)	(.196)	(.216)	(.195)	(.187)	(.158)	(.413)	(.305)	(.052)
Trait * Internship group	.077*	-.010	-.049	-.033	-.013	.199	-.644	.257	-.057
	(.046)	(.037)	(.037)	(.037)	(.035)	(.384)	(.640)	0.363)	(.073)
Constant	.683***	.486***	.223	.426***	.485***	.407***	.299	.502**	.490***
	(.126)	(.148)	(.152)	(.148)	(.148)	(.130)	(.324)	(.234)	(.054)
Observations	358	362	361	360	362	334	335	334	362
R-squared	.027	.019	.026	.020	0.019	.024	.019	.019	0.060
Mean (Standard Deviation)	3.54(1.16)	5.11(1.39)	5.68(1.35)	5.07(1.45)	5.36(1.35)	.396(.144)	.635(.083)	.826(.149)	.348(.477)

Notes: ***, **, * denote the significance level at 1%, 5%, and 10% respectively. Asset score is the sum of items owned out of improved toilet, refrigerator, and bicycle. For non-cognitive traits, a higher score indicates a stronger trait. Time preference, risk preference, and rationality are increasing functions of impatience, risk-lovingness, and rational decision making ability, respectively. A cognitive ability index is defined as the average z-score of the Raven's matrices test score, the math and English scores of the 2014 MSCE test, and the verbal and clerical ability test scores of the O*NET ability test. See the data appendix for the definitions of cognitive and non-cognitive trait variables.

Table 4 Training Performance

Dependent variable	Quiz score		Practice survey error rate	
	(1)	(2)	(3)	(4)
Panel A: 148 Trainee Sample				
Internship group	-2.01***	-1.84***	.104***	.089***
	(.344)	(.314)	(.026)	(.029)
Observations	148	147	148	147
R-squared	.228	.531	.111	.236
Wage Group Mean (Standard Deviation)	8.43 (1.82)		.272 (.142)	
Individual characteristics	X	O	X	O
Panel B: 137 Enumerator Sample				
Internship group	-1.44***	-1.41***	.093***	.080***
	(.329)	(.298)	(.028)	(.030)
Observations	137	136	137	136
R-squared	.163	.495	.094	.239
Wage Group Mean (Standard Deviation)	8.43 (1.82)		.272 (.142)	
Individual characteristics	X	O	X	O

Notes: Robust standard errors are reported in parentheses. ***, **, * denote the significance level at 1%, 5%, and 10% respectively. All specifications include number of siblings and binary indicator variables of whether an individual received a male circumcision offer, HIV/AIDS education, and/or he belonged to the same class of female scholarship beneficiaries, respectively. Individual characteristics include age, asset score, cognitive ability index, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items).

Table 5 Selection and Causal Effects of Work Incentives on Job Performance: Main Outcomes

VARIABLES	Survey quality (error rate)			Survey quantity (number of surveys per day)			Subjective performance evaluation (by survey respondents)			Subjective evaluation of work attitude (by supervisors)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Selection effect (G2 vs G3)											
G2	-.020*	-.020**	-.021**	1.48***	1.36***	1.41***	.783**	.708*	.691*	-.174*	-.137
	(.011)	(.008)	(.008)	(.516)	(.505)	(.486)	(.387)	(.372)	(.364)	(.100)	(.106)
Observations	11,130	11,130	11,130	1,003	1,003	1,003	6,473	6,473	6,473	65	65
R-squared	.156	.296	.302	.144	.092	.166	.443	.579	.592	.401	.606
Mean (SD) of G3	.077 (.078)			10.7 (5.45)			2.09 (1.65)			.850 (.163)	
Panel B: Causal effect of career incentives (G3 vs. G4)											
G3	.011	.005	.006	-.594	-.587	-.867	.095	.405	.391	.305***	.277***
	(.011)	(.010)	(.010)	(.602)	(.652)	(.623)	(.368)	(.378)	(.351)	(.038)	(.048)
Observations	11,775	11,775	11,775	1,063	1,063	1,063	7,233	7,233	7,233	74	74
R-squared	.181	.252	.265	.149	.115	.185	.379	.469	.492	.619	.681
Mean (SD) of G4	.082 (074)			11.5 (6.36)			2.08 (1.59)			.583 (.119)	
Panel C: Causal effect of wage incentives (G1 vs. G2)											
G2	-.028*	-.019*	-.019*	1.19*	.922	1.18	.276	.242	.237	-.134	-.151
	(.017)	(.011)	(.011)	(.619)	(.672)	(.735)	(.546)	(.613)	(.608)	(.155)	(.233)
Observations	9,779	9,779	9,779	914	914	914	4,516	4,516	4,516	63	63
R-squared	.167	.347	.354	.203	.169	.229	.389	.592	.607	.366	.502
Mean (SD) of G1	.075 (.068)			9.84 (5.19)			2.67 (1.66)			.803 (.162)	
Panel D: Combined effect (G1 vs. G4)											
G1	-.002	-.004	-.003	-1.45	-1.24	-1.35	-.269	-.005	-.042	.191***	.202**
	(.013)	(.013)	(.013)	(.984)	(1.03)	(1.05)	(.474)	(.491)	(.472)	(.067)	(.092)
Observations	10,424	10,424	10,424	974	974	974	5,276	5,276	5,276	72	72
R-squared	.194	.264	.276	.157	.163	.221	.517	.609	.623	.569	.627
Mean (SD) of G4	.082 (074)			11.5 (6.36)			2.08 (1.59)			.583 (.119)	
Survey date fixed effect	YES	NO	YES	YES	NO	YES	YES	NO	YES	NO	NO
Time trend	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	NO
Individual characteristics	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES

Notes: Standard errors clustered at enumerator level are reported in parentheses. ***, **, * denote the significance level at 1%, 5%, and 10% respectively. All specifications include number of siblings, catchment area control variables, supervisor fixed effect variables, and binary indicator variables of whether an individual received a male circumcision offer, HIV/AIDS education, and/or he belonged to the same class of female scholarship beneficiaries, respectively. Survey date fixed effect variable is dummy variable of each day from the beginning. Time trend control variable is an increasing constant variable starting from the beginning of the census. Individual characteristics include age, asset score, cognitive ability index, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items). Catchment area control variables include the total number of households, the number of family members, asset score (whether to own refrigerator, bicycle, and improved toilet), birth rate in the last 3 years, incidence of malaria among under 3, and deaths in the last 12 months. Supervisor fixed effect variable is multiplication of dummy variable for each supervision team and dummy variable for surveys after the supervision.

Table 6: Impacts of Supervisor Visits

Variable	Survey quality (error rate)		Survey quantity (number of surveys per day)		Subjective performance evaluation (by survey respondents)	
	(1)	(2)	(3)	(4)	(5)	(6)
First visit	-.015***	-.005	1.34**	.839	-.039	-.536**
(1 if surveyed from date of the first visit onward, 0 otherwise)	(.004)	(.006)	(.627)	(.606)	(.172)	(.251)
Second visit	-.004	.005	-1.63	-.954	.234	.149
(1 if surveyed from date of the second visit onward, 0 otherwise)	(.007)	(.008)	(1.34)	(1.44)	(.258)	(.310)
Observations	20,381	20,381	1,841	1,841	11,099	11,099
R-squared	.214	.228	.087	.125	.272	.296
Linear time trend	O	X	O	X	O	X
Survey date fixed effect	X	O	X	O	X	O
Mean (Standard Deviation) of the dependent variable	.074 (.071)		10.9 (5.70)		2.29 (1.69)	

Notes: Standard errors clustered at the enumerator level are reported in parentheses. ***, **, and * denote significance at 1%, 5%, and 10%, respectively. All specifications include binary indicator variables of whether an individual received a male circumcision offer, HIV/AIDS education, and/or he belonged to the same class of female scholarship beneficiaries, respectively. Catchment area control variables include the total number of households, the number of family members, asset score (whether to own refrigerator, bicycle, and improved toilet), birth rate in the last 3 years, incidence of malaria among under 3, and deaths in the last 12 months. Research group control is dummy variables for G1, G2, G3, and G4. Survey date fixed effect variable is dummy variable of each day from the beginning. Time trend control variable is an increasing constant variable starting from the beginning of the census. Individual characteristics include age, number of siblings, asset score, cognitive ability index, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items).

Table 7: Short-term Impacts of Job Experience on Employment

VARIABLES	Currently working for paid job	
	(1)	(2)
Panel A: Effect of short-term job experience (G1, G2, G3 and G4 vs. Control)		
Received a job offer (internship or wage)	-.042	-.064
	(.039)	(.042)
Observations	434	421
R-squared	.012	.051
Control Group Mean (SD)	.114 (.320)	
Panel B: Effect of career and wage incentives (Internship group, Wage group vs. Control)		
Internship offer	-0.015	-0.037
	(0.042)	(0.045)
Wage offer	-0.072*	-0.092**
	(0.040)	(0.043)
Observations	434	421
R-squared	0.021	0.059
Omitted group Mean (Standard Deviation)	.041 (.198)	
Panel C: Effect of career incentive (Internship group vs. Wage group)		
Received an internship offer	.054**	.048*
	(.027)	(.027)
Observations	355	349
R-squared	.029	.080
Wage Group Mean (SD)	.041 (.198)	
Panel D: Those accepted an internship offer vs others (those refused an internship offer + wage group)		
Accepted an internship offer	.099**	.091**
	(.045)	(.045)
Refused an internship offer	.025	.018
	(.029)	(.029)
Observations	355	349
R-squared	.038	.090
Omitted group Mean (Standard Deviation)	.041 (.198)	
Individual characteristics	NO	YES

Notes: ***, **, * denote the significance level at 1%, 5%, and 10% respectively. All specifications include binary indicator variables of whether an individual received a male circumcision offer, HIV/AIDS education, and/or he belonged to the same class of female scholarship beneficiaries, respectively. Individual characteristics include age, number of siblings, asset score, cognitive ability index, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items).

Online Appendix (not for publication)

Appendix Tables

Table A.1: Characteristics of baseline survey participants and non-participants

Variable	Participants	Non-participants	Mean difference between participants and non-participants (p-value)
	(1)	(2)	(3)
Height	164.5	164.5	.047
	(.367)	(.743)	(.955)
Weight	53.5	53.9	-.430
	(.342)	(.984)	(.680)
Age	16.1	16.0	.065
	(.070)	(.197)	(.758)
Living with a father	.639	.645	-.006
	(.023)	(.050)	(.908)
Living with a mother	.747	.667	.081
	(.021)	(.049)	(.134)
Asset score	1.17	1.41	-.240**
	(.042)	(.106)	(.037)
Subjective health is good or very good	.433	.538	.104*
	(.024)	(.052)	(.070)
Raven's matrices test score	20.0	18.7	1.32
	(.244)	(.696)	(.077)
Number of Observations	443	93	

Notes: ***, **, * denote the significance level at 1%, 5%, and 10%, respectively. The statistics are calculated from the 2011 secondary school census survey. Columns (1) and (2) show group-specific means and standard deviations. 536 male secondary school graduates were contacted and invited for the baseline survey, but 443 showed up on the survey date. Asset score is the sum of items owned out of improved toilet, refrigerator, bicycle, electricity, and car or truck.

Table A.2: Individual characteristics after job offer acceptance

Variable	Obs	Internship offer takers	Wage offer takers	Mean Difference	Standard Deviation
	(1)	(2)	(3)	(4)=(2)-(3)	(5)
Age	148	20.8	20.7	.162	1.46
BMI	148	19.9	19.5	.413	2.08
Number of siblings	148	4.86	4.46	.405	1.70
Asset score	148	.932	1.05	-.122	.804
Currently working	148	.081	.054	.027	.252
Self-esteem	148	19.1	18.6	.521	3.71
Intrinsic motivation	148	3.05	3.08	-.029	.326
Extrinsic motivation	148	2.78	2.83	-.046	.274
Extroversion	148	3.67	3.27	.405**	1.19
Agreeableness	148	5.03	5.10	-.074	1.44
Conscientiousness	148	5.67	5.87	-.196	1.26
Emotional stability	148	4.94	5.12	-.182	1.50
Openness to experiences	148	5.03	5.10	-.074	1.44
Time preference	137	.414	.411	.003	.136
Risk preference	137	.621	.645	-.024*	.079
Rationality	137	.831	.834	-.004	.139
Cognitive Ability Index	148	-.199	-.077	-.119	.591
Male circumcision treatment	148	.392	.338	.054	.483
HIV/AIDS education treatment	148	.473	.473	.000	.501
Scholarship treatment	148	.459	.473	-.013	.501
Transportation reimburse	148	1602.7	1652.7	-50.0	628.2
F-statistics (p-value)				.950 (.532)	
Number of Individuals		74	74	148	148

Notes: ***, **, * denote the significance level at 1%, 5%, and 10%, respectively. Asset score is the sum of items owned out of improved toilet, refrigerator, and bicycle. For non-cognitive traits, a higher score indicates a stronger trait. Time preference, risk preference, and rationality are increasing functions of impatience, risk-lovingness, and rational decision making ability, respectively. A cognitive ability index is defined as the average z-score of the Raven's matrices test score, the math and English scores of the 2014 MSCE test, and the verbal and clerical ability test scores of the O*NET ability test. See the data appendix for the definitions of cognitive and non-cognitive trait variables. Male circumcision, HIV/AIDS education treatment, and scholarship are binary indicator variables of the past eligibility status of AFF's corresponding interventions.

Table A.3: Selection and Causal Effects of Work Incentives on Job Performance: Additional Outcomes

VARIABLES	Survey quality				Survey quantity					
	Proportion of entries incorrectly entered		Proportion of entries incorrectly blank		Work hours (in mins)		Survey time per household (in mins)		Intermission time between surveys (in mins)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Selection effect (G2 vs G3)										
G2	.001	0.000	-.021**	-.021***	-5.00	-.871	-1.50	-1.19	-3.98**	-4.07***
	(.003)	(0.002)	(.009)	(.007)	(19.2)	(16.6)	(1.13)	(.935)	(1.67)	(1.32)
Observations	11,130	11,130	11,130	11,130	988	988	11,130	11,130	9,325	9,325
R-squared	.107	0.240	.148	.263	.139	.165	.276	.312	.018	.027
Mean (SD) of G3	.016 (.018)		.062 (.070)		429.3 (207.2)		25.5 (11.9)		19.2 (44.7)	
Panel B: Causal effect of career incentives (G3 vs. G4)										
G3	.002	-0.000	.010	.006	39.7**	28.0	1.09	1.36	6.72***	6.15***
	(.003)	(0.003)	(.009)	(.008)	(17.1)	(17.9)	(1.23)	(1.14)	(1.64)	(1.59)
Observations	11,775	11,775	11,775	11,775	1,054	1,054	11,775	11,775	10,179	10,179
R-squared	.161	0.290	.161	.217	.143	.158	.244	.261	.019	.026
Mean (SD) of G4	.019 (.021)		.063 (.066)		396.1 (201.1)		23.9 (11.2)		13.6 (33.7)	
Panel C: Causal effect of wage incentives (G1 vs. G2)										
G2	-.006	-0.008**	-.022	-.012	16.1	7.55	-2.94*	-2.39	.755	.591
	(.004)	(0.003)	(.014)	(.009)	(24.6)	(23.7)	(1.52)	(1.67)	(2.56)	(3.18)
Observations	9,779	9,779	9,779	9,779	889	889	9,780	9,780	8,184	8,184
R-squared	.102	0.230	.148	.297	.183	.210	.302	.334	.016	.022
Mean (SD) of G1	.019 (.019)		.056 (.061)		391.4 (199.3)		27.4 (12.1)		15.7 (35.7)	
Panel D: Combined effect (G1 vs. G4)										
G1	.007*	0.010**	-.009	-.012	-15.2	-17.6	2.17	1.66	1.96	2.29
	(.004)	(0.004)	(.011)	(.010)	(27.6)	(29.7)	(1.42)	(1.54)	(1.73)	(1.84)
Observations	10,424	10,424	10,424	10,424	955	955	10,425	10,425	9,038	9,038
R-squared	.158	0.258	.167	.237	.143	.166	.278	.317	.015	.022
Mean (SD) of G4	.019 (.021)		.063 (.066)		396.1 (201.1)		23.9 (11.2)		13.6 (33.7)	
Individual characteristics	NO	YES	NO	YES	NO	YES	NO	YES	NO	YES

Notes: Standard errors clustered at enumerator level are reported in parentheses. ***, **, * denote the significance level at 1%, 5%, and 10% respectively. All specifications include number of siblings, survey date fixed effect variable, catchment area control variables, supervisor fixed effect variables, and binary indicator variables of whether an individual received a male circumcision offer, HIV/AIDS education, and/or he belonged to the same class of female scholarship beneficiaries, respectively. Survey time outcome is imputed by set of control variables; catchment area control, survey date fixed effect, supervisor fixed effect, binary indicators of previous HIV/AIDS projects, and enumerator fixed effect. Individual characteristics include age, asset score, cognitive ability index, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items). Survey date fixed effect variable is dummy variable of each day from the beginning. Catchment area control variables include the total number of households, the number of family members, asset score (whether to own refrigerator, bicycle, and improved toilet), birth rate in the last 3 years, incidence of malaria among under 3, and deaths in the last 12 months. Supervisor fixed effect variable is multiplication of dummy variable for each supervision team and dummy variable for surveys after the supervision.

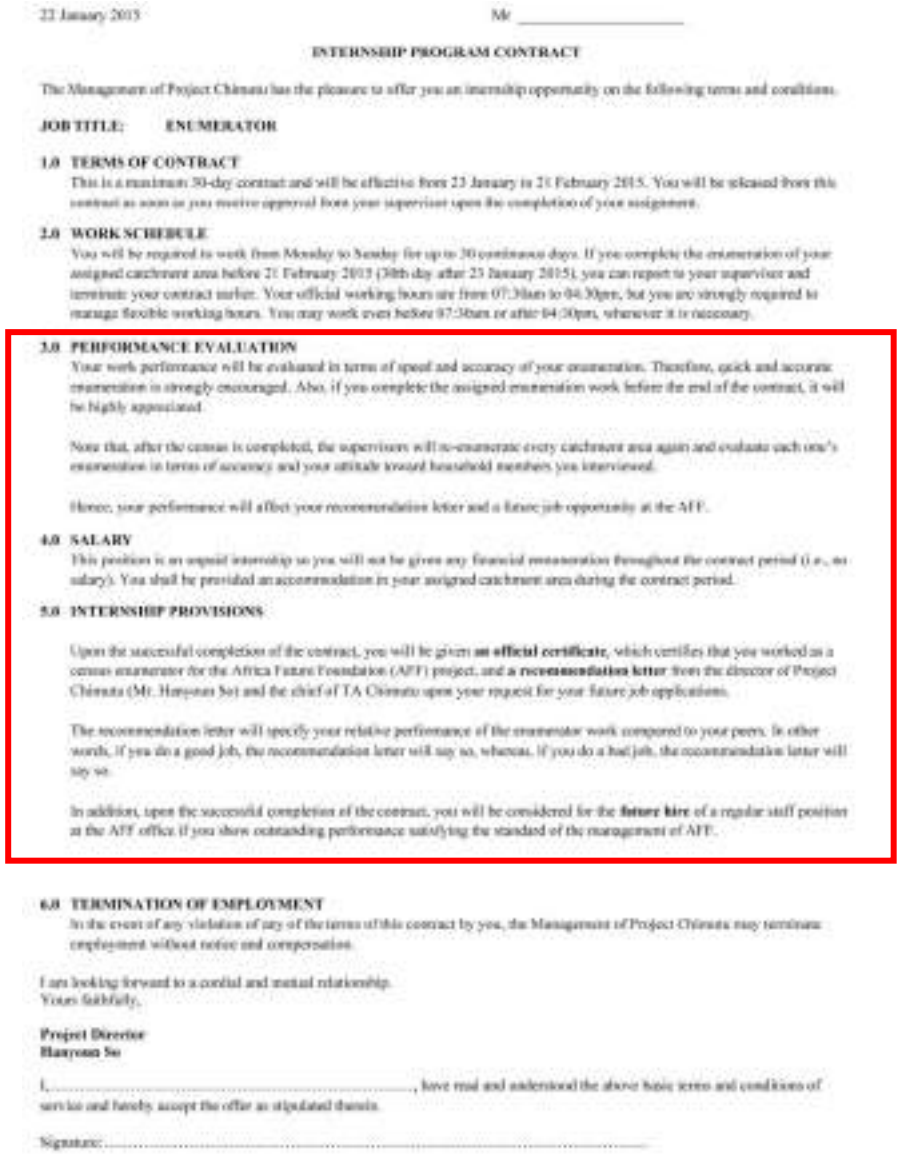
Table A.4: Selection and causal effects of work incentives on job performance after excluding 11 enumerators from the *Wage* group

VARIABLES	Survey quality (error rate)			Survey quantity (number of surveys)			Subjective performance evaluation (by survey respondents)			Subjective evaluation of work attitude (by supervisors)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Panel A: Selection effect (G2 vs G3)											
G2	-.005	-.014*	-.014*	1.82***	1.63***	1.70***	.843**	.826**	.814**	-.186*	-.156
	(.010)	(.007)	(.007)	(.540)	(.554)	(.530)	(.399)	(.388)	(.378)	(.104)	(.150)
Observations	10,150	10,150	10,150	917	917	917	5,906	5,906	5,906	59	59
R-squared	.165	.287	.293	.152	.100	.172	.446	.569	.584	.394	.617
Mean (SD) of G3	.067 (.064)			10.6 (5.60)			2.11 (1.66)			.845 (.169)	
Panel B: Causal effect of career incentives (G3 vs. G4)											
G3	.011	.012	.013	-1.30**	-1.48*	-1.82**	.342	.614	.594	.325***	.308***
	(.009)	(.011)	(.010)	(.624)	(.763)	(.764)	(.410)	(.404)	(.371)	(.052)	(.076)
Observations	9,666	9,666	9,666	876	876	876	5,983	5,983	5,983	63	63
R-squared	.197	.240	.258	.178	.138	.207	.348	.492	.518	.610	.692
Mean (SD) of G4	.085 (.076)			11.5 (6.47)			1.94 (1.52)			.596 (.123)	
Panel C: Causal effect of wage (G1 vs. G2)											
G2	-.028*	-.019*	-.019*	1.19*	.922	1.18	.276	.242	.237	-.129	-.151
	(.017)	(.011)	(.011)	(.619)	(.672)	(.735)	(.546)	(.613)	(.608)	(.130)	(.233)
Observations	9,779	9,779	9,779	914	914	914	4,516	4,516	4,516	63	63
R-squared	.167	.347	.354	.203	.169	.229	.389	.592	.607	.344	.502
Mean (SD) of G1	.075 (.068)			9.84 (5.19)			2.67 (1.66)			.803 (.162)	
Panel D: Combined effect (G1 vs. G4)											
G1	.000	.001	.002	-1.32	-1.33	-1.27	.013	.724	.666	.154**	.095
	(.013)	(.014)	(.013)	(1.02)	(1.18)	(1.23)	(.439)	(.468)	(.456)	(.063)	(.099)
Observations	9,295	9,295	9,295	873	873	873	4,593	4,593	4,593	67	67
R-squared	.196	.267	.282	.177	.179	.232	.574	.701	.710	.587	.723
Mean (SD) of G4	.085 (.076)			11.5 (6.47)			1.94 (1.52)			.596 (.123)	
Survey date fixed effect	YES	NO	YES	YES	NO	YES	YES	NO	YES	NO	NO
Time trend	NO	YES	NO	NO	YES	NO	NO	YES	NO	NO	NO
Individual characteristics	NO	YES	YES	NO	YES	YES	NO	YES	YES	NO	YES

Notes: 11 enumerators in the *Wage* group whose training performance is the lowest are excluded. Standard errors clustered at enumerator level are reported in parentheses. ***, **, * denote the significance level at 1%, 5%, and 10%, respectively. All specifications include i) binary indicatory variables of past eligibility status of AFF's intervention programs such as a male circumcision offer, HIV/AIDS education, and scholarship, ii) number of siblings, iii) supervisor-specific post-visit fixed effect, and iii) catchment area control variables which include the total number of households, the number of family members, asset score (whether subjects own refrigerator, bicycle, and improved toilet), birth rate in the last 3 years, incidence of malaria among under 3, and deaths in the last 12 months. Individual characteristics include age, asset score, cognitive ability index, and a set of non-cognitive traits (self-esteem, intrinsic and extrinsic motivation, and Big 5 personality items).

Appendix Figures

Figure A.1: Contract letter for Group 1 (G1)



22 January 2015 Mr _____

INTERNSHIP PROGRAM CONTRACT

The Management of Project Chimutu has the pleasure to offer you an internship opportunity on the following terms and conditions.

JOB TITLE: ENUMERATOR

1.0 TERMS OF CONTRACT
This is a maximum 30-day contract and will be effective from 23 January to 21 February 2015. You will be released from this contract as soon as you receive approval from your supervisor upon the completion of your assignment.

2.0 WORK SCHEDULE
You will be required to work from Monday to Sunday for up to 30 continuous days. If you complete the enumeration of your assigned catchment area before 21 February 2015 (30th day after 23 January 2015), you can report to your supervisor and terminate your contract earlier. Your official working hours are from 07:30am to 04:30pm, but you are strongly required to manage flexible working hours. You may work even before 07:30am or after 04:30pm, whenever it is necessary.

3.0 PERFORMANCE EVALUATION
Your work performance will be evaluated in terms of speed and accuracy of your enumeration. Therefore, quick and accurate enumeration is strongly encouraged. Also, if you complete the assigned enumeration work before the end of the contract, it will be highly appreciated.

Note that, after the census is completed, the supervisors will re-enumerate every catchment area again and evaluate each one's enumeration in terms of accuracy and your attitude toward household members you interviewed.

Hence, your performance will affect your recommendation letter and a future job opportunity at the AFF.

4.0 SALARY
This position is an unpaid internship so you will not be given any financial remuneration throughout the contract period (i.e., no salary). You shall be provided an accommodation in your assigned catchment area during the contract period.

5.0 INTERNSHIP PROVISIONS

Upon the successful completion of the contract, you will be given **an official certificate**, which certifies that you worked as a census enumerator for the Africa Future Foundation (AFF) project, and **a recommendation letter** from the director of Project Chimutu (Mr. Hanyoun So) and the chief of TA Chimutu upon your request for your future job applications.

The recommendation letter will specify your relative performance of the enumerator work compared to your peers. In other words, if you do a good job, the recommendation letter will say so, whereas, if you do a bad job, the recommendation letter will say so.

In addition, upon the successful completion of the contract, you will be considered for the **future hire** of a regular staff position at the AFF office if you show outstanding performance satisfying the standard of the management of AFF.

6.0 TERMINATION OF EMPLOYMENT
In the event of any violation of any of the terms of this contract by you, the Management of Project Chimutu may terminate employment without notice and compensation.

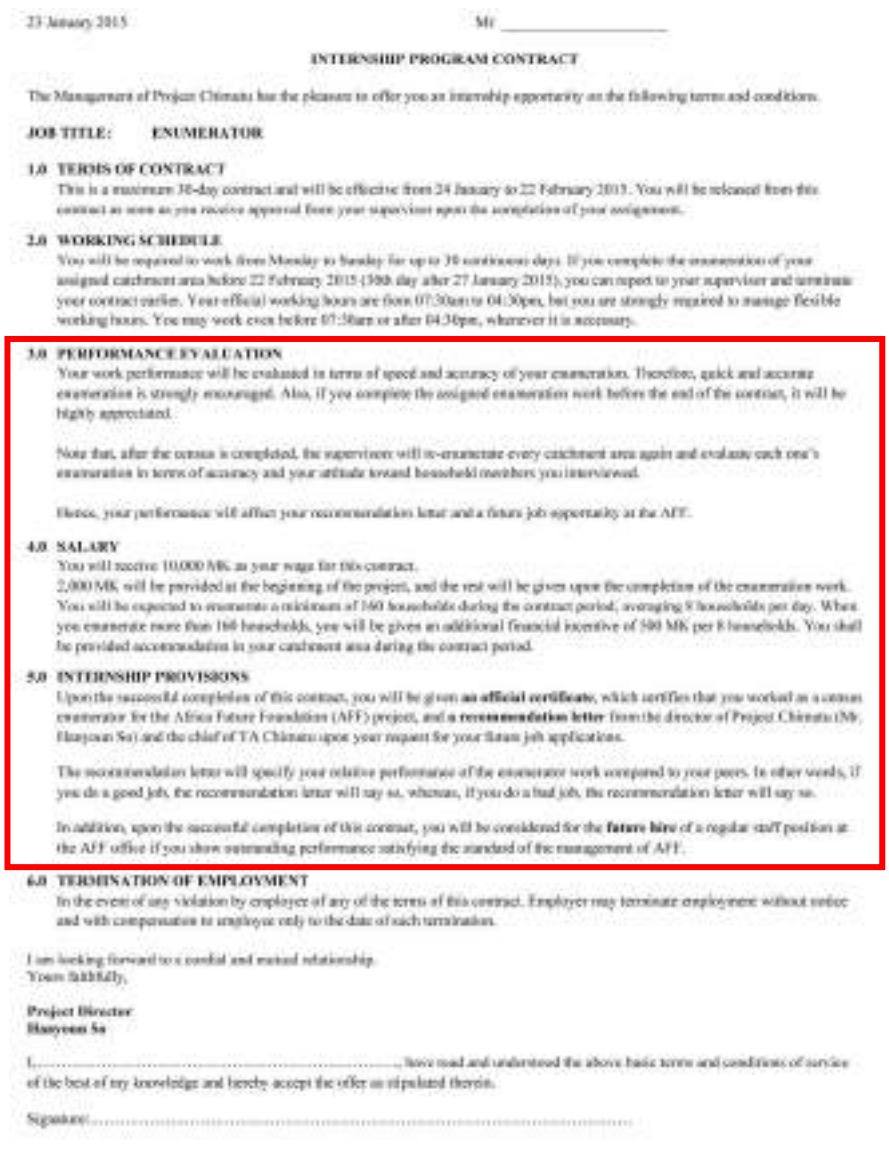
I am looking forward to a cordial and mutual relationship.
Yours faithfully,

Project Director
Hanyoun So

I,_____, have read and understood the above basic terms and conditions of service and hereby accept the offer as stipulated therein.

Signature:_____

Figure A.2: Contract letter for Group 2 (G2) and Group (G3) (the same contract letter for both groups)



23 January 2015 Mr _____

INTERNSHIP PROGRAM CONTRACT

The Management of Project Chimutu has the pleasure to offer you an internship opportunity on the following terms and conditions.

JOB TITLE: ENUMERATOR

1.0 TERMS OF CONTRACT
This is a maximum 30-day contract and will be effective from 24 January to 22 February 2015. You will be released from this contract as soon as you receive approval from your supervisor upon the completion of your assignment.

2.0 WORKING SCHEDULE
You will be required to work from Monday to Sunday for up to 30 continuous days. If you complete the enumeration of your assigned catchment area before 22 February 2015 (30th day after 27 January 2015), you can report to your supervisor and terminate your contract earlier. Your official working hours are from 07:30am to 04:30pm, but you are strongly required to manage flexible working hours. You may work even before 07:30am or after 04:30pm, whenever it is necessary.

3.0 PERFORMANCE EVALUATION
Your work performance will be evaluated in terms of speed and accuracy of your enumeration. Therefore, quick and accurate enumeration is strongly encouraged. Also, if you complete the assigned enumeration work before the end of the contract, it will be highly appreciated.

Note that, after the census is completed, the supervisors will re-enumerate every catchment area again and evaluate each one's enumeration in terms of accuracy and your attitude toward household members you interviewed.

Hence, your performance will affect your recommendation letter and a future job opportunity at the AFF.

4.0 SALARY
You will receive 10,000 MK as your wage for this contract.
2,000 MK will be provided at the beginning of the project, and the rest will be given upon the completion of the enumeration work.
You will be expected to enumerate a minimum of 160 households during the contract period, averaging 8 households per day. When you enumerate more than 160 households, you will be given an additional financial incentive of 500 MK per 8 households. You shall be provided accommodation in your catchment area during the contract period.

5.0 INTERNSHIP PROVISIONS
Upon the successful completion of this contract, you will be given **an official certificate**, which certifies that you worked as a census enumerator for the Africa Future Foundation (AFF) project, and **a recommendation letter** from the director of Project Chimutu (Mr. Hanyoun So) and the chief of TA Chimutu upon your request for your future job applications.

The recommendation letter will specify your relative performance of the enumerator work compared to your peers. In other words, if you do a good job, the recommendation letter will say so, whereas, if you do a bad job, the recommendation letter will say so.

In addition, upon the successful completion of this contract, you will be considered for the **future hire** of a regular staff position at the AFF office if you show outstanding performance satisfying the standard of the management of AFF.

6.0 TERMINATION OF EMPLOYMENT
In the event of any violation by employer of any of the terms of this contract, Employer may terminate employment without notice and with compensation to employee only to the date of such termination.

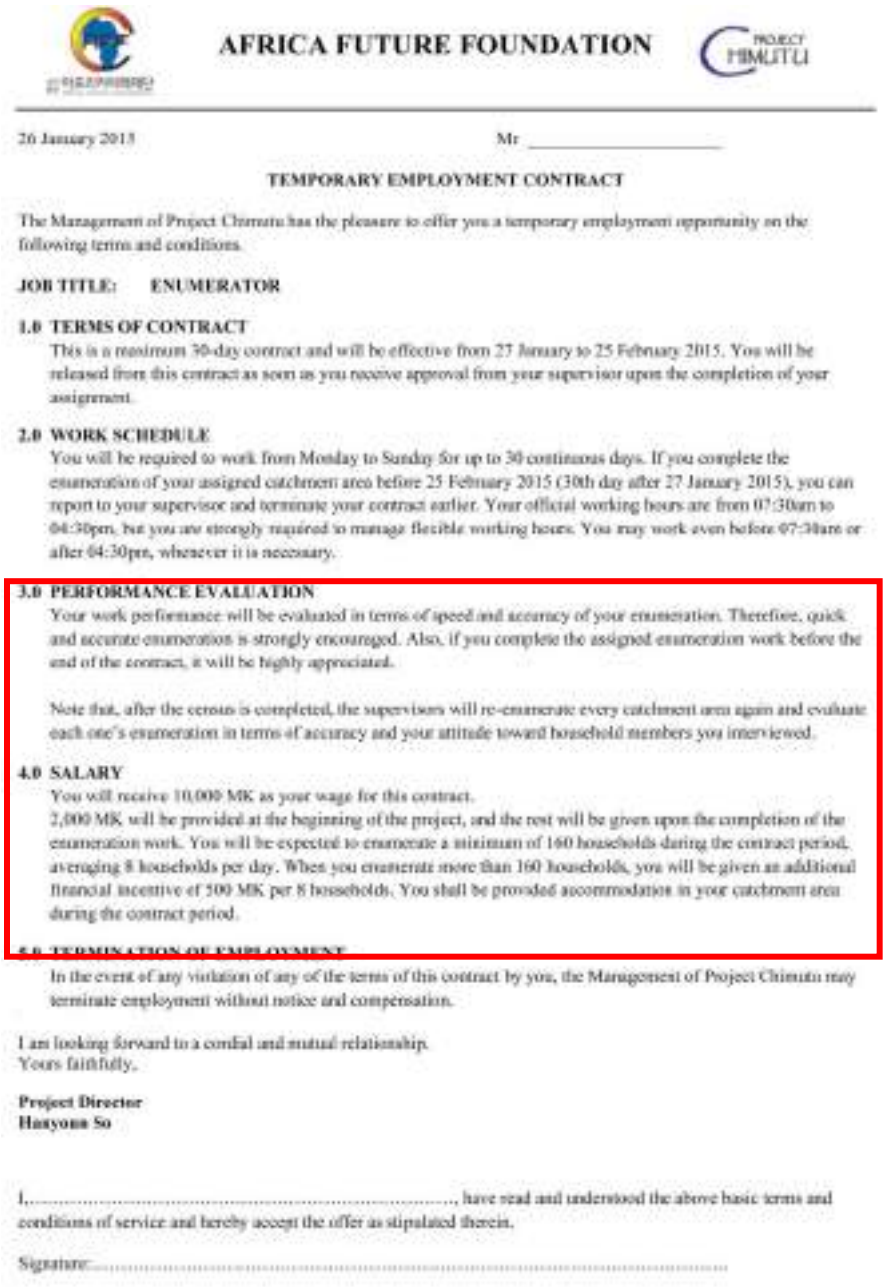
I am looking forward to a cordial and mutual relationship.
Yours faithfully,

Project Director
Hanyoun So

I,_____, have read and understood the above basic terms and conditions of service of the best of my knowledge and hereby accept the offer as stipulated therein.

Signature:_____

Figure A.3: Contract letter for Group 4 (G4)



AFRICA FUTURE FOUNDATION

PROJECT CHIMUTU

26 January 2015 Mr _____

TEMPORARY EMPLOYMENT CONTRACT

The Management of Project Chimutu has the pleasure to offer you a temporary employment opportunity on the following terms and conditions.

JOB TITLE: ENUMERATOR

1.0 TERMS OF CONTRACT

This is a maximum 30-day contract and will be effective from 27 January to 25 February 2015. You will be released from this contract as soon as you receive approval from your supervisor upon the completion of your assignment.

2.0 WORK SCHEDULE

You will be required to work from Monday to Sunday for up to 30 continuous days. If you complete the enumeration of your assigned catchment area before 25 February 2015 (30th day after 27 January 2015), you can report to your supervisor and terminate your contract earlier. Your official working hours are from 07:30am to 04:30pm, but you are strongly required to manage flexible working hours. You may work even before 07:30am or after 04:30pm, whenever it is necessary.

3.0 PERFORMANCE EVALUATION

Your work performance will be evaluated in terms of speed and accuracy of your enumeration. Therefore, quick and accurate enumeration is strongly encouraged. Also, if you complete the assigned enumeration work before the end of the contract, it will be highly appreciated.

Note that, after the census is completed, the supervisors will re-enumerate every catchment area again and evaluate each one's enumeration in terms of accuracy and your attitude toward household members you interviewed.

4.0 SALARY

You will receive 10,000 MK as your wage for this contract.
2,000 MK will be provided at the beginning of the project, and the rest will be given upon the completion of the enumeration work. You will be expected to enumerate a minimum of 160 households during the contract period, averaging 8 households per day. When you enumerate more than 160 households, you will be given an additional financial incentive of 500 MK per 8 households. You shall be provided accommodation in your catchment area during the contract period.

5.0 TERMINATION OF EMPLOYMENT

In the event of any violation of any of the terms of this contract by you, the Management of Project Chimutu may terminate employment without notice and compensation.

I am looking forward to a cordial and mutual relationship.
Yours faithfully,

Project Director
Haruyuu So

I,....., have read and understood the above basic terms and conditions of service and hereby accept the offer as stipulated therein.

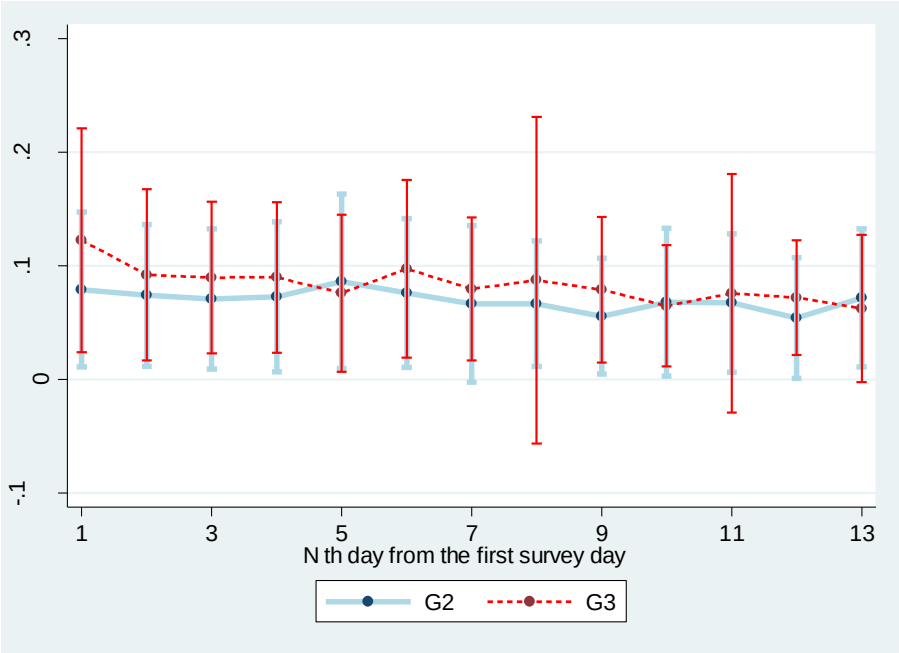
Signature:.....

Figure A.4: Training quiz questionnaire

No.	Question	Answer (Point)
1	An important reason for conducting the census is to achieve an improvement of overall quality of health in TA Chimutu. Describe the other two reasons why we conduct the census.	a. To make possible to reach out to every pregnant women who wanted to participate AFF MCH program. (0.5) b. To enrich the stock of socio-demographic data in T/A Chimutu that is necessary for elaboration of AFF MCH program. (0.5)
2	Regarding the roles of the enumerator, there are two functions you should NOT perform. Please fill them in the blank spaces below. A) Not to _____ B) Not to _____	a. Not to make any influence on answers (0.5) b. Not to change orders or words of questions (0.5)
3	What is the main standard required for households to be enumerated in the "2015 census of TA Chimutu," a modified version of the "population and housing census"?	Enumeration of all people, all housing units, and all other structures in TA Chimutu, who have stayed in TA Chimutu for more than 3 months during the past 12 months (1)
4	What is the name of the document that proves your eligibility to conduct the census?	Endorsement letter (1)
5	As what kind of structure would you categorize the following? *"A structure with sun-dried brick walls and asbestos roof"*	Semi-permanent (1)
6	Choose one that is <u>not</u> counted as a collective household. A) Hospitals, including three staff houses sharing food B) Lodge, including staff dwelling and sharing food C) Prison with many inmates' dwelling D) Store with owner's dwelling E) Military barracks with soldiers' dwelling	D (1)
7	What is the name of the document you have to sign before you start enumeration?	Consent form (1)
8	What are the three things you have to check before you leave the household?	Questionnaire, outbuildings, and Household ID number. (1, 0.5 point for partially correct)
9	What number do you put when you cannot meet any respondent from the household?	a. Do not put any number and just note down the household. (0.5) b. Put a latest number on it if you arrange to meet later. (0.5)
10	Your distributed alphabet is "C" and this household is the third household you enumerated in the catchment area. How did you place an ID number on the wall of the household?	0003C (1)
11	<u>True or false questions</u> A) It is okay if the questionnaire gets wet when there is heavy rain. B) You should not come to the completion meeting if you did not finish enumeration of your area. C) If you complete enumeration in your area, you should report to your supervisors immediately. D) You should bring all your housing necessities to the kickoff meeting.	A) False (0.5) B) False (0.5) C) True (0.5) D) True (0.5)

Figure A.5: Daily job performance trend

Panel A: Survey error rate



Panel B: Number of surveys per day

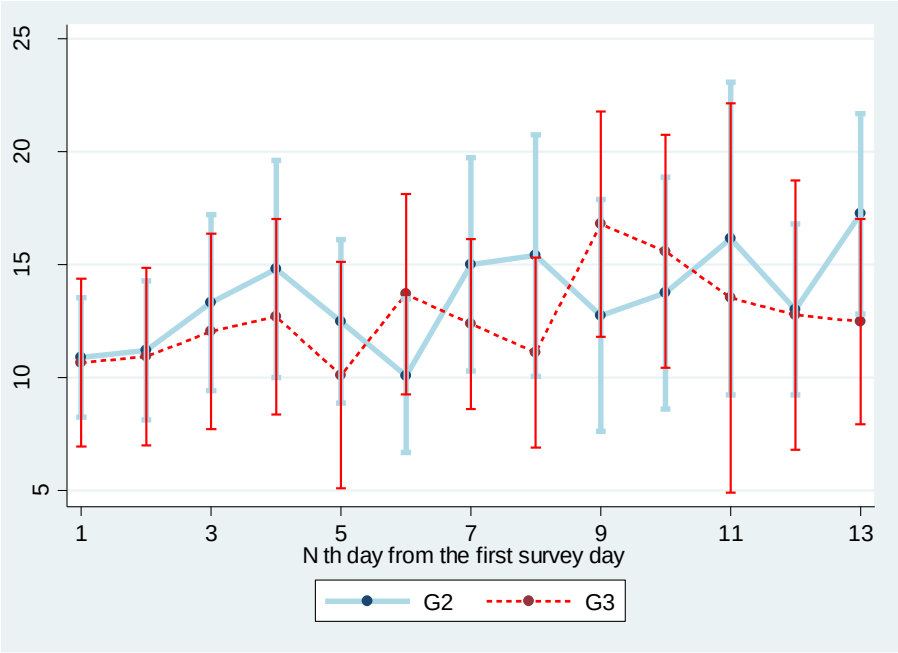
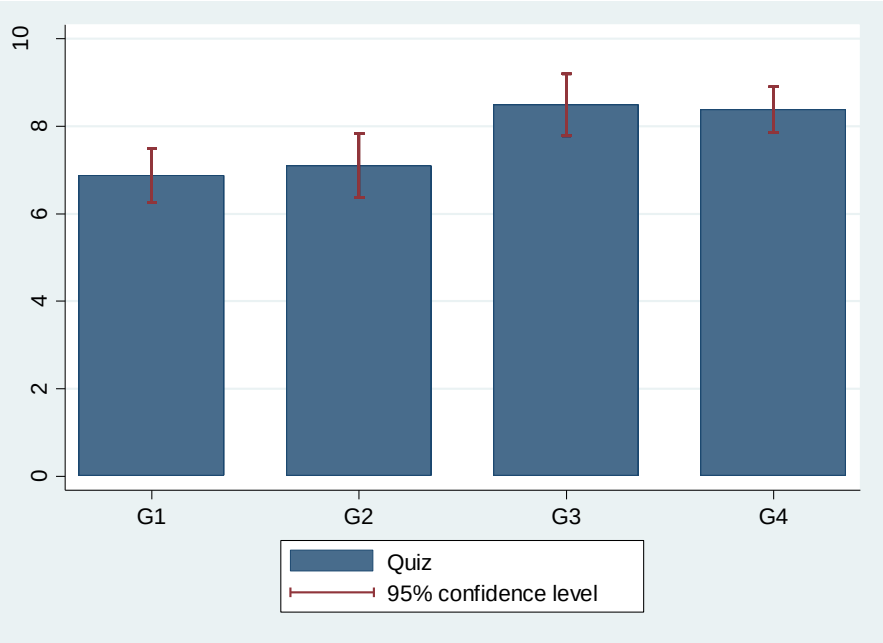
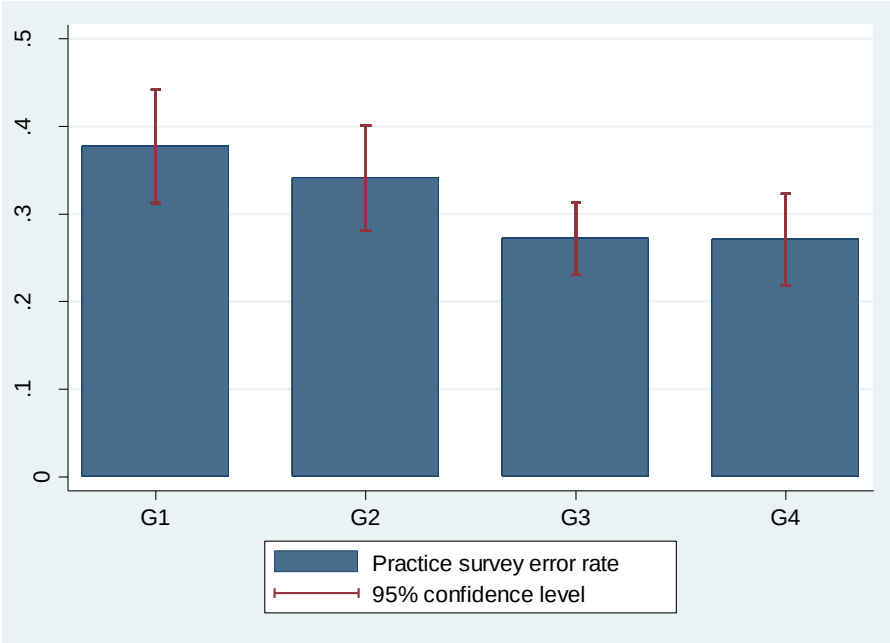


Figure A.6: Training outcomes

Panel A: Quiz



Panel B: Practice survey error rate



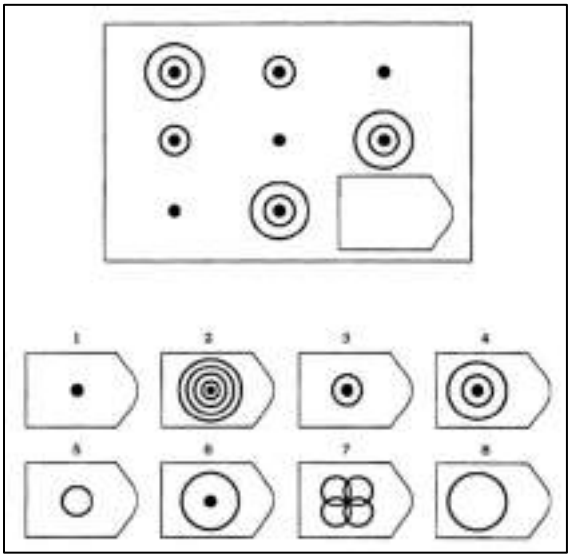
Data Appendix

Data A.1: Measurement of cognitive abilities and non-cognitive traits

A.1.1. Cognitive abilities

Raven's Progressive Matrices test

This is a widely used non-verbal test that evaluates "observation skills and clear-thinking ability" (Raven et al., 1998). Since it is independent of language skills, it is very easy to conduct in any setting including developing countries where the mother tongue is not English. The following figure is one example of the test questionnaire. In the test, a subject is required to choose one of eight options that match a missing pattern in the box. All questions follow similar visual patterns.



O*NET Ability Profiler

The O*NET Ability Profiler was originally developed by the United States Department of Labor as "a career exploration tool to help understand job seekers on their work skills" (O*NET Resource Center, 2010, p. 1). We use verbal and clerical ability tests of the Ability Profiler that are the most relevant skills for the enumerator job.

a. The verbal ability test measures how a test subject understands the definition of English words and properly uses them in conversation. Essentially, it is a

vocabulary test. The following is an example of the test questionnaire.

Choose the two words that are either most closely the same or most closely opposite in meaning

1. A. push
   B. dine
   C. nap
   D. eat

b. The clerical perception test measures an individual's "ability to see details in written materials quickly and correctly. It involves noticing if there are mistakes in the text and numbers, or if there are careless errors in working math problems. Many industrial occupations call for clerical perception even when the job does not require reading or math. This ability is measured by the Name Comparison exercise (O*NET Resource Center, 2010, p. 2)." The following is an example of the test questionnaire.

On the line in the middle, write S if the two names are exactly the same and write D if they are different.

1.	Paramore & Co.	—	Paramore & Co.
2.	Bimler	—	Binler
3.	E-Z Neon	—	E-Z Neon
4.	Blackstone	—	Blackstone
5.	Chris Brasch	—	Chris Grasch

Math and English scores of Malawi School Leaving Certificate Exam in 2014

All secondary school students in Malawi are required to take the Malawi School Leaving Certificate Exam during the third semester in Form 4 of secondary school (Grade 12 in the US) to achieve an official secondary school graduation status. The Malawi National Examination Board (MANEB) administers the whole process of the exam. Each student chooses 6–8 subjects out of about 20 subjects prepared by MANEB (MANEB, 2014). Math and English are mandatory subjects. The results of each subject are reported

in terms of a scale from 1 to 9. We use English and math test scores because they are mandatory subjects and thus, there are no missing values in the exam transcripts. We obtained the administrative record of the MSCE exam transcripts for all study participants through the Malawi Ministry of Education.

A.1.2. Non-cognitive traits

Rosenberg self-esteem scale

This is a 10-item scale developed by Rosenberg (1965) and is widely used to measure self-esteem by measuring positive and negative feelings about the self. All items are answered using a 4-point Likert scale format ranging from *strongly agree* to *strongly disagree*.

Intrinsic motivation

Intrinsic motivation is an individual's trait that captures whether the individual is motivated to do things by intrinsic rewards such as her own desire to pursue goals or challenges. It is the opposite of extrinsic motivation described below. We measure intrinsic motivation using a 15-item scale (Amabile et al., 1994). All items are answered using a 4-point Likert scale format ranging from *strongly agree* to *strongly disagree*.

Extrinsic motivation

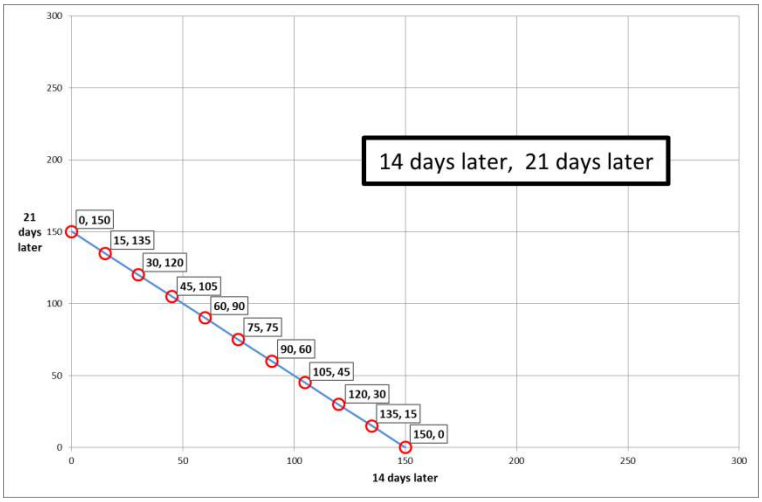
Extrinsic motivation is an individual's trait that captures whether the individual is motivated by external rewards, such as reputation, to do things. We use a 15-item scale to measure a level of motivation triggered by extrinsic values (Amabile et al., 1994). All items are answered using a 4-point Likert scale format ranging from *strongly agree* to *strongly disagree*.

Ten-item Big Five personality inventory (TIPI)

We measure an individual's personality types using a 10-item scale that assesses the respondent's characteristics based on traits commonly known as Big 5 personality traits (openness to experience, conscientiousness, extroversion, agreeableness, and emotional stability) (Gosling et al., 2003). All items are answered using a 7-point scale format (*Disagree strongly, Disagree moderately, Disagree a little, Neither agree nor disagree, Agree a little, Agree moderately, and Agree strongly*).

Time preference

In the 2015 secondary school student follow-up survey, AFF conducted an experiment measuring individual time preference with real monetary reward. Participants were given 20 decision problems. In each, they were asked to choose 1 out of 11 options on the line. Each option [X, Y] is a payoff set indicating the amount of money (X) they would receive 14 days later and the amount of money (Y) they would receive 21 days later (see the figure below). Participants were informed that AFF would randomly choose 1 out of 20 problems and would provide the amount of payoff the participants selected in the chosen decision problem according to the payoff rule.

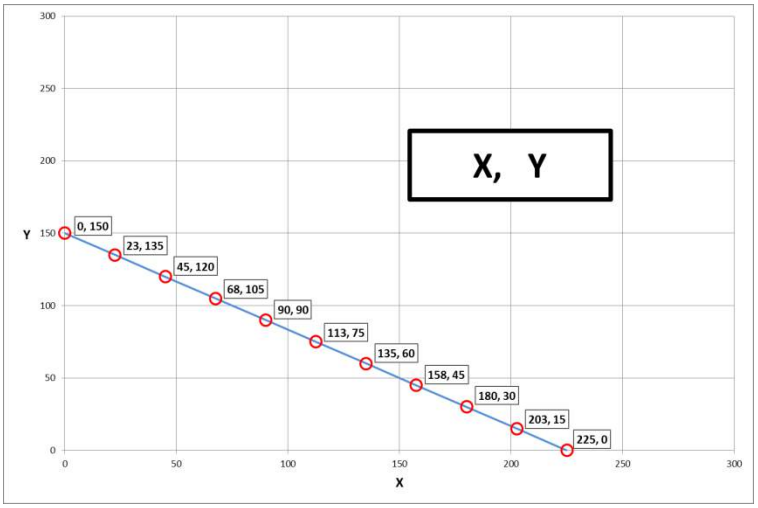


The choices individuals made through this experiment were used to infer their time preference, measured between 0 and 1 following the methodology proposed by Choi et al.

(2007). The closer the value is to 1, the more impatient a participant is, and the closer the value is to 0, the more patient the participant is.

Risk preference

In the 2015 secondary school student follow-up survey, AFF conducted another experiment measuring individual risk preference with real monetary reward. Participants were given 20 decision problems. In each, they were asked to choose 1 option out of 11 options on the line. An option [X, Y] indicates the amount of money a participant would earn if the X-axis (the horizontal axis) was chosen and the amount of money a participant would earn if the Y-axis (the vertical axis) was chosen (see the figure below). Participants were informed that AFF would randomly choose one out of 20 problems, and again randomly choose either X or Y with equal probability, and that the chosen payoff would be provided to the participants.



The choices made by individuals through this experiment were used to infer their individual-level risk preference measured between 0 and 1 following the methodology proposed by Choi et al. (2007). The closer the value is to 1, the more risk-taking a participant is, and the closer the value is to 0.5, the more risk-averse the participant is. Values lower than 0.5 reflect a violation of stochastic dominance and are excluded from the analysis (Choi et al, 2007).

Rationality

Using the Critical Cost Efficiency Index (CCEI; Afriat, 1972) we measured a level of consistency with the Generalized Axiom of Revealed Preference (GARP) based on the results from the time preference experiment. Considering all 20 decision problems in the time preference experiment, CCEI counts by how much the slope of the budget line in each problem should be adjusted to remove all violations of GARP. We took CCEI into account for the level of rationality (Choi et al, 2014). CCEI is measured between 0 and 1. The closer CCEI is to 1, the more a participant satisfies GARP overall, and the more rational (from an economic prospective) are the decisions made.

Center for Epidemiologic Studies Depression (CES-D)

The CES-D scale measures the extent to which the individuals experienced depression during the past week through items such as feeling lonely, having a poor appetite, and so on. We use a 20-item scale to measure a level of depression (Radloff, 1977). All items are answered using a 4-point Likert scale format ranging from *rarely or none of the time* to *most or all of the time*.

Pearlin Mastery scale

The Pearlin Mastery scale measures the extent to which the individuals consider that they have self-control over their own lives. We use a 7-item scale to measure a level of self-control (Pearlin et al., 1981). All items are answered using a 4-point Likert scale format ranging from s*trongly disagree* to *strongly agree*.

Accomplishment seeking

Accomplishment seeking is designed to measure the extent to which the individuals seek accomplishments. We use an 11-item scale to measure a level of accomplishment seeking (Barrick et al., 2002). All items are answered using a 4-point Likert scale format ranging from s*trongly disagree* to *strongly agree*.

Status seeking

Status seeking is designed to measure the extent to which the individuals seek a better status than others. We use an 11-item scale to measure a level of status seeking (Barrick et al., 2002). All items are answered using a 4-point Likert scale format ranging from s*trongly disagree* to *strongly agree*.

Internal motivation

Internal motivation measures the extent to which the individual is internally motivated. We use a 2-item scale to measure a level of internal motivation (Edmondson, 1999). All items are answered using a 4-point Likert scale format ranging from s*trongly disagree* to *strongly agree*.

Communion seeking

Communion seeking measures the extent to which the individuals seek communion in their lives. We use a 9-item scale to measure a level of seeking communion (Edmondson, 1999). All items are answered using a 4-point Likert scale format ranging from s*trongly disagree* to *strongly agree*.

Proactive personality

Proactive personality measures the extent to which the individual has a proactive personality. We use a 6-item scale to measure a level of proactive personality (Claes, 2005). All items are answered using a 4-point Likert scale format ranging from s*trongly disagree* to *strongly agree*.

Future orientation

Future orientation measures the extent to which the individual's behavior is future-oriented. We use a 3-item scale to measure a level of future orientation. Respondents are asked to choose one of two options in each of the three questions

1	A	You think a lot about things that might happen in the future.
	B	You usually just take things as they come.
2	A	You would rather spend your money and enjoy life today.
	B	You would save more for the future.
3	A	When you make plans ahead, you usually get to carry out things the way you expected.
	B	When you make plans ahead, things usually come up to make you change your plans.

Grit scale

The grit scale is designed to measure the extent to which the individual is passionate to finish a long-term goal. We use an 8-item scale to measure the level of grit an individual possesses (Duckworth and Quinn, 2009). All items are answered using a 5-point Likert scale format ranging from *not like me at all* to *very much like me*.

Job preference

The job preference scale consists of two questions asking about job characteristics. Each question has two options indicating two different kinds of jobs, and the individuals are asked to choose one.

1. Which job would you prefer?	A	A job you like even if the chances for a raise were small
	B	A job you do not like which offers a good chance for making more money
2. Which job would you prefer?	A	A job where you had a lot to say in what is going on
	B	A job where you had to think for yourself

Subjective expectation

We asked subjective probabilities of earning a four-year college degree and working for pay for more than 20 hours per week in the future.

		Percentage (0%-100%)
1	I earn a four-year college degree.	_____ **%**
2	I work for pay for more than 20 hours per week.	_____ **%**

Data A.2: Measurement of survey quality

AFF checked each questionnaire one by one and counted systematically inconsistent errors. First, project supervisors listed all possible systematic errors that could result from enumerators, not respondents. Second, data-entry clerks went through repeated training to catch those errors. Then, they started counting the number of systematic errors caused by enumerators for each sheet of the census survey.
Error collecting work was carried out in the following steps.

1. Two error-collecting data entry clerks checked one questionnaire separately.
2. They counted the total number of questions that must be answered.
3. 3. Three types of errors from each page of the questionnaire were counted, as follows.
   1) The total number of questions that must be answered but are blank.
   2) The total number of questions that must be answered but are wrongly answered.
   3) The total number of questions that must not be answered but are answered.
4. All the numbers on each page are added up and the total number of errors is recorded
5. The total number of errors independently counted by two clerks is compared.
6. If the difference between the total errors counted by two data entry clerks is larger than '5', a recount is undertaken.
7. The mean of the number of errors counted by two data entry clerks is recorded.

The following table provides the basic statistics of each number counted.

Index	Measurement	Mean (standard deviation)
A	The total number of all questions that must be answered	221.7 (61.8)
B	The total number of questions that must be answered but are blank	7.59 (10.3)
C	The total number of questions that must be answered but are wrongly answered	3.90 (4.26)
D	The total number of questions that must not be answered but are answered	5.53 (9.28)

Note: A could be different across households due to differences in household-specific characteristics, such as family structure.

Finally, the final outcome variable we use for survey quality (accuracy) in the analysis is as follows:

$$\text{error}_i = (B_i + C_i + D_i)/A_i$$

where error_i is the error rate of a specific census questionnaire i surveyed by an enumerator. A_i, B_i, C_i, and D_i are the corresponding numbers counted from the i-th census survey questionnaire by AFF's data clerks.

Data A.3: Imputation of missing survey beginning and end times

We find that there are significant missing values in the start time and end time on the surveyed census questionnaire due to the enumerators' mistakes. To preserve the sample size, we imputed either the survey beginning time or end time when only one of them is missing. Specifically, we first ran the regression of the questionnaire-specific length of survey.

$$\text{Surveytime}_{ijklt} = \alpha + \gamma \cdot H_i + \phi \cdot Z_k + V_{lt} + \sigma_t + \psi_{ijklt} \quad (A1)$$

$\text{Surveytime}_{ijktl}$ is survey time of household i by enumerator j whose supervisor is l, in catchment area k, surveyed on the t-th work day. H_i is a vector of respondents' household characteristics and Z_k is a vector of catchment area characteristics. σ_t is the survey date fixed effect. V_{lt} is the supervisor-specific post-visit effect. Standard errors are clustered at the enumerator level.

For surveyed census questionnaire sheets with either missing start time or end time, we imputed the missing time using the predicted length of a survey and non-missing beginning or ending time. Note that we could not use this method for an observation when both starting and ending times are missing. In this case, we did not make any changes and thus, the intermission time and survey length remain missing.

Data A.4: 2011 HIV/AIDS prevention programs of African Future Foundation

The HIV/AIDS prevention program of AFF covered 33 public schools in four districts in 2011: TA Chimutu, TA Chitukula, TA Tsabango, and TA Kalumba. In Table A.4, the experimental design of the study is summarized. The randomization process was implemented in two stages. Three types of interventions were randomly assigned to treatment groups independently. For HIV/AIDS education and male circumcision, classrooms were randomly assigned to one of three groups: 100% Treatment, 50% Treatment, and No Treatment classrooms. Treated students in 50% Treatment classrooms were randomly selected at the individual level. Treatments were given to everybody in 100% Treatment classrooms and to no one in No Treatment classrooms. For the girls' education support program, classrooms were randomly assigned either to the 100% Treatment or No Treatment group. AFF expected a minimal spill-over between classes because classrooms were self-contained, there were limited cross-classroom activities, and the majority (29 out of 33) of schools had only one class per grade.

The HIV/AIDS education intervention was designed to provide the most comprehensive HIV/AIDS education. On the top of the existing HIV/AIDS education curriculum, AFF additionally provided information on the medical benefits of male circumcision and the relative risk of cross-generational sexual relationships. The education was provided to both male and female students by trained staff members with a government certificate. The HIV/AIDS education was comprised of a 45-minute lecture and a 15-minute follow-up discussion. Study participants were assigned to one of four research groups: 100% Treatment (E1), Treated in 50% Treatment (E2), Untreated in 50% Treatment (E3), and No Treatment (E4).

The male circumcision offer consisted of free surgery at the assigned hospital, two complication check-ups (3-days and 1-week after surgery) at students' schools, and transportation support. Free surgery and complication check-ups were available for all study participants, but transportation support was randomly given. Selected students

could either choose a direct pick-up service or use a transportation voucher that is reimbursed after the circumcision surgery at the assigned hospital. The amount of a transportation voucher varied according to the distance between the hospital and a student's school. Study participants were also assigned to one of four research groups: 100% Treatment (C1), Treated in 50% Treatment (C2), Untreated in 50% Treatment (C3), and No Treatment (C4). Transportation support was given to groups C1 and C2 during the study period, and the remaining temporarily untreated group (groups C3 and C4) received the same treatment one year later.

Table A.4: Experimental Design

1) HIV/AIDS Education				
	Group	Assignment	Classrooms	Students
100% Treatment	E1	Treatment	41	2,480
50% Treatment	E2	Treatment	41	1,303
	E3	No treatment		1,263
No Treatment	E4	No treatment (Control)	42	2,925
Total			124	7,971
2) Male Circumcision				
	Group	Assignment	Classrooms	Students
100% Treatment	C1	Treatment	41	1,293
50% Treatment	C2	Treatment	41	679
	C3	No treatment		679
No Treatment	C4	No treatment (Control)	42	1,323
Total			124	3,974
3) Girls' Education Support				
	Group	Assignment	Classrooms	Students
100% Treatment	S1	Treatment	62	2,102
No Treatment	S2	No treatment (Control)	62	1,895
Total			124	3,997

Notes: For HIV/AIDS education and Male circumcision interventions, the randomizations were done in two stages. First, classrooms for each grade across 33 schools were randomly assigned to 100% treatment, 50% treatment, and no treatment. Then, within 50% treatment, only half of the students were randomly assigned to treatment.

The girls' education support program provided one-year school tuition and monthly cash stipends to female students in randomly selected classrooms (S1). School tuition and fees per semester (on average 21.2 USD) were directly deposited to each school's account and monthly cash stipends of 1.8 USD (300 MK) were distributed to treated students directly. The total amount of scholarship was approximately 70 USD per student during the study period.

References

Afriat, Sidney N. 1972. Efficiency Estimation of Production Function. *International Economic Review* 13 (3): 568–98.

Amabile, T.M., K.G. Hill, B.A. Hennessey, and E.M. Tighe. 1994. The Work Preference Inventory: Assessing Intrinsic and Extrinsic Motivational Orientations. *Journal of Personality and Social Psychology*, 66(5): 950.

Barrick, Murray R., Greg L. Stewart, and Mike Piotrowski, 2002. Personality and Job Performance: Test of the Mediating Effects of Motivation Among Sales Representatives, *Journal of Applied Psychology*, 87 (1), 43–51.

Choi, S., Fisman, R., Gale, D. and Kariv, S., 2007. Consistency and heterogeneity of individual behavior under uncertainty. *American Economic Review*, 97(5), pp.1921-1938.

Choi, S., Kariv, S., Müller, W. and Silverman, D., 2014. Who is (more) rational?. *American Economic Review*, 104(6), pp.1518-1550.

Claes, Rita, Colin Beheydt, and Björn Lemmens, 2005. Unidimensionality of Abbreviated Proactive Personality Scales Across Cultures, *Applied Psychology*, 54 (4), 476–489.

Duckworth, A.L, & Quinn, P.D, 2009. Development and validation of the Short Grit Scale (GritS). *Journal of Personality Assessment*, 91, 166-174

Edmondson, Amy, 1999. Psychological Safety and Learning Behavior inWork Teams, *Administrative Science Quarterly*, 44 (2), 350–383.

Gosling, S.D., P.J. Rentfrow, and W.B. Swann. 2003. A Very Brief Measure of the Big-Five Personality Domains. *Journal of Research in Personality*, 37(6): 504–528.

Malawi National Examination Board (MANEB). 2014. Malawi School Certificate of Education Examination—Grades and Awards for Candidates, Malawi.

O*NET Resource Center. 2010. *O*NET Ability Profiler™ Score Report*, pp. 1–2

Pearlin, Leonard I.; Lieberman, Morton A.; Menaghan, Elizabeth G.; and Joseph T. Mullan. 1981. The Stress Process. *Journal of Health and Social Behavior*, V.22, No. 4 (December): 337-356.

Radloff, L. S. 1977. The CES-D scale: A self report depression scale for research in the general population. *Applied Psychological Measurements*, 1, 385-401.

Raven, J., J.C. Raven, and J.H. Court. 1998. Manual for Raven's Progressive Matrices and Vocabulary Scales. The Standard Progressive Matrices, Section 3. Oxford Psychologists Press: Oxford, England/The Psychological Corporation: San Antonio, TX.

Rosenberg, Morris. 1965. Society and the adolescent self-image. Princeton, NJ: Princeton University Press.